

UNIVERSIDAD NACIONAL AUTONOMA DE MEXICO

FACULTAD DE MEDICINA

Análisis y rediseño in silico de la afinidad de las proteínas
periplásmicas de unión a aminoácidos

T E S I S

QUE PARA OBTENER EL TITULO DE
Licenciado en Investigación
Biomédica Básica

PRESENTA

BERNARDO OCHOA MONTAÑO

ASESOR: DR ALEJANDRO SOSA PEINADO

MÉXICO D.F

ABRIL 2006



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.

DEDICATORIAS

A mi familia, por su cariño y apoyo incondicionales.

A mis maestros, por compartir conmigo los pedazos de verdad que cada uno ha ido
recogiendo en el camino y por nutrir mi intereses.

A mis amigos, por los incontables momentos de alegría, confianza y aprendizaje.

To Zsuzsi, for having been my encouragement and motivation to reach out and
become better.

AGRADECIMIENTOS

A mi tutor, Dr. Alejandro Sosa Peinado, por su apoyo y guía durante la elaboración de este trabajo.

A los miembros de mi jurado y al Dr. Alejandro Fernández Velasco, por sus valiosos comentarios y sugerencias.

Al Dr. Lenin Domínguez Ramírez, por su ayuda en la reconstrucción/corrección de la estructura del PDB 1WDN, así como por sus explicaciones, sugerencias y discusiones.

A todos los miembros del Lab. de Físicoquímica e Ingeniería de Proteínas de la Facultad de Medicina, por juntos integrar el laboratorio más divertido y agradable que he conocido.

A mis padres, por la importancia e impulso que siempre le han dado a mi educación.

TABLA DE CONTENIDOS

ÍNDICE DE FIGURAS	4
ÍNDICE DE TABLAS	5
ABREVIATURAS	6
RESUMEN	7
INTRODUCCIÓN	8
FUNDAMENTACIÓN	9
La relación estructura-función de las proteínas	9
Interacciones proteína-ligando	10
Bioinformática estructural	11
Rediseño computacional con Rosetta	13
<i>Docking</i> y predicción de afinidades de unión	15
La superfamilia de las proteínas periplásmicas de unión	16
LAO, QBP y otras proteínas periplásmicas de unión a aminoácidos	18
HIPÓTESIS	23
OBJETIVOS	23
MÉTODOS	23
Estrategia general del rediseño	23
PDB: Estructuras cristalográficas	24
AutoDock: Generación de estructuras iniciales mediante <i>docking</i>	26
RosettaDock: selección de estructuras base y evaluaciones de energía nativas	28
LPC: Análisis de contactos ligando-proteína	29
RosettaDesign: Rediseño de afinidades	30
Funciones de evaluación independientes y clasificación por consenso	32
RESULTADOS Y DISCUSIÓN	34
Generación de estructuras	34
Evaluación de Rosetta de estructuras nativas de PBPs	37
LPC: Análisis de contactos	40
RosettaDesign y evaluaciones por consenso	44
Análisis de las estructuras seleccionadas por consenso	52
CONCLUSIONES Y PERSPECTIVAS	58
REFERENCIAS	60
Apéndice I : Opciones de Rosetta	64
Apéndice II: <i>Resfiles</i>	66

ÍNDICE DE FIGURAS

FIGURA 1: Diagramas topológicos de algunas proteínas periplásmicas de unión

FIGURA 2: a) Representaciones de caricatura de HBP, QBP y LAO con lisina.
b) Estructuras de LAO y LIV

FIGURA 3: Representaciones de Ligplot de las interacciones de unión del ligando en LAO con lisina y ornitina, QBP con glutamina, LIV con valina y HBP con histidina.

FIGURA 4: Diagrama de flujo resumiendo la metodología del presente trabajo

FIGURA 5: a) Estructuras de las conformaciones de AutoDock preseleccionadas para Rosetta
b) Agrupaciones de resultados en AutoDock

FIGURA 6: Representación del sitio de unión a glutamina para los dos rediseños mejor clasificados por puntuación consenso realizados mediante la estrategia libre

FIGURA 7: Representación del sitio de unión a glutamina para los dos rediseños mejor clasificados por puntuación consenso realizados mediante la estrategia conservativa

FIGURA 8: Representación del sitio de unión a glutamina para los dos rediseños mejor clasificados por puntuación consenso realizados mediante la estrategia guiada.

FIGURA 9: Representación del sitio de unión a valina para los dos rediseños mejor clasificados por puntuación consenso.

ÍNDICE DE TABLAS

- TABLA 1:** Afinidades de unión de PBPs de aminoácidos
- TABLA 2:** A) Cálculo de RosettaDock de energías de unión (“*lig_sum*”) para selección de conformaciones
B) Suma de parámetros de la función de energía global de Rosetta Dock (“*bk_tot*”)
- TABLA 3:** Valores de energía de unión a ligando “*lig_sum*” para estructuras nativas de PBPs
- TABLA 4:** Tablas de LPC de contactos atómicos ligando-proteína para las construcciones LAO-Q1, LAO-Q2 y LAO-V, QBP, LAO-K y LIV-V
- TABLA 5:** 25 mejores resultados por puntuación promedio para estrategia de rediseño libre
- TABLA 6:** 25 mejores resultados por puntuación promedio para estrategia de rediseño conservativa
- TABLA 7:** 25 mejores resultados por puntuación promedio para estrategia de rediseño guiada
- TABLA 8:** 25 mejores resultados por puntuación promedio para todas las estrategias
- TABLA 9:** 25 mejores resultados globales ordenados por resultados de BAPPL para todas las estrategias
- TABLA 10:** 25 mejores resultados globales ordenados por resultados de BAPPL para todas las estrategias

RESUMEN

Utilizando la nueva modalidad de ligando del programa Rosetta y en combinación con otras herramientas bioinformáticas, se realizaron propuestas *in silico* para aumentar la afinidad de la proteína periplásmica de unión a lisina, ornitina e histidina (LAO) a los aminoácidos glutamina y valina. Se generaron modelos de los complejos de LAO-Glu y LAO-Val mediante los programas AutoDock y RosettaDock y sus sitios de unión a ligando fueron rediseñados por cuatro estrategias mediante el programa RosettaDesign para optimizar su energética de unión. Las afinidades de todos los complejos rediseñados fueron estimadas también por las funciones de evaluación X-Score, DrugScore y BAPPL. Se seleccionaron siete propuestas finales a través del consenso de las evaluaciones de las distintas funciones.

Los resultados presentaron mutaciones frecuentes, dependiendo de la estrategia, caracterizadas en general por promover principalmente la formación nuevos puentes de hidrógeno, ya sea con el ligando o intramoleculares, lo que favorecería la estabilidad estructural de la proteína. En la selección por consenso existió cierta complementariedad entre las distintas evaluaciones y sobresalieron diseños tanto con las mutaciones más frecuentes como con otras menos comunes.

INTRODUCCIÓN

La gran mayoría de las funciones biológicas son realizadas por proteínas. Su importancia primordial para la vida en nuestro planeta radica en la inmensa versatilidad funcional de estas macromoléculas. Las proteínas son polímeros lineales de α -L-aminoácidos de longitud variable, desde 51 residuos como en la insulina, hasta varios miles de como en la titina, de 27,000, aunque la gran mayoría de ellas suelen poseer entre 100 y 2000 aminoácidos. [3, 4]. Como ejemplo ilustrativo de su variedad potencial, si consideramos que existen 20 aminoácidos diferentes, para una proteína típica de unos 100 residuos podrían existir hasta 20^{100} ó 1.268×10^{130} permutaciones distintas, un número mucho mayor que la cantidad de átomos en el universo visible. No obstante, la funcionalidad de las proteínas no proviene simplemente de la inmensa variedad posible de su composición, la cual es imposible de explorar en su totalidad, sino de la diversidad de propiedades químicas de los aminoácidos que las conforman. También, de la gran complejidad de sus interacciones, lo que da origen a que las cadenas polipeptídicas lineales puedan adquirir una estructura tridimensional definida en el proceso llamado plegamiento y, en muchos casos, asociarse con otras cadenas iguales o diferentes para integrar, funcionalmente, una sola proteína multimérica.

Con la predicción de la formación de estructuras secundarias en polipéptidos en 1951 por Linus Pauling y la determinación por cristalografía de rayos X de la estructura tridimensional de la mioglobina por John Kendrew en 1958, comenzó a develarse la importancia de la conformación espacial de la cadena polipeptídica para el funcionamiento de las proteínas. Asimismo, el descubrimiento por Vernon Ingram en 1957 que un solo cambio de aminoácido en la hemoglobina humana era responsable de la enfermedad conocida como anemia falciforme y, más tarde, la elucidación de la estructura de la hemoglobina por Max Perutz, ilustraron la crucial relevancia de las propiedades químicas y posicionamiento sutil de ciertos residuos clave en las proteínas.

A casi medio siglo y con alrededor de 35,000 estructuras proteicas conocidas en la base de datos del Protein Data Bank (PDB)[5], en numerosos casos ya somos capaces de comprender cómo mutaciones en ciertos residuos afectan a una proteína estructural y funcionalmente y, en otros tantos, de predecir las consecuencias de algunas mutaciones, provocarlas y analizarlas en el laboratorio. A consecuencia de estos avances, el diseño de proteínas[6], que busca la manipulación racional de proteínas para introducir cambios en su función, estructura o estabilidad, ha cobrado auge recientemente por la gran potencialidad que esto representa para la medicina y la biotecnología.

El diseño de proteínas representa un reto a nuestro entendimiento a nivel molecular de su estructura y función. Debido a la gran complejidad y número de interacciones relevantes para la funcionalidad, plegamiento y estabilidad de una proteína, resulta de gran importancia y utilidad tener la capacidad de analizar grandes cantidades de datos estructurales y procesarlos eficientemente de acuerdo a modelos derivados del estudio y comprensión de procesos. El gran progreso reciente de la capacidad de cómputo y en los algoritmos para utilizarla han convertido a la bioinformática estructural en una bienvenida herramienta de gran flexibilidad y potencial para el estudio y manipulación eficiente de la vasta información relacionada con el funcionamiento molecular de las proteínas. Con el desarrollo de métodos nuevos y más eficientes, la bioinformática estructural promete ser capaz de generar modelos *in silico* cada vez más exactos del comportamiento fisicoquímico de las proteínas para estudiarlas o manipularlas con una fracción del costo y tiempo asociados a la combinatoria experimental requerida para abordar los problemas desde un enfoque meramente empírico.

El presente trabajo pretende hacer uso de estos avances recientes y metodologías nuevas para estudiar una de las áreas más relevantes de la bioquímica que es la unión de ligandos a las proteínas.

FUNDAMENTACIÓN

La relación estructura-función de las proteínas

A través del proceso llamado plegamiento, las proteínas adquieren una o varias conformaciones tridimensionales consideradas más o menos estables en solución las cuales se encuentran determinadas, en primer lugar, por su secuencia primaria de aminoácidos. Estas estructuras están asociadas a la función de la proteína, ya sea enzimática, mecánica, de señalización, transporte, etc. y que, por lo general, se encuentra altamente especializada para una elevada especificidad o eficiencia.

Muchas de las funciones biológicas son parecidas entre sí, lo cual se evidencia normalmente en la similitud de las secuencias de las proteínas involucradas o, más esencialmente, de sus estructuras, que han sido preservadas y diversificadas por la evolución debido a su utilidad. A pesar de la semejanza de las proteínas dentro de las distintas familias, éstas difieren entre sí por variaciones, algunas de las cuales afectan su funcionalidad y han sido seleccionadas por proveer modificaciones útiles o nuevas cualidades ventajosas para la célula o el organismo.

Todos los estudios realizados sobre los mecanismos moleculares de las funciones específicas de las proteínas han revelado que éstas dependen en mayor o menor grado de que ciertos residuos tengan propiedades y posiciones precisas, al mismo tiempo de que la estructura global reúna ciertos requisitos de estabilidad y flexibilidad. Por lo tanto, desde el nivel global hasta el sutil, la estructura tridimensional de las proteínas es de una relevancia esencial para su funcionamiento.

Aunque se ha aprendido suficiente acerca de numerosas proteínas como para derivar algunos principios básicos de la correlación estructura-función, las interacciones inter- e intramoleculares en las proteínas generalmente son demasiado complejas como para poder obtener un entendimiento simple de cómo un cambio, o una serie de ellos, afectaría exactamente las distintas propiedades de una proteína. Por tal motivo, muchas veces resulta necesario realizar estudios especializados o bien desarrollar herramientas avanzadas que permitan simplificar los problemas a nuestras capacidades de la forma más adecuada posible.

Finalmente, dada la importancia de las proteínas, no es de extrañar la frecuencia con que alteraciones en éstas están involucradas directamente en enfermedades: desde el caso clásico de la anemia falciforme, donde un residuo de glutamato de la hemoglobina es sustituido por una valina hidrofóbica, que provoca la generación de fibras por agregación y polimerización; hasta la infección del VIH-1, iniciada por la interacción de la glucoproteína viral gp120 con el receptor CD4 en células

del sistema inmune. Afortunadamente, el estudio de las proteínas también ha comenzado a rendir frutos en esta área aplicada de gran relevancia, y para ambas enfermedades se han desarrollado recientemente fármacos[7] o posibles tratamientos[8] gracias al entendimiento existente acerca de la estructura de las proteínas relevantes.

Interacciones proteína-ligando

Una de las propiedades más básicas e importantes de las proteínas es la unión de diversas moléculas pequeñas o ligandos. Gracias a la diversidad de propiedades químicas y la capacidad de las cadenas de aminoácidos de adquirir una inmensa variedad de estructuras, las proteínas tienen la potencialidad de unir prácticamente cualquier molécula orgánica con alta especificidad, y sus ligandos naturales van desde iones hasta grandes compuestos orgánicos como polinucleótidos y polipéptidos.

La afinidad y especificidad de la unión de un ligando con una proteína depende de las interacciones de éste con los residuos expuestos de la proteína. Las interacciones se clasifican químicamente como débiles y comprenden fuerzas hidrofóbicas, interacciones coulombicas entre cargas, carga-dipolo, dipolo-dipolo (que incluye el caso especial del puente de hidrógeno), dipolo inducido-dipolo y dipolo inducido-dipolo inducido[9]. Entre las más comunes y determinantes están el efecto hidrofóbico, causado por la estructura del agua que excluye moléculas hidrofóbicas y favorece la unión de un ligando no polar a un sitio no polar en la proteína; las interacciones iónicas entre grupos cargados y los puentes de hidrógeno. El posicionamiento espacial y accesibilidad de residuos capaces de tener dichas interacciones con otras moléculas determinan la especificidad y la afinidad de la unión. Esta última se describe mediante la constante de disociación K_d , que representa la constante de equilibrio de la disociación de un complejo proteína-ligando:



$$K_d = \frac{[\text{Proteína}][\text{Ligando}]}{[\text{Complejo}]} \quad \text{Ec. 1}$$

Las constantes de disociación en sistemas biológicos pueden variar en varios órdenes de magnitud y el rango relevante para su función biológica puede ir desde milimolar hasta picomolar en el caso de algunos anticuerpos [3].

La función correcta de muchas proteínas depende íntimamente de su afinidad por las diversas moléculas con las que interactúan y un cambio estructural que afecte negativamente la unión de alguna proteína puede dar origen a una deficiencia metabólica en el organismo. La toxicidad y peligrosidad de muchos compuestos puede ser rastreada a su interacción con las proteínas del organismo y el trastorno de su función natural. Asimismo, la unión de ligandos o receptores en los procesos de infección de un gran número de agentes patógenos hace altamente relevante el estudio y manipulación de afinidades para la creación de fármacos terapéuticos.

Bioinformática estructural

El gran crecimiento en el número de las estructuras tridimensionales depositadas en el PDB en la última década ha sextuplicado el número de estructuras disponibles en esta base de datos en los últimos diez años y casi duplicado en los últimos cuatro [10]. Este auge también ha coincidido con el desarrollo de métodos computacionales diseñados para extraer información relevante tanto de manera general y global, como topología y similitud estructural; como específica del funcionamiento de determinadas proteínas, tales como estabilidad general o interacciones particulares.

El reciente aumento explosivo en la capacidad de cómputo ha permitido que estos métodos proliferen y sean capaces hoy en día de abordar problemas complejos y tomar en cuenta gran número de procesos que se han ido descubriendo y caracterizando en las últimas décadas acerca de las proteínas. Estos métodos requerían, hasta hace poco, de recursos costosos y poco accesibles para una gran parte de los grupos de investigación, pero hoy en día es relativamente sencillo establecer sistemas capaces de realizar muchos de los métodos eficientemente, al menos a pequeña escala.

La bioinformática estructural comprende una gran variedad de aplicaciones desde la visualización molecular; numerosas bases de datos entre las que destaca el Protein Data Bank (PDB); alineación, clasificación y comparación de estructuras; identificación de dominios o motivos estructurales, “folds”, estructura secundaria o posible función, con base en similitudes estructurales; así como predicción de interacciones proteína-proteína o proteína-ligando e incluso de estructura, a partir de varias estrategias [11]. Estas aplicaciones se encuentran en diversas etapas de desarrollo y en la mayoría de ellas aún hay mucho por mejorar. No obstante, incluso en uno de los campos más desafiantes, como el de la predicción de estructura terciaria, se han hecho significativos avances en la última década, como han demostrado los experimentos comunitarios “Critical Assessment of Structure Prediction” (CASP) [12, 13], los cuales se realizan cada dos años entre la comunidad

científica de predicción de estructura y en los cuales los grupos de investigación participantes buscan realizar predicciones de la estructura de moléculas blanco antes de que éstas sean determinadas experimentalmente o publicadas. Otra prueba comunitaria de este estilo es CAPRI [14] o Critical Assessment of PRedicted Interactions, donde el objetivo es predecir las interacciones entre proteínas a través del llamado “docking”.

A pesar de los logros alcanzados, muchas de las herramientas continúan siendo principalmente aproximativas; aun así, la creciente precisión de los métodos ha alcanzado un punto de utilidad real para realizar predicciones independientes de manera convincente, como lo han demostrado los CASP. Debido a lo anterior, en el presente trabajo decidimos hacer uso de estos avances para abordar el tema de las bases estructurales de la afinidad proteína-ligando y evaluar el potencial de los métodos *in silico* para la manipulación racional de ésta, en lo que se conoce como rediseño de proteínas.

Rediseño computacional con Rosetta

Desde que se reconoce la importancia de las proteínas en todos los procesos biológicos, el hombre ha tenido el deseo de comprender y aprovechar su inmensa potencialidad, tanto para ser capaz de investigar más profundamente la naturaleza, como para beneficiarse por la estrecha relevancia que tienen en nuestras vidas, combatiendo así enfermedades e impulsando el progreso tecnológico. La modificación racional de las funciones y propiedades de proteínas existentes es uno de los enfoques para ello y se le conoce como rediseño.

El progreso de la bioinformática y su capacidad para realizar simulaciones y aproximaciones cada vez más precisas con una cantidad relativamente reducida de recursos, la ha llevado a ser ahora una herramienta crítica para los esfuerzos de rediseño. Uno de los programas recientes que, en distintas versiones o modalidades, ha participado con bastante éxito tanto en el CASP[15] como en el CAPRI[16], es el Rosetta[17]/RosettaDock[18], desarrollados originalmente en el laboratorio del Dr. David Baker en la Universidad de Washington en Estados Unidos.

En su versión original, Rosetta fue concebido para la predicción *de novo* de la estructura de proteínas, pero el éxito que ha tenido lo ha llevado a diversificarse hacia otros problemas como el de las interacciones proteína-proteína con su modalidad de *docking* “RosettaDock”, y en la de diseño de secuencias compatibles con una estructura, RosettaDesign[19], con la que se ha logrado rediseñar completamente las secuencias de proteínas naturales y lograr un plegamiento con la misma topología y con estabilidad similar (o superior aun) [20] e incluso, crear una secuencia de aminoácidos que se pliega en una forma completamente nueva[21]. No obstante, estos estudios se encuentran limitados a proteínas pequeñas (~ 100 a.a.), debido a la complejidad y al poder y tiempo de cómputo requeridos.

El enfoque general de Rosetta pretende simular el efecto de las interacciones locales y globales en la determinación de la estructura de una proteína y para ello se basa en métodos Monte Carlo, los cuales incorporan el uso de números aleatorios y probabilidades estadísticas para hacer un muestreo parcial pero representativo relevante al problema en cuestión. Rosetta maneja bibliotecas de pequeños fragmentos proteicos, de tres o nueve aminoácidos, y de rotámeros (es decir, conformaciones de cadenas laterales recopiladas a partir estructuras cristalográficas y de NMR con base a su frecuencia) como parte de su estrategia para reducir la búsqueda computacional y acelerar así el proceso de cómputo, al igual que aproximaciones en los potenciales a escalas “gruesas” para simplificar eventos globales. No obstante, también implementa funciones de potencial o campos de

fuerza a nivel atómico para el afinamiento a niveles detallados y puede incluir pequeñas perturbaciones de posiciones y ángulos para hacer un muestreo más robusto y flexible.

Rosetta evalúa las energías de interacción entre átomos mediante una suma ponderada de diversos parámetros :

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

Ec. 2

donde las W representan los distintos coeficientes de ponderación y las E los valores de energía de diversas fuentes, que incluyen potenciales de Lennard-Jones para interacciones de van der Waals (E_{atr} y E_{rep}), modelos de solvatación (E_{solv}), puentes de hidrógeno (E_{bb_hbond} , E_{sc_hbond} , $E_{sc_bb_hbond}$), potenciales empíricos entre pares de residuos (E_{pair}) y probabilidades de rotámeros (E_{rot}) y ángulos phi-psi ($E_{aa\backslash\phi,\psi}$, E_{rama})

Una de las extensiones más recientes a Rosetta, tanto en sus modalidades de diseño como de *docking*, es el modo ligando, desarrollado para habilitar la evaluación energética de las interacciones entre la proteína y moléculas pequeñas, de manera que posibilita el rediseño de la unión de la proteína con un ligando o bien el *docking* de ligando, que permite considerar posibles efectos de ajuste inducido. La implementación de este modo aún no ha sido publicada formalmente y una versión de desarrollo fue gentilmente proporcionada a nuestro laboratorio por el Dr. Jens Meiler del laboratorio del Dr. David Baker.

A través del modo diseño de tipo ligando, Rosetta es capaz de modificar tanto la identidad como la conformación de los residuos que se especifique, ya sea con el fin de crear un nuevo sitio de unión para un ligando, o bien, para optimizar la unión con un ligando conocido. Como parte del algoritmo, Rosetta realiza un estimado interno de la energía de interacción de la proteína con el ligando en cada iteración y, al final, éste sirve para indicar los mejores resultados del proceso. La misma evaluación se realiza en modo *docking* para buscar la conformación de unión más probable y precisa.

Dado el estado novedoso y no público de esta función de Rosetta, aún no existen reportes de la efectividad de éste método para realizar predicciones acertadas; no obstante, procedimientos computacionales para diseñar sitios de unión ya han sido utilizados recientemente con resultados impresionantes, con Homme Hellinga de la Universidad de Duke como pionero. A través de un método basado en algoritmos de eliminación de callejones sin salida o “dead-end elimination”

(DEE)[22], Hellinga y sus colaboradores han logrado transformar la afinidad de la proteína de unión a maltosa en un biosensor de zinc[23], dar actividad de triosafosfato isomerasa a la proteína de unión a ribosa[24] y rediseñar en varias proteínas la capacidad de unir ligandos como el trinitrotolueno (TNT), L-lactato y serotonina[25] o el producto hidrolítico de un gas nervioso[26].

Aunque el funcionamiento de Rosetta opera bajo principios matemáticamente distintos al método de Hellinga, como ya se vio (Monte Carlo vs. eliminación de callejones sin salida) sus resultados positivos en las áreas del plegamiento y *docking* de proteína a proteína resultan alentadores para su incursión en el terreno del rediseño de unión a ligandos.

Docking y predicción de afinidades de unión

Si bien los principios básicos de la formación de complejos entre proteínas y ligandos son relativamente bien conocidos, la predicción detallada de ellos a partir de sus estructuras y de las repercusiones de su interacción a nivel bioquímico son problemas complejos que se han intentado abordar de diversas maneras. La primera parte de la cuestión consiste en encontrar el sitio y modo de unión correctos entre la proteína y el ligando, el cual se ha intentado resolver mediante los llamados métodos de “docking”[27, 28]. El problema del *docking* se compone de dos partes. La primera consta de la búsqueda de conformaciones espaciales y fisicoquímicas favorables de la interacción del receptor con el ligando, mientras que la segunda radica en la evaluación del complejo potencial para estimar su energética (ΔG) y, por lo tanto, su posible afinidad. Aunque puede hacerse la distinción entre ambas partes, éstas se encuentran en mayor o menor medida interrelacionadas, ya que muchos algoritmos de búsqueda dependen de una función integrada para evaluar su progreso y cualquier función de evaluación depende de un método para generar interacciones potenciales.

Los métodos computacionales de búsqueda para *docking* que se han desarrollado son diversos y comprenden métodos por fuerza bruta, de complementariedad de forma o estocásticos en forma de algoritmos genéticos o Monte Carlo. También existen distintos tipos de funciones de evaluación (“*scoring functions*”) y generalmente se clasifican como basadas en campo de fuerza explícito, empíricas o como “basadas en conocimiento” (*knowledge-based*) [29]. Las primeras estiman la energía libre de unión a partir de campos de fuerza como el de AMBER o CHARMM para calcular las interacciones directas, mientras que las empíricas o basadas en regresión, lo hacen a través de una suma ponderada de su valoración de diversos tipos de interacciones, como formación de puentes de hidrógeno, efectos de solvatación, etc., cuyos pesos relativos son derivados mediante regresiones de

conjuntos de datos empíricos de “calibración” o de “entrenamiento” con afinidades experimentales conocidas. Las funciones basadas en conocimiento calculan en primer lugar pseudo-potenciales entre pares de átomos o residuos a partir de análisis estadísticos de un gran número de interfases y, basados en ellos, favorecen o penalizan los contactos del complejo para realizar la estimación.

Dada la inmensa importancia para la industria farmacéutica de poder examinar eficientemente un gran número de fármacos potencialmente terapéuticos previamente a la fase experimental, existe un gran interés en que las funciones de evaluación sean capaces de generar información relevante y útil. Existen varios criterios por los que se puede juzgar una función, dependiendo de los objetivos personales, que son su competencia para distinguir el modo correcto de unión entre varias conformaciones alternativas; su capacidad para seleccionar ligandos potenciales entre una serie de compuestos inactivos y la de generar una jerarquización o clasificación entre éstos con base en un estimado de la afinidad de unión[30].

En el caso del rediseño racional de la unión de una proteína con un ligando, la función más relevante es la de generar una predicción de la afinidad de unión. Se han hecho algunos estudios comparativos[31-34] de diversas funciones, pero aunque se ha observado progreso en la precisión en varios aspectos de sus objetivos, también se ha encontrado que la correlación entre afinidad de unión experimental y la predicha por las funciones no siempre es tan alta como se desearía. La evaluación por consenso, es decir, que toma en consideración los resultados de diversas funciones, es capaz por razones estadísticas de mejorar la correlación entre las predicciones y los valores experimentales[35], aunque probablemente no lo suficiente para lograr un estimado cuantitativo confiable de la afinidad de unión absoluta. No obstante, cuando el objetivo es mejorar una afinidad incipiente de un receptor por un ligando similar al nativo, consideramos que la capacidad de las funciones de evaluación para establecer una predicción relativa del cambio de afinidad, puede ser suficiente para discriminar resultados promisorios de los que no lo son.

La superfamilia de las proteínas periplásmicas de unión

Las proteínas periplásmicas de unión (PBP por “periplasmic binding proteins”) constituyen una superfamilia bacteriana involucrada en las funciones de quimiorrecepción y transporte transmembranal[36]. Realizan estas funciones de manera acoplada con proteínas membranales de quimiorrecepción, de usualmente dos cruces transmembranales, o de transporte, que cruzan la membrana cinco o seis veces. Estas últimas, los transportadores llamados de tipo cassette de unión a

ATP (*ABC-type* por sus siglas en inglés) interactúan con los complejos de una o dos proteínas solubles de unión para transportar al ligando a través de la membrana. Existen proteínas periplásmicas de unión a una gran variedad de compuestos que incluyen mono- y disacáridos, aminoácidos, péptidos y iones, entre otros.

Las PBPs son monoméricas y se caracterizan por consistir de dos dominios globulares de tipo principalmente α/β , conectados de manera flexible a través de una región de bisagra, y por unir sus ligandos en la hendidura entre los dominios. La unión se ve acompañada por un desplazamiento del equilibrio de la proteína de una conformación abierta hacia una cerrada, por lo que el ligando queda envuelto entre los dominios. Aunque la identidad de secuencia entre la mayoría de los pares de PBPs

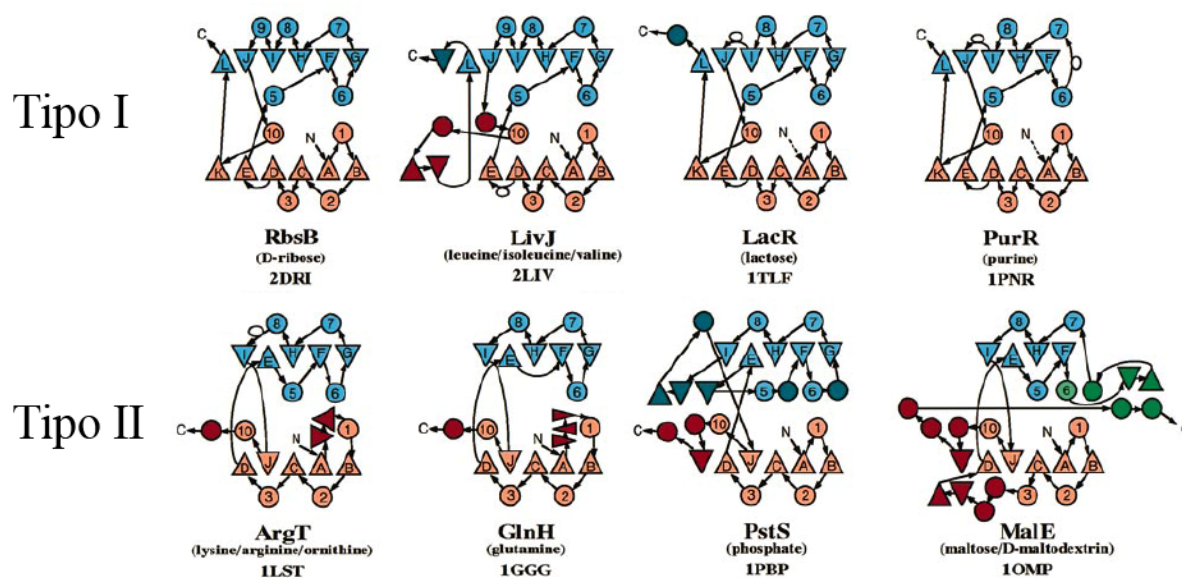


Figura 1: Diagramas topológicos de algunas proteínas periplásmicas de unión. Los dominios 1, 2 y 3 se representan de color rojo, azul y verde, respectivamente. Las hélices se muestran como círculos y las hebras- β como triángulos, cuya orientación entre indica si son paralelas o antiparalelas. Las letras y números se asignan a las estructuras en posición estructural equivalente. Los bucles en las líneas indican hélices 3_{10} . (Adaptado de [37])

es baja e insuficiente para inferir relaciones filogenéticas precisas, se han podido clasificar en ocho conjuntos por su secuencia y función [36]. A través del estudio de su topología estructural, las PBPs se han podido agrupar en dos tipos denominados I y II [37]. Éstos se diferencian en el arreglo de las hojas β en cada dominio su estructura, que en el tipo I adoptan la forma $\beta_2\beta_1\beta_3\beta_4\beta_5$ mientras que en el tipo II la topología es de la forma $\beta_2\beta_1\beta_3\beta_n\beta_4$, donde β_n es la hebra inmediatamente posterior al cruce de la cadena un dominio al otro (Figura 1).

Las características de las PBPs las hacen un excelente modelo para la ingeniería de proteínas como lo demuestran los exitosos rediseños computacionales descritos realizados en el laboratorio de

Hellinga [23-26], los cuales fueron llevados a cabo en miembros de esta superfamilia. Su gran versatilidad radica en las condiciones de unión al ligando [38]. El posicionamiento del sitio entre los dominios simula el interior de una proteína, con las ventajas que otorga un ambiente de baja constante dieléctrica sin exposición al solvente, al favorecer la desolvatación del ligando y facilitar su unión con la proteína, pero al mismo tiempo conservando la adaptabilidad de poseer los residuos de interacción situados en la superficie de los dominios. Asimismo, el gran cambio conformacional favorecido por la unión del ligando se presta a su vez a su explotación para el acoplamiento de funciones, lo cual provee a estas proteínas de un gran potencial para el diseño de biosensores [39].

Debido a lo anterior y también a su fácil manejabilidad a nivel bioquímico, decidimos utilizar a proteínas de esta superfamilia como modelo para la prueba de rediseño con Rosetta.

LAO, OBP y otras proteínas periplásmicas de unión a aminoácidos

Un tipo de ligando que varias PBPs se han especializado en unir son los aminoácidos. Actualmente se conocen las estructuras cristalográficas de alta resolución ($< 2.0 \text{ \AA}$) de PBPs de *Escherichia coli* o *Salmonella typhimurium* que unen aminoácidos alifáticos como leucina, isoleucina y valina (LivJ – *E. coli*)[40] o polares como lisina, arginina, ornitina (ArgT – *S. typhimurium*)[41, 42]; histidina (HisJ – *S. typhimurium*)[43, 44]; glutamina (GlnH – *E. coli*)[45, 46] y, para una PBP de *Thermus thermophilus*, glutamato o glutamina (*ttZBP*)[47]. La proteína de unión a lisina, arginina y ornitina (LAO) fue una de las primeras en ser cristalizadas y tener sus estructuras resueltas a alta resolución con sus cuatro posibles ligandos de alta afinidad. Es una PBP de tipo II, de 238 aminoácidos y 26 kDa, que como todas las PBPs, consiste de dos dominios tipo α/β . El grande, o lóbulo I, comprende los residuos 1 al 88 y 195 al 238, mientras que el lóbulo II, más pequeño, incluye los residuos 93 al 185[48]. La identidad de secuencia con su homólogo más cercano, la proteína de unión a histidina (HBP), es del 70% y con la proteína de unión a glutamina (QBP), de 30%. No obstante, la similitud estructural es muy alta para ambos casos, como se puede observar en la figura 2a. Su similitud con la proteína de unión a leucina, isoleucina y valina (LIV), una PBP de tipo I, es considerablemente menor, aunque aun puede apreciarse, como se presenta en la figura 2b, una topología general en común.

El sitio de unión a ligando se encuentra relativamente bien conservado en las proteínas cercanas a LAO, aunque con ciertas diferencias que probablemente dan origen a las distintas especificidades y afinidades. Dada la alta similitud estructural, no sorprende que los ligandos se unan

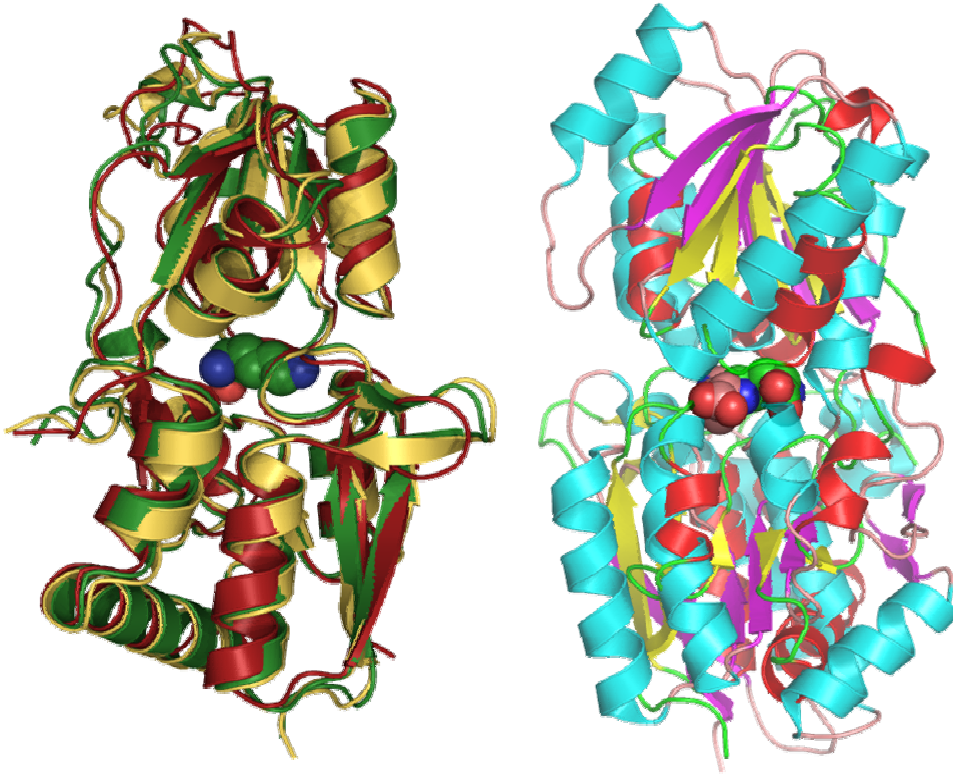


Figura 2: a) Representaciones de caricatura de HBP (amarillo), QBP (rojo) y LAO (verde) con lisina en el sitio de unión. La alineación fue hecha por CE-MC[1]. b) Estructuras de LAO (hélices- α rojas y hojas- β amarillas, lisina unida en verde) y LIV (hélices- α en cian y hojas- β en magenta, valina unida en rosa) alineadas mediante CE[2]. Todas las visualizaciones generadas en PyMol.

en conformacionales orientaciones espaciales, o “poses”, muy similares. En LAO, los grupos carboxilo y amino de sus ligandos son estabilizados por puentes de hidrógeno con los residuos Ser-70, Ser-72, Arg-77, Thr-121 y Asp-161. Las cadenas laterales, por su parte, se mantienen en el sitio de unión por las interacciones de van der Waals de sus carbonos con los residuos hidrofóbicos Tyr-14, Phe-52 y Leu-117, mientras que los distintos nitrógenos polares de los ligandos forman puentes de hidrógeno con los residuos Asp-11 y Ser-69 de la proteína, al igual que con una o dos aguas estructurales. La esquematización bidimensional de las interacciones de unión del ligando se presenta en la figura 3, conjuntamente con los esquemas de las otras proteínas mencionadas por comparación. En el caso de LIV, por ser evolutivamente

Tabla 1 – Afinidades de unión de PBPs de aminoácidos

	K_D	Referencia
LAO	Lisina: 15 nM	Nikaido & Ames, 1992 [41]
	Arginina: 14 nM	
	Ornitina: 29 nM	
	Histidina: 0.5 μ M	
HBP	Histidina: 40 nM	Oh et al., 1994 [43]
	Arginina: 0.7 μ M	
QBP	Glutamina: 0.5 μ M	Sun et al., 1998 [46]
LIV	Leucina: 2.3 μ M	Trakhanov et al., 2005 [40]
	Isoleucina: 0.9 μ M	
	Valina: 4 μ M	

más distante a LAO, incluso de otro tipo, y tener sólo un 10% de identidad en aminoácidos, era de esperarse la completa disimilitud en los residuos involucrados en la unión, a pesar de que existe cierta equivalencia en algunas interacciones respecto a la unión de LAO por sus ligandos. Asimismo, puede apreciarse en la figura 2b la distinta orientación del ligando en LIV respecto a LAO (que es representativa también de HBP y QBP).

En lo que respecta a la afinidad de estas proteínas de unión por sus ligandos respectivos, los valores experimentales determinados (excepto para *ttZBP*, para la que no existen reportes) se esquematizan en la tabla 1. La afinidad es alta, con constantes de disociación desde el rango micromolar hasta el nanomolar. Como puede observarse en esta tabla, LAO y HBP presentan también una afinidad “cruzada”, en la que ambas son capaces de unir los mismos ligandos, pero con afinidades diferentes.

No obstante, no existe una afinidad considerable de LAO para glutamina o los aminoácidos alifáticos de LIV. El único estimado documentado, hasta donde sabemos, de la afinidad de glutamina por el sitio activo de LAO fue realizado como un estudio por competencia y se encontró un 15% de inhibición de Arg a 10 nM con Gln a 1 mM, sugiriendo que existen varios órdenes de magnitud de diferencia en sus afinidades [41].

Con tal diferencia de afinidades y siendo moléculas similares que se unen a sitios de unión relativamente conservados, resulta una prueba atractiva para Rosetta el intentar manipular los aminoácidos del sitio de unión para mejorar la afinidad por glutamina mediante las mutaciones que considere más favorable. Trabajo previo en nuestro grupo ha mostrado que la electrostática puede jugar un papel importante en la unión de ligandos cargados en PBPs, por lo que la glutamina, al ser eléctricamente neutra, tiene la ventaja de no requerir un análisis en este aspecto. Igualmente, la ausencia de moléculas de agua involucradas en la unión de Gln en QBP, a diferencia de LAO, resultó un factor en su elección.

Por otra parte, los aminoácidos alifáticos unidos por LIV, aunque aún estructuralmente parecidos por ser aminoácidos, son químicamente aún más distintos en su cadena lateral que la glutamina (apolar en vez de polar), y en el caso de la valina, también de tamaño y forma considerablemente distintos. Para observar el desempeño de Rosetta en un rediseño probablemente un poco más drástico del sitio, decidimos también probar con valina, el ligando de menor afinidad a LIV, para aumentar su afinidad a LAO.

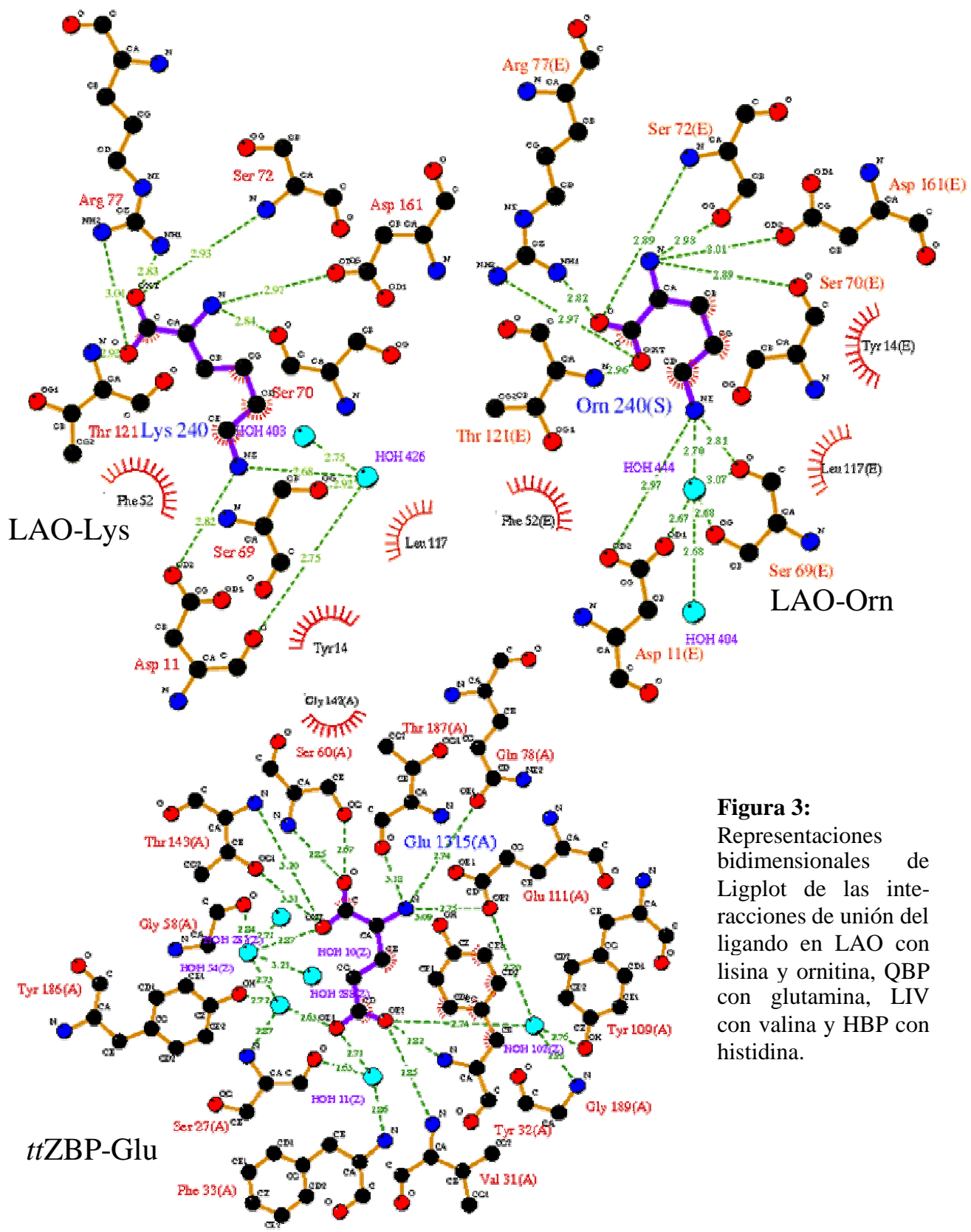
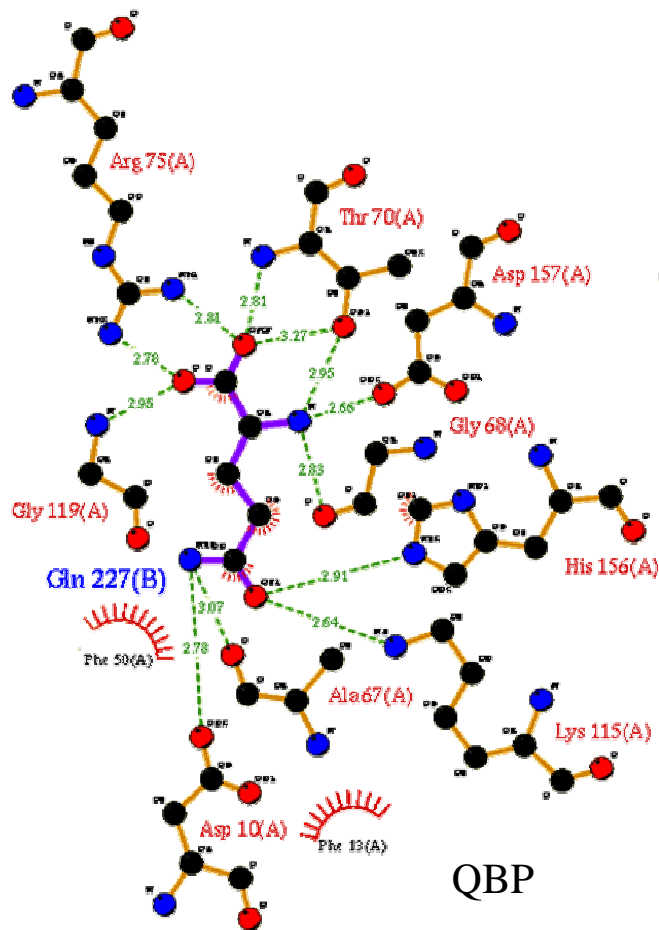
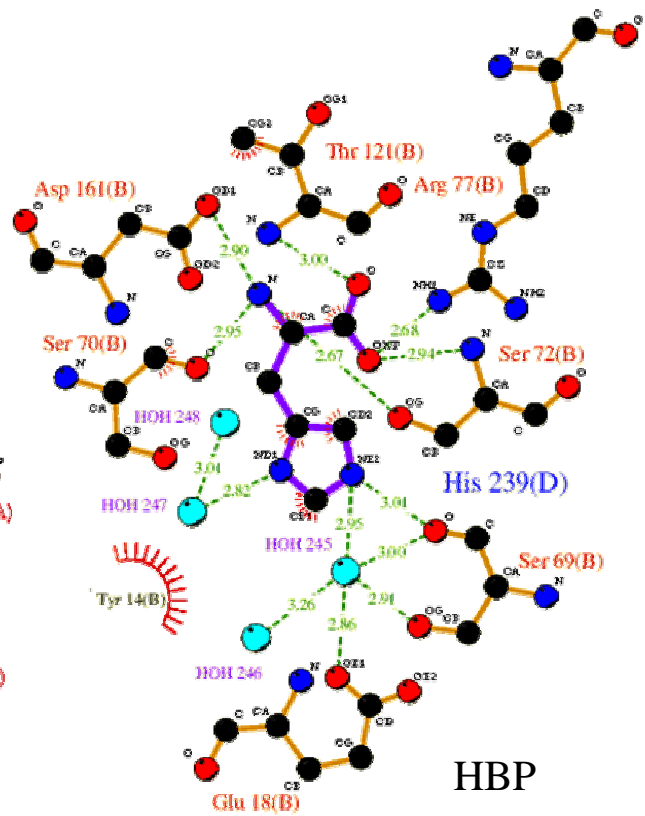







Figura 3: Representaciones bidimensionales de Ligplot de las interacciones de unión del ligando en LAO con lisina y ornitina, QBP con glutamina, LIV con valina y HBP con histidina.



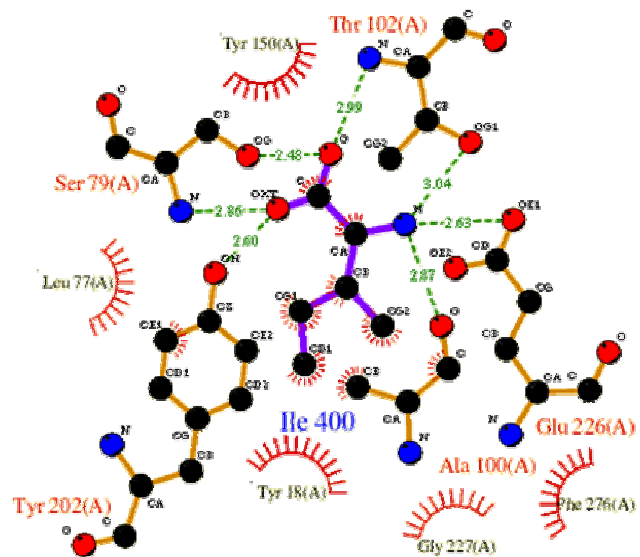
QBP



HBP

-  Enlace del ligando
-  Enlace de la proteína
-  Puente de hidrógeno y su longitud
-  Residuos involucrados en contactos hidrofóbicos
-  Átomos del ligando con contactos hidrofóbicos.

LIV-Val



HIPÓTESIS

A través de la modalidad de ligando del programa Rosetta, es posible rediseñar las interfases de unión de complejos LAO-Gln y LAO-Val generados por *docking* para mejorar la estimación de la energética de unión realizada por un consenso de cuatro funciones de evaluación. Mediante del análisis de las estructuras generadas, pueden identificarse tendencias del programa en sus rediseños favorables para este modelo.

OBJETIVOS

Entre los objetivos de este trabajo están, en primer lugar, generar modelos de las estructuras de los complejos LAO-ligando que sirvan de base para el rediseño con el programa Rosetta. También, establecer una serie de estrategias para evaluar distintos modos en que Rosetta, a través de sus diversas opciones, puede realizar los rediseños. En segundo lugar, establecer una metodología de selección de las mejores estructuras candidatas a través de una clasificación por consenso con otras funciones de evaluación independientes. Finalmente, observar y determinar sesgos o limitaciones en los rediseños de Rosetta para contribuir a la modificación y al desarrollo ulterior del programa

MÉTODOS

Estrategia general del rediseño

La figura 4 presenta un esquema general de la metodología utilizada para este trabajo, que se basa en la utilización de tres programas centrales para crear los rediseños: AutoDock, para generar estructuras del ligando en el sitio de unión de LAO; RosettaDock, para refinar y evaluar inicialmente dichas estructuras y RosettaDesign, para modificar sus residuos y rotámeros y evaluar la nueva energía de interacción proteína-ligando.

En todos los casos, se partió de estructuras generadas por el programa AutoDock, que realiza *docking* de moléculas, es decir, simula la interacción entre ellas para generar el complejo más estable de acuerdo a su modelo. Para el caso de glutamina, estas estructuras sirvieron de preselección para un segundo filtro, realizado por el programa Rosetta en su modalidad dock. En cualquier caso, Rosetta fue utilizado para generar una evaluación inicial de las energías de unión y para minimizar las estructuras para ser usadas en el paso de rediseño con Rosetta en modalidad diseño, que fue realizado con base en diversas estrategias detalladas más adelante. Simultáneamente, las estructuras cristalográficas de otras proteínas en el PDB fueron sometidas al mismo procesamiento por RosettaDock con fines comparativos y se realizaron análisis de LPC para obtener información detallada acerca de las interacciones de los complejos. Posteriormente, los resultados generados por RosettaDesign fueron procesados para utilizarse en una evaluación alternativa con otras funciones independientes, tras lo cual se calculó un consenso entre todas las evaluaciones para generar una clasificación final.

PDB: Estructuras cristalográficas

Todas las estructuras tridimensionales de las proteínas fueron obtenidas del RCSB Protein Data Bank (<http://www.rcsb.org/>). Los códigos PDB para LAO en estado abierto sin ligando y en estado cerrado unida a lisina, arginina, ornitina e histidina son 2LAO, 1LST, 1LAF, 1LAH y 1LAG [42, 48], respectivamente. Para las demás proteínas los códigos de PDB relevantes son 1HPB (HBP en conformación abierta), 1HSL (HBP con histidina) [43], 1GGG (QBP abierta)[45], 1WDN (QBP con glutamina) [46], 1Z15 (LIV superabierta), 1Z16 (LIV con leucina, LIV-L), 1Z17 (LIV con isoleucina, LIV-I) y 1Z18 (LIV con valina, LIV-V). En los archivos PDB donde el ligando venía

designado como ATOM (1LST, 1LAF, 1LAG, 1LAH, 1HSL), éste fue cambiado por la designación de heteroátomo (HETATM), debido muchas de las herramientas usadas requieren que el ligando sea definido así para reconocerlo. En el caso de 1HSL, la unidad asimétrica del cristal es un dímero, por lo se presentan dos monómeros independientes en el mismo archivo, que fue modificado para incluir únicamente uno de ellos. El PDB 1WDN para QBP presentó el problema de que el ligando glutamina carecía de uno de los oxígenos de su grupo carboxilo. El oxígeno faltante pudo ser reconstruido mediante una re-afinación del mapa de densidad electrónica original de la estructura, obtenido del *Electron Density Server* de la Universidad de Uppsala (<http://fsrv1.bmc.uu.se/eds/>) [49], mediante el programa Refmac 5.2 [50]. Adicionalmente, para los programas que consideran la estructura completa como en *docking* y rediseño, las cadenas laterales faltantes en la estructura fueron modeladas en base a la secuencia mediante el servidor de WHAT IF (<http://swift.cmbi.kun.nl/WIWWWI/>) [51].

Como estructura base de la proteína LAO para el rediseño se tomó la estructura PDB 1LST con el ligando lisina removido, por ser la estructura unida de LAO resuelta a mayor resolución.

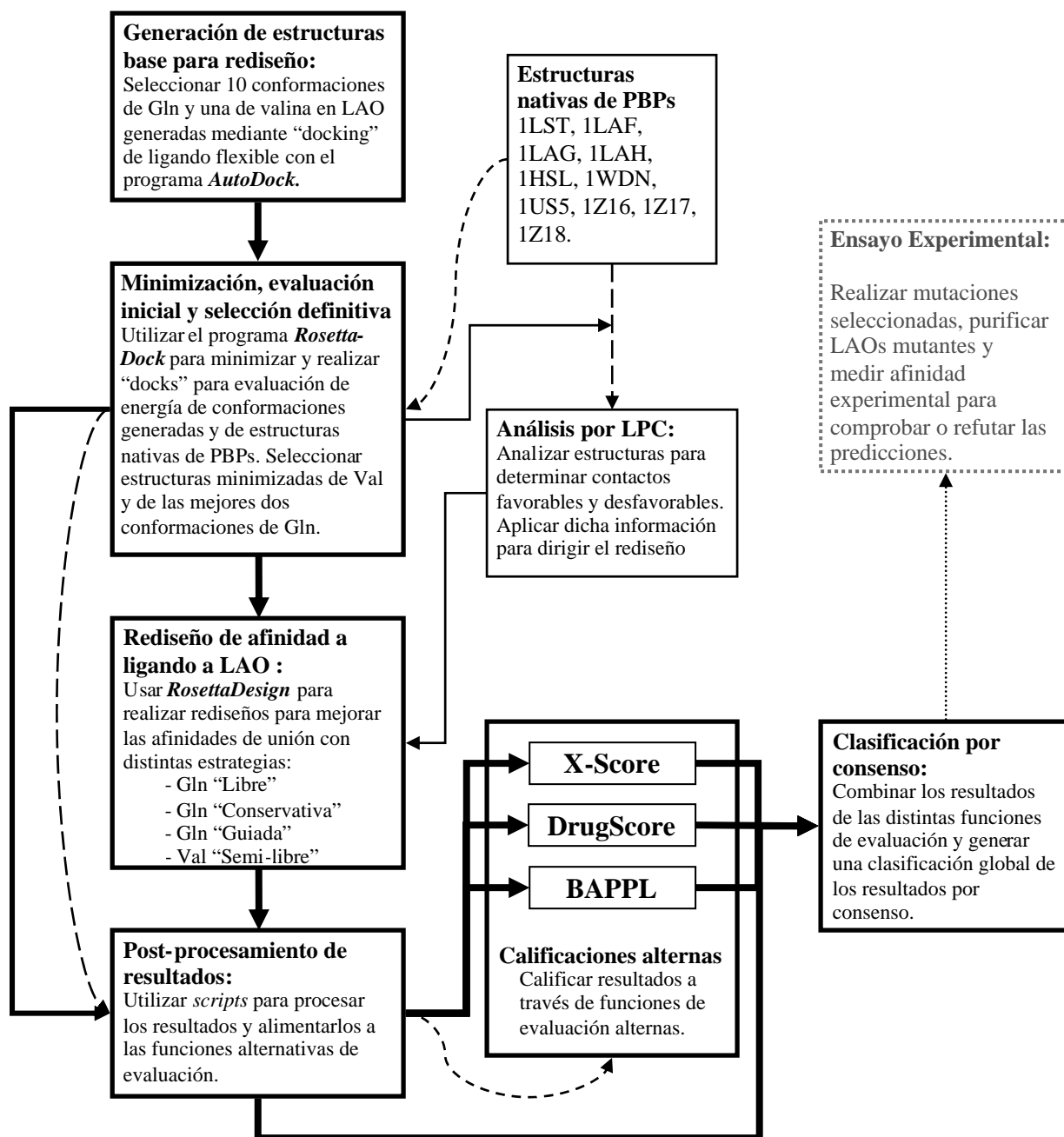


Figura 4: Diagrama de flujo resumiendo la metodología del presente trabajo. La línea gruesa muestra los pasos principales del proceso de rediseño de LAO y la discontinua y curva sigue los pasos que fueron realizados sobre las estructuras nativas de las PBPs usadas en este estudio con fines comparativos. En líneas punteadas y en gris se muestra el trabajo futuro experimental para corroborar los resultados y propuestas.

AutoDock: Generación de estructuras iniciales mediante *docking*

Con el propósito de generar una estructura inicial para el rediseño con Rosetta, decidimos aprovechar la similitud entre los ligandos y la capacidad general de LAO para unir los grupos α -carboxilo y α -amino de sus aminoácidos, por lo que se realizó un proceso de *docking* mediante el programa AutoDock [52] para generar una estructura de LAO con el nuevo ligando insertado en las posiciones energéticamente más favorable de acuerdo al programa. Este paso fue necesario ya que una de las características limitantes de Rosetta en su estado actual es que tanto en su modo de *docking* como en el de rediseño, el ligando siempre es considerado de manera rígida, es decir, sin tomar en cuenta la posible rotación de enlaces internos, por lo que Rosetta tratará de acomodar el ligando a partir de la conformación dada originalmente, independientemente de la flexibilidad y energética del ligando. Esta restricción en modo *docking* de Rosetta con ligandos de alta flexibilidad conformacional traería consigo una importante restricción en su eficiencia, puesto que el programa estaría ciego a una gran parte del conjunto de posibles interacciones entre la proteína y el ligando que se posibilitan mediante la rotación de enlaces en este último. En el modo de rediseño, por otra parte, el efecto repercute en la geometría que debe acomodar mediante mutaciones y prueba de rotámeros, lo cual es parte natural de la tarea del programa cuando los ligandos difieren considerablemente entre sí y, por lo tanto, el efecto de la selección del ligando no debe ser tan crítico para el funcionamiento. No obstante, utilizar una conformación del ligando nuevo que difiera mucho de la óptima entre éste y la proteína nativa daría origen a un estimado inferior de la afinidad real, por lo que sería difícil hacer una comparación válida de este parámetro entre el rediseño y la estructura original. Asimismo, las moléculas de ligando flexibles por sí mismas pueden tener conformaciones preferenciales debido a sus propias interacciones intramoleculares. Si se utilizara alguna conformación desfavorable como base para el rediseño, la estabilización del ligando por parte de los residuos de unión estaría también cubriendo este “déficit” energético no considerado en los cálculos de Rosetta, por lo que probablemente se estaría sobreestimando la energética de unión.

Por otro lado, aunque AutoDock considera la flexibilidad del ligando y es capaz de reproducir frecuentemente la conformación y orientación nativas (que, en combinación, se denominan “pose”) con relativa precisión, su función de evaluación de energía no es lo suficientemente fina como para poder predecir que sus mejores resultados son también las aproximaciones más cercanas a la estructura nativa [31]. Por lo tanto, para no sesgar la conformación del ligando hacia la preferencia de AutoDock, en el caso del rediseño para glutamina se eligieron diez posibles conformaciones a

partir distintas agrupaciones producidas por el programa. Para la generación de nueve de estas conformaciones, se ajustó el parámetro de AutoDock del número máximo de evaluaciones de energía de manera que no se permitió la convergencia completa de los todos resultados hacia una sola agrupación (RMSD menor a 0.5 Å del mejor resultado) cercana al mínimo global de AutoDock, como ocurre si se selecciona un mayor número de evaluaciones de energía por corrida. En cambio, se manipuló dicho valor para generar nueve agrupaciones finales de RMSDs < 0.75 Å con más de una conformación de población y fueron seleccionadas las de menor energía de cada una de ellas para procesamiento posterior. Adicionalmente, se realizó un *docking* de ligando rígido con la glutamina extraída de la estructura de QBP para generar una décima estructura inicial del complejo previo a la minimización y *docking* con Rosetta.

En el caso del rediseño para valina, no existe una similitud significativa entre los residuos del sitio de unión de LAO con LIV y la valina es químicamente más distinta que la glutamina a los ligandos nativos de LAO, por lo que se esperan cambios más drásticos en la interfase a partir del rediseño, por lo que las diferencias en la pose de unión no son tan significativas. Por lo tanto, para seleccionar la conformación de valina para el rediseño se dejó converger a Autodock en una agrupación principal con RMSD < 0.5 y, entre los cinco con mejores energías de unión, se eligió aquél con el parámetro de energía interna más favorable.

Tanto para glutamina como para valina, los *dockings* fueron preparados inicialmente mediante la interfaz gráfica AutoDockTools (ADT). Las estructuras iniciales de los aminoácidos fueron tomadas directamente de los archivos de PDB de las PBPs correspondientes (1WDN para glutamina y 1Z18 para valina) y fueron protonadas por el programa PRODRG [53] (<http://davapc1.bioch.dundee.ac.uk/programs/prodrg/>). El “grid” de *docking* para LAO, el cual sirve de representación de la proteína para la búsqueda, fue generado para una caja de 20 Å × 20 Å × 20 Å alrededor del ligando con un espaciamiento de 0.25 Å. Como algoritmo de búsqueda se seleccionó el algoritmo genético lamarckiano, el cual incluye la posibilidad de realizar búsquedas locales de minimización en los resultados intermedios, lo que proporciona una mayor eficiencia[52]. Salvo en los casos descritos a continuación, se utilizaron los parámetros predeterminados del programa. Para ambos aminoácidos, el número de individuos para el algoritmo genético se estableció en 250, lo que mejora el rendimiento del programa, y el número máximo de generaciones en un millón, lo cual asegura simplemente que la duración de las corridas sea determinada siempre por el número máximo de evaluaciones de energía. Para la glutamina se realizaron 50 corridas de un millón de evaluaciones

de energía para los *dockings* de ligando flexible y 10 corridas para el de ligando rígido, mientras que para valina se realizaron 25 corridas de cinco millones de evaluaciones cada una.

RosettaDock: selección de estructuras base y evaluaciones de energía nativas

Para generar las estructuras definitivas del ligando dentro del sitio de unión de LAO para el rediseño, los resultados seleccionados de AutoDock fueron sometidos a una ronda de diez *dockings* por perturbación de RosettaDock en su modo ligando. El *docking* de perturbación consiste en alterar la posición inicial del ligando mediante un desplazamiento y rotación pequeños y aleatorios antes de llevar a cabo la búsqueda dada por el algoritmo. Rosetta requiere que los ligandos estén en un formato especial, así que éstos fueron transformados previamente a través del programa CMolecule proporcionado por Dr. Jens Meiler de la Universidad de Vanderbilt, E.U.A. Como parte del procedimiento, RosettaDock también realiza un reempacamiento (es decir, reemplaza las cadenas laterales originales de la estructura por aquéllas más favorables de su propia biblioteca de rotámeros, evitando en lo posible choques y espacios vacíos), realiza dos formas de minimización (descenso máximo o *steepest descent* y Monte Carlo) y evalúa las estructuras de entrada de *docking* de igual manera que en el proceso de refinamiento en el *docking*.

Este paso fue realizado para las conformaciones preseleccionadas por AutoDock con el propósito de evaluar su energía con la función de Rosetta e identificar así a las más favorables, así como para obtener un estimado de la afinidad inicial del complejo en términos de Rosetta y construir las estructuras minimizadas requeridas para el rediseño. Se seleccionaron dos conformaciones para el rediseño basadas en sus estimaciones de energía y, en particular la de sus minimizaciones Monte Carlo, que fueron finalmente seleccionadas como estructuras iniciales para el rediseño. Estas estructuras serán denominadas Q1 y Q2. Asimismo, por consistencia, se eligió la minimización Monte Carlo del complejo LAO-Val generado por AutoDock para el rediseño con valina.

Por otra parte, también se realizó el mismo procedimiento sobre las estructuras originales de las PBPs de aminoácidos con sus ligandos, que se llaman “nativas”, con el propósito, por un lado, de obtener valores de referencia para comparar los del rediseño y, por el otro, de observar la correlación del parámetro energético de Rosetta con las afinidades reportadas para esos complejos e identificar problemas potenciales.

Los opciones para Rosetta relevantes para este paso fueron `-dock -ligand -use_input_sc -find_disulf -norepack_disulf -ex1 -ex2 -dock_mcm -dock_pert 0.5 0.5 5.0 -nstruct 10 -`

output_input_pdb. De manera breve, estas opciones indican a Rosetta realizar diez *dockings* del ligando, perturbando aleatoriamente la estructura inicial con desplazamientos 0.5 Å y rotaciones de 5° y minimizando la estructura final por el método Monte Carlo. También se le da la instrucción a Rosetta de escribir en el disco la estructura inicial así como sus procesamientos previos, que incluyen el reempacamiento y las minimizaciones. Una descripción más detallada de las opciones individuales de Rosetta utilizadas en este trabajo se encuentra en el apéndice I.

LPC: Análisis de contactos ligando-proteína

Los diagramas de Ligplot de la figura 3 y la inspección tridimensional de las estructuras permiten conocer y entender visualmente la unión de un ligando a su sitio, pero son de naturaleza inherentemente cualitativa, por lo que no resultan siempre suficientes cuando se requiere información más cuantitativa acerca de la unión del ligando a su sitio. Por otra parte, las funciones de evaluación pueden describir la energética general de un sitio en una escala numérica, pero no suelen desglosar sus términos para explicar cada uno de los contactos del complejo, por lo que es difícil obtener información detallada a partir de ellos. El *software* Ligand-Protein Contacts (LPC) analiza los contactos de cada uno de los átomos del ligando y sus contactos con la proteína, expresando la superficie de éstos y el tipo de interacciones que representan. De esta manera, pueden identificarse los contactos más importantes tanto favorables como desestabilizantes, con los cuales puede generarse un valor aproximado de complementariedad.

A través del servidor del programa LPC (<http://bip.weizmann.ac.il/oca-bin/lpccsu>) se analizaron las estructuras nativas de LAO y QBP, al igual que las de LAO con la glutamina y la valina acopladas que serían usadas como base del rediseño. Su uso es simple, aunque fue necesario utilizar las estructuras modificadas de LAO y QBP con el ligando como heteroátomo. Su utilización no está, sin embargo, libre de errores o “bugs”, por lo que es posible encontrarse con contactos considerados como desestabilizantes cuando no lo son. Igualmente, tiene la limitante de no ser capaz de considerar las moléculas de agua estructurales que intervienen en la unión, lo cual puede llevar a subestimar la superficie de contacto favorable.

RosettaDesign: Rediseño de afinidades

Aunque la modalidad de diseño para ligandos aún se encuentra en fase preliminar, cuenta ya con diversas opciones, algunas nuevas y otras adaptadas, para realizar su función. Una de ellas, la capacidad de especificar a través de un archivo, referido como *resfile*, exactamente qué aminoácidos de la proteína van a ser sujetos a rediseño y cómo. Para cada aminoácido se le puede especificar si va a ser dejado intacto, reempacado o cambiado, y en tal caso, si puede ser cambiado por cualquier aminoácido, un tipo de ellos o por uno o varios en particular. Esta característica proporciona una flexibilidad y variedad inmensa de posibilidades de diseño, desde el completo rediseño de la secuencia a discreción del programa, hasta la selección individual de cada mutación.

Si el método de evaluación fuera altamente confiable, al realizar un rediseño libre de la secuencia completa se contaría, en principio, con una mayor probabilidad de encontrar una combinación óptima global para la unión, debido a un mayor espacio de búsqueda. Los cambios en regiones lejanas no resultarían beneficiosos y serían descartados. No obstante, esta estrategia implicaría tanto un requerimiento muy alto, tanto en la memoria necesaria para considerar el número de rotámeros posibles involucrados, como en el tiempo de cómputo para muestrear el gran número de posibilidades, sólo para descartar una inmensa mayoría de ellas. Por otra parte, si la función de evaluación tuviera tendencias o desviaciones considerables, dejar a su criterio la secuencia completa o partes grandes de ella podría también resultar en cambios inútiles, producto del ruido o, peor aún, alguna desestabilización no prevista por el algoritmo. Resulta, por lo tanto, más práctico y cauto restringir los cambios en regiones que, o bien no tienen por qué influir en la afinidad del ligando, o si influyesen, Rosetta podría detectarlas en forma confiable. No hay que olvidar, sin embargo, la posibilidad de que un cambio favorable genere al mismo tiempo un choque estérico que podría ser eliminado rotando una cadena o, más drásticamente, cambiándola. Tomando en cuenta ambas consideraciones y procurando mantener la mayor simplicidad que sea compatible con una automatización, una parte de los rediseños se realizaron considerando dos esferas de cambio: la primera, de 5 Å que incluye a los residuos en contacto con el ligando y una segunda, de 8 Å, donde a los residuos nativos se les permite reacomodarse. Dichas mediciones de distancias, realizadas mediante el programa de visualización Pymol (<http://pymol.sourceforge.net/>), fueron hechas sobre las estructuras previas a la minimización, ya que ésta, en respuesta a los contactos con el nuevo ligando, genera la retracción de algunas cadenas laterales relevantes fuera de las esferas de cambio, lo que puede restringir importantemente las posibilidades de rediseño.

Aun con la limitación anterior, al combinar las demás opciones de rediseño, se cuenta con un gran número de variaciones factibles. Para abarcar un intervalo amplio de posibilidades de rediseño, se decidió adoptar, para el caso de la glutamina, tres estrategias distintas. En la primera, que denominamos “libre” se dejó a Rosetta rediseñar a su criterio simplemente con el espacio de búsqueda restringido como se describió y sin ningún sesgo impuesto en sus opciones de comando. En la segunda, referida como “conservativa”, se permitió, igualmente, cualquier cambio en los residuos permitidos, pero se aplicó un sesgo a favor de los aminoácidos originales, para limitar las mutaciones cuyo cambio en energía respecto al residuo nativo no sea superior a cierto valor, filtrando así cambios de impacto marginal o dudoso. Finalmente, en una tercera estrategia que se llamó “guiada”, a diferencia de las anteriores, el archivo de mutaciones permitidas fue elaborado específicamente, tomando en consideración información acerca del sitio de unión asimilada a partir de los resultados de LPC y de la inspección visual de las estructuras, así como de la conservación de residuos en otras PBP's homólogas, particularmente QBP. De esta manera, se propusieron cambios adicionales considerados como plausibles y se protegieron algunos aminoácidos conservados, amortiguando posibles desbalances o sesgos de la evaluación de Rosetta.

Para el caso de la valina, para la que se esperaban cambios más drásticos para poder acomodarla favorablemente, se adoptó una única estrategia de rediseño “semi-libre”, con la aplicación de algunos ajustes como un ligero sesgo hacia los residuos nativos y hacia los no polares. Esto último a consecuencia de la tendencia de Rosetta de aparentemente subvalorar las interacciones hidrofóbicas respecto a contactos polares como puentes de hidrógeno. Asimismo, se extendió la esfera de rediseño a todos los residuos con átomos de su cadena lateral a una distancia fija de 8 Å, en vez de solamente a los aminoácidos de contacto, para aumentar las posibilidades de acomodar el nuevo ligando.

Las opciones de comando relevantes de Rosetta para las corridas de rediseño, con las diferencias notables entre estrategias marcadas en negrita, fueron:

- Gln “Libre”: *-design -ligand -fixbb -use_input_sc -find_disulf -norepack_disulf -ex1 -ex2 -design_dock -dock_pert 0.5 0.5 -ndruns 50 -resfile LaoGln-redesign.resfile -ligand_filter_score -12 -ligand_filter_count 100*
- Gln “Conservativa”: *-design -ligand -fixbb -use_input_sc -find_disulf -norepack_disulf -ex1 -ex2 -design_dock -dock_pert 0.5 0.5 5 -resfile LaoGln-redesign.resfile -ndruns 25(×4) -ligand_filter_score -20 -ligand_filter_count 1000 -**favor_native_residue -1.5***

- Gln “Guiada”: *-design -ligand -fixbb -use_input_sc -find_disulf -norepack_disulf -ex1 -ex2 -design_dock -dock_pert 0.5 0.5 5 -ndruns 25(×4) -ligand_filter_score -18 -ligand_filter_count 1000 -resfile LaoGln-ratdesign.resfile -favor_native_residue -1.0*
- Val “Semi-libre”: *-design -ligand -fixbb -use_input_sc -find_disulf -norepack_disulf -ex1 -ex2 -design_dock -dock_pert 0.5 0.5 5 -s LaoVal.pdb -ndruns 50 -resfile LaoVal-design.resfile -ligand_filter_score -15 -ligand_filter_count 200 resfile -favor_native_residue -0.5 -favor_nonpolar -0.5*

La explicación de las opciones individuales se presenta nuevamente en el apéndice I. Asimismo, los archivos *resfile* utilizados para el rediseño, comentados donde resulte relevante, pueden consultarse en el apéndice II

Funciones de evaluación independientes y clasificación por consenso

Dado el estado preliminar y sin validaciones publicadas de la función de Rosetta en su modo ligando, resultaba sensato realizar una corroboración y comparación independiente de sus estimaciones de energía y, por tanto, de su afinidad. Aunque el desarrollo de las funciones de evaluación aún no permite una predicción precisa de afinidades reales de complejos [33], algunas de ellas han logrado ya mostrar niveles útiles de correlación. Mediante la comparación con las estimaciones de las estructuras nativas y combinando las distintas evaluaciones entre sí a través de una calificación por consenso se espera, en principio, obtener predicciones lo más aproximadas posible respecto al cambio relativo de afinidad originado por el rediseño.

Para realizar esta evaluación independiente, se utilizaron tres funciones de evaluación de acceso académico gratuito publicadas recientemente, las cuales han demostrado un rendimiento respetable en pruebas de validación o comparación: DrugScore^{ONLINE} V0.9 [54] (<http://pc1664.pharmazie.uni-marburg.de/drugscore/introduction.php>), X-Score 1.2 [55] (<http://sw16.im.med.umich.edu/software/xtool/>) y BAPPL [56] (<http://www.scfbio-iitd.res.in/software/drugdesign/bappl.jsp>)

DrugScore es una función desarrollada en la Universidad de Marburg, Alemania por el grupo de Gerhard Klebe. Es una función basada en conocimiento, lo que significa que se basa en el análisis estadísticos de un gran número de estructuras para establecer reglas acerca de las interacciones preferidas, mismas que se transforman en “pseudo-potenciales” que se pueden aplicar para evaluar complejos generados por computadora [29]. DrugScore reporta un coeficiente de correlación de Spearman de 0.62 entre la puntuación del programa y las afinidades experimentales de su conjunto

de prueba. El programa se encuentra libremente en la red como el servidor web DrugScore^{ONLINE}. Las opciones usadas fueron utilizar únicamente los potenciales de pares para la evaluación (PAIR) y usar los derivados del Cambridge Structural Database (CSD).

X-Score es una función desarrollada por el grupo de Shaomeng Wang de la Universidad de Michigan, E.U.A. Es de tipo empírico, por lo que funciona bajo la suposición de que la energía de unión puede separarse en distintos términos responsables por un tipo específico de interacciones, como van der Waals o puentes de hidrógeno, y que luego son sumados de forma ponderada, con los pesos relativos establecidos a partir de la regresión de conjuntos de estructuras de afinidad experimental conocida. Técnicamente podría considerársele también como método de consenso, ya que su resultado es el promedio de tres estimaciones individuales, aunque calibradas con las mismas estructuras. En pruebas comparativas realizadas posteriormente por los autores [31, 33], X-Score ha obtenido coeficientes de correlación en la prueba de $R_S = 0.660$ y $R_S = 0.603$.

BAPPL (por *Binding Affinity Prediction of Protein-Ligand*), por su parte, es una función que se clasifica a sí misma como “de campo de fuerza/empírica” y fue desarrollada en el laboratorio de B. Jayaram en el Instituto Hindú de Tecnología, India. La función muestra un coeficiente de correlación de $r = 0.92$ en su estudio de validación[56]. Para la utilización de BAPPL en este trabajo, las estructuras fueron evaluadas por el Método 2.

Todas las funciones tienen requerimientos especiales en cuanto al formato de entrada que aceptan, por lo que todos los resultados de Rosetta debían de ser procesados y adaptados de acuerdo a ellos. Por otra parte, dos de ellas (DrugScore, BAPPL) están basadas en servidores web y no son viables para el procesamiento manual de muchos resultados. Aunque DrugScore es capaz de realizar automáticamente múltiples evaluaciones, este modo está diseñado para filtrar diversos ligandos para una misma estructura, por lo que no es de utilidad en el caso inverso. Debido a lo anterior, se programaron scripts en Perl tanto para la traducción de formatos como para el envío automático de resultados a los servidores web. En algunos de estos scripts se recurre al uso de algunos programas externos para la realización de algunas funciones. Para la conversión básica de formato PDB a MOL2, requerida tanto por X-Score como DrugScore, se utilizó el programa OpenBabel 2.0 (<http://openbabel.sourceforge.net/>). Algunas otras modificaciones a los archivos PDB como renombrar cadenas y reenumerar los átomos fueron realizadas por la herramienta PDBSUM del paquete CCP4 [57] (<http://www.ccp4.ac.uk/main.html>).

Para la evaluación consenso de los resultados se adoptaron dos estrategias de comparación: por puntuación promedio y por clasificación promedio. La primera se basa en el promedio de las

evaluaciones normalizadas, o puntuaciones, para cada función y la segunda se calcula como el promedio de las clasificaciones o posiciones de determinada estructura en un ordenamiento de acuerdo a sus valores para cada función.

Para el cálculo de la puntuación por consenso, dadas las escalas distintas, se realizó una normalización de los resultados de cada función como se describe en [58]:

$$x_{norm} = \frac{x - x_{min}}{x_{max} - x_{min}} \quad Ec. 3$$

donde x_{norm} es la puntuación normalizada, x_{min} es el valor de la peor evaluación en el conjunto dado de resultados y x_{max} es la mejor evaluación. Esto da origen a un intervalo de 0 a 1, donde 1 representa la mejor evaluación de energía y 0, la peor. No obstante, debido a la presencia de algunos puntos con evaluaciones particularmente bajas y con el objetivo de evitar la “compresión” de la puntuación y consecuente sesgo al fijar el 0 en un valor demasiado bajo, se estableció un límite mínimo absoluto en el valor correspondiente al promedio menos tres desviaciones estándar.

RESULTADOS Y DISCUSIÓN

A continuación se presentan los resultados obtenidos para los distintos pasos del proceso de rediseño y su discusión. En primer lugar, se muestran los de la generación de los complejos de LAO con sus nuevos ligandos mediante el programa AutoDock y posteriormente los de las minimizaciones, evaluaciones y selección final de modelos a través de la modalidad *dock* de Rosetta y el análisis de interacciones con LPC. Finalmente, se presentan y discuten los resultados de los rediseños mediante RosettaDesign con diferentes estrategias y sus evaluaciones con las funciones independientes así como su clasificación final por consenso.

Generación de Estructuras

En la figura 5 se presentan las diez conformaciones de glutamina en el sitio de unión de LAO generadas por AutoDock con la idea de contar con una muestra representativa de las posibles poses de unión del ligando en el sitio de LAO, así como la estructura de valina generada por el programa, seleccionada directamente para el rediseño. Igualmente, se presentan las agrupaciones de AutoDock de las cuales fueron seleccionados los nueve modelos flexibles de glutamina.

Las estructuras de glutamina generadas por AutoDock, aun en distintas agrupaciones, muestran una similitud considerable con la conformación del aminoácido en la estructura cristalográfica de QBP, lo cual es consistente con el conocimiento que se tiene acerca de la alta similitud estructural del sitio activo entre estas proteínas, tanto a nivel del volumen como los contactos que permiten el acomodo del ligando. Lo poco que LAO une a glutamina[41], probablemente lo hace de manera similar a QBP.

La tabla 2A, presenta las evaluaciones de energía de RosettaDock, referidas como “*lig_sum*”, para las distintas conformaciones, sus minimizaciones y diez *dockings* independientes. El valor de *lig_sum* representa la suma de distintos términos correspondientes a las diversas interacciones que Rosetta evalúa como relevantes para la afinidad por el ligando. Éstos son *lig_rep*, *lig_atr*, *lig_sol*, *lig_cou*, *lig_hb*, *lig_vir*, *lig_h2o* y *lig_sas*. Ya que no existe aún una publicación acerca del modo ligando de Rosetta, el origen de cada uno de estos términos no está disponible, aunque es evidente que varios de ellos son análogos a los presentes en la ecuación 2. Considerando los valores de la tabla conjuntamente con lo obtenido por AutoDock, se puede observar que, aunque AutoDock es capaz de generar poses casi “nativas” para glutamina, a niveles finos su estimación de energía de unión no concuerda exactamente con la de Rosetta. Incluso parece a veces estar en conflicto directo al evaluar con energías de unión altamente desfavorables a algunas de las conformaciones generadas por Rosetta. No obstante, el reempacamiento con Rosetta mejora considerablemente los valores de *lig_sum* y tras las minimizaciones, éstos se encuentran ya en rangos razonables. Este efecto probablemente pueda deberse a la mayor “resolución” de la evaluación de átomos completos por Rosetta respecto a la obtenida por AutoDock a través del “grid” o malla con que simplifica la superficie de la proteína. En la tabla 2B se presentan los resultados para un segundo valor con el que Rosetta puede analizar las estructuras, denominado *bk_tot*, el cual consiste en la suma de los distintos términos de interacciones entre residuos de la proteína, correspondientes a la ecuación 2, y representa una medida de la estabilidad de la proteína. Aunque no proporciona información directa acerca de la interacción con el ligando y las pequeñas variaciones probablemente no resultan significativas, es notable el hecho de aquellas conformaciones que presentaban *lig_sum* muy desfavorables inicialmente, también demuestran *bk_tots* considerablemente inferiores, lo cual apoya la noción de que en estos casos AutoDock generó choques estéricos que no fue capaz de detectar en su representación de “grid”, pero que Rosetta sí, debido a la representación de átomos completos que utiliza en esta situación.

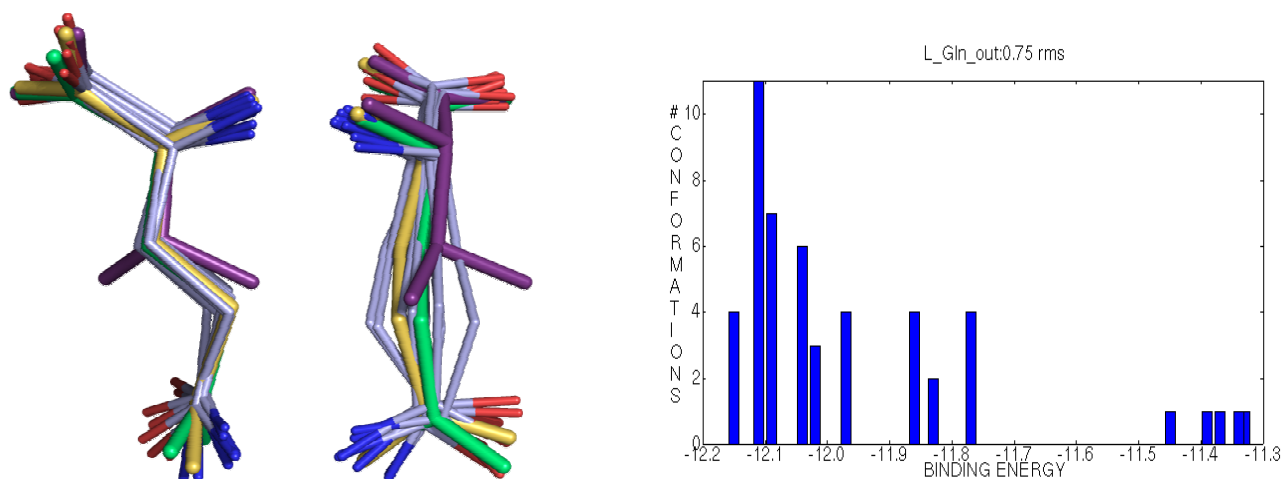


Figura 5: a) Estructuras de las conformaciones de Autodock preseleccionadas para Rosetta (desde dos ángulos). En verde: Conformación nativa de glutamina en QBP acoplada mediante Autodock a LAO; amarillo: Conformación más favorable según RosettaDock (mínimo de la agrupación 6 de izq. a der.); morado: conformación de valina utilizada.

b) Agrupaciones de resultados en Autodock. Tomando el mejor valor como referencia inicial, las conformaciones con un RMSD < 0.75 Å respecto a ésta se agrupan en una sola barra, centrada en el valor de energía de unión de la primera. Las demás barras se construyen sucesivamente tomando la siguiente mejor conformación que aún no pertenezca a un grupo.

Tabla 2A – Cálculo de RosettaDock de energías de unión (“lig_sum”) para selección de conformaciones

	Procedimiento o No. de Docking														
	Auto ¹	Rep ²	Min ³	Mcm ⁴	1	2	3	4	5	6	7	8	9	10	Prom
QBP	-11.79	-13.19	-18.39	-18.41	-17.27	-18.80	-13.78	-15.96	-17.36	-17.49	-17.87	-18.17	-17.26	-17.43	-17.14
Conf1	12.81	-2.37	-11.37	-15.76	-15.77	-14.34	-14.72	-15.08	-14.60	-14.65	-13.88	-14.54	-12.66	-10.38	-14.06
Conf2	34.78	-8.00	-14.87	-14.83	-14.12	-14.64	-15.47	-13.92	-13.37	-15.88	-14.88	-6.39	-13.82	-14.87	-13.74
Conf3	38.13	-6.10	-13.88	-16.32	-14.12	-14.64	-15.47	-13.92	-13.37	-15.88	-14.88	-6.39	-13.82	-14.87	-13.74
Conf4	20.22	-5.74	-11.04	-12.04	-14.56	-4.97	-17.22	-17.26	-17.03	-16.83	-8.12	-17.27	-13.73	-16.81	-14.38
Conf5	-8.84	-10.99	-13.78	-15.49	-16.71	-15.18	-13.96	-16.33	-16.67	-15.84	-13.51	-15.09	-15.43	-15.59	-15.43
Conf6	-9.60	-11.94	-18.32	-19.73	-15.97	-14.94	-19.66	-18.41	-17.50	-17.73	-19.45	-17.76	-16.46	-18.61	-17.65
Conf7	29.64	-9.14	-13.26	-16.19	-16.02	-14.38	-14.33	-13.98	-12.50	-15.22	-13.67	-13.78	-14.29	-14.5	-14.27
Conf8	21.48	5.59	-11.33	-12.64	-15.88	-12.09	-14.68	-14.86	-14.38	-14.99	-15.18	-14.76	-14.64	-15.88	-14.73
Conf9	5.86	-5.22	-11.52	-16.79	-18.81	-15.22	-16.88	-16.57	-8.17	-16.42	-14.66	-16.65	-15.77	-13.82	-15.30
Val	-9.54	-10.63	-14.43	-14.52	-15.35	-10.5	-12.97	-15.35	-14.07	-15.53	-14.9	-14.26	-15.23	-13.89	-14.21

Tabla 2B – Suma de parámetros de la función de energía global de Rosetta Dock (“bk_tot”)

	Procedimiento o No. de Docking														
	Auto ¹	Rep ²	Min ³	Mcm ⁴	1	2	3	4	5	6	7	8	9	10	Prom
QBP	-756.5	-763.4	-768.6	-768.7	-767.0	-767.1	-750.3	-765.0	-769.7	-769.3	-770.7	-767.9	-770.5	-772.0	-767.0
Conf1	-733.3	-751.8	-760.8	-765.4	-765.6	-767.9	-763.3	-767.6	-770.3	-766.4	-765.8	-766.5	-758.0	-745.1	-763.7
Conf2	-711.3	-757.6	-764.4	-765.8	-748.0	-764.7	-764.8	-766.3	-768.0	-749.2	-769.7	-766.4	-766.7	-765.9	-763.0
Conf3	-708.0	-755.3	-763.0	-768.0	-767.4	-763.3	-765.1	-766.4	-763.2	-767.3	-763.9	-740.2	-763.6	-768.0	-762.8
Conf4	-725.9	-737.2	-742.5	-748.9	-748.4	-742.6	-769.4	-767.8	-768.1	-768.8	-739.3	-769.6	-764.9	-768.6	-760.7
Conf5	-754.9	-761.1	-763.8	-765.5	-769.8	-767.1	-766.2	-766.3	-769.6	-766.1	-767.6	-766.9	-769.7	-765.0	-767.4
Conf6	-755.7	-761.7	-768.0	-768.0	-765.4	-746.2	-769.3	-771.6	-768.1	-770.5	-769.3	-767.2	-767.0	-769.8	-766.4
Conf7	-716.5	-759.0	-763.1	-766.5	-764.7	-765.3	-766.3	-765.3	-764.3	-768.5	-766.0	-766.4	-745.9	-763.4	-763.6
Conf8	-723.2	-744.0	-760.9	-763.1	-765.5	-760.1	-765.4	-763.7	-763.2	-766.5	-766.7	-765.3	-766.2	-752.8	-763.5
Conf9	-740.2	-755.7	-762.0	-768.6	-752.8	-765.1	-767.8	-767.7	-758.1	-768.5	-769.2	-768.7	-767.8	-767.3	-765.3
Val	-755.6	-760.4	-764.2	-764.5	-767.0	-749.1	-766.2	-770.3	-763.5	-747.3	-767.3	-769.2	-766.9	-767.8	-763.5

¹Auto: Estructura salida directamente de AutoDock. ²Rep: Estructura “reempacada” (ver texto). ³Min: Minimización de descenso máximo (steepest descent). ⁴Mcm: Minimización Monte Carlo.

Evaluación de Rosetta de estructuras nativas de PBPs

En la tabla 3 se presentan los valores de *lig_sum* para las minimizaciones y *dockings* de prueba con RosettaDock para las estructuras de las PBPs nativas. Los resultados obtenidos para las diversas PBPs revelan ciertas tendencias en la evaluación de Rosetta, que al tomarse en conjunto con las afinidades conocidas de los complejos (también incluidos en la tabla por conveniencia) presentan ciertas consideraciones para el análisis de los resultados del rediseño. Igualmente se incluyen los valores de *lig_sum* para las estructuras iniciales de LAO para el rediseño tanto con glutamina como con valina.

LAO con sus ligandos nativos lisina y ornitina presenta *lig_sums* en sus estructuras minimizadas alrededor de -18, mostrando valores visiblemente mayores con arginina (así como con la energía libre de Gibbs ΔG , entre más negativos sean los valores, mejor es la afinidad estimada). Con histidina, su ligando de menor afinidad, muestra valores ligeramente inferiores, de manera consistente con lo esperado. HBP, de mayor afinidad nativa por histidina, muestra por su parte una afinidad estimada superior a la de LAO en sus estructuras minimizadas, en concordancia con las

Tabla 3 – Valores de energía de unión a ligando “*lig_sum*” para estructuras nativas de PBPs

	Procedimiento o No. de <i>Docking</i>						
	Original	Repack ¹	Min ²	Mcm ³	Promedio 10 docks	Desv. Est. Docks	Afinidad Exp. (K_D)
LAO-K	-11.08	-17.26	-18.22	-19.47	-17.76	0.66	1.5×10^{-8} M
LAO-R	-19.76	-20.13	-20.21	-20.19	-20.87	0.52	1.4×10^{-8} M
LAO-O	-17.77	-17.91	-18.68	-18.49	-17.70	0.96	2.9×10^{-8} M
LAO-H	-13.64	-7.24	-14.53	-17.21	-14.18	2.68	5×10^{-7} M
HBP-H	-11.08	-17.26	-18.22	-19.47	-17.76	2.85	4×10^{-8} M
QBP-Q	-25.27	-24.49	-26.65	-26.32	-25.85	0.71	5×10^{-7} M
ttZBP-E	-14.18	-14.41	-15.81	-15.33	-16.05	1.96	N/A
LIV-L	-14.04	-14.24	-14.33	-15.71	-14.61	1.23	2.3×10^{-6} M
LIV-I	-7.29	-7.61	-10.77	-10.77	-12.04	1.42	9×10^{-7} M
LIV-V	-13.79	-16.06	-17.10	-17.13	-16.24	1.29	4×10^{-6} M
LAO-V	-9.54	-10.63	-14.43	-14.52	-14.21	1.54	¿Insignificante?
LAO-Q2	-9.60	-11.94	-18.32	-19.73	-17.65	1.51	Muy pobre [41]

¹Repack: Estructura “reempacada” (ver texto). ²Min: Minimización de descenso máximo (*steepest descent*). ³Mcm: Minimización Monte Carlo.

expectativas, aunque en su estructura cristalográfica original presenta un estimado considerablemente más pobre, incluso inferior a su contraparte. Podría especularse en torno a esta discrepancia que podría haber sido influida por el hecho de que la cristalización de HBP fue realizada a un pH de 9.5, lejos de 5.5 [43], que fue el usado para su medición de afinidad, en contraste a LAO, que fue cristalizada a pH 6.5 [48] y sus afinidades medidas a 7.0 [41]. Por otro lado, QBP, con una K_D de 0.5 μM , presenta *lig_sums* muy superiores, alrededor de -26, en clara discordancia con los valores de afinidad experimentales, si se toman en relación con los de LAO. Respecto a esto, resulta esencial mencionar una diferencia importante a la que podría atribuirse tal diferencia, y que es la participación de aguas estructurales en la unión de ligando para LAO, las cuales son omitidas en el análisis de Rosetta. Dichas aguas median algunos puentes de hidrógeno en la interacción con los ligandos, como puede observarse en los diagramas de Ligplot en la figura 3, por lo que su omisión probablemente llevar a una subestimación de la afinidad. QBP, por su parte, aunque también presenta algunas aguas en la estructura de su sitio de unión, éstas no participan en la unión del ligando, que es estabilizado directamente por los residuos de la proteína. La comparación con *ttZBP* resultaría interesante, ya que se presume que ésta tiene un ácido glutámico y/o glutamina también, pero en un sitio de unión distinto en el cual sí intervienen aguas. Sin embargo, la carencia de una medición experimental de afinidad excluye la posibilidad de asociar los estimados de Rosetta, más similares en magnitud a los de LAO que a QBP, a la presencia de aguas y no simplemente a una afinidad menor por el ligando. Cabe añadir que Rosetta recientemente implementó la posibilidad de utilizar un modelo de rotámeros que es capaz de tomar en cuenta puentes de hidrógeno mediados por aguas en interfases proteína-proteína [59]. Aunque la opción es elegible en modo ligando en la versión con la que contamos, esta opción no se considera adecuada todavía (Jens Meiler, comunicación personal), además de que eleva los costos computacionales por encima de nuestra infraestructura actual.

En cuanto a la interacción de LIV con aminoácidos alifáticos, los resultados de Rosetta posicionan sus energías a niveles ligeramente menores, en conformidad con las afinidades un par de órdenes de magnitud inferiores a las de LAO/HBP, aunque con las afinidades relativas entre ligandos invertidas y acentuadas. No obstante, los resultados más discrepantes son los de los *dockings* de LAO con glutamina y valina para el rediseño. A pesar de que LAO no tiene afinidad significativa ni para glutamina ni, presumiblemente, para valina; después de las minimizaciones, Rosetta le atribuye a estos complejos “artificiales” *lig_sums* comparables e incluso superiores a las de LAO y LIV por sus ligandos originales. Es quizá notable que este fenómeno no ocurre para las estructuras originales

generadas con AutoDock, que cuentan con valores considerablemente inferiores, pero es fácil argumentar, como se hizo anteriormente para la selección de rotámeros, que las distintas representaciones de la proteína que emplean ambos programas pueden generar discordancias en las superficies, lo cual se manifiesta en choques estéricos y, por tanto, en energías desfavorables de interacción, las cuales pueden ser compensadas por reempacamiento o minimizaciones. De cualquier manera, el hecho de que Rosetta atribuya a los *dockings* previos al rediseño energías de interacción comparables a las de los ligandos nativos sugiere una desviación o limitante en la función de evaluación. Al respecto resulta relevante el hecho de que en un sistema de bisagras y envolvimiento de ligando como las PBP, las interacciones directas entre ligando y proteína no son las únicas que pueden influir en la afinidad, sino también aquéllas entre los dominios de la proteína o las involucradas en el movimiento de bisagras. Un ejemplo de esto se da en un estudio de rediseño de afinidad en el que, sin modificar los aminoácidos de unión, se logró mejorar considerablemente las afinidades mediante la explotación del acoplamiento conformacional, es decir, modificando el estado de equilibrio entre los estados abierto y cerrado de la proteína mediante la modificación selectiva de sitios fuera del sitio de unión que varían en los distintos estados, estabilizando así uno u otro[60]. Se ha observado que la unión del ligando a QBP por sí misma tiende a estabilizar el estado cerrado [61-63] (lo que presumiblemente es cierto en otras PBP también), por lo que es concebible que ciertas interacciones particulares entre el ligando y la proteína tiendan a favorecer este desplazamiento en el equilibrio. Dado que las construcciones por AutoDock fueron “forzadas” sobre el estado cerrado y evaluadas de esta manera por todos los programas, la consideración de posibles efectos conformacionales no se toma en cuenta. Así, la afinidad, que en realidad es resultado de más de un factor, es evaluada por Rosetta con base en uno solo de ellos (las interacciones directas), causando discrepancias sin que éstas necesariamente impliquen un problema con los cálculos o parámetros de la función.

A primera vista, los resultados de la tabla parecen manifestar cierta dificultad por parte de la función de Rosetta para mostrar una correlación inequívoca con las afinidades experimentales conocidas, lo cual no es del todo inesperado dados los estudios que han demostrado que la capacidad de las funciones de evaluación para la predicción confiable de afinidades es en varios casos apenas modesta. [32-34]. No obstante, en contexto con el conocimiento de trasfondo de las proteínas involucradas y su estructura, estos resultados pueden entenderse de manera que no minan la suposición de que Rosetta es capaz de proporcionar información útil mediante su función de energía. Aun así, la ambigüedad de los valores pone de manifiesto la necesidad de contar con criterios

alternos para el discernimiento de interacciones favorables de unión, lo cual fue tratado a través de las funciones de evaluación alternativas, cuyos resultados se exponen más adelante.

LPC: análisis de contactos

Las distintas partes de la tabla 4 presentan una síntesis de los resultados obtenidos por LPC respecto al análisis de contactos en diferentes PBPs así como en las construcciones tomadas como base para el rediseño. Las tablas muestran los contactos determinados por el programa entre cada átomo del ligando y los de la proteína, así como su distancia y su superficie. Asimismo, se presentan las clasificaciones que se le asignan a cada átomo en base a su naturaleza química, lo cual determina su carácter favorable o desestabilizante. El parámetro de superficie de contacto ayuda a identificar los principales contactos, aunque la información es probablemente demasiado cruda para definir con confianza la importancia energética de sus contribuciones; no obstante, los valores obtenidos pueden ayudar a establecer una lista de residuos de la proteína conocidos por su interacción con el ligando y, por tanto, posibles blancos del rediseño. Asimismo, la determinación de su naturaleza favorable o desfavorable resulta útil para descubrir contactos particularmente atractivos como objetivos para el rediseño y, como se verá más adelante, para interpretar los resultados generados por Rosetta.

El punto más sobresaliente del análisis de las estructuras base para el rediseño con glutamina, LAO-Q1 y LAO-Q2, resulta ser el contacto desfavorable del oxígeno OE1 de la cadena lateral con la cadena lateral del residuo Leu-117 de la proteína, con una superficie de contacto de alrededor de 27 Å² en promedio. Por tanto, esta interacción de tipo aceptor-hidrofóbico se presenta claramente como un blanco probable para el rediseño. Además de este contacto negativo, el análisis no presenta otros con superficies de contacto comparables, aunque dados los resultados del rediseño con Rosetta, discutidos más adelante, resulta relevante señalar los contactos de los átomos polares de la cadena lateral del ligando OE1 y NE1, con los residuos aromáticos Tyr-14 y Phe-52, respectivamente, cada uno con superficies mayores a 10 Å². Estos contactos no son considerados desfavorables por LPC, aunque, como se verá en los rediseños por Rosetta, constituyen los residuos modificados más frecuentemente.

Las estructuras nativas de LAO y QBP presentan también algunos contactos desestabilizantes, aunque en general tienden a ser menores. Los más notables son los del CB de la lisina en LAO, que forman contactos desfavorables con los oxígenos de Tyr-14 y Gln-122. No obstante, estos residuos a su vez se encuentran formando puentes de hidrógeno con una molécula de agua estructural, lo cual

sirve para neutralizar sus cargas parciales y aminorar así el efecto desestabilizante. Por su parte, los principales contactos desfavorables en LIV con valina también se encuentran amortiguados por interacciones con átomos neutralizando las cargas parciales responsables.

En general, aunque los resultados de LPC resultan útiles para obtener un panorama de las interacciones, como puede verse de lo anterior, también tiene ciertas limitaciones, entre ellas la de la falta de consideración de moléculas de agua estructurales.

LAO-Q1

Ligando		Proteína			Contacto	
Atom	Cl	Residuo	Atom	Cl	Dist.	Sup.
CA	VI	GLN 122	OE1	II	3.9	14.1
C	VI	SER 72	OG	I	3.6	2.9
C	VI	ARG 77	NH1	III	3.6	2.9
C	VI	THR 121	OG1	I	3.7	6.5
O	II	ARG 77	NH2	III	2.9	21.0
O	II	THR 121	N	III	2.9	17.5
O	II	SER 120	CA	VII	3.4	4.0
OXT	II	SER 72	N	III	2.8	30.7
OXT	II	SER 72	OG	I	2.8	4.2
OXT	II	ARG 77	NH1	III	2.9	6.9
N	I	ASP 161	OD2	II	2.7	28.1
N	I	ASP 161	OD1	II	2.9	9.5
N	I	SER 70	O	II	3.1	6.9
N	I	SER 72	OG	I	3.4	7.7
CB	IV	SER 70	O	II	3.3	5.6 *
CB	IV	PHE 52	CZ	V	3.3	10.1
CB	IV	SER 120	CB	VI	4.7	1.1
CG	IV	TYR 14	CZ	V	3.2	19.3
CG	IV	TYR 14	CE2	V	3.3	3.4
CG	IV	SER 70	O	II	4.0	2.0 *
CG	IV	LEU 117	CD2	IV	4.6	1.8
CD	VI	TYR 14	CZ	V	3.5	5.6
CD	VI	TYR 14	CE2	V	3.6	1.6
CD	VI	PHE 52	CE1	V	3.7	3.8
CD	VI	LEU 117	CD2	IV	3.8	1.1
OE1	II	LEU 117	CD2	IV	2.8	25.6 *
OE1	II	TYR 14	CE1	V	3.1	11.4
OE1	II	TYR 14	CD1	V	3.2	2.9
OE1	II	ASP 11	OD2	II	5.2	1.0 *
NE2	I	SER 69	O	II	2.7	25.9
NE2	I	PHE 52	CE1	V	3.1	10.7
NE2	I	PHE 52	CD1	V	3.2	3.4
NE2	I	SER 69	C	VI	3.8	2.0
NE2	I	SER 69	OG	I	3.8	3.6
NE2	I	SER 70	CB	VI	4.1	1.8
NE2	I	TYR 14	CD2	V	4.1	2.8
NE2	I	TYR 14	CG	V	4.2	1.2
NE2	I	TYR 14	CB	IV	4.7	2.6 *
NE2	I	ASP 11	CB	IV	5.0	2.2 *

LAO-Q2

Ligando		Proteína			Contacto	
Atom	Cl	Residuo	Atom	Cl	Dist.	Sup.
CA	VI	GLN 122	OE1	II	4.0	13.7
C	VI	ARG 77	NH1	III	3.6	3.1
C	VI	ARG 77	NH2	III	3.6	1.3
C	VI	SER 72	OG	I	3.7	1.3
C	VI	THR 121	N	III	3.9	2.9
C	VI	THR 121	OG1	I	3.9	3.1
O	II	SER 72	OG	I	2.8	21.0
O	II	ARG 77	NH1	III	2.8	17.5
O	II	SER 72	N	III	2.9	4.9
OXT	II	THR 121	N	III	2.9	26.5
OXT	II	ARG 77	NH2	III	2.9	11.6
OXT	II	SER 120	CA	VII	3.2	4.5
N	I	ASP 161	OD2	II	2.6	26.9
N	I	ASP 161	OD1	II	3.0	8.9
N	I	SER 70	O	II	3.1	7.7
N	I	SER 72	OG	I	3.5	7.3
CB	IV	SER 70	O	II	3.2	9.4 *
CB	IV	PHE 52	CZ	V	3.2	13.0
CG	IV	TYR 14	CE2	V	3.3	19.5
CG	IV	TYR 14	CZ	V	3.4	2.9
CG	IV	SER 70	O	II	3.8	3.4 *
CD	VI	PHE 52	CE1	V	3.6	3.1
CD	VI	TYR 14	CZ	V	3.7	6.7
CD	VI	TYR 14	CE1	V	3.8	1.1
OE1	II	LEU 117	CD1	IV	3.5	27.7 *
OE1	II	TYR 14	CE1	V	3.5	2.3
OE1	II	PHE 52	CD1	V	3.7	6.8
OE1	II	PHE 52	CG	V	3.8	1.4
NE2	I	SER 69	O	II	2.8	22.2
NE2	I	PHE 52	CD1	V	3.4	10.3
NE2	I	SER 69	OG	I	3.7	6.5
NE2	I	SER 69	C	VI	3.8	1.2
NE2	I	TYR 14	CD2	V	4.0	7.7
NE2	I	TYR 14	CG	V	4.0	2.4
NE2	I	TYR 14	CB	IV	4.4	3.2 *
NE2	I	ASP 11	CB	IV	4.8	3.6 *

I	Hidrofílico	Átomos de N u O que pueden donar y aceptar puentes de hidrógeno.
II	Aceptor	Átomos de N u O que sólo puede aceptar puentes de hidrógeno
III	Donador	Átomos de N u O que sólo pueden donar puentes de hidrógeno
IV	Hidrofóbico	Cl, B, I y todos los átomos de C que no formen parte de anillos aromáticos o estén unidos covalentemente con N u O
V	Aromático	Átomos de C en anillos aromáticos independientemente de otros enlaces.
VI	Neutro	Átomos de C con enlaces covalente con al menos un átomo de clase I o 2 ó mas de clase II. S, F, P y átomos metálicos en todo caso.
VII	Neutro-donador	Átomos de C con un solo enlace a un átomo de clase III.
VIII	Neutro-aceptor	Átomos de C con un solo enlace a un átomo de clase III.

QBP (1WDN)

Ligando		Proteína		Contacto	
Atom	Cl	Residuo	Atom	Cl	Sup.
N	I	ASP 157	OD2	II	23.6
N	I	GLY 68	O	II	15.0
N	I	THR 70	OG1	I	13.3
N	I	GLN 183	OE1	II	1.2
CA	VI	ASP 157	OD2	II	1.6
CA	VI	ASP 157	OD1	II	3.4
CA	VI	THR 118	CB	VI	4.7
CA	VI	GLY 119	N	III	1.6
C	VI	ARG 75	NH2	III	2.7
C	VI	PHE 50	CZ	V	2.2
C	VI	ARG 75	NH1	III	1.8
C	VI	THR 70	OG1	I	3.4
C	VI	GLY 119	N	III	2.5
O	II	THR 70	N	III	27.0
O	II	ARG 75	NH1	III	10.0
O	II	PHE 50	CZ	V	4.2
CB	IV	GLY 68	O	II	4.5 *
CB	IV	PHE 50	CE1	V	9.4
CB	IV	THR 118	CG2	IV	3.6
CG	IV	PHE 13	CE1	V	15.9
CG	IV	PHE 13	CZ	V	2.9
CG	IV	HIS 156	NE2	I	7.6 *
CG	IV	ASP 157	OD1	II	1.1 *
CD	VI	PHE 13	CD1	V	8.1
CD	VI	LYS 115	NZ	III	1.6
CD	VI	PHE 50	CD1	V	1.3
OE1	II	LYS 115	NZ	III	21.8
OE1	II	HIS 156	NE2	I	16.3
OE1	II	LYS 115	CE	VII	1.9
OE1	II	PHE 13	CD1	V	2.9
NE2	I	ASP 10	OD2	II	24.4
NE2	I	ALA 67	O	II	17.0
NE2	I	PHE 50	CD1	V	3.2
NE2	I	PHE 13	CG	V	8.5
OXT	II	ARG 75	NH2	III	21.1
OXT	II	GLY 119	N	III	16.3
OXT	II	THR 118	CA	VII	5.0

LAO-K (1LST)

Ligando		Proteína		Contacto	
Atom	Cl	Residuo	Atom	Cl	Sup.
N	I	SER 70	O	II	16.0
N	I	ASP 161	OD2	II	15.6
N	I	ASP 161	OD1	II	3.8
N	I	SER 72	OG	I	12.9
CA	VI	ASP 161	OD1	II	5.6
CA	VI	GLN 122	OE1	II	7.0
CA	VI	THR 121	OG1	I	5.2
C	VI	ARG 77	NH1	III	3.1
C	VI	THR 121	OG1	I	7.0
C	VI	PHE 52	CZ	V	1.6
C	VI	THR 121	CG2	IV	1.1
O	II	ARG 77	NH1	III	22.5
O	II	SER 72	N	III	10.9
O	II	SER 70	O	II	2.1 *
O	II	LEU 71	CA	VII	1.6
CB	IV	GLN 122	OE1	II	13.2 *
CB	IV	TYR 14	OH	I	10.1 *
CB	IV	SER 120	CB	VI	3.1
CG	IV	SER 70	O	II	9.9 *
CG	IV	PHE 52	CZ	V	3.8
CG	IV	TYR 14	OH	I	1.8 *
CG	IV	TYR 14	CE2	V	3.1
CD	IV	TYR 14	CZ	V	11.0
CD	IV	LEU 117	CD2	IV	14.1
CD	IV	PHE 52	CE1	V	4.5
CE	VI	SER 69	O	II	20.2
CE	VI	PHE 52	CE1	V	5.2
CE	VI	TYR 14	CD2	V	13.7
NZ	I	ASP 11	OD2	II	29.1
NZ	I	TYR 14	CG	V	15.4
NZ	I	TYR 14	CD1	V	2.8
NZ	I	PHE 52	CD1	V	4.4
NZ	I	LEU 117	CD2	IV	2.6 *
NZ	I	ASP 11	CG	VI	2.4
NZ	I	TYR 14	CB	IV	1.4 *
NZ	I	ASP 11	CB	IV	1.4 *
OXT	II	THR 121	N	III	23.0
OXT	II	ARG 77	NH2	III	10.7
OXT	II	SER 120	CA	VII	5.5
OXT	II	PHE 52	CZ	V	1.4

Clase Atómica		I	II	III	IV	V	VI	VII	VIII
I	Hidrofílico	+	+	+	-	+	+	+	+
II	Aceptor	+	-	+	-	+	+	+	-
III	Donador	+	+	-	-	+	+	-	+
IV	Hidrofóbico	-	-	-	+	+	+	+	+
V	Aromático	+	+	+	+	+	+	+	+
VI	Neutro	+	+	+	+	+	+	+	+
VII	Neutro-donador	+	+	-	+	+	+	-	+
VIII	Neutro-aceptor	+	-	+	+	+	+	+	-

Tabla 4: Tablas de LPC de contactos atómicos ligando-proteína par alas construcciones LAO-Q1, LAO-Q2 y LAO-V, así como para las estructuras nativas de QBP, LAO-K y LIV-V. *Dist.*: Distancia de la interacción en Å. *Sup.*: Superficie de contacto en Å². Los asteriscos (*) indican los contactos considerados como desestabilizantes. Por razones de espacio sólo se presentan los contactos de superficie mayor a 1 Å. Se incluyen también la guía de las clases atómicas así como la de contactos válidos y desestabilizantes

LAO-V

Ligando		Proteína			Contacto	
Atom	Cl	Residuo	Atom	Cl	Dist.	Sup.
CA	VI	GLN 122	OE1	II	3.8	11.4
CB	IV	SER 70	O	II	3.4	5.6 *
CB	IV	PHE 52	CZ	V	3.6	4.9
CG1	IV	PHE 52	CZ	V	3.2	22.2
CG1	IV	PHE 52	CE2	V	3.3	6.3
CG1	IV	SER 120	CB	VI	3.5	14.8
CG1	IV	LEU 117	CD2	IV	3.6	17.3
CG1	IV	PHE 52	CE1	V	3.6	1.3
CG1	IV	GLN 122	OE1	II	4.3	1.6 *
CG2	IV	TYR 14	CE2	V	3.1	36.3
CG2	IV	TYR 14	OH	I	3.2	4.7 *
CG2	IV	ASP 161	OD2	II	3.4	1.8 *
CG2	IV	SER 70	O	II	3.5	6.3 *
CG2	IV	SER 70	CB	VI	3.7	6.3
CG2	IV	SER 69	O	II	4.3	2.0 *
C	VI	ARG 77	NH1	III	3.6	4.0
C	VI	THR 121	OG1	I	3.6	8.7
C	VI	PHE 52	CZ	V	4.0	1.6
O	II	SER 72	N	III	2.8	25.8
O	II	ARG 77	NH1	III	2.9	6.9
O	II	SER 72	OG	I	2.9	4.5
OXT	II	THR 121	N	III	2.9	24.8
OXT	II	ARG 77	NH2	III	2.9	12.6
N	I	ASP 161	OD2	II	2.7	24.7
N	I	SER 70	O	II	3.0	7.7
N	I	ASP 161	OD1	II	3.1	6.3
N	I	SER 72	OG	I	3.2	9.9

LIV-V (1Z18)

Ligando		Proteína			Contacto	
Atom	Cl	Residuo	Atom	Cl	Dist.	Sup.
N	I	GLU 226	OE2	II	2.7	22.0
N	I	ALA 100	O	II	2.8	15.0
N	I	THR 102	OG1	I	3.1	12.7
CA	VI	TYR 150	CE1	V	3.3	11.7
C	VI	TYR 150	CD1	V	3.4	9.4
C	VI	SER 79	N	III	3.6	4.3
O	II	SER 79	OG	I	2.7	22.0
O	II	THR 102	N	III	3.1	10.4
O	II	ALA 101	CA	VII	3.4	2.6
O	II	THR 102	OG1	I	3.5	1.9
CB	IV	TYR 202	CE1	V	3.9	6.5
CB	IV	TYR 202	OH	I	4.1	1.1 *
CB	IV	TYR 18	CE2	V	4.4	1.8
CG1	IV	GLU 226	OE2	II	3.3	12.3 *
CG1	IV	GLY 227	CA	VI	3.8	24.2
CG1	IV	TYR 18	CZ	V	3.9	17.0
CG1	IV	PHE 276	CZ	V	3.9	9.2
CG1	IV	TYR 202	CE1	V	4.1	2.5
CG2	IV	LEU 77	O	II	3.1	26.5 *
CG2	IV	ALA 100	O	II	3.3	7.6 *
CG2	IV	ALA 100	CB	IV	3.5	13.2
CG2	IV	LEU 77	C	VI	3.7	4.3
CG2	IV	PHE 276	CZ	V	4.1	2.2
CG2	IV	PHE 276	CE2	V	4.4	1.8
CG2	IV	LEU 77	CB	IV	4.4	2.7
CG2	IV	TYR 18	CD2	V	4.8	3.1
OXT	II	TYR 202	OH	I	2.6	25.3
OXT	II	SER 79	N	III	2.9	13.7
OXT	II	TYR 150	CB	IV	3.7	3.8 *

RosettaDesign y evaluaciones por consenso

En las tablas 5 a 7 se presentan los 25 mejores resultados de las tres estrategias de rediseño para glutamina, ordenados por puntaje consenso. La tabla 5 muestra los resultados y mutaciones de la estrategia libre, combinando los resultados para ambos conformeros de la glutamina. Para esta estrategia, el número de mutaciones propuestas van de tres hasta diez y consistentemente se concentran en torno a los residuos Tyr-14, Phe-52 y Leu-117. Para Tyr-14, la mutación más usual fue por ácido glutámico, mientras que para la Phe-52 fuer principalmente serina y, en menor grado, histidina. Las mutaciones preferidas para Leu-117 fueron generalmente por ácido aspártico o asparagina. Otros residuos que presentaron mutaciones frecuentes fueron la Gln-122, predominante-

Tabla 5 – 25 mejores resultados por puntuación promedio para estrategia de rediseño libre

	<i>BK</i>	<i>LS</i>	<i>XS</i>	<i>DS</i>	<i>BAP</i>	<i>Punt. Prom.</i>	<i>Clas. Prom.</i>	<i>Mutaciones</i>
LIBRE-Q1-38	-783.72	-21.1	4.98	-158027	-0.32	0.8791	11.3	(5) Y14E F52H L117N T121S Q122E
LIBRE-Q2-22	-787.97	-22.14	4.93	-156382	-0.54	0.877	7.8	(7) Y14E F52H S70D L117N Q122E D161T T195C
LIBRE-Q1-42	-786.05	-20.36	4.93	-168454	0.71	0.8506	21.8	(4) Y14E F52H S72N L117N
LIBRE-Q1-36	-783.71	-20.38	5.00	-158701	0.78	0.849	21.0	(5) Y14E F52H S72N L117N T121S
LIBRE-Q2-31	-784.78	-22.7	4.92	-138844	-1.18	0.8362	13.0	(8) Y14E F52H S70D S72T L117C T121S D161T T195C
LIBRE-Q1-45	-784.42	-20.88	4.90	-159894	0.29	0.8264	17.8	(4) Y14E F52H S72T L117N
LIBRE-Q1-41	-786.38	-20.34	5.00	-156736	1.77	0.8175	29.8	(8) Y14E F52H S70D S72T L117D Q122E D161T T195C
LIBRE-Q2-37	-789.38	-22.25	4.87	-146343	-0.44	0.8107	14.3	(8) D11N Y14N F52T S70D L117I Q122E D161T T195C
LIBRE-Q2-25	-785.84	-21.74	4.89	-146210	0.74	0.7806	19.8	(7) Y14E F52H S70D L117N Q122D D161T T195C
LIBRE-Q1-34	-787.28	-20.81	4.92	-150575	1.31	0.7767	27.3	(7) Y14E F52S S70D L117D Q122E D161T T195C
LIBRE-Q2-48	-786.75	-22.38	4.81	-137553	-1.43	0.7752	23.3	(7) Y14S F52S S70D L117D Q122T D161T T195C
LIBRE-Q2-38	-785.07	-21.35	4.88	-135878	-1.04	0.7715	22.8	(8) Y14E F52H S70D S72T L117C Q122E D161T T195S
LIBRE-Q1-24	-786.07	-21.68	4.83	-149537	0.42	0.7689	21.0	(4) Y14E F52S L117N T121S
LIBRE-Q2-46	-786.17	-21.34	4.86	-142671	-0.16	0.765	21.0	(7) Y14D F52H S69D S70D L117V D161T T195C
LIBRE-Q2-32	-785.06	-21.93	4.85	-130839	-1.57	0.7639	26.8	(8) Y14S F52S S70D L71I L117N Q122T D161T T195S
LIBRE-Q2-23	-785.88	-19.93	4.87	-147387	-0.48	0.7627	26.5	(8) Y14E F52S S70D L71I L117N Q122E D161T T195C
LIBRE-Q1-22	-783.56	-21.03	4.88	-142635	0.08	0.7621	22.0	(7) Y14E F52S S70D L117N Q122T D161T T195C
LIBRE-Q1-46	-787.09	-21.09	4.85	-146400	0.09	0.7619	22.0	(8) Y14E F52S S70D S72T L117N Q122E D161T T195S
LIBRE-Q1-25	-787.70	-21.04	4.88	-149747	1.32	0.7587	27.0	(7) Y14E F52S S70D L117D Q122E D161T T195C
LIBRE-Q1-32	-787.00	-20.56	4.88	-155374	1.93	0.7538	33.5	(4) Y14E F52S S72N L117N
LIBRE-Q2-33	-787.21	-22.17	4.85	-144536	1.55	0.745	29.8	(8) Y14E F52T S70D S72T L117I Q122E D161T T195C
LIBRE-Q2-24	-784.43	-22.93	4.80	-142490	1.1	0.7406	29.5	(4) Y14E F52S L117N Q122T
LIBRE-Q1-30	-783.59	-22.26	4.83	-143432	1.26	0.74	28.0	(4) Y14E F52H L117D Q122T
LIBRE-Q2-30	-788.26	-22.63	4.76	-146900	0.86	0.7355	30.0	(3) Y14E F52S L117N
LIBRE-Q1-31	-784.77	-20.21	4.83	-150701	0.88	0.7288	34.8	(9) Y14N F52H S69D S70D L71I L117N Q122E D161T T195S

BK: BK_TOT, LS: Ligsum, XS: X-Score, DS: DrugScore, BAP: BAPPL, Punt. Prom: Consenso por puntuación promedio, Clas. Prom.: Consenso por clasificación promedio.

mente por glutámico; Ser-72, por asparagina, y Asp-161, por treonina. Los dos mejores resultados por consenso, en un caso por puntuación y en otro por clasificación promedio, de cinco y siete mutaciones, respectivamente, poseen cuatro de las mutaciones más frecuentes (Y14E, F52H, L117N, Q122E) en común, lo que podría sugerir cierta correlación entre la frecuencia global de mutaciones y afinidad de unión por consenso.

Para la estrategia conservativa, cuyos mejores 25 resultados se muestran en la tabla 6, el número de mutaciones oscila entre una y cuatro, demostrando que la compensación energética sobre los aminoácidos nativos que caracteriza a esta estrategia, se manifiesta efectivamente sobre el número de

Tabla 6 – 25 mejores resultados por puntuación promedio para estrategia de rediseño conservativa

	<i>BK</i>	<i>LS</i>	<i>XS</i>	<i>DS</i>	<i>BAP</i>	<i>Punt. Prom.</i>	<i>Clas. Prom.</i>	<i>Mutaciones</i>
CONSER-Q1-V1-09	-784.69	-21.32	5.00	-163308	-1.47	0.894	5.5	(1) Y14T
CONSER-Q1-V1-07	-784.12	-20.42	5.00	-167025	-0.40	0.802	15.5	(3) Y14E F52H L117N
CONSER-Q1-V2-19	-784.06	-20.39	4.97	-160963	-0.97	0.787	17.0	(2) Y14E L117N
CONSER-Q2-V1-17	-785.52	-21.28	4.92	-160200	-0.74	0.784	14.0	(3) Y14E F52H L117N
CONSER-Q1-V1-01	-784.31	-20.46	4.97	-167872	-0.23	0.780	17.0	(2) Y14E L117N
CONSER-Q1-V1-10	-785.26	-20.65	4.90	-164208	-0.93	0.772	16.8	(3) Y14E F52H L117N
CONSER-Q2-V1-07	-784.83	-21.80	4.90	-158680	-0.11	0.753	19.5	(2) Y14E L117N
CONSER-Q1-V1-04	-784.81	-20.97	4.90	-160872	-0.56	0.749	16.3	(3) Y14E F52H L117N
CONSER-Q2-V1-03	-783.94	-21.89	4.87	-157303	-0.10	0.732	21.8	(3) Y14E F52H L117C
CONSER-Q1-V1-15	-784.12	-20.56	4.87	-155331	-1.24	0.724	21.8	(3) Y14E F52H L117C
CONSER-Q1-V1-20	-782.55	-20.34	5.01	-155095	0.28	0.704	27.8	(2) Y14E L117N
CONSER-Q1-V2-07	-786.58	-20.66	5.01	-158948	0.98	0.697	27.8	(1) Y14E
CONSER-Q1-V1-05	-785.17	-21.71	4.85	-158474	0.55	0.678	28.3	(3) Y14E F52H L117N
CONSER-Q1-V1-08	-784.61	-20.75	4.90	-156950	0.14	0.677	26.8	(2) Y14E L117N
CONSER-Q1-V2-04	-786.30	-20.53	4.95	-159372	0.83	0.666	30.0	(1) Y14E
CONSER-Q1-V1-19	-784.38	-20.18	4.86	-162888	-0.03	0.664	32.3	(2) Y14E L117N
CONSER-Q1-V1-24	-784.86	-20.24	4.85	-155503	-0.66	0.662	31.8	(3) Y14E F52H L117N
CONSER-Q1-V2-18	-785.86	-20.17	4.94	-165101	1.05	0.657	36.3	(1) Y14E
CONSER-Q1-V2-10	-781.66	-19.12	4.96	-155254	-0.32	0.650	38.0	(2) Y14E L117N
CONSER-Q1-V1-23	-787.07	-21.02	4.90	-157415	0.88	0.648	30.0	(2) Y14E F52H
CONSER-Q1-V1-12	-783.82	-19.73	4.84	-154279	-0.99	0.644	38.0	(3) Y14E F52H L117N
CONSER-Q1-V2-11	-786.65	-21.22	4.94	-152767	1.23	0.638	33.0	(1) Y14E
CONSER-Q2-V2-21	-785.17	-21.44	4.90	-151213	0.93	0.635	33.8	(2) Y14E L117N
CONSER-Q1-V2-22	-786.37	-20.46	4.96	-168442	2.14	0.634	34.0	(1) Y14E
CONSER-Q2-V1-14	-785.24	-21.37	4.85	-151048	0.40	0.633	33.0	(3) Y14E F52H L117N

BK: BK_TOT, LS: Ligsum, XS: X-Score, DS: DrugScore, BAP: BAPPL, Punt. Prom: Consenso por puntuación promedio, Clas. Prom.: Consenso por clasificación promedio. Q1/2: Confórmero de glutamina; V1/2, vuelta o corrida de rediseño.

mutaciones. Nuevamente, las mutaciones más comunes fueron sobre Tyr-14, Phe-52 y Leu-117, manteniendo asimismo las preferencias relativas de la estrategia libre. Otras mutaciones con apariciones repetidas ocurrieron para Gln-122 y Thr-121. Interesantemente, la propuesta con más alta clasificación por ambos métodos por consenso para esta estrategia (y para todas en general, como puede observarse en la tabla 8), presenta una sola mutación, que además es única entre todos los resultados producidos por Rosetta para el rediseño de glutamina: Y14T. Este resultado, así como otros importantes, serán discutidos más adelante.

Tabla 7 – 25 mejores resultados por puntuación promedio para estrategia de rediseño guiada

	BK	LS	XS	DS	BAP	Punt. Prom.	Clas. Prom.	Mutaciones
GUIADA-Q2-V2-22	-782.26	-18.48	5.01	-150760	-1.79	0.7995	5.8	(3) L117K Q143G Q160D
GUIADA-Q1-V1-09	-749.75	-17.65	4.89	-158221	-3.40	0.7863	15.5	(5) Y14F L117T Q122T Q160T D161E
GUIADA-Q2-V2-17	-781.81	-18.02	5.06	-147025	-1.35	0.7527	13.0	(3) L117K Q143G Q160D
GUIADA-Q2-V2-25	-779.41	-19.29	4.94	-135993	-1.41	0.6945	25.5	(4) L117T Q122T Q143T Q160D
GUIADA-Q2-V2-05	-782.18	-18.38	4.91	-143657	-1.39	0.6679	20.5	(4) F52L L117K Q143G Q160D
GUIADA-Q2-V1-01	-779.22	-18.73	4.94	-144088	-0.11	0.6646	24.3	(4) F52L L117T Q122T Q160S
GUIADA-Q1-V1-22	-778.65	-18.07	4.96	-140734	-1.13	0.6443	22.5	(3) L117N T121S Q160S
GUIADA-Q2-V2-11	-782.80	-17.83	4.91	-142590	-1.37	0.6209	26.3	(4) F52L L117K Q143G Q160D
GUIADA-Q2-V2-07	-779.88	-16.62	5.02	-149967	-0.42	0.6175	29.5	(2) L117T Q160S
GUIADA-Q1-V1-15	-780.37	-18.02	4.79	-154290	-0.37	0.6111	31.8	(4) F52C L117D T121S Q160S
GUIADA-Q1-V2-23	-780.59	-17.18	4.86	-152158	-1.13	0.6063	27.8	(4) F52L L117T Q143A Q160D
GUIADA-Q2-V2-06	-776.19	-17.83	4.98	-132215	-1.69	0.5988	30.8	(4) L117N Q122T Q143T Q160D
GUIADA-Q2-V1-03	-777.78	-17.97	4.84	-142481	-1.48	0.5934	29.8	(3) F52G L117K Q160S
GUIADA-Q2-V2-03	-781.57	-17.21	4.92	-150157	-0.13	0.5911	30.0	(4) F52L L117T Q143T Q160D
GUIADA-Q2-V1-06	-781.88	-17.89	4.90	-148111	0.62	0.5839	36.8	(4) F52L L117T Q122E Q160S
GUIADA-Q2-V2-01	-777.99	-18.29	4.92	-131774	-1.34	0.5797	33.8	(3) L117T Q122T Q160S
GUIADA-Q2-V2-21	-779.25	-18.48	4.95	-129194	-0.76	0.5699	34.5	(4) L117T Q122T Q143T Q160D
GUIADA-Q1-V2-08	-751.22	-16.55	4.83	-143831	-3.32	0.5675	41.0	(6) Y14F L117T Q122T Q143N Q160D D161E
GUIADA-Q2-V2-15	-779.86	-18.95	4.88	-143219	1.78	0.5673	43.8	(5) F52L L117T Q122T Q143T Q160D
GUIADA-Q2-V2-23	-782.85	-18.81	4.81	-142000	0.33	0.5629	45.5	(3) F52C L117D Q160S
GUIADA-Q2-V2-02	-781.53	-16.74	4.98	-141919	-0.76	0.5574	37.0	(4) F52L L117K Q143G Q160D
GUIADA-Q2-V2-10	-779.99	-17.23	4.92	-142764	-0.20	0.5417	36.8	(4) L117T Q122D Q143T Q160D
GUIADA-Q2-V1-22	-781.89	-17.81	4.87	-140215	-0.28	0.5379	40.5	(3) F52L L117T Q160S
GUIADA-Q2-V1-10	-777.32	-17.64	4.92	-134123	-0.94	0.5360	37.8	(3) L117T Q122T Q160S
GUIADA-Q1-V1-21	-779.79	-17.06	4.91	-137704	-1.45	0.5348	39.0	(2) L117N Q160S

BK: BK_TOT, LS: Ligsum, XS: X-Score, DS: DrugScore, BAP: BAPPL, Punt. Prom: Consenso por puntuación promedio, Clas. Prom.: Consenso por clasificación promedio. Q1/2: Confórmero de glutamina; V1/2, vuelta o corrida de rediseño.

Cabe mencionar que para las estrategias anteriores es posible notar ciertas tendencias en los resultados respecto a los distintos conformeros de glutamina. Para el conformero 1, es decir, el presente en la estructura cristalográfica de QBP, las mutaciones en Leu-117 son menos frecuentes que para el conformero 2, generado por AutoDock. Asimismo, existe una mayor tendencia a favorecer serina como mutación para Phe-52 para el conformero 2, en comparación al conformero 1, que tiene una inclinación relativamente mayor por histidina.

Tabla 8 – 25 mejores resultados globales por puntuación promedio para todas las estrategias

	<i>BK</i>	<i>LS</i>	<i>XS</i>	<i>DS</i>	<i>BAP</i>	<i>Punt. Prom.</i>	<i>Clas. Prom.</i>	<i>Mutaciones</i>
CONSER-Q1-V1-09	-784.69	-21.32	5.00	-163308	-1.47	0.8649	15.25	(1) Y14T
CONSER-Q1-V1-07	-784.12	-20.42	5.00	-167025	-0.40	0.8319	40	(3) Y14E F52H L117N
CONSER-Q1-V1-01	-784.31	-20.46	4.97	-167872	-0.23	0.8167	45	(2) Y14E L117N
CONSER-Q1-V2-19	-784.06	-20.39	4.97	-160963	-0.97	0.8077	37	(2) Y14E L117N
LIBRE-Q1-38	-783.72	-21.10	4.98	-158027	-0.32	0.8019	36.75	(5) Y14E F52H L117N T121S Q122E
LIBRE-Q2-22	-787.97	-22.14	4.93	-156382	-0.54	0.7999	33.25	(7) Y14E F52H L117N S70D Q122E D161T T195C
CONSER-Q2-V1-17	-785.52	-21.28	4.92	-160200	-0.74	0.7945	37	(3) Y14E F52H L117N
CONSER-Q1-V1-10	-785.26	-20.65	4.90	-164208	-0.93	0.7897	45.5	(3) Y14E F52H L117N
CONSER-Q1-V2-07	-786.58	-20.66	5.01	-158948	0.98	0.7783	72	(1) Y14E
CONSER-Q1-V1-04	-784.81	-20.97	4.90	-160872	-0.56	0.7755	45.5	(3) Y14E F52H L117N
CONSER-Q2-V1-07	-784.83	-21.80	4.90	-158680	-0.11	0.7751	51	(2) Y14E L117N
CONSER-Q1-V1-20	-782.55	-20.34	5.01	-155095	0.28	0.7740	64.5	(2) Y14E L117N
LIBRE-Q1-42	-786.05	-20.36	4.93	-168454	0.71	0.7734	74.25	(4) Y14E F52H L117N S72N
LIBRE-Q1-36	-783.71	-20.38	5.00	-158701	0.78	0.7710	72.5	(5) Y14E F52H L117N S72N T121S
GUIADA-Q2-V2-22	-782.26	-18.48	5.01	-150760	-1.79	0.7667	57.5	(3) L117K Q143G Q160D
LIBRE-Q2-31	-784.78	-22.70	4.92	-138844	-1.18	0.7583	58.75	(8) Y14E F52H S70D S72T L117C T121S D161T T195C
CONSER-Q2-V1-03	-783.94	-21.89	4.87	-157303	-0.10	0.7568	59	(3) Y14E F52H L117C
GUIADA-Q2-V2-17	-781.81	-18.02	5.06	-147025	-1.35	0.7565	69.5	(3) L117K Q143G Q160D
CONSER-Q1-V2-22	-786.37	-20.46	4.96	-168442	2.14	0.7554	91.25	(1) Y14E
GUIADA-Q1-V1-09	-749.75	-17.65	4.89	-158221	-3.40	0.7552	74.25	(5) Y14F L117T Q122T Q160T D161E
CONSER-Q1-V2-18	-785.86	-20.17	4.94	-165101	1.05	0.7533	89.25	(1) Y14E
CONSER-Q1-V2-04	-786.30	-20.53	4.95	-159372	0.83	0.7506	76.75	(1) Y14E
NATIVA Q2	-768.00	-19.73	5.00	-148523	-0.52	0.7504	68.0	(0)
LIBRE-Q1-45	-784.42	-20.88	4.90	-159894	0.29	0.7488	68.0	(4) Y14E F52H L117N S72T
CONSER-Q1-V1-15	-784.12	-20.56	4.87	-155331	-1.24	0.7473	54.0	(3) Y14E F52H L117C

BK: BK_TOT, LS: Ligsum, XS: X-Score, DS: DrugScore, BAP: BAPPL, Punt. Prom: Consenso por puntuación promedio, Clas. Prom.: Consenso por clasificación promedio. Q1/2: Conformero de glutamina; V1/2, vuelta o corrida de rediseño. En negrita se muestran las mutaciones más frecuentes

Los resultados de la estrategia guiada, expuestos en la tabla 7, presentan entre dos y cinco mutaciones. Dadas las restricciones impuestas para esta estrategia (la tirosina 14 no se permitió cambiar más que por fenilalanina, como en QBP) este residuo ya no resultó ser de los consistentemente mutados, que en este caso fueron Leu-117 y Gln-160, seguidos de Phe-52 y Gln-143. Leu-117 fue mutado primordialmente por treonina, aspártico o asparagina, mientras que Gln-160 lo fue casi exclusivamente por serina. Las mutaciones de Phe-52 fueron la mayor parte hacia leucina o cisteína y las de Gln-143 principalmente a treonina. Por otro lado, puede observarse también que para esta estrategia los *lig_sums* tienden a ser menores en general que en las estrategias anteriores. Esto es en parte debido al filtro menor usado para Rosetta, pero éste es a su vez consecuencia lógica de las restricciones, ya que limita a Rosetta de posibilidades que podrían ser consideradas positivas, como es el caso, por ejemplo, de Y14E.

En la tabla 8 puede verse la clasificación de los primeros 25 resultados de las tres estrategias calificadas entre sí, e incluyendo también las estructuras minimizadas iniciales. Como puede verse, la estructura del confórmero 2 incluso logró colocarse en el lugar 23 por puntuación. El confórmero 1 estuvo en el lugar 115 (fuera de lo que muestra la tabla), aunque más de la mitad de los 25 modelos mejor clasificados fueron rediseños en torno al mismo. Los primeros lugares estuvieron dominados por la estrategia conservativa sobre la libre, a pesar de la restricción de la primera, lo que sugiere que permitir un mayor número de mutaciones no necesariamente resulta en mejores resultados. La estrategia guiada quedó relegada con únicamente tres lugares entre los 25 superiores. No obstante, considerando que no se habrían posicionado ni siquiera en los primeros 50 basados bajo el criterio único de *ligsum*, es evidente que otras funciones evalúan los resultados de esta estrategia de manera más favorable que Rosetta. De hecho, cabe hacer notar que el ordenamiento de los resultados basado en las evaluaciones de la función BAPPL, como se muestra en la tabla 9 resulta en un importante predominio de los resultados de la estrategia guiada. De particular notoriedad son los primeros dos, que se separan del resto con una estimación de energía sobresaliente y con la mutación Y14F en común. No obstante, muestran *bk_tots* notablemente inferiores, debido a un aumento en el parámetro repulsivo a consecuencia del choque entre Phe-14 y la Ala-164. Resulta imposible determinar si este impedimento resultaría experimentalmente en un efecto perjudicial para la proteína, como incapacidad para plegarse, o simplemente una adaptación de la cadena principal, cuya flexibilidad no es tomada en cuenta por Rosetta en sus rediseños.

Tabla 9 – 25 mejores resultados globales ordenados por resultados de BAPPL para todas las estrategias

	<i>BK</i>	<i>LS</i>	<i>XS</i>	<i>DS</i>	<i>BAP</i>	<i>Punt. Prom.</i>	<i>Clas. Prom.</i>	<i>Mutaciones</i>
GUIADA-Q1-V1-09	-749.75	-17.65	4.89	-158221	-3.40	0.7552	72.8	(5) Y14F L117T Q122T Q160T D161E
GUIADA-Q1-V2-08	-751.22	-16.55	4.83	-143831	-3.32	0.6448	120.5	(6) Y14F L117T Q122T Q143N Q160D D161E
GUIADA-Q2-V1-09	-780.91	-17.86	4.79	-134521	-2.05	0.5887	144.3	(5) F52C L117D Q122T Q160S D161T
GUIADA-Q1-V1-23	-779.44	-16.56	4.95	-134125	-1.87	0.6332	121.8	(2) L117N Q160S
GUIADA-Q2-V2-22	-782.26	-18.48	5.01	-150760	-1.79	0.7667	57.3	(3) L117K Q143G Q160D
GUIADA-Q2-V2-06	-776.19	-17.83	4.98	-132215	-1.69	0.6656	110.0	(4) L117N Q122T Q143T Q160D
GUIADA-Q1-V1-06	-778.75	-16.52	4.96	-136818	-1.62	0.6412	115.5	(2) L117N Q160S
LIBRE-Q2-32	-785.06	-21.93	4.85	-130839	-1.57	0.6857	95.3	(8) Y14S F52S S70D L71I L117N Q122T D161T T195S
GUIADA-Q1-V1-17	-779.81	-16.91	4.88	-138935	-1.49	0.6147	121.0	(2) L117N Q160S
GUIADA-Q2-V1-03	-777.78	-17.97	4.84	-142481	-1.48	0.6317	109.8	(3) F52G L117K Q160S
GUIADA-Q2-V1-13	-776.40	-17.50	4.82	-137076	-1.47	0.5907	134.8	(3) F52G L117K Q160D
CONSER-Q1-V1-09	-784.69	-21.32	5.00	-163308	-1.47	0.8649	14.3	(1) Y14T
GUIADA-Q1-V1-21	-779.79	-17.06	4.91	-137704	-1.45	0.6275	117.0	(2) L117N Q160S
LIBRE-Q2-48	-786.75	-22.38	4.81	-137553	-1.43	0.6974	91.3	(7) Y14S F52S S70D L117D Q122T D161T T195C
GUIADA-Q2-V2-25	-779.41	-19.29	4.94	-135993	-1.41	0.6859	98.8	(4) L117T Q122T Q143T Q160D
GUIADA-Q2-V2-05	-782.18	-18.38	4.91	-143657	-1.39	0.6781	87.8	(4) F52L L117K Q143G Q160D
GUIADA-Q2-V2-11	-782.80	-17.83	4.91	-142590	-1.37	0.6612	96.0	(4) F52L L117K Q143G Q160D
GUIADA-Q2-V2-17	-781.81	-18.02	5.06	-147025	-1.35	0.7565	68.8	(3) L117K Q143G Q160D
GUIADA-Q2-V2-01	-777.99	-18.29	4.92	-131774	-1.34	0.6357	118.5	(3) L117T Q122T Q160S
CONSER-Q1-V1-15	-784.12	-20.56	4.87	-155331	-1.24	0.7473	54.0	(3) Y14E F52H L117C
LIBRE-Q2-31	-784.78	-22.70	4.92	-138844	-1.18	0.7583	59.8	(8) Y14E F52H S70D S72T L117C T121S D161T T195C
GUIADA-Q1-V1-22	-778.65	-18.07	4.96	-140734	-1.13	0.6788	92.3	(3) L117N T121S Q160S
GUIADA-Q1-V2-23	-780.59	-17.18	4.86	-152158	-1.13	0.651	94.8	(4) F52L L117T Q143A Q160D
CONSER-Q2-V2-06	-785.08	-19.01	4.66	-133911	-1.12	0.5246	171.8	(3) Y14N F52S L117N
GUIADA-Q1-V2-07	-779.93	-17.14	4.74	-148660	-1.11	0.5766	136.3	(3) F52A L117N Q160S

BK: BK_TOT, LS: Ligsun, XS: X-Score, DS: DrugScore, BAP: BAPPL, Punt. Prom: Consenso por puntuación promedio, Clas. Prom.: Consenso por clasificación promedio. Q1/2: Conórmoro de glutamina; V1/2, vuelta o corrida de rediseño.

Los resultados del rediseño para valina se exponen en la tabla 10. El número de mutaciones va de 4 a 11, entre ellas se presentan consistentemente a los residuos Tyr-14, Gln-143, Gln-160 y Ala-164. Para Tyr-14, en esta ocasión la mutación más frecuente fue por glutamina, mientras que para Gln-

143, Gln-160 y Ala-164 fueron treonina, ácido aspártico e isoleucina, respectivamente. Otros residuos mutados con cierta frecuencia fueron Phe-17, Phe-52, Ala-90, Leu-117 y Gln-122. El mejor resultado por consenso presenta cinco mutaciones, Y14E, F17H, Q143L, Q160D y A164I. Las mutaciones F17H, Q160D y A164I se cuentan entre las más comunes para cada residuo, no obstante, Y14E y Q143L no representan las mutaciones más frecuentes para estas posiciones. La estructura usada como base para el rediseño con valina se encuentra en el lugar 55 de la tabla completa.

Tabla 10 – 25 mejores resultados por puntuación promedio para el rediseño para valina

	<i>BK</i>	<i>LS</i>	<i>XS</i>	<i>DS</i>	<i>BAP</i>	<i>Punt. Prom.</i>	<i>Clas. Prom.</i>	<i>Mutaciones</i>
VAL-V1-09	-786.29	-16.83	5.22	-138286	-3.43	0.8152	5.0	(5) Y14E F17H Q143L Q160D A164I
VAL-V1-47	-783.29	-17.85	5.21	-117874	-3.79	0.7828	23.0	(10) Y14I F17H S70D Y88H Q122T Q143V Q160D D161T A164I T195C
VAL-V2-28	-786.21	-16.63	5.17	-138488	-3.30	0.7617	13.8	(5) Y14E F17H Q143L Q160D A164I
VAL-V2-49	-785.20	-16.60	5.20	-140687	-2.71	0.7503	17.5	(6) Y14E F17H Q122E Q143L Q160D A164I
VAL-V1-43	-786.16	-15.71	5.23	-139088	-3.18	0.7437	14.5	(6) Y14E F17H A90S Q143L Q160D A164I
VAL-V1-27	-783.92	-15.52	5.20	-143874	-3.08	0.7312	20.0	(5) Y14E F17H Q143L Q160D A164I
VAL-V1-37	-787.21	-16.00	5.20	-137633	-3.16	0.7310	15.0	(6) Y14N F17H A90S Q143T Q160D A164I
VAL-V2-37	-786.10	-15.94	5.20	-134133	-3.44	0.7285	17.3	(5) Y14N A90S Q143T Q160D A164I
VAL-V1-30	-785.38	-15.77	5.21	-141265	-2.86	0.7228	19.8	(5) Y14E F17H Q143L Q160D A164I
VAL-V1-33	-784.10	-16.21	5.08	-129846	-4.59	0.7213	30.0	(8) Y14N F17H A90S Q122T Q143T Q160D D161S A164I
VAL-V2-31	-786.20	-16.05	5.21	-135739	-2.85	0.7085	21.0	(6) Y14E F17H A90S Q143L Q160D A164I
VAL-V2-03	-785.36	-16.06	5.16	-135835	-3.27	0.7055	21.0	(4) Y14N Q143T Q160D A164I
VAL-V2-22	-784.67	-15.37	5.19	-136062	-3.46	0.7005	24.8	(4) Y14N Q143T Q160D A164I
VAL-V2-41	-785.33	-15.96	5.18	-135316	-2.91	0.6849	26.5	(5) Y14E F17H Q143L Q160D A164I
VAL-V1-48	-783.69	-16.48	5.19	-128752	-2.88	0.6843	25.3	(5) Y14I F17H Q143L Q160N A164L
VAL-V2-45	-785.26	-16.51	5.18	-125642	-3.02	0.6724	28.8	(8) Y14I A90S S92Q Q122S Q143V Q160D A164I F191N
VAL-V2-42	-786.22	-16.33	5.19	-131096	-2.49	0.6606	31.5	(7) Y14E F17H A90S Q122D Q143L Q160D A164I
VAL-V1-49	-784.69	-16.09	5.14	-135461	-2.79	0.6577	31.8	(5) Y14I F17H Q143V Q160D A164I
VAL-V1-19	-785.97	-15.56	5.16	-135800	-2.91	0.6509	33.0	(5) Y14N F17H Q143T Q160D A164I
VAL-V2-05	-785.60	-16.08	5.15	-129983	-3.00	0.6492	31.3	(5) Y14E F17H Q143L Q160D A164I
VAL-V2-43	-783.95	-15.62	5.15	-132256	-3.17	0.6469	32.5	(3) Y14N Q143T Q160D
VAL-V1-38	-784.79	-15.61	5.14	-131718	-3.30	0.6459	33.3	(4) Y14N Q143T Q160D A164I
VAL-V1-22	-784.11	-15.63	5.16	-139635	-2.48	0.6452	35.8	(4) Y14E Q143L Q160D A164I
VAL-V1-12	-785.45	-16.38	5.18	-132877	-2.13	0.6409	34.0	(6) Y14E F17H Q122E Q143L Q160D A164I
VAL-V2-02	-786.74	-16.25	5.19	-131326	-2.24	0.6395	34.5	(7) Y14E F17H A90S Q122E Q143L Q160D A164I

BK: BK_TOT, LS: Lignum, XS: X-Score, DS: DrugScore, BAP: BAPPL, Punt. Prom: Consenso por puntuación promedio, Clas. Prom.: Consenso por clasificación promedio. V1/2, vuelta o corrida de rediseño.

Resulta interesante hacer notar que, para el rediseño de valina, las funciones X-Score y BAPPL parecen asignar evaluaciones más favorables que para el caso de la glutamina, mientras que ocurre lo opuesto para el *lig_sum* y, ligeramente, para DrugScore. Dadas las diferentes naturalezas químicas entre lisina o arginina con valina y su mayor similitud con glutamina, se esperaría que la evaluación fuera más favorable con este último ligando. La discordancia probablemente refleja las diferencias en la estimación y en la importancia asignada a los diferentes tipos de interacciones entre las diferentes funciones de evaluación, pero al mismo tiempo proporciona robustez al método de evaluación por consenso, al sugerir cierta complementariedad entre las diversas funciones.

Análisis de las estructuras seleccionadas por consenso

A continuación se presentan y discuten las estructuras de los resultados más sobresalientes para cada estrategia, de acuerdo a los resultados por consenso descritos en la sección anterior.

En la figura 6 se muestra la representación tridimensional del sitio de unión desde dos perspectivas distintas para los dos rediseños más sobresalientes generados mediante la estrategia libre (tabla 5), ya sea por puntuación o clasificación promedio, los cuales presentan cinco o siete mutaciones, respectivamente. El rediseño con la mejor puntuación promedio (LIBRE-Q1-38) contiene las mutaciones Y14E F52H L117N T121S Q122E, mientras que aquél con la mejor clasificación promedio (LIBRE-Q2-22) cuenta con las mutaciones Y14E, F52H, S70D, L117N, Q122E, D161T, T195C. De las cuatro que tienen en común, Y14E F52H L117N Q122E, las tres primeras se enfocan en aumentar los puentes de hidrógeno entre la proteína y la cadena lateral de la glutamina. Tyr-14, ahora Glu-14, junto con la nueva His-52 forman puentes de hidrógeno con el nitrógeno de la glutamina, mientras que Asn-117, estabiliza al oxígeno mediante el mismo topo de interacción. La cuarta mutación en común, Q122E, no interviene directamente con la unión del ligando, pero resulta en el fortalecimiento de un puente de hidrógeno con Ser-92, al ser realizado con el carboxilo del ácido glutámico en vez de la amina de la glutamina original. No es inesperado que Rosetta realice este tipo de mutaciones, ya que está programado tanto para mejorar la unión al ligando (*lig_sum*) como para optimizar en lo posible la estabilidad de la proteína (*bk_tot*). No obstante, es incierto el efecto que esto pudiera tener en la afinidad experimental de la proteína. Cabe hacer notar que, adicionalmente, de manera común para ambas estructuras rediseñadas, la mutación Y14E permite a su vez un cambio de conformación en Gln-143, que se posiciona para formar un puente de unión a través de su grupo amino con el oxígeno de la cadena lateral del ligando.

Por otra parte, la mutación adicional a estas cuatro que mostró el rediseño LIBRE-Q1-38, fue T121S, que tiene el efecto de crear un nuevo puente de hidrógeno entre el oxígeno de la cadena lateral de serina con uno de los oxígenos del carboxilo. Por parte de LIBRE-Q2-22, dos de sus mutaciones únicas S70D y D161T forman puentes de hidrógeno con el nitrógeno del grupo amino principal, la primera siendo habilitada por la segunda, al despejar ésta el impedimento estérico que tendría con ácido aspártico, pero no con treonina. A consecuencia de este cambio, el puente de hidrógeno entre este residuo y el nitrógeno del ligando se debilita respecto al del ácido aspártico original, pero con el beneficio de un puente de hidrógeno adicional con el residuo 70. Finalmente, la última proteína de este rediseño, T195C, está fuera del sitio de unión. Se encuentra en contacto principalmente con Asp-70 y con Ala-90. Siendo un aminoácido neutral, Rosetta probablemente determinó la interacción del grupo sulfhidrilo con el nuevo Asp-70 como más favorable que con el carbono original de Thr-195, sin perturbar la interacción hidrofóbica con Ala-90.

En general, a partir de los resultados de esta estrategia, puede observarse que Rosetta, con pocas restricciones, muestra una fuerte tendencia a maximizar la formación de puentes de hidrógeno para estabilizar el ligando, aún a costa de perder las interacciones hidrofóbicas de “*stacking*” que Tyr-14 y Phe-52 mantienen con el ligando y que se encuentran conservadas entre las proteínas similares a LAO, incluyendo QBP. Rosetta ya ha demostrado que rediseños distintos pueden llegar a estructuras similares, incluso con mayor estabilidad [20], por lo que no es imposible que se puedan encontrar maneras de unir el ligando distintas a las que la naturaleza halló para QBP.

En la figura 7 se presenta el sitio de unión para la estructura más sobresaliente de la estrategia conservativa y con mejor puntuación de manera global para la glutamina, CONSER-Q1-V1-09, que interesantemente presenta una sola mutación: Y14T. Esta mutación se posiciona de manera que el hidroxilo de la treonina forma un puente de hidrógeno con el oxígeno de la cadena lateral del ligando y habilitando al mismo tiempo, como en las mutaciones de la estrategia libre, la reorientación de la cadena lateral de la Gln-143 para formar un puente de hidrógeno con el mismo oxígeno. En la figura puede observarse que la treonina se posiciona a una distancia marginalmente viable (3.63 Å) para formar también un puente de hidrógeno con el nitrógeno del ligando, la cual podría hacerse más significativa mediante un reacomodo de la cadena lateral. Esta sencilla mutación, realizada a cambio de las posibles interacciones que la tirosina pudiera tener con el ligando, hace posible dos o hasta tres posibles puentes de hidrógeno adicionales con el ligando y es vista favorablemente por todas las funciones de evaluación utilizadas en este trabajo.

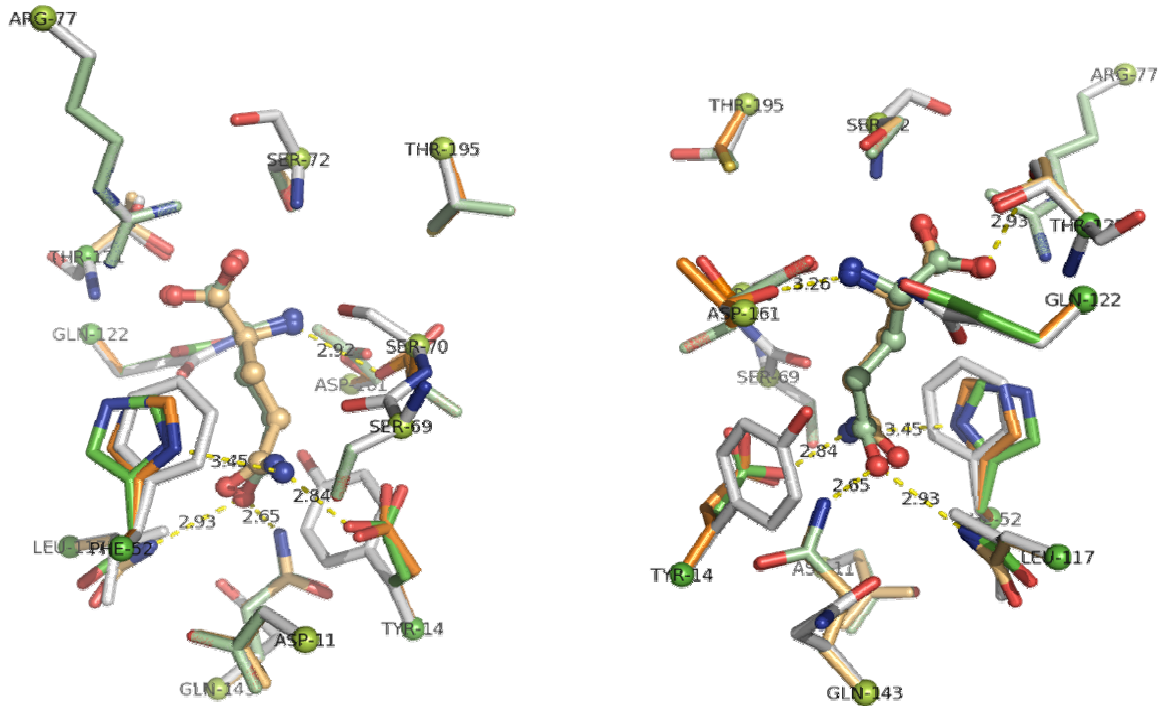


Figura 6: Representación del sitio de unión a glutamina desde dos perspectivas para los dos rediseños mejor clasificados por puntuación consenso y realizados mediante la estrategia libre. En colores verdes se muestran las cadenas laterales de los residuos para el rediseño LIBRE-Q1-38 (Y14E F52H L117N T121S Q122E) y en tonos anaranjados para LIBRE-Q2-22 (Y14E F52H S70D L117N Q122E D161T T195C). Los tonos claros corresponden a los residuos que no fueron mutados y al ligando, mientras que las mutaciones realizadas se presentan en tonos más intensos. Las estructuras en blanco muestran la estructura base para el rediseño con el conformero 1 (Q1). Las esferas representan los carbonos alfa de los residuos o los átomos del ligando. Los rótulos se refieren al aminoácido original de la estructura. En líneas punteadas amarillas se muestran los nuevos puentes de hidrógeno en el rediseño.

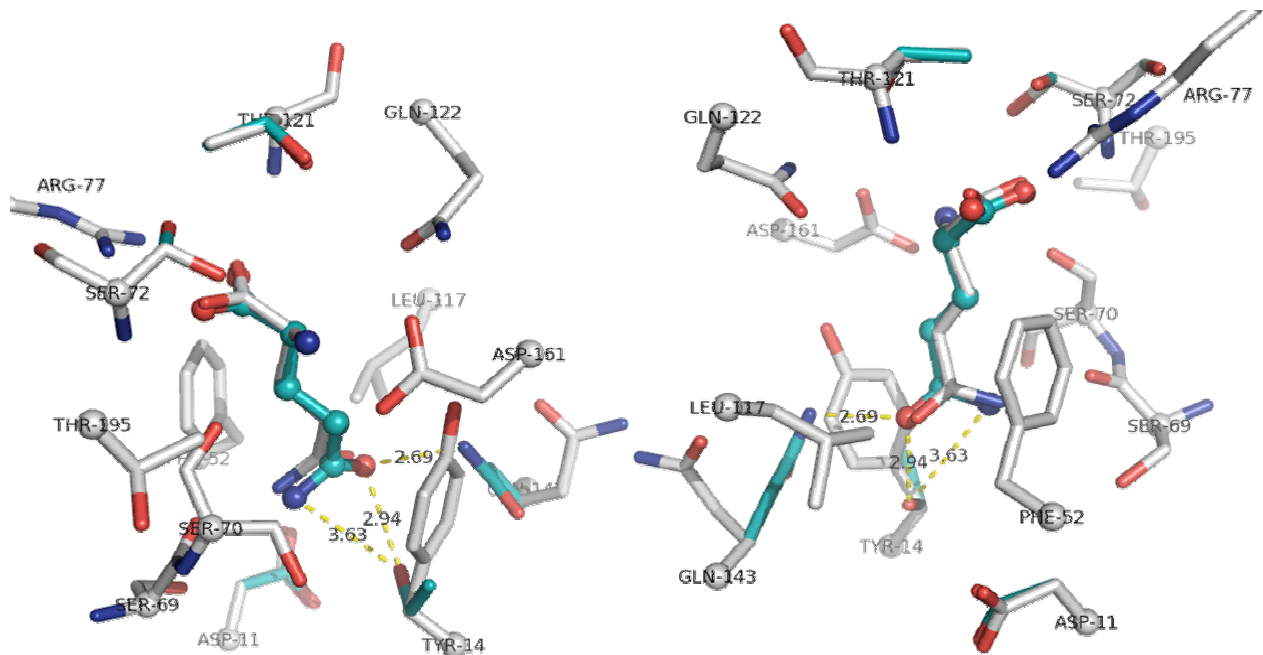


Figura 7: Representación del sitio de unión a glutamina desde dos perspectivas para el rediseño mejor clasificado por puntuación consenso realizado mediante la estrategia conservativa, CONSER-Q1-V1-09 (Y14T). Los residuos en blanco representan la estructura base para el rediseño con el conformero 1 (Q1), mientras que aquéllos en cian corresponden al rediseño. Las esferas representan los carbonos alfa de los residuos. Los rótulos se refieren al aminoácido original de la estructura. Las líneas punteadas amarillas muestran los posibles nuevos puentes de hidrógeno generados en el rediseño.

Fuera de este resultado, que, como se mencionó anteriormente, fue único en cuanto a la mutación realizada; los resultados de la estrategia conservativa se asemejan, como se esperaba, a los resultados de la estrategia libre en cuanto a las mutaciones preferidas, aunque presentando un número menor de ellas en el rediseño. Otros resultados importantes de la estrategia conservativa, incluso a nivel global, presentan las mutaciones Y14E, F52H y L117N. Estos modelos no se muestran explícitamente ya que sus mutaciones se encuentran representadas entre los resultados de la estrategia libre, en la figura 5, y se discuten con ellos.

La figura 8 muestra los rediseños más notables realizados bajo la estrategia guiada, GUIADA-Q2-V2-22 y GUIADA-Q1-V1-09 con las mutaciones L117K, Q143G, Q160D para la primera y Y14F, L117T, Q122T, Q160T, D161E para la segunda. La mutación L117K crea un puente de hidrógeno a partir de Leu-117, que de acuerdo al análisis por LPC, representaba la principal interacción desestabilizante de los complejos LAO-Gln generado por *docking*. La mutación Q143G elimina la cadena lateral de Gln-143, que de otra manera presentaría un impedimento estérico para la mutación anterior. Q160D, por su parte, no está involucrada en el sitio de unión. Como puede verse en esta figura, la conformación del ácido aspártico apunta su cadena lateral hacia fuera del sitio de unión e interacciona con la Tyr-147 (no mostrada), formando un puente de hidrógeno con su hidroxilo. Esto probablemente fue juzgado por Rosetta como estabilizante para la proteína en general, aunque, nuevamente, su efecto sobre la afinidad del ligando no puede precisarse.

El rediseño GUIADA-Q1-V1-09 resulta peculiar en que ninguna de sus mutaciones genera un nuevo puente de hidrógeno con el ligando. Asimismo, es la única estructura que presenta un decremento significativo en el parámetro bk_{tot} de estabilidad. Este efecto se le atribuye a la mutación Y14F, que coloca a la fenilalanina de manera retraída hacia fuera del sitio respecto a la tirosina original, debido a la posición más cercana del ligando en esa estructura. Este cambio tiene como consecuencia un choque estérico con la Ala-164, que se encuentra por detrás. Al estar fija la cadena principal y con el residuo sin poderse acomodar ni mutar, esto genera la reducción en estabilidad observada en el bk_{tot} . En la realidad, el ligando mismo podría reacomodarse ligeramente para aliviar el impedimento estérico, o bien, la cadena principal de la proteína podría hacerlo. No obstante, no resulta posible descartar la posibilidad de que el plegamiento de la enzima se vea afectado. Por otra parte, L117T, genera un puente de hidrógeno con Gln-143, presumiblemente estabilizando la estructura, pero aparentemente sin influir en la unión del ligando. La mutación Q122T tampoco genera un contacto favorable con el ligando, sino que forma un puente

de hidrógeno con Ser-120, que no muta pero que cambia su conformación para permitir la formación de un puente con uno de los oxígenos del carboxilo del ligando. Las mutaciones Q160T y D161E igualmente no involucran la formación de una interacción con el ligando, sino la formación de puentes con el oxígeno de la Gln-147 y el nitrógeno de la Val-143, respectivamente. Dado que el rediseño no parece haber afectado la unión directa del ligando, resulta curiosa la evaluación tan favorable por parte de la función BAPPL.

Las representaciones de las dos estructuras mejor clasificadas por consenso para el rediseño con valina se muestran en la figura 9, VAL-V1-09, de cinco mutaciones (Y14E, F17H, Q143L, Q160D, A164I) y VAL-V1-47 de diez (Y14I, F17H, S70D, Y88H, Q122T, Q143V, Q160D, D161T, A164I, T195C). Para VAL-V1-09, únicamente las mutaciones Q143L y A164I pretenden contribuir a la estabilización de la cadena lateral alifática de valina, mientras que el resto busca crear interacciones estabilizantes internas de la proteína. VAL-V1-47, como LIBRE-Q2-22, presenta las mutaciones S70D y Q122T que, como ya se vio, forman puentes de hidrógeno con el nitrógeno del ligando. F17H y Y88H, por su parte, posibilitan igualmente la mutación S70D, al estabilizar con sus nitrógenos el otro oxígeno del carboxilo mediante puentes de hidrógeno. Otras de sus mutaciones, como Y14I, Q143V, A164I, intentan construir un ambiente hidrofóbico en torno a la cadena lateral del ligando.

Un efecto no tan evidente en esta figura pero considerable para ambos modelos, es la presencia de espacio vacío en el sitio de unión, causado por el hecho de que valina es un aminoácido más pequeño que los originales de LAO. Aunque dicho espacio vacío podría reducirse concebiblemente (por ejemplo, mediante la utilización de residuos más voluminosos), Rosetta decidió, en cambio, construir puentes de hidrógeno internos. Dado que el proceso de unión del ligando en las PBP no sólo implica la interacción ligando-proteína, sino también la estabilización del estado cerrado [64], resulta incierto si el rediseño, que sólo puede tomar en cuenta las interacciones en estado cerrado, pueda llegar a generar un estado cerrado estable, a pesar del espacio vacío en la interfase con el ligando.

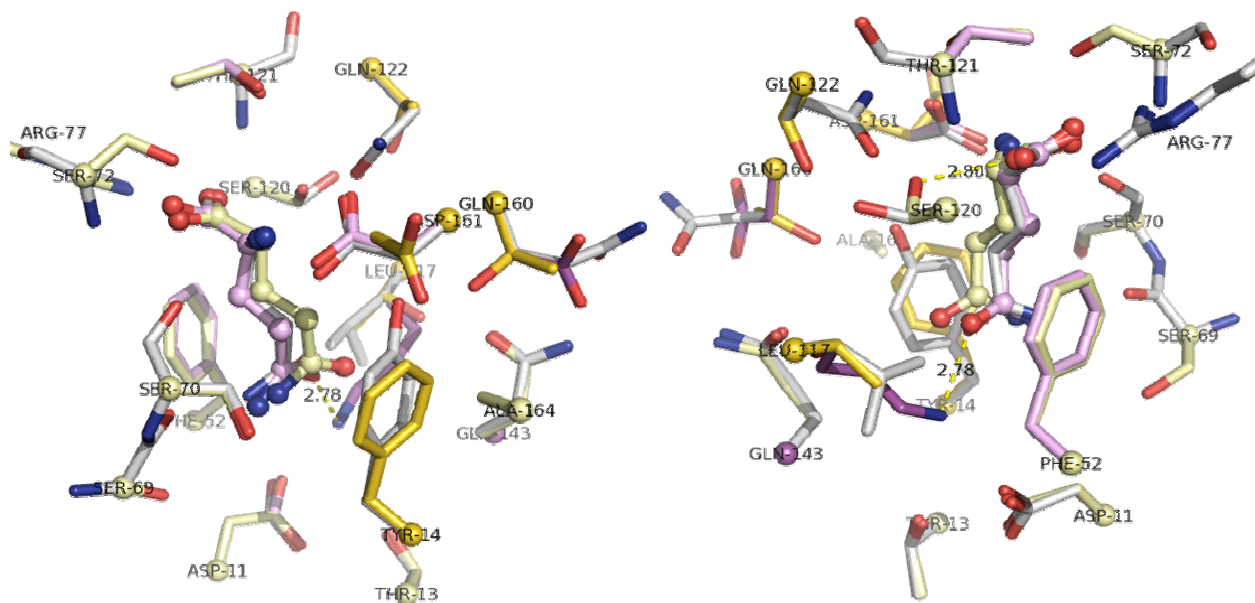


Figura 8: Representación del sitio de unión a glutamina desde dos perspectivas para los dos rediseños realizados mediante la estrategia guiada mejor clasificados por consenso. En colores rosas se muestran las cadenas laterales de los residuos para el rediseño GUIADA-Q2-V2-22 (L117K Q143G Q160D) y en amarillos para GUIADA-Q1-V1-09 (Y14F L117T Q122T Q160T D161E). Los tonos claros corresponden a los residuos que no fueron mutados y al ligando, mientras que las mutaciones realizadas se presentan en tonos más intensos. Las estructuras en blanco muestran la estructura base para el rediseño con el confórmero 1 (Q1). Las esferas representan los carbonos alfa de los residuos. Los rótulos se refieren al aminoácido original de la estructura. Las líneas punteadas amarillas muestran los nuevos puentes de hidrógeno causados por el rediseño.

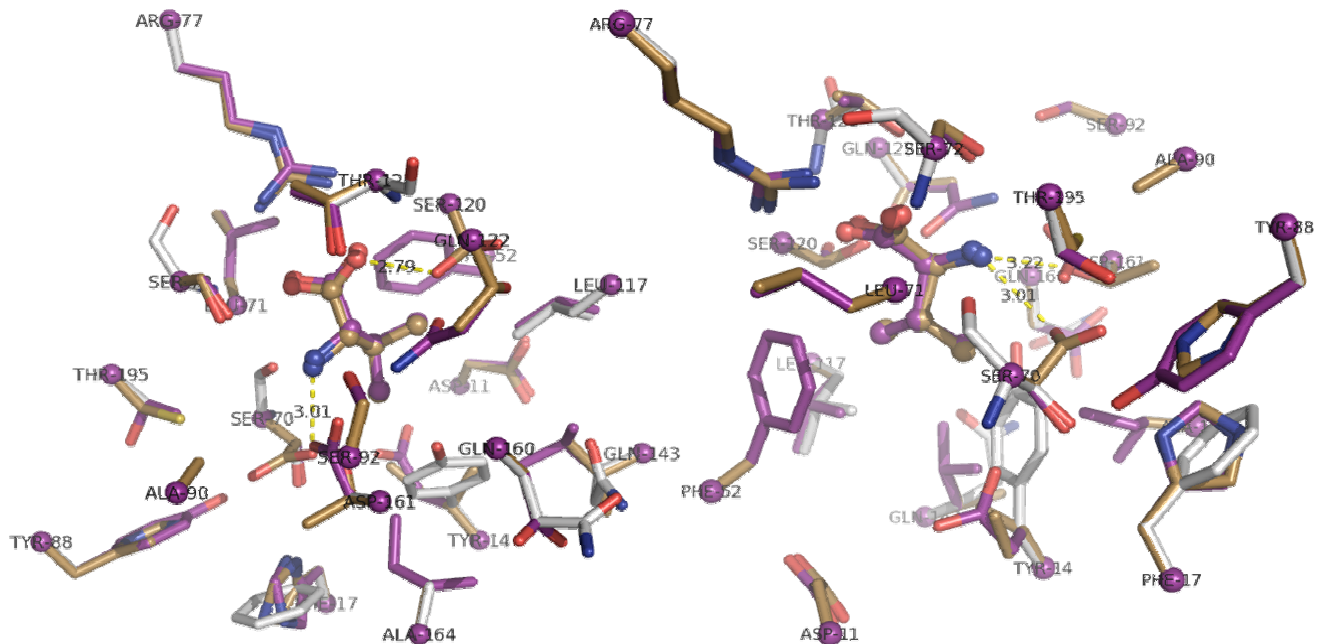


Figura 9: Representación del sitio de unión a valina desde dos perspectivas para los dos rediseños mejor clasificados por consenso. En violeta se presentan las cadenas laterales de los residuos para el rediseño VAL-V1-09 (Y14E F17H Q143L Q160D A164I) y en café para VAL-V1-47 (Y14I F17H S70D Y88H Q122T Q143V Q160D D161T A164I T195C). Las estructuras en blanco muestran la estructura base para el rediseño. Las esferas representan los carbonos alfa de los residuos y a los átomos de la valina. Los rótulos se refieren al aminoácido original de la estructura. Las líneas punteadas amarillas representan los nuevos puentes de hidrógeno generados por el rediseño.

CONCLUSIONES Y PERSPECTIVAS

Los resultados expuestos mostraron que el programa Rosetta puede ser bastante flexible respecto a las restricciones impuestas y llegar a generar resultados aparentemente viables. No obstante, también se hicieron evidentes ciertas tendencias del programa en sus rediseños. Una de ellas es la de favorecer los puentes de hidrógeno como interacción de estabilización, a juzgar por la consistencia con que Rosetta decidía mutar los aminoácidos aromáticos Tyr-14 y Phe-52, los cuales intervienen en la estabilización de los carbonos hidrofóbicos de las cadenas laterales de los ligandos. Las posiciones se encuentran relativamente conservadas en las PBP's similares a LAO, incluyendo QBP, que une glutamina nativamente. Otra observación que puede hacerse es la propensión de Rosetta a generar mutaciones de estabilidad, en algunos casos aun a costa de sacrificar enlaces potenciales con el ligando. Es sabido que antes de que Rosetta cuidara de la energética general de la proteína en modo ligando, las proteínas resultantes solían tener problemas de plegamiento; por ello, posteriormente se permitió que, al mismo tiempo de tratar de unir al ligando, Rosetta optimizara las interacciones internas de la proteína (Dr. Jens Meiler, comunicación personal). No obstante, en algunas ocasiones las elecciones parecen cuestionables, ya que, a pesar de tener la posibilidad de generar interacciones propicias con el ligando, Rosetta optó por favorecer la estabilización estructural global. Quizá haría falta un ajuste en su modelo, o bien, hacerlo ajustable.

Los resultados de este trabajo también ponen en manifiesto algunas de las limitaciones de Rosetta. Una de ellas, evidente desde un principio, es la incapacidad de considerar flexibilidad en los ligandos. Aunque se intentó aminorar el impacto de esta restricción mediante la preselección de dos conformaciones, resulta inevitable que exista una reducción grande del espacio de búsqueda disponible para encontrar mínimos de energía y por tanto, cierta incertidumbre respecto a la relevancia de la evaluación. Aunque la introducción de flexibilidad en el ligando en Rosetta implicaría un aumento en complejidad que impactaría en la eficiencia computacional, es probable que la calidad de resultados se beneficiaría.

Otra limitación importante del programa, aunque es común para muchas herramientas bioinformáticas, es la omisión de moléculas de agua estructural en las evaluaciones. Esto provoca que Rosetta, rediseñe las interacciones para suplir la pérdida de posibles contactos del ligando y la proteína con las moléculas de agua; no obstante, no es posible predecir que ocurriría una vez que el agua entrara en contacto con la superficie de unión rediseñada, ni cómo influiría ésta en la afinidad al ligando.

Adicionalmente, es importante notar otra restricción con un posible impacto en los resultados, en este caso referente al diseño mismo de experimento, la cual es la consideración en todo caso de la unión de la proteína en su estado cerrado. No se espera un impacto negativo serio, ya que la unión en el estado cerrado es, últimamente, la más relevante para la afinidad real. Por otro lado, esto tiene como consecuencia la posible omisión de criterios que, en el estado abierto, podrían dificultar la interacción y por lo tanto, tener un impacto negativo no previsto en la afinidad. Uno de ellos podría ser la electrostática, que puede tener importancia en la unión inicial del ligando al sitio [65]. El impacto de esto, no obstante, se espera que sea menor dada la consideración en el diseño de utilizar aminoácidos no cargados.

Finalmente, queda mencionar otra limitación en el procedimiento, también consecuencia de la naturaleza de la unión en LAO y las PBPs en general, que es la influencia de la dinámica entre estados abierto y cerrado. Como ya ha sido mencionado anteriormente, la afinidad de unión en LAO no sólo depende de las interacciones directas con el ligando, sino también de la estabilidad relativa del estado cerrado unido respecto al estado abierto[60]. Este efecto lamentablemente no puede ser considerado bajo el presente modelo, aunque se buscó minimizar un posible impacto indirecto sobre el mismo al restringir las mutaciones al sitio de unión.

A pesar de todo lo anterior, Rosetta fue capaz de generar resultados evaluados favorablemente respecto a la estructura del complejo base, tanto por Rosetta como por las demás funciones. La clasificación por consenso, por su parte, también pareció ser una buena herramienta, señalando como importantes a resultados que, sin destacar particularmente bajo ninguna de las funciones, muestran evaluaciones consistentemente favorables entre todas las usadas. Sin embargo, la utilidad real de esta metodología no podrá ser evaluada hasta tener una medición de la afinidad experimental de los resultados. Las diferencias relativas en la evaluación de los resultados por cada función, sugieren que cada una tiene criterios distintos potencialmente complementarios, lo cual fortalece a la selección por consenso.

En suma, con base a los criterios del presente trabajo, se sugieren cinco rediseños de glutamina como candidatos interesantes a una corroboración experimental de afinidad: dos provenientes de la estrategia libre (LIBRE-Q1-38 y LIBRE-Q2-22), uno para la conservativa (CONSER-Q1-V1-09) y dos para la guiada (GUIADA-Q2-V2-22 y GUIADA-Q1-V1-09). Asimismo, se proponen dos rediseños sobresalientes para valina, VAL-V1-09 y VAL-V1-47.

Una perspectiva y trabajo futuro obvios resultantes de este estudio es la construcción de las mutantes relevantes y la medición experimental de sus afinidades por glutamina y valina, lo cual

servirá para determinar la efectividad, por un lado, de la metodología propuesta para la selección de candidatos y, por otro, la capacidad del modo diseño de ligando de Rosetta de cumplir con su función y generar estructuras viables para el modelo de LAO utilizado. Los resultados de esto, independientemente si son favorables o no, proveerán información valiosa acerca tanto del proceso de rediseño y selección, como del mecanismo de unión al ligando en las interesantes proteínas que son las PBPs.

ABREVIATURAS

CE-MC	<i>Combinatorial Extension – Monte Carlo</i> [1]
CE	<i>Combinatorial Extension</i> [2]
HBP	Proteína de unión a histidina (<i>Histidine-Binding Protein</i>)
LAO	Proteína de unión a lisina, arginina y ornitina (<i>Lysine-, Arginine-, Ornithine-Binding Protein</i>)
LIV	Proteína de unión a leucina, isoleucina y valina (<i>Leucine-, Isoleucine-, Valine-Binding Protein</i>)
LPC	<i>Ligand-Protein Contacts</i>
PBP	Proteína periplásmica de unión (<i>Periplasmic binding protein</i>)
PDB	Protein Data Bank
QBP	Proteína de unión a glutamina (<i>Glutamine-Binding Protein</i>)
RMSD	<i>Root Mean Square Deviation</i>
<i>tt</i> ZBP	Proteína de unión a glutamato/glutamina de <i>Thermus thermophilus</i>

Apéndice I – Opciones de Rosetta

A continuación se explican las funciones de cada una de las opciones de línea de comando utilizadas en este trabajo, clasificadas por tipo:

Generales

-dock : Habilita el modo de *docking* de Rosetta, haciendo que opere cualquier opción relevante para *docking*, las cuales se especifican como tales. Esta opción es mutuamente excluyente con *-design*.

-design : Activa la modalidad de diseño de Rosetta, permitiendo la aplicación de las opciones especificadas como de diseño. Esta opción es mutuamente excluyente con *-dock*.

-ligand : Especifica el uso del modo ligando para Rosetta. Esta opción no es excluyente con los modos diseño y *dock*. En modo diseño, *-ligand* da la instrucción a Rosetta de usar su evaluación para ligando para rediseñar la proteína en base, principalmente, a la interacción con el ligando.

-s archivo.pdb : Especifica el archivo PDB de entrada para Rosetta.

-inform : Aumenta la cantidad de información presentada por el programa durante la corrida.

-timer : No funcional. Supuestamente habilita la escritura del tiempo por “señuelo” (resultado).

Docking

Las siguientes opciones sólo entran en funcionamiento cuando se tiene el modo *dock* habilitado, salvo por las indicadas con *, que también pueden ser utilizados

-dock_mcm : Indica a Rosetta realizar una minimización Monte Carlo a la estructura generada por el algoritmo.

*-dock_pert x y z** : Activa el modo de perturbación inicial. Este modo genera una perturbación perpendicular de $x \text{ \AA}$, una paralela de $y \text{ \AA}$ y una rotacional de z° . Esto tiene como consecuencia una exploración del espacio inmediato en torno a la posición original del ligando. En modo *docking* esto facilita el encontrar distintos mínimos en torno a la posición de referencia y en modo diseño permite al algoritmo hallar interacciones favorables que de otra manera no sería posible

-nstruct : Especifica el número de *dockings* a realizar.

Diseño

Funciones posibles bajo el modo diseño de ligando (*-design*)

- fixbb* : Establece el rediseño de la secuencia sin alterar la estructura de la cadena principal.
- design_dock* : Permite el uso de la opción *-dock_pert* en modo de diseño.
- favor_native_residue x* : Indica a Rosetta favorecer los aminoácidos nativos por x “kcal/mol”
- favor_nonpolar x* : Favorece los aminoácidos no polares por x “kcal/mol”.
- resfile* : Especifica el archivo de rediseño, en el cual se determinan las mutaciones posibles.
- ndruns x* : Número de rediseños requeridos.
- ligand_filter_score x* : Valor mínimo de *lig_sum* del rediseño para ser aceptado.
- ligand_filter_count x* : Número de intentos de rediseño tras los cuales el valor de *ligand_filter_score* es modificado dinámicamente al 90% del mejor valor visto hasta entonces.

Empacamiento

Estas opciones son de aplicación general, pero se detallan aparte ya que se refieren específicamente al comportamiento de Rosetta respecto al manejo de estructuras y rotámeros.

- use_input_sc* : Indica a Rosetta que considere los rotámeros nativos de la estructura en el proceso de reempacamiento
- find_disulf* : Ordena a Rosetta hacer una búsqueda de puentes disulfuro presentes en la estructura.
- norepack_disulf* : Hace que Rosetta no intente reempacar los puentes disulfuro que halle.
- ex1* : Indica a Rosetta utilizar sub-rotámeros adicionales del enlace CA-CB (*chi1*)
- ex2* : Permite la utilización de sub-rotámeros extra para el enlace *chi2*.

Apéndice II – Resfiles

Aquí se presentan los archivos *resfile* utilizados, los cuales determinan qué residuos de la estructura se permite mutar y, si se desea, específicamente por cuáles aminoácidos. Asimismo, se presenta una breve guía de la terminología utilizada en este archivo. Aunque los archivos en sí son lineales, se presentan aquí en columnas por ahorro de espacio.

Columna 2: Cadena
Columna 4-7: Número de residuo secuencial
Columna 9-12: Número de residuo en el PDB
Columna 14-18: ID (descrito a continuación)
Columna 21-40: amino acids to be used

NATAA => usar el aminoácido nativo
ALLAA => todos los aminoácidos
NATRO => usar aminoácido y rotámero nativos.
PIKAA => seleccionar aminoácidos específicos
POLAR => aminoácidos polares.
APOLA => aminoácidos apolares.

LaoGln-redesign.resfile (usado para estrategias libre y conservativa):

```
start      _ 25 25 NATRO      _ 50 50 NATAA
_ 1 1 NATRO   _ 26 26 NATRO   _ 51 51 NATAA
_ 2 2 NATRO   _ 27 27 NATRO   _ 52 52 ALLAA
_ 3 3 NATRO   _ 28 28 NATRO   _ 53 53 NATAA
_ 4 4 NATRO   _ 29 29 NATRO   _ 54 54 NATRO
_ 5 5 NATRO   _ 30 30 NATAA   _ 55 55 NATAA
_ 6 6 NATRO   _ 31 31 NATRO   _ 56 56 NATAA
_ 7 7 NATRO   _ 32 32 NATRO   _ 57 57 NATRO
_ 8 8 NATRO   _ 33 33 NATRO   _ 58 58 NATRO
_ 9 9 NATAA   _ 34 34 NATRO   _ 59 59 NATRO
_ 10 10 NATAA _ 35 35 NATRO   _ 60 60 NATRO
_ 11 11 ALLAA _ 36 36 NATRO   _ 61 61 NATRO
_ 12 12 NATAA _ 37 37 NATRO   _ 62 62 NATRO
_ 13 13 NATAA _ 38 38 NATRO   _ 63 63 NATRO
_ 14 14 ALLAA _ 39 39 NATRO   _ 64 64 NATRO
_ 15 15 NATAA _ 40 40 NATRO   _ 65 65 NATRO
_ 16 16 NATRO _ 41 41 NATRO   _ 66 66 NATRO
_ 17 17 NATAA _ 42 42 NATRO   _ 67 67 NATAA
_ 18 18 NATAA _ 43 43 NATRO   _ 68 68 NATAA
_ 19 19 NATRO _ 44 44 NATRO   _ 69 69 ALLAA
_ 20 20 NATRO _ 45 45 NATRO   _ 70 70 ALLAA
_ 21 21 NATRO _ 46 46 NATRO   _ 71 71 ALLAA
_ 22 22 NATRO _ 47 47 NATRO   _ 72 72 ALLAA
_ 23 23 NATRO _ 48 48 NATRO   _ 73 73 NATAA
_ 24 24 NATRO _ 49 49 NATRO   _ 74 74 NATAA
```

_ 75	75 NATRO	_ 130	130 NATRO	_ 185	185 NATRO
_ 76	76 NATAA	_ 131	131 NATRO	_ 186	186 NATRO
_ 77	77 ALLAA	_ 132	132 NATRO	_ 187	187 NATRO
_ 78	78 NATAA	_ 133	133 NATRO	_ 188	188 NATRO
_ 79	79 NATRO	_ 134	134 NATRO	_ 189	189 NATRO
_ 80	80 NATRO	_ 135	135 NATRO	_ 190	190 NATAA
_ 81	81 NATRO	_ 136	136 NATRO	_ 191	191 NATAA
_ 82	82 NATRO	_ 137	137 NATRO	_ 192	192 NATAA
_ 83	83 NATAA	_ 138	138 NATRO	_ 193	193 NATAA
_ 84	84 NATRO	_ 139	139 NATRO	_ 194	194 NATAA
_ 85	85 NATRO	_ 140	140 NATAA	_ 195	195 ALLAA
_ 86	86 NATRO	_ 141	141 NATAA	_ 196	196 NATAA
_ 87	87 NATRO	_ 142	142 NATAA	_ 197	197 NATRO
_ 88	88 NATAA	_ 143	143 NATAA	_ 198	198 NATRO
_ 89	89 NATAA	_ 144	144 NATRO	_ 199	199 NATRO
_ 90	90 NATAA	_ 145	145 NATRO	_ 200	200 NATRO
_ 91	91 NATAA	_ 146	146 NATAA	_ 201	201 NATRO
_ 92	92 NATAA	_ 147	147 NATAA	_ 202	202 NATRO
_ 93	93 NATAA	_ 148	148 NATRO	_ 203	203 NATRO
_ 94	94 NATRO	_ 149	149 NATRO	_ 204	204 NATRO
_ 95	95 NATRO	_ 150	150 NATRO	_ 205	205 NATRO
_ 96	96 NATRO	_ 151	151 NATRO	_ 206	206 NATRO
_ 97	97 NATRO	_ 152	152 NATRO	_ 207	207 NATRO
_ 98	98 NATRO	_ 153	153 NATRO	_ 208	208 NATRO
_ 99	99 NATRO	_ 154	154 NATRO	_ 209	209 NATRO
_ 100	100 NATRO	_ 155	155 NATRO	_ 210	210 NATRO
_ 101	101 NATRO	_ 156	156 NATRO	_ 211	211 NATRO
_ 102	102 NATRO	_ 157	157 NATRO	_ 212	212 NATRO
_ 103	103 NATRO	_ 158	158 NATRO	_ 213	213 NATRO
_ 104	104 NATRO	_ 159	159 NATAA	_ 214	214 NATRO
_ 105	105 NATRO	_ 160	160 NATAA	_ 215	215 NATRO
_ 106	106 NATRO	_ 161	161 ALLAA	_ 216	216 NATRO
_ 107	107 NATRO	_ 162	162 NATAA	_ 217	217 NATRO
_ 108	108 NATRO	_ 163	163 NATAA	_ 218	218 NATRO
_ 109	109 NATRO	_ 164	164 NATAA	_ 219	219 NATRO
_ 110	110 NATRO	_ 165	165 NATAA	_ 220	220 NATRO
_ 111	111 NATRO	_ 166	166 NATRO	_ 221	221 NATRO
_ 112	112 NATRO	_ 167	167 NATRO	_ 222	222 NATRO
_ 113	113 NATRO	_ 168	168 NATRO	_ 223	223 NATRO
_ 114	114 NATRO	_ 169	169 NATRO	_ 224	224 NATRO
_ 115	115 NATRO	_ 170	170 NATRO	_ 225	225 NATRO
_ 116	116 NATAA	_ 171	171 NATRO	_ 226	226 NATRO
_ 117	117 ALLAA	_ 172	172 NATRO	_ 227	227 NATRO
_ 118	118 NATAA	_ 173	173 NATRO	_ 228	228 NATRO
_ 119	119 NATAA	_ 174	174 NATRO	_ 229	229 NATRO
_ 120	120 ALLAA	_ 175	175 NATRO	_ 230	230 NATRO
_ 121	121 ALLAA	_ 176	176 NATRO	_ 231	231 NATRO
_ 122	122 ALLAA	_ 177	177 NATRO	_ 232	232 NATRO
_ 123	123 NATAA	_ 178	178 NATRO	_ 233	233 NATRO
_ 124	124 NATAA	_ 179	179 NATRO	_ 234	234 NATRO
_ 125	125 NATRO	_ 180	180 NATRO	_ 235	235 NATRO
_ 126	126 NATRO	_ 181	181 NATRO	_ 236	236 NATRO
_ 127	127 NATRO	_ 182	182 NATRO	_ 237	237 NATRO
_ 128	128 NATRO	_ 183	183 NATRO	_ 238	238 NATRO
_ 129	129 NATRO	_ 184	184 NATRO		

LaoGln-ratdesign.resfile (usado para estrategia guiada):

```
start
- 1 1 NATRO
- 2 2 NATRO
- 3 3 NATRO
- 4 4 NATRO
- 5 5 NATRO
- 6 6 NATRO
- 7 7 NATRO
- 8 8 NATRO
- 9 9 NATAA
- 10 10 NATAA
- 11 11 PIKAA DENQ
- 12 12 NATAA
- 13 13 NATAA
- 14 14 PIKAA YF
- 15 15 NATAA
- 16 16 NATRO
- 17 17 NATAA
- 18 18 NATAA
- 19 19 NATRO
- 20 20 NATRO
- 21 21 NATRO
- 22 22 NATRO
- 23 23 NATRO
- 24 24 NATRO
- 25 25 NATRO
- 26 26 NATRO
- 27 27 NATRO
- 28 28 NATRO
- 29 29 NATRO
- 30 30 NATAA
- 31 31 NATRO
- 32 32 NATRO
- 33 33 NATRO
- 34 34 NATRO
- 35 35 NATRO
- 36 36 NATRO
- 37 37 NATRO
- 38 38 NATRO
- 39 39 NATRO
- 40 40 NATRO
- 41 41 NATRO
- 42 42 NATRO
- 43 43 NATRO
- 44 44 NATRO
- 45 45 NATRO
- 46 46 NATRO
- 47 47 NATRO
- 48 48 NATRO
- 49 49 NATRO
- 50 50 NATAA
- 51 51 NATAA
- 52 52 APOLA
- 53 53 NATAA
- 54 54 NATRO
- 55 55 NATAA
- 56 56 NATAA
- 57 57 NATRO
- 58 58 NATRO
- 59 59 NATRO
- 60 60 NATRO
- 61 61 NATRO
- 62 62 NATRO
- 63 63 NATRO
- 64 64 NATRO
- 65 65 NATRO
- 66 66 NATRO
- 67 67 NATAA
- 68 68 NATAA
- 69 69 ALLAA
- 70 70 ALLAA
- 71 71 ALLAA
- 72 72 POLAR
- 73 73 NATAA
- 74 74 NATAA
- 75 75 NATRO
- 76 76 NATAA
- 77 77 NATAA
- 78 78 NATAA
- 79 79 NATRO
- 80 80 NATRO
- 81 81 NATRO
- 82 82 NATRO
- 83 83 NATAA
- 84 84 NATRO
- 85 85 NATRO
- 86 86 NATRO
- 87 87 NATRO
- 88 88 NATAA
- 89 89 NATAA
- 90 90 NATAA
- 91 91 NATAA
- 92 92 NATAA
- 93 93 NATAA
- 94 94 NATRO
- 95 95 NATRO
- 96 96 NATRO
- 97 97 NATRO
- 98 98 NATRO
- 99 99 NATRO
- 100 100 NATRO
- 101 101 NATRO
- 102 102 NATRO
- 103 103 NATRO
- 104 104 NATRO
- 105 105 NATRO
- 106 106 NATRO
- 107 107 NATRO
- 108 108 NATRO
- 109 109 NATRO
- 110 110 NATRO
- 111 111 NATRO
- 112 112 NATRO
- 113 113 NATRO
- 114 114 NATRO
- 115 115 NATRO
- 116 116 NATAA
- 117 117 POLAR
- 118 118 NATAA
- 119 119 NATAA
- 120 120 ALLAA
- 121 121 ALLAA
- 122 122 ALLAA
- 123 123 NATAA
- 124 124 NATAA
- 125 125 NATRO
- 126 126 NATRO
- 127 127 NATRO
- 128 128 NATRO
- 129 129 NATRO
- 130 130 NATRO
- 131 131 NATRO
- 132 132 NATRO
- 133 133 NATRO
- 134 134 NATRO
- 135 135 NATRO
- 136 136 NATRO
- 137 137 NATRO
- 138 138 NATRO
- 139 139 NATRO
- 140 140 NATAA
- 141 141 NATAA
- 142 142 NATAA
- 143 143 ALLAA
- 144 144 NATAA
- 145 145 NATRO
- 146 146 NATAA
- 147 147 NATAA
- 148 148 NATRO
- 149 149 NATRO
- 150 150 NATRO
- 151 151 NATRO
- 152 152 NATRO
- 153 153 NATRO
- 154 154 NATRO
- 155 155 NATRO
- 156 156 NATRO
- 157 157 NATRO
- 158 158 NATRO
- 159 159 ALLAA
- 160 160 POLAR
- 161 161 POLAR
```

_ 162 162 NATAA	_ 188 188 NATRO	_ 214 214 NATRO
_ 163 163 NATAA	_ 189 189 NATRO	_ 215 215 NATRO
_ 164 164 NATAA	_ 190 190 NATAA	_ 216 216 NATRO
_ 165 165 NATAA	_ 191 191 NATAA	_ 217 217 NATRO
_ 166 166 NATRO	_ 192 192 NATAA	_ 218 218 NATRO
_ 167 167 NATRO	_ 193 193 NATAA	_ 219 219 NATRO
_ 168 168 NATRO	_ 194 194 NATAA	_ 220 220 NATRO
_ 169 169 NATRO	_ 195 195 ALLAA	_ 221 221 NATRO
_ 170 170 NATRO	_ 196 196 NATAA	_ 222 222 NATRO
_ 171 171 NATRO	_ 197 197 NATRO	_ 223 223 NATRO
_ 172 172 NATRO	_ 198 198 NATRO	_ 224 224 NATRO
_ 173 173 NATRO	_ 199 199 NATRO	_ 225 225 NATRO
_ 174 174 NATRO	_ 200 200 NATRO	_ 226 226 NATRO
_ 175 175 NATRO	_ 201 201 NATRO	_ 227 227 NATRO
_ 176 176 NATRO	_ 202 202 NATRO	_ 228 228 NATRO
_ 177 177 NATRO	_ 203 203 NATRO	_ 229 229 NATRO
_ 178 178 NATRO	_ 204 204 NATRO	_ 230 230 NATRO
_ 179 179 NATRO	_ 205 205 NATRO	_ 231 231 NATRO
_ 180 180 NATRO	_ 206 206 NATRO	_ 232 232 NATRO
_ 181 181 NATRO	_ 207 207 NATRO	_ 233 233 NATRO
_ 182 182 NATRO	_ 208 208 NATRO	_ 234 234 NATRO
_ 183 183 NATRO	_ 209 209 NATRO	_ 235 235 NATRO
_ 184 184 NATRO	_ 210 210 NATRO	_ 236 236 NATRO
_ 185 185 NATRO	_ 211 211 NATRO	_ 237 237 NATRO
_ 186 186 NATRO	_ 212 212 NATRO	_ 238 238 NATRO
_ 187 187 NATRO	_ 213 213 NATRO	

LaoVal-design.resfile (usado para el rediseño con valina):

start	_ 26 26 NATRO	_ 52 52 ALLAA
_ 1 1 NATRO	_ 27 27 NATRO	_ 53 53 NATAA
_ 2 2 NATRO	_ 28 28 NATRO	_ 54 54 NATAA
_ 3 3 NATRO	_ 29 29 NATAA	_ 55 55 ALLAA
_ 4 4 NATRO	_ 30 30 ALLAA	_ 56 56 ALLAA
_ 5 5 NATRO	_ 31 31 NATAA	_ 57 57 NATRO
_ 6 6 NATRO	_ 32 32 NATRO	_ 58 58 NATRO
_ 7 7 NATRO	_ 33 33 NATRO	_ 59 59 NATRO
_ 8 8 NATRO	_ 34 34 NATRO	_ 60 60 NATRO
_ 9 9 NATRO	_ 35 35 NATRO	_ 61 61 NATRO
_ 10 10 NATAA	_ 36 36 NATRO	_ 62 62 NATRO
_ 11 11 ALLAA	_ 37 37 NATRO	_ 63 63 NATRO
_ 12 12 NATAA	_ 38 38 NATRO	_ 64 64 NATRO
_ 13 13 NATAA	_ 39 39 NATRO	_ 65 65 NATRO
_ 14 14 ALLAA	_ 40 40 NATRO	_ 66 66 NATRO
_ 15 15 NATAA	_ 41 41 NATRO	_ 67 67 NATAA
_ 16 16 NATAA	_ 42 42 NATRO	_ 68 68 NATAA
_ 17 17 ALLAA	_ 43 43 NATRO	_ 69 69 ALLAA
_ 18 18 NATAA	_ 44 44 NATRO	_ 70 70 ALLAA
_ 19 19 NATRO	_ 45 45 NATRO	_ 71 71 ALLAA
_ 20 20 NATRO	_ 46 46 NATRO	_ 72 72 ALLAA
_ 21 21 NATRO	_ 47 47 NATRO	_ 73 73 ALLAA
_ 22 22 NATRO	_ 48 48 NATRO	_ 74 74 ALLAA
_ 23 23 NATRO	_ 49 49 NATRO	_ 75 75 NATAA
_ 24 24 NATRO	_ 50 50 NATRO	_ 76 76 NATRO
_ 25 25 NATRO	_ 51 51 NATAA	_ 77 77 ALLAA

_ 78	78 NATRO	_ 133	133 NATRO	_ 188	188 NATRO
_ 79	79 NATRO	_ 134	134 NATRO	_ 189	189 NATRO
_ 80	80 NATRO	_ 135	135 NATRO	_ 190	190 NATAA
_ 81	81 NATRO	_ 136	136 NATRO	_ 191	191 ALLAA
_ 82	82 NATRO	_ 137	137 NATRO	_ 192	192 NATAA
_ 83	83 ALLAA	_ 138	138 NATRO	_ 193	193 NATAA
_ 84	84 NATRO	_ 139	139 NATRO	_ 194	194 ALLAA
_ 85	85 NATRO	_ 140	140 NATRO	_ 195	195 ALLAA
_ 86	86 NATRO	_ 141	141 NATRO	_ 196	196 NATRO
_ 87	87 NATRO	_ 142	142 NATAA	_ 197	197 NATRO
_ 88	88 ALLAA	_ 143	143 ALLAA	_ 198	198 NATRO
_ 89	89 NATAA	_ 144	144 NATAA	_ 199	199 NATRO
_ 90	90 ALLAA	_ 145	145 NATRO	_ 200	200 NATRO
_ 91	91 ALLAA	_ 146	146 NATAA	_ 201	201 NATRO
_ 92	92 ALLAA	_ 147	147 NATAA	_ 202	202 NATRO
_ 93	93 NATAA	_ 148	148 NATRO	_ 203	203 NATRO
_ 94	94 NATRO	_ 149	149 NATRO	_ 204	204 NATRO
_ 95	95 NATRO	_ 150	150 NATRO	_ 205	205 NATRO
_ 96	96 NATRO	_ 151	151 NATRO	_ 206	206 NATRO
_ 97	97 NATRO	_ 152	152 NATRO	_ 207	207 NATRO
_ 98	98 NATRO	_ 153	153 NATRO	_ 208	208 NATRO
_ 99	99 NATRO	_ 154	154 NATRO	_ 209	209 NATRO
_ 100	100 NATRO	_ 155	155 NATRO	_ 210	210 NATRO
_ 101	101 NATRO	_ 156	156 NATRO	_ 211	211 NATRO
_ 102	102 NATRO	_ 157	157 NATRO	_ 212	212 NATRO
_ 103	103 NATRO	_ 158	158 NATRO	_ 213	213 NATRO
_ 104	104 NATRO	_ 159	159 NATAA	_ 214	214 NATRO
_ 105	105 NATRO	_ 160	160 ALLAA	_ 215	215 NATRO
_ 106	106 NATRO	_ 161	161 ALLAA	_ 216	216 NATRO
_ 107	107 NATRO	_ 162	162 NATAA	_ 217	217 NATRO
_ 108	108 NATRO	_ 163	163 NATAA	_ 218	218 NATRO
_ 109	109 NATRO	_ 164	164 ALLAA	_ 219	219 NATRO
_ 110	110 NATRO	_ 165	165 NATAA	_ 220	220 NATRO
_ 111	111 NATRO	_ 166	166 NATRO	_ 221	221 NATRO
_ 112	112 NATRO	_ 167	167 NATRO	_ 222	222 NATRO
_ 113	113 NATRO	_ 168	168 NATRO	_ 223	223 NATRO
_ 114	114 NATRO	_ 169	169 NATRO	_ 224	224 NATRO
_ 115	115 NATRO	_ 170	170 NATRO	_ 225	225 NATRO
_ 116	116 NATAA	_ 171	171 NATRO	_ 226	226 NATRO
_ 117	117 ALLAA	_ 172	172 NATRO	_ 227	227 NATRO
_ 118	118 NATAA	_ 173	173 NATRO	_ 228	228 NATRO
_ 119	119 ALLAA	_ 174	174 NATRO	_ 229	229 NATRO
_ 120	120 ALLAA	_ 175	175 NATRO	_ 230	230 NATRO
_ 121	121 ALLAA	_ 176	176 NATRO	_ 231	231 NATRO
_ 122	122 ALLAA	_ 177	177 NATRO	_ 232	232 NATRO
_ 123	123 ALLAA	_ 178	178 NATRO	_ 233	233 NATRO
_ 124	124 NATAA	_ 179	179 NATRO	_ 234	234 NATRO
_ 125	125 NATAA	_ 180	180 NATRO	_ 235	235 NATRO
_ 126	126 NATRO	_ 181	181 NATRO	_ 236	236 NATRO
_ 127	127 NATRO	_ 182	182 NATRO	_ 237	237 NATRO
_ 128	128 NATRO	_ 183	183 NATRO	_ 238	238 NATRO
_ 129	129 NATRO	_ 184	184 NATRO		
_ 130	130 NATRO	_ 185	185 NATRO		
_ 131	131 NATRO	_ 186	186 NATRO		
_ 132	132 NATRO	_ 187	187 NATRO		

REFERENCIAS

1. Guda, C., Lu, S., Scheeff, E.D., Bourne, P.E., Shindyalov, I.N., *CE-MC: a multiple protein structure alignment server*. Nucleic Acids Res., 2004. **1**(32): p. W100-3.
2. Shindyalov, I.N., Bourne, P.E., *Protein structure alignment by incremental combinatorial extension (CE) of the optimal path*. Protein Eng., 1998. **11**(9): p. 739-47.
3. Nelson, D.L., Cox, M.M., *Lehninger Principles of Biochemistry*. 3rd ed. 2000: Worth Publishers Inc.
4. Cagney, G., Amiri, S., Premawaradana, T., Lindo, M., Emili, A., *In silico proteome analysis to facilitate proteomics experiments using mass spectrometry*. Proteome Sci., 2003. **1**(1): p. 5.
5. Berman H.M., W., J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., Bourne, P.E., *The Protein Data Bank*. Nucleic Acids Res., 2000. **28**(1): p. 235-42.
6. Regan, L., *Protein redesign*. Curr Opin Struct Biol, 1999. **9**(4): p. 494-9.
7. Rusconi, S., Scozzafava, A., Mastrolorenzo, A., Supuran, C.T., *New advances in HIV entry inhibitors development*. Curr Drug Targets Infect Disord., 2004. **4**(4): p. 339-55.
8. Levasseur, D.N., Ryan, T.M., Reilly, M.P., McCune, S.L., Asakura, T., Townes, T.M., *A recombinant human hemoglobin with anti-sickling properties greater than fetal hemoglobin*. J Biol Chem, 2004. **279**(26): p. 27518-24.
9. Dunn, M.L., *Protein-Ligand Interactions: General Description*, in *Encyclopedia of Life Sciences*. 2000, John Wiley & Sons, Ltd.: Chichester.
10. Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., Bourne, P.E., *The Protein Data Bank*. Nucleic Acids Res., 2000. **28**(1): p. 235-42.
11. Bourne, P.E., Weissig, H., ed. *Structural Bioinformatics*. 2003, Wiley-Liss, Inc.: Hoboken, NJ, USA. 649 pp.
12. Moulton, J., *A decade of CASP: progress, bottlenecks and prognosis in protein structure prediction*. Curr Opin Struct Biol, 2005. **15**(3): p. 285-9.
13. Bourne, P.E., *CASP and CAFASP experiments and their findings*, in *Structural Bioinformatics*, P.E. Bourne, Weissig, H., Editor. 2003, Wiley-Liss: Hoboken, NJ, USA. p. 501-7.
14. Mendez, R., Leplae, R., Lensink, M.F., Wodak, S.J., *Assessment of CAPRI predictions in rounds 3-5 shows progress in docking procedures*. Proteins., 2005. **60**(2): p. 150-69.
15. Bradley, P., Malmstrom, L., Qian, B., Schonbrun, J., Chivian, D., Kim, D.E., Meiler, J., Misura, K.M., Baker, D., *Free modeling with Rosetta in CASP6*. Proteins, 2005. **61 Suppl 7**: p. 128-34.
16. Daily, M.D., Masica, D., Sivasubramanian, A., Somarouthu, S., Gray, J.J., *CAPRI rounds 3-5 reveal promising successes and future challenges for RosettaDock*. Proteins, 2005. **60**(2): p. 181-6.
17. Rohl, C.A., Strauss, C.E., Misura, K.M., Baker, D., *Protein structure prediction using Rosetta*. Methods Enzymol., 2004. **383**: p. 66-93.
18. Gray, J.J., Moughon, S., Wang, C., Schueler-Furman, O., Kuhlman, B., Rohl, C.A., Baker, D., *Protein-protein docking with simultaneous optimization of rigid-body displacement and side-chain conformations*. J Mol Biol, 2003. **331**(1): p. 281-99.
19. Kuhlman, B., Baker, D., *Native protein sequences are close to optimal for their structures*. Proc Natl Acad Sci U S A, 2000. **97**(19): p. 10383-8.
20. Dantas, G., Kuhlman, B., Callender, D., Wong, M., Baker, D., *A large scale test of computational protein design: folding and stability of nine completely redesigned globular proteins*. J Mol Biol, 2003. **332**(2): p. 449-60.

21. Kuhlman, B., Dantas, G., Ireton, G.C., Varani, G., Stoddard, B.L., Baker, D., *Design of a novel globular protein fold with atomic-level accuracy*. Science, 2003. **302**(5649): p. 1364-8.
22. Looger, L.L., Hellinga, H.W., *Generalized dead-end elimination algorithms make large-scale protein side-chain structure prediction tractable: implications for protein design and structural genomics*. J Mol Biol, 2001. **307**(1): p. 429-45.
23. Marvin, J.S., Hellinga, H.W., *Conversion of a maltose receptor into a zinc biosensor by computational design*. Proc Natl Acad Sci U S A, 2001. **98**(9): p. 4955-60.
24. Dwyer, M.A., Looger, L.L., Hellinga, H.W., *Computational design of a biologically active enzyme*. Science, 2004. **304**(5679): p. 1967-71.
25. Looger, L.L., Dwyer, M.A., Smith, J.J., Hellinga, H.W., *Computational design of receptor and sensor proteins with novel functions*. Nature, 2003. **423**(6936): p. 185-90.
26. Allert, M., Rizk, S.S., Looger, L.L., Hellinga, H.W., *Computational design of receptors for an organophosphate surrogate of the nerve agent soman*. Proc Natl Acad Sci U S A, 2004. **101**(21): p. 7907-12.
27. Halperin, I., Ma, B., Wolfson, H., Nussinov, R., *Principles of docking: An overview of search algorithms and a guide to scoring functions*. Proteins, 2002. **47**(4): p. 409-43.
28. Schneidman-Duhovny, D., Nussinov, R., Wolfson, H.J., *Predicting molecular interactions in silico: II. Protein-protein and protein-drug docking*. Curr Med Chem, 2004. **11**(1): p. 107.
29. Gohlke, H., Klebe, G., *Statistical potentials and scoring functions applied to protein-ligand binding*. Curr Opin Struct Biol, 2001. **11**(2): p. 231-5.
30. Stahl, M., Rarey, M., *Detailed analysis of scoring functions for virtual screening*. J Med Chem, 2001. **44**(7): p. 1035-42.
31. Wang, R., Lu, Y., Wang, S., *Comparative evaluation of 11 scoring functions for molecular docking*. J Med Chem, 2003. **46**(12): p. 2287-303.
32. Ferrarak, P., Gohlke, H., Price, D.J., Klebe, G., Brooks, C.L. 3rd., *Assessing scoring functions for protein-ligand interactions*. J Med Chem, 2004. **47**(12): p. 3032-47.
33. Wang, R., Lu, Y., Fang, X., Wang, S., *An extensive test of 14 scoring functions using the PDBbind refined set of 800 protein-ligand complexes*. J Chem Inf Comput Sci, 2004. **44**(6): p. 2114-25.
34. Marsden, P.M., Puvanendrapillai, D., Mitchell, J.B., Glen, R.C., *Predicting protein-ligand binding affinities: a low scoring game?* Org Biomol Chem, 2004. **2**(22): p. 3267-73.
35. Wang, R., Wang, S., *How does consensus scoring work for virtual library screening? An idealized computer experiment*. J Chem Inf Comput Sci, 2001. **41**(5): p. 1422-6.
36. Tam, R., Saier, M.H. Jr., *Structural, functional, and evolutionary relationships among extracellular solute-binding receptors of bacteria*. Microbiol Rev., 1993. **57**(2): p. 320-46.
37. Fukami-Kobayashi, K., Tateno, Y., Nishikawa, K., *Domain dislocation: a change of core structure in periplasmic binding proteins in their evolutionary history*. J Mol Biol, 1999. **286**(1): p. 279-90.
38. Dwyer, M.A., Hellinga, H.W., *Periplasmic binding proteins: a versatile superfamily for protein engineering*. Curr Opin Struct Biol, 2004. **14**(4): p. 495-504.
39. de Lorimier, R.M., Smith, J.J., Dwyer, M.A., Looger, L.L., Sali, K.M., Paavola, C.D., Rizk, S.S., Sadigov, S., Conrad, D.W., Loew, L., Hellinga, H.W., *Construction of a fluorescent biosensor family*. Protein Sci., 2002. **11**(11): p. 2655-75.
40. Trakhanov, S., Vyas, N.K., Luecke, H., Kristensen, D.M., Ma, J., Quijcho, F.A., *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, 2005. **44**(17): p. 6597-608.

41. Nikaido, K., Ames, G.F., *Purification and characterization of the periplasmic lysine-, arginine-, ornithine-binding protein (LAO) from Salmonella typhimurium.* J Biol Chem, 1992. **267**(29): p. 20706-12.
42. Oh, B.H., Ames, G.F., Kim, S.H., *Structural basis for multiple ligand specificity of the periplasmic lysine-, arginine-, ornithine-binding protein.* J Biol Chem, 1994. **269**(42): p. 26323-30.
43. Oh, B.H., Kang, C.H., De Bondt, H., Kim, S.H., Nikaido, K., Joshi, A.K., Ames, G.F., *The bacterial periplasmic histidine-binding protein. Structure/function analysis of the ligand-binding site and comparison with related proteins.* J Biol Chem, 1994. **269**(6): p. 4135-43.
44. Wolf, A., Shaw, E.W., Oh, B.H., De Bondt, H., Joshi, A.K., Ames, G.F., *Structure/function analysis of the periplasmic histidine-binding protein. Mutations decreasing ligand binding alter the properties of the conformational change and of the closed form.* J Biol Chem, 1995. **270**(27): p. 16097-106.
45. Hsiao, C.D., Sun, Y.J., Rose, J., Wang, B.C., *The crystal structure of glutamine-binding protein from Escherichia coli.* J Mol Biol, 1996. **262**(2): p. 225-42.
46. Sun, Y.J., Rose, J., Wang, B.C., Hsiao, C.D., *The structure of glutamine-binding protein complexed with glutamine at 1.94 Å resolution: comparisons with other amino acid binding proteins.* J Mol Biol, 1998. **278**(1): p. 219-29.
47. Takahashi, H., Inagaki, E., Kuroishi, C., Tahirov, T.H., *Structure of the Thermus thermophilus putative periplasmic glutamate/glutamine-binding protein.* Acta Crystallogr D Biol Crystallogr., 2004. **60**(Pt 10): p. 1846-54.
48. Oh, B.H., Pandit, J., Kang, C.H., Nikaido, K., Gokcen, S., Ames, G.F., Kim, S.H., *Three-dimensional structures of the periplasmic lysine/arginine/ornithine-binding protein with and without a ligand.* J Biol Chem, 1993. **268**(15): p. 11348-55.
49. Kleywegt, G.J., Harris, M.R., Zou, J.Y., Taylor, T.C., Wahlby, A., Jones, T.A., *The Uppsala Electron-Density Server.* Acta Crystallogr D Biol Crystallogr., 2004. **60**(Pt 12 Pt 1): p. 2240-9.
50. Murshudov, G.N., Vagin, A.A., Dodson, E.J., *Refinement of macromolecules structures by the maximum-likelihood method.* Acta Crystallogr D Biol Crystallogr., 1997. **53**(Pt. 3): p. 240-55.
51. Vriend, G., *WHAT IF: A molecular modeling and drug design program.* J. Mol. Graph, 1990. **8**: p. 52-6.
52. Morris, G.M., Goodsell, D.S.; Halliday, R.S.; Huey, R., and W.E. Hart, Belew, R.K., Olson, A.J., *Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function.* J. Comput. Chem., 1998. **19**(14): p. 1639-62.
53. Schuettelkopf, A.W., van Aalten, D.M., *PRODRG: a tool for high-throughput crystallography of protein-ligand complexes.* Acta Crystallogr D Biol Crystallogr., 2004. **60**(Pt. 8): p. 1355-63.
54. Velec, H.F., Gohlke, H., Klebe, G., *DrugScore(CSD)-knowledge-based scoring function derived from small molecule crystal data with superior recognition rate of near-native ligand poses and better affinity prediction.* J Med Chem, 2005. **48**(20): p. 6296-303.
55. Wang, R., Lai, L., Wang, S., *Further development and validation of empirical scoring functions for structure-based binding affinity prediction.* J Comput Aided Mol Des., 2002. **16**(1): p. 11-26.
56. Jain, T., Jayaram. B., *An all atom energy based computational protocol for predicting binding affinities of protein-ligand complexes.* FEBS Lett., 2005. **579**(29): p. 6659-66.
57. Collaborative Computational Project, N., *The CCP4 suite: programs for protein crystallography.* Acta Crystallogr D Biol Crystallogr., 1994. **50**(Pt. 5): p. 760-3.

58. Oda, A., Tsuchida, K., Takakura, T., Yamaotsu, N., Hirono, S., *Comparison of consensus scoring strategies for evaluating computational models of protein-ligand complexes*. J Chem Inf Model, 2006. **46**(1): p. 380-91.
59. Jiang, L., Kuhlman, B., Kortemme, T., Baker, D., *A "solvated rotamer" approach to modeling water-mediated hydrogen bonds at protein-protein interfaces*. Proteins, 2005. **58**(4): p. 893-904.
60. Marvin, J.S., Hellinga, H.W., *Manipulation of ligand binding affinity by exploitation of conformational coupling*. Nat Struct Biol, 2001. **8**(9): p. 795-8.
61. Pang, A., Arinaminpathy, Y., Sansom, M.S. & Biggin, P.C., *Interdomain dynamics and ligand binding: molecular dynamics simulations of glutamine binding protein*. FEBS Lett., 2005. **550**(1-3): p. 168-74.
62. Herman, P., Vecer, J., Scognamiglio, V., Staiano, M., Rossi, M., D'Auria, S., *A recombinant glutamine-binding protein from Escherichia coli: effect of ligand-binding on protein conformational dynamics*. 2004, 2004. **20**(6): p. 1847-54.
63. D'Auria, S., Scire, A., Varriale, A., Scognamiglio, V., Staiano, M., Ausili, A., Marabotti, A., Rossi, M., Tanfani, F., *Binding of glutamine to glutamine-binding protein from Escherichia coli induces changes in protein structure and increases protein stability*. Proteins, 2005. **58**(1): p. 80-7.
64. Telmer, P.G., Shilton, B.H., *Insights into the conformational equilibria of maltose-binding protein by analysis of high affinity mutants*. J Biol Chem, 2003. **278**(36): p. 34555-67.
65. Wade, R.C., Gabdouliline, R.R., Ludemann, S.K., Lounnas, V., *Electrostatic steering and ionic tethering in enzyme-ligand binding: insights from simulations*. Proc Natl Acad Sci U S A, 1998. **95**(11): p. 5942-9.