



# UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

## Maestría y Doctorado en Ciencias Bioquímicas

CODEPENDENCIA EVOLUTIVA ENTRE RESIDUOS COMO DELIMITADORES  
DURANTE EL REDISEÑO DE UNIÓN A LIGANDOS EN LA PROTEÍNA LAO

TESIS

QUE PARA OPTAR POR EL GRADO DE:

Doctor en Ciencias

PRESENTA:

JESÚS AGUSTÍN BANDA VÁZQUEZ

TUTOR PRINCIPAL

Dr. Alejandro Sosa Peinado  
[Facultad de Medicina, UNAM](#)

MIEMBROS DEL COMITÉ TUTOR

Dr. Alfredo Torres Larios  
[Instituto de Fisiología Celular, UNAM](#)  
Dr. León Patricio Martínez Castilla  
[Facultad de Química, UNAM](#)  
[Universidad Autónoma Metropolitana](#)

Ciudad de México. Octubre, 2018



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.

PMDCB/926/2018

**Banda Vázquez Jesús Agustín**  
**Estudiante de Doctorado en Ciencias Bioquímicas**  
**P r e s e n t e**

Los miembros del Subcomité Académico en reunión ordinaria del día 16 de abril del presente año, conocieron su solicitud de asignación de **JURADO DE EXAMEN** para optar por el grado de **DOCTORADO EN CIENCIAS**, con la réplica de la tesis "**Codependencia evolutiva entre residuos como delimitadores durante el rediseño de unión a ligandos en la proteína LAO**", dirigida por el/la Dr(a). **Sosa Peinado Alejandro**.

De su análisis se acordó nombrar el siguiente jurado integrado por los doctores:

PRESIDENTE	Garza-Ramos Martínez Georgina Regina
VOCAL	Peimbert Torres Mariana
VOCAL	González Andrade Martín
VOCAL	Domínguez Dueñas Laura
SECRETARIO	Saab Rincón Gloria

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

**Atentamente**  
**"POR MI RAZA, HABLARÁ EL ESPÍRITU"**  
Cd. Universitaria, Cd. Mx., a 16 de abril de 2018.  
**COORDINADORA**



**Dra. ANA BRÍGIDA CLORINDA ARIAS ÁLVAREZ**

# ÍNDICE

Versión breve de esta tesis ... ..	i
Introducción ... ..	1
Antecedentes ... ..	4
Posibles motivos por los que el diseño de proteínas es un problema sin resolver ... ..	5
Herramientas exploradas durante este proyecto... ..	7
Injerto de cavidad de unión ( <i>Binding Pocket Grafting</i> )... ..	7
SCA: Análisis de Acoplamiento Estadístico... ..	9
Software “Rosetta suite”... ..	10
Acoplamiento molecular ( <i>Docking</i> , del inglés <i>molecular docking</i> ) ... ..	11
Estrategias combinadas ... ..	12
Objetivo General ... ..	13
Hipótesis ... ..	13
Metodología. ....	13
Resultados y Discusión ... ..	17
Comparación entre dos cavidades de unión parecidas pero con diferente especificidad ... ..	17
Enfoque de Binding Pocket Grafting ... ..	17
Expresión de LAOBP_graft y comportamiento en presencia de Gln ... ..	19
Enfoque de Acoplamiento Estadístico ... ..	21
Construcción y caracterización de LAOBP_sca ... ..	24
Un análisis de mutaciones compartidas en ambos enfoques nos revela que un sólo cambio es suficiente para cambiar el perfil de especificidad ... ..	25
LAOBP_L117K une Gln con casi la misma afinidad que QBP ... ..	26
Estructura Cristalográfica ... ..	30
Otras mutantes intentadas durante este proyecto ... ..	31
El sitio 117 siempre fue mutado ... ..	34
Haciendo .un SCA-grafting de únicamente el sector rojo es posible unir Gln con	

una afinidad comparable a Arg ... ..	35
Mutar el sitio 220 tiene un impacto en la estabilidad ... ..	35
La proteína LAOBP_graft es soluble pero su comportamiento con L-Glutamina es incierto ...	36
LAOBP_sca une L-Glutamina, pero necesita tratamiento para plegarse ... ..	36
Una comparación de las mutaciones compartidas por ambos enfoques de diseño mutacionl nos reveló un cambio clave: La mutación L117K a partir de LAOBP permite la unión de L-Glutamina ... ..	38
LAOBP_L117K presenta un perfil de especificidad diferente comparado con LAOBP o QBP ... ..	38
Posibles explicaciones por las que la mutación L117K impacta directamente en la unión de L-Glutamina ... ..	38
Comparación con otros trabajos ... ..	39
Conclusiones ... ..	40
Perspectivas ... ..	41
Agradecimientos ... ..	47
Bibliografía ... ..	48
Apéndice ... ..	52

## **Dedicatoria:**

*A tí, lector, quien tal vez estés leyendo esto por curiosidad y no por error u obligación, o en un afán de citar a algún incauto que se puso a mutar proteínas para tu propio trabajo escrito.*

*A tí en especial, si tienes una idea radical y sigues adelante con ella, nutriéndola cada día, aunque siempre te encuentres con quien diga que tu trabajo no es relevante.*

*A tí, quien no se hace pendejo en su proyecto.*

## Versión breve de esta tesis

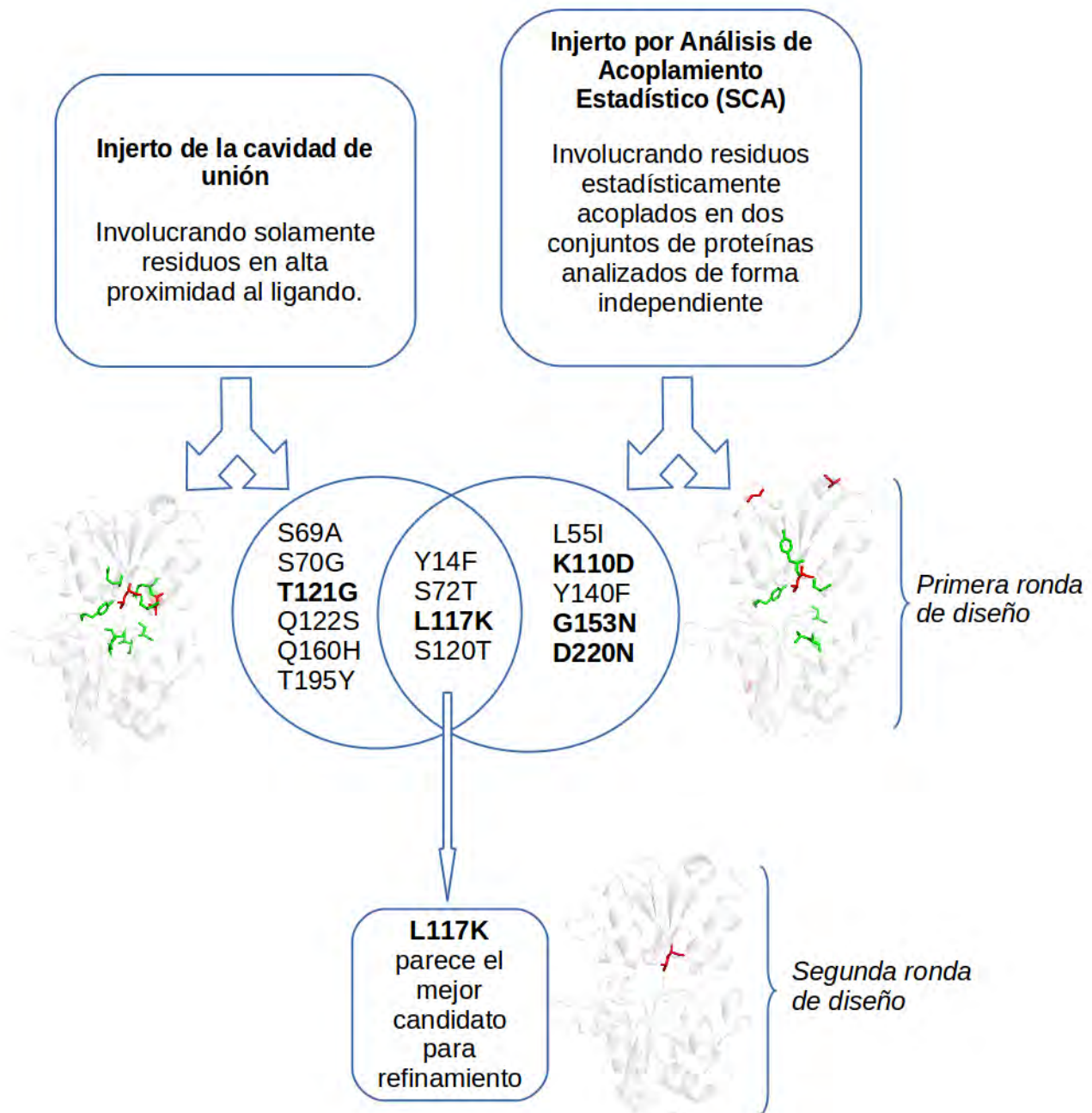
Durante este proyecto de doctorado se abordó la problemática de lograr la unión de un ligando particular en una proteína usando enfoques meramente racionales. Como caso de estudio, usamos a la proteína LAOBP de *Salmonella typhimurium*, la cual une específicamente a los aminoácidos positivos Lisina, Arginina, Ornitina e Histidina. El reto fue lograr, por medio de la propuesta de mutaciones, la unión del aminoácido Glutamina, el cual no puede ser unido por la proteína silvestre. Esto se hizo para determinar si es posible aplicar una o más estrategias para rediseñar la especificidad de la proteína LAOBP, con el menor número de mutaciones posibles para evitar la generación de miles o millones de proteínas mutantes a probar, así como evitar que la nueva mutante sea especialmente inestable.

Una de las estrategias utilizadas fue la del Grafting (injerto o trasplante) de la cavidad de unión a ligando (en inglés: *Binding Pocket Grafting*), utilizada ya previamente en PBP por el grupo de la Dra. Birte Höcker en Alemania con quien colaboramos (Scheib *et al.*, 2013). En nuestro caso, comparamos directamente la cavidad de unión de LAOBP con la cavidad análoga de la proteína de unión a L-Glutamina (Gln) de *Escherichia coli* (*E. coli*), la cual de ahora en adelante será referida como QBP y cuya  $K_d$  por Gln es 500 nM (Sun *et al.*, 1998). De esta manera, creamos a la proteína mutante llamada LAOBP\_graft.

La otra estrategia utilizada, fue rediseñar la proteína LAOBP sin limitarnos a la cavidad de unión, de acuerdo a lo que nos dictó la comparación de dos análisis de acoplamiento estadístico (mejor conocido como SCA, *Statistical Coupling Analysis*, Halabi *et al.*, 2009), uno para un alineamiento múltiple de secuencias (MSA, por sus siglas en inglés) de proteínas con estructura similar a LAOBP y otro MSA para secuencias con estructura similar a QBP. Al diseño concebido de esta forma la llamamos LAOBP\_sca.

Aunque introdujimos una cantidad similar de mutaciones con ambas estrategias, 10 mutaciones puntuales para LAOBP\_graft y 9 para LAOBP\_sca, dichas mutaciones no coinciden más que en 4 posiciones. Fue un análisis racional de la naturaleza bioquímica de las sustituciones en estas posiciones, la que nos llevó a probar, en una nueva ronda de diseño, si era posible obtener unión de Gln por LAOBP al mutar la leucina 117 por una lisina, dando lugar al diseño LAOBP\_L117K (ver Figura 0). LAOBP\_L117K presenta una  $K_d$  promedio de 1.6  $\mu$ M por Gln, además de ser una proteína de expresión soluble, con un rendimiento promedio de 100 mg/L de cultivo y su estructura cristalográfica en presencia de Gln fue depositada en la base de datos PDB con el identificador 5owf, como parte de los resultados obtenidos durante mi doctorado y publicado recientemente: Banda-Vázquez *et al.*, 2018.





**Figura 0. Combinación de estrategias para lograr que LAOBP uniera Gln.** Las mutaciones no conservativas se muestran en *negritas* en el diagrama de Venn del centro y en rojo en las imágenes de proteínas, las mutaciones conservativas están en verde. Por medio de la comparación de las mutaciones coincidentes entre las dos estrategias abordadas en la primera ronda fue posible delimitar el problema a una única mutación.



Sin embargo, dado que no estábamos seguros que la combinación de dos estrategias fuera suficiente para lograr afinidad por Gln en la proteína LAOBP, otras mutantes fueron probadas experimentalmente, las cuales fueron intelectualmente concebidas por diversos enfoques. Tales mutantes fueron seleccionadas de entre decenas de miles de proteínas diseñadas computacionalmente. Si bien los experimentales obtenidos a partir de ellas no fueron publicados en un revista científica, los resultados obtenidos reforzaron nuestras conclusiones con respecto a la proteína LAOBP\_L117K.

**Codependencia evolutiva entre residuos como delimitadores durante el rediseño de  
unión a ligandos en la proteína LAO**

# INTRODUCCIÓN

Un aspecto fundamental para muchos procesos biológicos es el balance adecuado entre afinidad<sup>1</sup> a un ligando<sup>2</sup> y la especificidad<sup>3</sup> a dicha molécula. Tal proceso incluye el reconocimiento inmunológico, metabolismo celular, regulación génica y señalización celular, entre otros. Debido a ésto, la habilidad de predecir de forma precisa y conocer las causas moleculares para la unión de ligandos es fundamental para entender tal fenómeno biológico (Stoddard, 2016) y la potencial aplicación en aspectos médicos, como lo son terapias y diagnóstico; diseño de enzimas y biosensores<sup>4</sup> en la industria; o como herramientas para campos de investigación en diversos aspectos de la biología (Yang & Lai, 2017).

El interés por el entendimiento de estos procesos de reconocimiento molecular están inspirados en la naturaleza, en donde la evolución ha favorecido que proteínas homólogas<sup>5</sup> discriminen ligandos estructuralmente similares. Dicho aspecto ofrece la oportunidad de usar tales proteínas como modelos para estudiar las bases moleculares para la unión a ligandos específicos, lo cual puede ayudar al diseño de proteínas con una especificidad deseada (Scheib *et al.*, 2013).

Un ejemplo de lo mencionado anteriormente es el caso de las proteínas de unión a periplasma<sup>6</sup> (PBPs por el término en inglés *Periplasmic Binding Proteins*), concepto ampliado por otros autores a proteínas de unión a sustratos (SBPs por *Substrate Binding Proteins*, Pulido *et al.*, 2014). Dichas proteínas reconocen pequeñas moléculas y comparten prácticamente la misma estructura de su esqueleto<sup>7</sup> (Scheib *et al.*, 2013). Las PBPs se pliegan<sup>8</sup> en dos lóbulos<sup>9</sup> flexibles. Cada uno de esos lóbulos tiene tres capas  $\alpha/\beta/\alpha^{10}$ , separadas por una hendidura

---

1 En el aspecto bioquímico, la afinidad es el grado al cual una sustancia tiende a combinarse o unirse con otra.

Fuente: <https://en.oxforddictionaries.com/definition/affinity>

2 Para fines de este trabajo, un ligando puede entenderse como una molécula que se une a una proteína. Fuente: [https://www.rosettacommons.org/docs/latest/rosetta\\_basics/Glossary/Glossary](https://www.rosettacommons.org/docs/latest/rosetta_basics/Glossary/Glossary)

3 En química de proteínas, y aunque nunca he visto que algún autor la defina propiamente, podemos decir que la especificidad es la habilidad que un sitio de unión de una proteína tiene para unir únicamente ciertos ligandos. Entre menos ligandos sea una proteína capaz de unir, mayor es su especificidad.

4 Un biosensor es un dispositivo que utiliza materiales biológicos para monitorear la presencia de ciertos químicos en una sustancia. Fuente: *The American Heritage® Science Dictionary*.

5 En biología, homología se define como la existencia de ancestría compartida entre un par de estructuras, genes, o secuencias de DNA o proteína. (Pearson, 2014)

6 El periplasma es el espacio entre las membranas interna y externa en las bacterias Gram-negativas. En bacterias Gram-positivas, un espacio periplásmico más pequeño se encuentra entre la membrana interna y la capa de peptidoglicano. Fuente <http://www.uniprot.org/locations/SL-0200>

7 Para el caso de las proteínas, el esqueleto (del inglés *backbone*) consiste en la cadena lineal de aminoácidos conectados por medio de enlaces covalentes particulares denominados enlaces peptídicos, los cuales se forman por la reacción de condensación entre el grupo amino de un aminoácido y el grupo carboxilo de otro. Fuente: Lodish *et al.*, 2000. *Molecular Cell Biology*. 4Th edition. (Section 3.1 Hierarchical Structure of Proteins) New York: W. H. Freeman.

8 El plegamiento es la manera en que una proteína alcanza su forma característica. Existe un problema referente a esto llamado “El problema del plegamiento de proteínas” (en inglés *The protein folding problem*) el cual es concerniente a cómo la secuencia determina la forma tridimensional y cómo es que las proteínas alcanzan su forma característica. (Zhuravleva & Korzhnev, 2016)

9 Un lóbulo, en su sentido general, es una de las partes, a manera de ondas, que sobresale en el borde de una cosa; como en la hoja de una planta, en un órgano, o para nuestro caso, en una proteína. Fuente: Diccionario de la Real Academia Española, [dle.rae.es](http://dle.rae.es)

10 La estructura secundaria de una molécula polimérica biológica está caracterizada por el plegamiento local dentro de sí misma, como resultado de los enlaces de hidrógeno. En las proteínas, la estructura secundaria de una proteína pertenece al plegamiento de la cadena polipeptídica resultando en  $\alpha$ -hélice, hoja- $\beta$  o lazo aleatorio (en inglés *random coil*). La estructura secundaria se refiere a la organización localizada de parte de una cadena de proteína, la cual asume diferentes

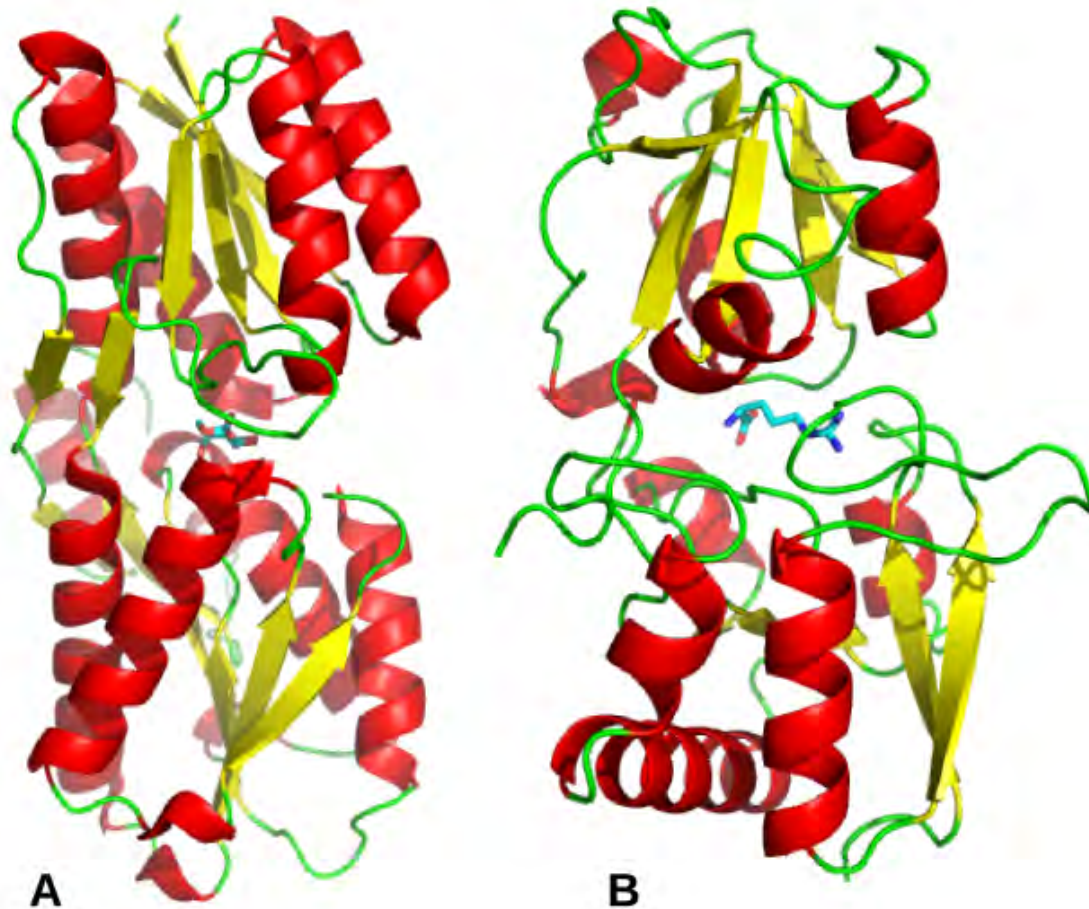
en la que se une el ligando (Planamente *et al.*, 2010) (ver Figura 1). Tal plegamiento estructural se presenta en todos los reinos de la vida (Scheib *et al.*, 2013). Una vez que el proceso de unión al ligando se ha iniciado, las PBPs sufren un cambio conformacional asociado al reconocimiento de tal ligando, pasando de un estado abierto a uno cerrado, en el cual el ligando es atrapado (Oh *et al.*, 1994). Debido a eso, las PBPs pueden ser utilizadas como un andamiaje potencial para la ingeniería de interacciones proteína-ligando (Scheib *et al.*, 2013), lo que puede emplearse con propósitos tecnológicos, como puede ser el diseño de biosensores (Jeffery *et al.*, 2011). Sin embargo, intentos previos para diseñar<sup>11</sup> computacionalmente PBPs que unan nuevos o diferentes ligandos (Looger *et al.*, 2003) no han resuelto el problema del diseño computacional de proteínas para la unión a ligandos<sup>12</sup>, pues tales intentos han dado lugar a importantes discusiones a causa de fallas en la reproducibilidad de los datos publicados (Schreier *et al.*, 2009). De hecho, el problema que encierra diseñar sobre andamiajes de PBPs no está delimitado a encontrar posiciones<sup>13</sup> sobre los cuales inducir mutaciones<sup>14</sup> adecuadas para favorecer la unión a ligandos diferentes o novedosos, sino que también es necesario evitar que tales mutaciones comprometan la estabilidad<sup>15</sup> y/o los cambios conformacionales de la proteína, siendo estos últimos aspectos clave para el adecuado cerrado de la PBP rediseñada al momento de atrapar al ligando. Tal problema está ligado al balance de la robustez a mutaciones aleatorias, dado que las proteínas, en general, toleran perturbaciones en sitios de su secuencia sin alteraciones evidentes en su función; y sin embargo, las proteínas son también suficientemente plásticas como para adaptarse a presiones selectivas<sup>16</sup> cambiantes al permitir variaciones específicas en algunas de sus posiciones para alterar profundamente su función (Reynolds *et al.*, 2013). Por lo cual, pienso que encontrar tal balance entre robustez y plasticidad mutacional para cada proteína es,

---

arreglos espaciales. Una sola proteína puede presentar todos los tipos de estructura secundaria. Sin ningún tipo de interacciones estabilizantes, una cadena de proteína adquiere una estructura de lazo aleatorio. Sin embargo, cuando enlaces de hidrógeno se forman entre ciertos residuos, la cadena se estabiliza y entonces el esqueleto se pliega de forma periódica en uno o dos arreglos geométricos: una  $\alpha$ -hélice, la cual es una espiral en forma de bastón, o una hoja- $\beta$ , la cual es una estructura plana compuesta por alineamientos de dos o más hebra- $\beta$ , quienes son segmentos del esqueleto relativamente cortos y completamente extendidos. Fuente: Lodish *et al.*, 2000. Molecular Cell Biology. 4Th edition. (Section 3.1 Hierarchical Structure of Proteins) New York: W. H. Freeman.

- 11 Con diseño de proteínas, nos referimos al hecho de optimizar la secuencia de aminoácidos de tal proteína para un fin particular. Fuente: [https://www.rosettacommons.org/docs/latest/rosetta\\_basics/Glossary/Glossary](https://www.rosettacommons.org/docs/latest/rosetta_basics/Glossary/Glossary)
- 12 Aunque nunca he visto una definición formal que sea satisfactoria entre las personas que dicen hacer diseño computacional de proteínas, para los fines de este trabajo podemos entenderlo como el uso de herramientas informáticas para la optimización de la secuencia de aminoácidos de una o varias proteínas para un fin particular.
- 13 Una posición en secuencia es un monómero de dicha secuencia. En analogía a las palabras, una posición es cada una de las letras que la componen. Por ejemplo, la palabra *calor* presenta una *c* en la primera posición, una *a* en la segunda, y así sucesivamente. En el contexto de las proteínas, una posición en la secuencia puede, en principio, presentar cualquiera de los 20 aminoácidos que forman parte del alfabeto de las proteínas. Visión personal.
- 14 Una mutación, en el sentido biológico, es una desviación repentina con respecto al tipo parental en una o más características hereditarias, causada por un cambio en un gen o un cromosoma. Fuente: NIH; U.S. National Library of Medicine. En lo referente a este trabajo, el cambio se limita a una posición particular de una secuencia de proteína. Visión personal
- 15 La estabilidad, en lo referente a una proteína, podría definirse como el cambio en la resistencia al desplegamiento de la misma ante la influencia de factores del medio en que se encuentra, sean estos físicos o químicos (Kelly & Price, 1997). Por ejemplo, en la desnaturalización térmica, la temperatura de desnaturalización es definida como aquella temperatura a la cual el 50% de la proteína en solución se mantiene en su estado nativo (Sanchez-Ruiz, 2007). En este sentido, podemos decir que una proteína es más estable (particularmente “termoestable”) cuando presenta una temperatura de desnaturalización superior a otra bajo las mismas circunstancias.
- 16 Una presión selectiva es el grado en el que un organismo que posee una característica particular es eliminado o favorecido por las demandas que le impone el medio ambiente en el que se encuentra. Fuente: (Coleman & Chisholm, 2010). En el contexto de este trabajo, podríamos adaptar esta definición (con ciertas reservas) a moléculas particulares, como las proteínas, que sean capaces de mantener ciertas características, como estabilidad o función, aunque su impacto sea limitado o neutro para la supervivencia del organismo en que se encuentran.

posiblemente, un obstáculo importante para eventualmente resolver el problema del diseño computacional de unión a ligandos.



**Figura 1. Ejemplos de PBPs.** A) Clase I, representada por la proteína de unión a arabinosa. Esta clase presenta seis hebras  $\beta$  en orden 213456. B) Clase II, representada por la proteína LAOBP en complejo con Arg. En este caso existen cinco hebras  $\beta$  en un orden 21354. *Imagen adaptada de Dwyer & Hellinga, 2004.*

Durante este proyecto, tomamos a la proteína LAO PBP (LAOBP) de *Salmonella typhimurium* como un modelo de andamiaje para rediseñar la unión por el aminoácido L-glutamina (Gln), quien es un ligando no reconocido por LAOBP (Clifton & Jackson, 2016). LAOBP es altamente específica para los L-aminoácidos positivos lisina (Lys;  $K_d=15$  nM<sup>17</sup>), arginina (Arg;  $K_d=14$  nM), ornitina (Orn, un aminoácido no proteínico involucrado en el ciclo de la urea;  $K_d=30$  nM) e histidina (His;  $K_d=500$  nM, Oh *et al.*, 1994). Aunque los valores pueden variar dependiendo de la técnica de medición y las condiciones experimentales, Pulido *et al.*, 2014). Por medio de una combinación de estrategias durante mi doctorado, logramos un cambio deseado del perfil de

17 La constante de disociación (denotada por  $K_d$ ) es una cantidad que expresa el grado al cual, una sustancia particular en solución, es disociada en sus componentes (como iones; o para nuestro caso: proteína y ligando). Esta constante se expresa como el producto de las concentraciones de los respectivos componentes, divididos por la concentración de la sustancia (o en nuestro caso, el complejo proteína-ligando) no disociada. Fuente: [https://en.oxforddictionaries.com/definition/dissociation\\_constant](https://en.oxforddictionaries.com/definition/dissociation_constant).

especificidad de LAOBP hacia el reconocimiento de (Gln) usando mutagénesis dirigida, y sin recurrir a técnicas de mutagénesis aleatoria.

## Antecedentes

Se sabe que las PBP se han diversificado para unir una amplia variedad de ligandos, incluyendo carbohidratos, aminoácidos, iones e incluso oligopéptidos. En bacterias, las PBP han sido identificadas como mediadoras en quimiotaxis<sup>18</sup> y absorción de solutos, mientras que dominios con un plegamiento del mismo estilo se han encontrado en receptores eucarióticos, represores de DNA y enzimas. Aunque el plegamiento estructural de estas proteínas está conservado, las secuencias de esta superfamilia son muy diversas (Dwyer & Hellinga, 2004), y en algunas comparaciones, como en nuestro caso, presentan menos del 30% de identidad en secuencia. Se ha propuesto que los lóbulos de las PBP forman un ambiente similar al interior de la proteína cuando se cierran en presencia del ligando específico, haciendo que los residuos que anteriormente estaban en la superficie de los lóbulos formen ahora la cavidad de unión. Esto combina las características fisicoquímicas deseables para que el ligando se localice en regiones excluidas del solvente<sup>19</sup>, dieléctricamente bajas<sup>20</sup> y con la evolvabilidad<sup>21</sup> de los sitios en la superficie de tales lóbulos para alcanzar nuevas adaptaciones. Además, es justo el cambio conformacional mediado por la unión del ligando es lo que permite el desarrollo de funciones acopladas con un potencial para la ingeniería de biosensores, elementos de control alostérico, receptores biológicamente activos, e incluso la posibilidad de crear catalizadores a partir de PBP, pues se los movimientos inducidos por la unión del ligando son similares a las encontradas en algunas enzimas, como la hexocinasa y la fosfoglicerato cinasa (Dwyer & Hellinga, 2004).

Sin embargo, aunque se ha reportado la supuesta predicción exitosa de unión a ligandos usando como andamios de diseño a las PBP (Looger *et al.*, 2003 y Allert *et al.*, 2004), las predicciones de interacción distan mucho de lo esperado, puesto que se ha encontrado que estas proteínas diseñadas no son adecuadas para estudios estructurales, lo cual puede deberse a efectos inesperados como inestabilidad y agregación de las mismas. Alternativamente, se ha observado una total ausencia de unión por los nuevos ligandos, por lo que se ha declarado que el diseño computacional de unión a ligandos es un problema aún sin resolver (Schreier *et al.*, 2009).

Anteriormente y durante mi doctorado, los grupos de investigación de los doctores Alejandro Sosa Peinado y Alejandro Fernández Velasco, ambos en la Facultad de Medicina de la Universidad Nacional Autónoma de México (UNAM), han hecho estudios sobre LAOBP, partiendo de la estructura cristalográfica en estado abierto (Oh *et al.*, 1993) y cerrado (Oh *et al.*, 1994). LAOBP ha sido un buen modelo para estudiar termodinámicamente las bases moleculares de unión de forma experimental (Pulido *et al.*, 2015), por medio de simulaciones de dinámica molecular (MD, por sus siglas en inglés, Silva *et al.*, 2011), así como por barrido de

---

18 La quimiotaxis es la movilidad dirigida, por medio de la cual los microbios sensan señales químicas y se relocalizan hacia ambientes más favorables. (Fuente: Ud-Din & Roujeinikova, 2017. Véase bibliografía).

19 Un solvente es un líquido que contiene a otra u otras sustancias, llamándose estas otras, por conveniencia, solutos. En este trabajo, los solventes por excelencia son soluciones amortiguadoras las cuales se prepararon usando como líquido agua (ver Métodos). Fuente: [goldbook.iupac.org/html/S/S05746.html](http://goldbook.iupac.org/html/S/S05746.html)

20 Un dieléctrico es un material que es mal conductor de la electricidad. Fuente: <http://dle.rae.es/?id=DiB8uC1> (Real Academia Española).

21 La evolvabilidad es definida como la capacidad de generar variación benéfica heredable en los sistemas biológicos. (Stiffler *et al.*, 2015, véase bibliografía).

alaninas<sup>22</sup> para intentar identificar la contribución puntual de residuos de la cavidad de unión en la unión de los ligandos originales (Vergara, Tesis de doctorado en proceso) y el diseño de la especificidad (este trabajo).

Aunque aún existen diferentes aspectos que no entendemos del todo sobre el funcionamiento de LAOBP, pensamos que ya contábamos con información suficiente para intentar modificar su especificidad hacia la unión de un aminoácido no básico sin carga, como Gln, al combinar diferentes estrategias de diseño.

### Puntos clave a recordar:

-Las PBP, están presentes en todos los dominios de la naturaleza y son versátiles en cuanto a los diferentes tipos de ligandos que pueden unir de forma específica. Dado que sufren un cambio conformacional importante en el momento de la unión de su ligando respectivo, se ha propuesto usarlas como andamios para el diseño de proteínas con uniones o capacidades novedosas, como biosensores o enzimas.

-Sin embargo, algunos casos de diseño computacional basados en las PBP han sido infructuosos o carecen de reproducibilidad, por lo que el diseño computacional de proteínas sigue siendo un problema sin resolver.

-En nuestro grupo de investigación, pensamos que con el conocimiento acumulado de la proteína LAOBP, quien naturalmente une aminoácidos positivos (Lys, Arg, Orn y His), es posible diseñar una nueva especificidad en esta proteína.

### **Posibles motivos por los que el diseño de proteínas es un problema sin resolver**

Considerando a una proteína en su totalidad, es importante tener en cuenta que la conservación estructural es un concepto que suele ser vagamente definido, puesto que incluso un único cambio de aminoácido puede llevar a pequeñas o incluso drásticas diferencias en la estructura, dinámica y función de proteínas (Najmanovich 2017). Esta observación es familiar para muchos de nosotros, y fue en gran parte inspiración para este proyecto.

De hecho, el grupo de Najmanovich reportó en 2008 el papel de los residuos evolutivamente conservados para reconocer el mismo ligando entre diferentes familias, lo cual lograron a través de la detección de similitudes entre cavidades de unión, llegando a la conclusión de que tales residuos conservados no son buenos predictores de la unión a ligando, sugiriendo así que los residuos sólo son conservados en proteínas con un mismo plegamiento pero diferentes al comparar distintos plegamientos de proteínas, las cuales han evolucionado para unir a la misma molécula. Incluso se han generado grandes bases de datos de mutaciones neutrales (con poco o ningún efecto en la función) a partir de las estructuras de proteína depositadas en el PDB<sup>23</sup> (Rose *et al.*, 2015), lo que ha permitido observar que una o más mutaciones de residuos altamente conservados y que contactan al ligando no evitan su unión. De hecho, aunque la conservación evolutiva de los residuos del sitio de unión que están en contacto con un ligando implica un papel funcional y estructural, debe siempre recordarse que el reconocimiento molecular no es un proceso que ocurre de forma aislada, sino en presencia de otros ligandos y proteínas; por lo que es recomendable integrar la biología estructural con la biología de sistemas para un mejor entendimiento de dicho proceso (Samish *et al.*, 2015). Ante ello, los residuos evolutivamente conservados del sitio de unión en proteínas con un mismo plegamiento pueden ser importantes no sólo para la

---

22 El barrido de alaninas, también llamado en inglés *Alanine Scanning* o *Comparative Mutational Scanning*, es una técnica para examinar los efectos que una mutación hacia el residuo alanina (quien es un residuo pequeño y químicamente inerte) provoca en diferentes contextos genéticos. Esta técnica se emplea para determinar la contribución de un residuo en particular en un sitio en particular a la estabilidad y/o función de una proteína. (Yang *et al.*, 2016).

23 PDB es la siglas para Protein Data Bank, que es un sistema de bases de datos sobre las formas tridimensionales de proteínas, ácidos nucleicos y ensamblajes complejos. Fuente: [www.rcsb.org](http://www.rcsb.org)



unión de un ligando, sino para evitar la unión de otras moléculas (Najmanovich 2017).

Es así que para diseñar una nueva especificidad en una cavidad de unión es lógico que nos inspiremos en la cavidad de otra proteína con estructura similar y que ya presente la capacidad de unir la molécula de interés. Sin embargo, como indico más adelante, tal estrategia no es del todo funcional en todos los casos.

Inducir unas cuantas mutaciones en la cavidad de unión no siempre ofrece una solución inmediata o efectos del todo predecibles, pues aunque en experimentos de evolución dirigida se ha observado que es posible obtener o aumentar la promiscuidad catalítica de enzimas induciendo tan sólo unas cuantas mutaciones, éstas son altamente contingentes (es decir, dependen del orden específico de aparición) a lo largo de la historia evolutiva. Ahondando en este aspecto, los autores de tales experimentos (Baier & Tokuriki, 2014) observaron muchas mutaciones neutrales en el contexto de la proteína silvestre, pero que se vuelven benéficas cuando ocurren después, como parte de una secuencia de mutaciones consecutivas adquiridas a lo largo del experimento de evolución dirigida. Estas observaciones fueron ampliadas posteriormente por Miton & Tokuriki 2016, observando que la frecuencia de reversión se va reduciendo en cuanto se alcanza una estabilización con respecto a la contribución de mutaciones adicionales al nuevo fenotipo. Curiosamente, muchas mutaciones no suceden en la cavidad de unión, sino en los residuos de la segunda esfera (residuos secundarios), los cuales ayudan a estabilizar conformaciones particulares de los residuos del sitio activo necesarias para el cambio de funcionalidad (Yang *et al.*, 2016). A este fenómeno, en el cual el fenotipo observado es diferente a lo esperado de acuerdo a la suma de los efectos de las mutaciones particulares, se le conoce como epistasis. El término se acuñó para designar ciertas interacciones genéticas debidas a mutaciones en diferentes *loci*<sup>24</sup> en un organismo, pero es perfectamente aplicable al contexto genético en una única proteína con respecto a mutaciones puntuales en sitios distintos de la secuencia codificante (Starr & Thornton 2016).

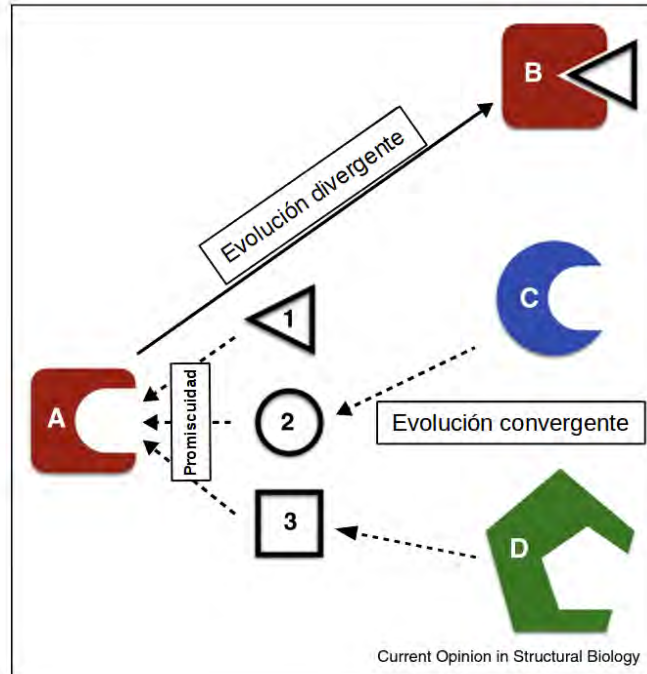
Como mencioné anteriormente, también está la convergencia evolutiva para la unión de ligandos específicos (ver Figura 2), puesto que diferentes patrones de residuos son capaces de reconocer ligandos idénticos, aún entre proteínas tan distantes como para no compartir ni siquiera un mismo plegamiento (Barelrier *et al.*, 2015). En el estudio de Skolnick *et al.*, 2015, en donde se aborda la problemática de que a menudo el diseño de fármacos que debieran unirse a una proteína específica terminan uniéndose a otras proteínas que ni si siquiera comparten una relación evolutiva con la proteína de interés, se sugiere que el universo de cavidades que dan lugar a sitios de unión ya está explorado prácticamente en su totalidad por la naturaleza, y que tales cavidades surgen, posiblemente, como consecuencia de imperfecciones en el empaquetamiento geométrico de proteínas.

En resumen, aunque los métodos para la detección de similitudes de cavidades son útiles para estudiar la evolución de sitios de unión, existen limitaciones impuestas por la naturaleza de los eventos de reconocimiento molecular (Najmanovich 2017).

Por si eso fuera poco el problema de rediseñar una proteína para el reconocimiento de un ligando específico no está delimitado a un entendimiento puntual de las interacciones de la proteína con dicho ligando, pues existe el problema de que descripciones químicas más precisas del complejo proteína-ligando no suelen llevar a predicciones más precisas de la afinidad de unión (Ballester *et al.*, 2014) debido a aspectos como la dinámica intrínseca y los grandes cambios conformacionales de las proteínas, los cuales pueden impactar las propiedades de la cavidad de unión a ligandos (Yang & Lain, 2017).

---

24 *Loci* es el plural de locus. En el contexto genético, un locus se refiere al lugar específico en un cromosoma en el que un gen (la unidad física y funcional de herencia biológica) se localiza. Fuente: Griffiths *et al.*, 2000. An Introduction to Genetic Analysis. 7th edition (Glosario). New York: W. H. Freeman. Disponible en <https://www.ncbi.nlm.nih.gov/books/NBK21878>



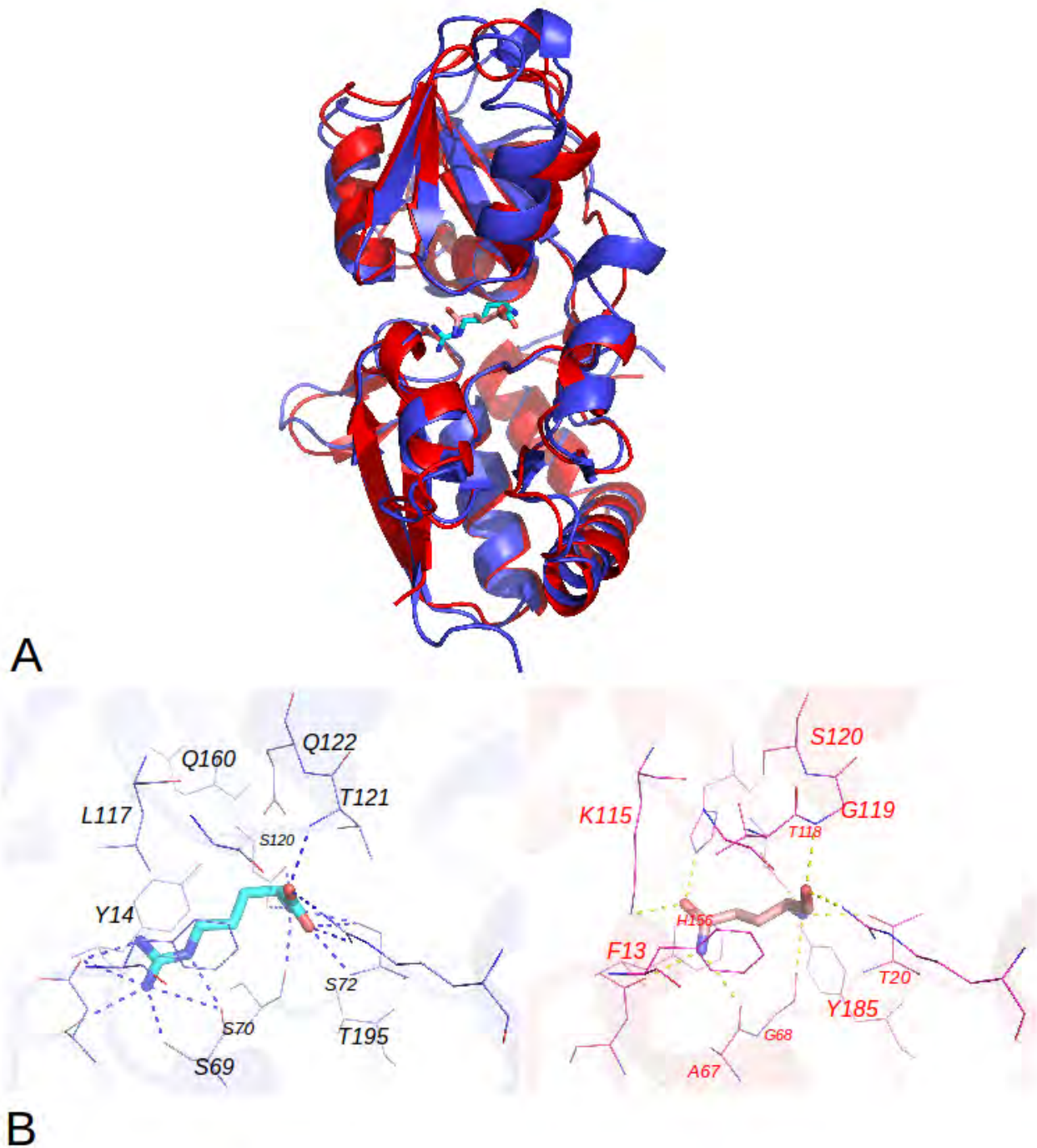
**Figura 2. Diversificación de la especificidad a ligandos.** Los eventos moleculares que provocan que proteínas homólogas sean capaces de reconocer ligandos distintos (divergencia evolutiva), o que proteínas que no comparten un ancestro común reciente sean capaces de unir un mismo ligando (convergencia evolutiva), así como la capacidad de una proteína de unir diferentes moléculas (promiscuidad), dificultan el entendimiento para realizar predicciones adecuadas que permitan el diseño exitoso de proteínas con nuevas uniones a ligandos. *Imagen tomada y adaptada de Najmanovich 2017.*

## Herramientas exploradas durante este proyecto

### Injerto de cavidad de unión (*Binding Pocket Grafting*)

La técnica de *Binding Pocket Grafting* (Injerto de cavidad de unión) consiste en intercambiar, por medio de mutagénesis dirigida, los residuos que son diferentes dentro de la primera esfera de unión a ligando en dos proteínas, después de que éstas han sido alineadas estructuralmente (ver Figura 3). Si bien es cierto que los principales sitios sospechosos implicados en la especificidad de una proteína son aquellos que se encuentran en la primera esfera de unión, es decir, en contacto directo con el ligando (particularmente, aquellos que hacen enlaces de hidrógeno con el ligando), la energía de la interacción puede tener una contribución significativa de residuos adicionales que no contactan al ligando. Aunque los resultados casi siempre conllevan una modificación dramática de la especificidad original, la nueva especificidad puede no ser la deseada, sea porque no se observa unión al ligando objetivo, o bien, si la proteína original ya presentaba alguna afinidad por este último, el cambio podría no ser significativo (Scheib *et al.*, 2013). Es por ello y por lo mencionado anteriormente que es relevante considerar la participación de residuos de contacto primario con el ligando y aquellos que a su vez contactan a los de contacto primario (residuos secundarios).

Durante el desarrollo de este trabajo, decidimos utilizar el enfoque de *binding pocket grafting* como punto de comparación con la metodología alternativa que se describe más abajo, la cual es capaz de considerar sitios que se encuentran más allá del contacto primario.



**Figura 3. Superposición de LAOBP en complejo con L-arginina (PDB: 1laf) y QBP en complejo con L-glutamina (PDB: 1wdn).** A) Esqueletos en representación caricaturizada: LAOBP en azul, QBP en rojo. Arg es mostrada como bastones cian, QBP y Gln como bastones rosas. B) Contactos detallados del ligando en cada PBP en colores como (A). Las líneas punteadas azules indican interacciones polares LAOBP con Arg en el lado izquierdo, mientras que las amarillas a la derecha indican interacciones de QBP con Gln. Algunos residuos están etiquetados. Por simplicidad, las moléculas de agua no se muestran.

### SCA: Análisis de Acoplamiento Estadístico

De acuerdo a Mclaughlin *et al.*, 2012, SCA (*Statistical Coupling Analysis*) consiste en un conjunto de cálculos y subanálisis realizados a partir de un MSA de proteínas. Dicho análisis aprovecha la conservación evolutiva, pero más allá de la mera conservación puntual su mayor ventaja radica en encontrar correlación en la conservación conjunta entre diferentes sitios, y agrupar aquellas posiciones con una conservación más correlacionadas entre sí en lo que se ha llamado “sectores de proteína”, los cuales han sido descritos como grupos de sitios con un papel estructural, bioquímico y/o de otro tipo que no necesariamente guarda una correlación con la jerarquía de estructura primaria, secundaria, terciaria ni cuaternaria (Halabi *et al.*, 2009).

Aunque se ha observado que en los casos en que solo se encuentra un sector de proteína por medio de SCA, la información obtenida no es significativamente diferente a la obtenida por medio de conservación por sitio independiente, en las proteínas con más de un sector, los análisis experimentales son deseables precisamente por ofrecer información más compleja (Teşileanu *et al.*, 2014, ver Tabla 1).

En el más trivial de los casos, los sectores de proteína podrían no tener ningún significado más allá del meramente estadístico, siendo simplemente clusters<sup>25</sup> de mutaciones correlacionadas en el MSA y cuya relación puede analizarse con más detalle (ver Tabla 1). Ahora bien, un sector también podría interpretarse como un grupo de residuos que implican un papel funcional o estructural, y el mutar posiciones del sector implicaría cambiar las propiedades funcionales o estructurales de la proteína. Es en este último aspecto en el que nos hemos basado para lograr el cambio de especificidad de LAOBP comparándolo con un SCA de una proteína de unión a Gln. Siguiendo las observaciones que reportó Redelings en 2014 para generar MSAs, yo consideré que la interpretación de un sector en particular depende, primeramente, de la forma en la que fueron elegidas las secuencias que formarán parte del alineamiento, así como la forma en la que estas secuencias son alineadas. Es decir, para poder identificar aquellos sitios que pudieran ser directamente útiles para intercambiar los fenotipos entre dos proteínas A y B (quienes podrán no compartir función, pero sí compartir otras características, como puede ser el plegamiento), necesitamos construir dos bases de datos de secuencias de proteínas que habrán de ser alineadas por separado, con características funcionales a y b, respectivamente. Para aumentar nuestro aprendizaje a partir de las diferencias, ambos alineamientos deben de ser exclusivos, es decir, no deben compartir miembros. De acuerdo al primer trabajo en el que se acuñó el término “sectores de proteína” (Halabi *et al.*, 2009), para poder realizar estadísticas confiables, cada uno de los alineamientos debe contener, por lo menos, centenares de secuencias, y dichas secuencias no deberían formar subgrupos de secuencias significativamente más parecidas entre sí que con otras del mismo alineamiento.

**Tabla 1. Diferentes significados que pueden asociarse a los sectores de proteínas.** Si bien las propiedades características y formas de exploración son variadas, estas interpretaciones no son excluyentes entre sí. *Tabla adaptada de Teşileanu et al., 2015.*

Interpretación del sector	Propiedad característica	Posibles maneras de exploración cuantitativa
<i>Estadística</i>	Clusters de mutaciones correlacionadas en un MSA	Análisis estadístico del MSA
<i>Evolutiva</i>	Identidad bajo dinámica evolutiva	Experimentos de evolución artificial
<i>Estructural</i>	Propiedades físicas distintas con respecto a los alrededores	NMR, cristalografía de rayos X, simulaciones de Dinámica Molecular
<i>Funcional</i>	La alteración de las posiciones del sector cambia las propiedades	Estudios de mutagénesis

25 Un cluster es difícil de definir, aún en el contexto científico (Henning, 2015), pero para los fines de esta tesis podemos entenderlo como un agrupamiento de cosas por alguna propiedad común. En nuestro caso, sitios de una proteína.

### Software “*Rosetta suite*”

Por otra parte, hoy en día podemos usar el software Rosetta, quien representa una compilación de herramientas computacionales para modelar estructuras de proteínas, así como sus interacciones con otras proteínas, pequeñas moléculas, RNA y DNA. Rosetta ha contribuido al avance de la Biología estructural al abordar retos en el diseño de proteínas *de novo*<sup>26</sup>, modelado comparativo, diseño de proteínas, *docking* proteína-proteína, y *docking* de proteína-molécula pequeña (Bender *et al.*, 2016). Sin embargo, en la práctica, es necesario hacer miles de diseños computacionales, a menudo con decenas de miles de secuencias alternativas. Estos diseños, muchas veces, necesitan revisarse “manualmente” con ayuda de visualizadores de coordenadas, como lo es el programa PyMOL, para eventualmente elegir candidatos a probar experimentalmente. Se ha estimado que el porcentaje de éxito en estas empresas es del 10% (comunicación personal con algunos desarrolladores del software Rosetta). Por una parte, para proponer mutaciones sin alterar el esqueleto de la proteína original, usamos el algoritmo *fixbb* de Rosetta para diseñar una cavidad de unión que pudiera adecuarse a la presencia de Gln, pero por otra, aplicando la metodología Rd.HMM (Martínez-Castilla & Rodríguez-Sotres, 2010), usamos Rosetta para generar una pequeña base de datos de secuencias artificiales, a partir de la cual fue posible construir un modelo oculto de Markov<sup>27</sup>, el cual fue un señuelo para buscar en bases de datos y guiar el alineamiento de los resultados encontrados, el cual ya consistió de secuencias naturales que, plausiblemente, eran capaces de plegarse con el mismo esqueleto que nuestra proteína inicial. Esta estrategia fue muy poderosa para alimentar a SCA al realizar el presente trabajo (ver Figura 4).

26 El problema del diseño de proteínas *de novo* consiste en la generación de nuevas proteínas basándose en principios físicos pero usando secuencias de proteína que no se encuentran en la naturaleza. (Huang *et al.*, 2016)

27 Un modelo oculto de Markov, abreviado como HMM (por sus siglas en inglés) es un modelo finito que describe una distribución de probabilidad sobre un número infinito de secuencias posibles. El HMM está compuesto por un número de estados, los cuáles podrían corresponder a un alineamiento múltiple de secuencias, incluyendo secuencias de proteínas. Cada estado “emite” símbolos (residuos) de acuerdo a probabilidades de emisión de tales símbolos específicas para dicho estado, mientras que los estados están interconectados por probabilidades de transición de estado. Comenzando desde algún estado inicial, una secuencia de estados es generada al moverse de estado en estado de acuerdo a las probabilidades de transición, hasta que un estado final es alcanzado. Los modelos ocultos de Markov reciben tal nombre debido a que, por una parte, la secuencia de estados es una cadena en la que la probabilidad de ocupar el siguiente estado sólo depende del estado actual. A esta propiedad de carecer de memoria se le llama “Markoviana”. Por otra parte, los estados no son observados, pues están ocultos. Sólo la secuencia de símbolos que dichos estados generan es lo que podemos observar. La secuencia de estados más verosímil debe ser inferida de un alineamiento del HMM con la secuencia observada. (Eddy, 1996)



estas interacciones es aproximada por una calificación (score), la cual representa la potencialidad de unión. Sin embargo, en una primera instancia, la precisión de los resultados, así como la velocidad de cómputo, dependen de si la búsqueda se realiza asumiendo cuerpos rígidos (mayor rapidez, menor precisión) o flexibles (menor rapidez, mayor precisión), ya sea de ligando y/o proteína receptora (Pagadala *et al.*, 2017). Es así que una fuente importante de error en las funciones que determinan las calificaciones en *docking* es la descripción limitada de la flexibilidad de la proteína, además de un tratamiento implícito del solvente. Por otra parte, las funciones empleadas son empíricas casi siempre y se parametrizan usando un conjunto de complejos de referencia, por lo que aquellos complejos que se desvían del conjunto de referencia no serán evaluados correctamente, además de que el modelado debe garantizar que las contribuciones energéticas se comporten de forma aditiva (Baum *et al.*, 2010, Ballester *et al.*, 2017), lo cual está directamente relacionado con el problema de la epistasis discutido anteriormente, en el sentido, pienso yo, de que las interacciones energéticas entre el ligando de interés y los residuos más proximales son evaluados durante un momento dado de la simulación, pero no el impacto a largo alcance de la o las mutaciones con el resto de la proteína, por lo que el efecto de las mutaciones sigue siendo no predecible para alcanzar el objetivo general: lograr que la proteína una un nuevo ligando de forma detectable. Es muy importante considerar también que las funciones para calificar de los algoritmos de *docking* sólo explotan la información contenida en el estado proteína-ligando, como se representa en una estructura cristalográfica. Sin embargo, la afinidad también depende de las contribuciones energéticas del ligando con la desolvatación de la proteína y del ligando, así como del ajuste inducido una vez llevada a cabo la unión, y es por eso que existe la incertidumbre de qué tan bien una descripción particular del complejo de unión, ya formado, está describiendo el sistema justo antes de que la desolvatación y el ajuste inducido se lleven a cabo (Ballester *et al.*, 2017). Justo en este punto, las PBPs con su gran cambio conformacional en ausencia y presencia de ligando, son un claro ejemplo. Estas problemáticas nos obligaron a no necesariamente considerar los resultados de *docking* para decidir si un diseño pudiera o no unir de forma exitosa un nuevo ligando antes de probarlo experimentalmente.

### Estrategias combinadas

Es evidente que cada estrategia presenta información valiosa para nuestro objetivo que otras no presentan. En este sentido, combinar por lo menos dos de ellas ayudó a converger a un diseño exitoso.

### Puntos a recordar:

- Injertar una cavidad de unión de una proteína en otra (*binding pocket grafting*) pudiera no resultar en el fenotipo deseado debido a que se han ignorado a los residuos secundarios, los cuales no contactan directamente al ligando.
- Un obstáculo que consideramos importante para el éxito en el diseño de la especificidad de proteínas es la epistasis, fenómeno que involucra interacciones difíciles de predecir entre sitios puntuales de una proteína. Dichas interacciones no pueden ser delatadas a partir de análisis por sitio independiente, como lo es la conservación por sitio en un alineamiento.
- Los métodos de análisis capaces de detectar dependencia evolutiva entre los sitios de una proteína pueden ser una herramienta poderosa para delatar posiciones que tengan efectos clave para la modificación de un fenotipo particular.
- Aunque actualmente existen diferentes herramientas computacionales para explorar secuencias y estructuras de proteínas, los resultados de cada una de esas herramientas ofrecen información útil pero parcial para lograr el diseño exitoso de proteínas con un fenotipo de unión deseado, por lo que recomendamos combinar diferentes enfoques.



## Objetivo General

Cambiar la especificidad de LAOBP para lograr la unión de un ligando particular y previamente no reconocido, como es Gln.

## Hipótesis

La dependencia de cambio entre sitios en una secuencia de proteínas, estadísticamente delatado a partir de las secuencias en una colección de miembros de una familia de proteínas de unión a periplasma, es suficiente información para rediseñar la especificidad de unión a ligandos de, por lo menos, uno de sus miembros, como LAOBP.

## Metodología

Figuras estructurales. Todas las imágenes de estructura de proteína fueron realizadas usando Pymol 1.6.x.

Búsqueda de secuencias. Tomamos el enfoque Rosetta-Hmmer (Rd.HMM, Martínez-Castilla & Rodríguez-Sotres, 2010) usando como semilla el PDB: 1laf (estructura cerrada de LAOBP con L-arginina; Oh *et al.*, 1994). Después de remover las moléculas de agua del archivo, el esqueleto de la proteína fue rediseñado 146 veces con el algoritmo *fixbb* de Rosetta v3.4 (Bender *et al.*, 2016), disponible en <https://www.rosettacommons.org/software>. Las secuencias diseñadas fueron recolectadas en un mismo archivo FASTA, y el Modelo Oculto de Markov (hmm) fue construido usando *hmmbuild* de HMMER v3.0 (Eddy, 2011), disponible en <http://hmmer.org/>. Tal hmm fue usado como anzuelo para buscar con *hmmsearch*, también de HMMER v3.0, en la base de datos UniRef90 (Suzek *et al.*, 2015), disponible en <http://www.uniprot.org/help/uniref> en Mayo 2014. De forma independiente, el mismo enfoque fue tomado con el archivo PDB: 1wdn (QBP en complejo con L-glutamina; Sun *et al.*, 1998) y el PDB: 1xt8 (Una PBP de unión a cisteína en complejo con L-cisteína; Muller *et al.*, 2005). Los identificadores de los resultados en las tres búsquedas fueron comparados por medio de diagramas de Venn usando venny (Oliveros, 2007-2015). Las secuencias exclusivas en cada conjunto (Figura S3) fueron alineadas usando la herramienta *hmmalign* de HMMER v3.0 contra el hmm apropiado en cada caso. Cada alineamiento fue transformado a formato FASTA usando scripts en bash que yo codifiqué, disponibles a petición.

Análisis de Acoplamiento Estadístico (SCA). Usé SCA v5.0 (McLaughlin *et al.*, 2012) disponible en [http://systems.swmed.edu/rr\\_lab/sca.html](http://systems.swmed.edu/rr_lab/sca.html), el cual es un conjunto de herramientas para MATLAB (<https://www.mathworks.com>, versión R2012B). Hice tres análisis independientes usando el script “Tutorial\_pdz\_v5.m”, el cual fue modificado para tomar el alineamiento de interés en cada caso, el archivo PDB correspondiente para mantener una numeración de los sitios consistente con lo reportado, un punto de corte de gaps máximo de 0.6. y un punto de corte mínimo de 0.95 para considerar sitios acoplados por sector. Los tres análisis sugirieron la presencia de múltiples sectores, así que los scripts correspondientes siempre fueron modificados para mostrar tres grupos de sitios estadísticamente acoplados.

Localización de los sitios a mutar. Al alinear LAOBP usando *hmmalign* y LAO hmm, se eligieron los sitios simultáneamente acoplados tanto en a algún sector de LAOBP como en algún sector de QBP.

OTRAS MUTANTES ABORDADAS DURANTE ESTE PROYECTO (La lista de las mutaciones puntuales están en la Tabla S2):

- LAOBP\_sca\_consenso. Se mutó hacia el residuo más común en cada una de las posiciones mutadas en el diseño LAO\_sca de acuerdo al alineamiento QBP\_like.
- LAOBP\_sca\_rosetta. Se le dió total libertad a Rosetta *fixbb* de decidir qué residuo usar en cada uno de los 23 sitios mutables. Nos quedamos con el diseño que, entre ~10000 diseños, mostró el score total más negativo, lo cual es sinónimo de menor energía (Kauffmann *et al.*, 2010).
- LAOBP\_bindingpocket\_rosetta. Se le dio total libertad a Rosetta *fixbb* de decidir qué residuo usar en cada uno de los sitios que se encuentran a, por lo menos, 5 Å de cualquier átomo de la Arg ligando en la estructura 1laf. Nos quedamos con el diseño que, entre ~10000 diseños, mostró el score más negativo.
- LAOBP\_sphere\_consenso. Se mutaron las posiciones que se presentaban a, por lo menos, 5 Å de cualquier átomo de la Gln ligando en la estructura 1wdn, hacia el residuo más común en el alineamiento QBP-like.
- LAOBP\_sca\_redsector\_rosetta. Se le dio total libertad a Rosetta *fixbb* de decidir qué residuo usar en cada uno de los sitios del sector rojo de LAOBP. Nos quedamos con el diseño que, entre ~10000 diseños, mostró el score más negativo.
- LAOBP\_sca\_redsector\_consenso. Se mutó hacia el residuo más común en cada una de los sitios mutables de, únicamente, el sector rojo de LAOBP, de acuerdo al alineamiento QBP-like.
- LAOBP\_sca\_redsector. Se mutó hacia el residuo equivalente en QBP de, únicamente, el sector rojo de LAOBP.
- LAOBP\_sca\_nat1 y nat2. Basándome únicamente en los 2 residuos más comunes en el alineamiento de QBP en cada uno de 22 de los 23 sitios mutables de la figura 5 (pues el sitio 28 es una invariable glicina), evalué las  $2^{22}$  combinaciones posibles con respecto a las secuencias de aminoácidos de las proteínas del alineamiento del universo Gln-like usando scripts en bash y python propios. De estas 4194304 secuencias exploradas, 72 presentaron combinaciones que existen en la naturaleza (de algunas de las secuencias de Gln-like) y, simultáneamente, en los 3 sectores de LAOBP. Por medio de Rosetta *fixbb* se realizó un modelo estructural de cada una de estas 72 secuencias usando como molde 1laf y Gln como ligando. Posteriormente, cada uno de estos modelos fue relajado 1000 veces con Rosetta *relax*, y escogí aquellas dos secuencias que consistentemente presentaron el score más bajo.

Docking. Por medio de Rosetta 3.4 se utilizó el protocolo RosettaLigand tal como es descrito por Combs *et al.*, 2013, el cual ofrece la posibilidad de flexibilizar tanto el ligando como los residuos presentes a una cercanía definida del ligando aun cuando éste pueda desplazarse entre simulaciones. Realicé  $10^3$  simulaciones por proteína indicada en la Tabla S2 (excepto para Ored\_mejorado y LAOBP\_sca\_220D, utilizando Gln como ligando. Dado que previamente se había hecho la superposición de todos los modelos usados antes de realizar el *docking*, se calculó el RMSD entre todos los átomos del ligando Gln en todas las parejas posibles de las  $10^4$  simulaciones. Esta medida delataría diferencias importantes en cuando a la conformación del ligando y su posición en el espacio entre cada una de las 49990000 comparaciones posibles. Sin embargo, dado que los responsables del protocolo RosettaLigand nunca nos compartieron las librerías para realizar las comparaciones de RMSD<sup>28</sup> entre los ligandos, decidimos usar un script en python propio para llevar a cabo las evaluaciones de posición y configuración del ligando en las distintas conformaciones (499000) exploradas por modelo.

---

28 La raíz de la desviación media cuadrática (RMSD por sus siglas en inglés), es la medida cuantitativa más comúnmente usada para determinar la similitud entre dos coordenadas atómicas superpuestas. Detalles en Kufareva & Abagyan, 2015.

Expresión y purificación de proteína. Se transformaron células BL21 DE3 de *E. coli* con el plásmido pET28b+, conteniendo ya sea la construcción LAOBP, QBP o LAOBP\_graft, cuyas secuencias fueron sintetizadas por la compañía GeneScript. 2 L de medio LB fueron inducidos con 1 mM de IPTG a 0.75 OD<sub>600</sub> y agitados a 180 rpm por 4 horas. Las pastillas precipitado celulares fueron colectadas por centrifugación a 4000 rpm durante 15 minutos, lavadas y suspendidas en solución amortiguadora (*buffer*) de fosfato de potasio (KP) pH 8 al 50 mM con cloruro de potasio (KCl) al 300 mM. Esta solución fue sonicada con pulsos de 5 segundos con descansos de 12 segundos durante 25 minutos. Después de una centrifugación de la muestra a 18000 rpm durante una hora, la fracción soluble del extracto fue filtrada a través de una membrana con poros de 0.22 µm. La muestra filtrada fue purificada primeramente con una columna de níquel HiTrap de 5 ml y posteriormente con una S75 High Prep por medio del sistema Äkta.

Geles SDS-PAGE. Se hicieron geles de SDS-PAGE al 12% de acrilamida. La intensidad de las bandas de proteína cargadas por pozo fueron usadas durante los ensayos de expresión para estimar de forma meramente visual el porcentaje de expresión soluble con respecto a los cuerpos de inclusión.

Solubilización y replegamiento de cuerpos de inclusión de LAOBP sca y purificación. La pastilla precipitado de un cultivo celular de 2 L fue suspendida en 30 mL de buffer (50 mM KP pH 8, 300 mM KCl), lo cual fue seguido de sonificado y centrifugado como se ha especificado anteriormente. El sobrenadante fue desechado. El nuevo precipitado fue suspendido en *buffer* y centrifugado de nuevo a 18000 rpm, 4 °C por 30 minutos. Nuevamente, se deshecho el sobrenadante y el pellet fue resuspendido por completo en 10 mL de Cloruro de Guanidinio (GdHCl) 6 M. La suspensión fue agitada delicadamente durante 1 hora a 4 °C. Se añadieron 10 mL de GdHCl 1 M y esta nueva solución se agitó delicadamente 1 hora adicional a 4 °C. La muestra fue centrifugada a 18000 rpm, 4 °C por 30 minutos. El sobrenadante se recuperó y se añadió el volumen necesario de GdHCl 2 M hasta alcanzar 50 mL de volumen final. La muestra fue dializada en una membrana de celulosa con poros de 10 kDa por 3 veces en recambios de 4 L de *buffer* (50 mM KP pH 8, 300 mM KCl) a 4° C y agitado delicadamente por 4 horas en cada recambio. La muestra fue centrifugada y el sobrenadante purificado como se ha especificado anteriormente.

Ensayos de Fluorescencia. Todos los ensayos fueron realizados en equipos JASCO FP6500, en una celda tubular de 0.5 cm. En todo momento, la solución amortiguadora (KP) pH 8 50 mM con KCl 300 mM fue utilizada. La concentración de proteína por celda fue de 10 µM con LAO\_L117K, y de 7 µM en todos los demás casos. La concentración de ligando fue 100 veces la de la proteína usada.

Ensayos de Dicroísmo Circular (CD). Todos los ensayos fueron realizados en equipos JASCO J-715, en el rango de 200 a 240 nm, en una celda de 0.1 cm. En todo momento, la solución amortiguadora (KP) pH 8 50 mM con KCl 300 mM fue utilizada. La concentración de proteína por celda fue de 10 µM con LAO\_L117K, y de 7 µM en todos los demás casos. La concentración de ligando fue 100 veces la de la proteína usada.

Ensayos de Desnaturalización Térmica (*Thermal Melting*). Todos los ensayos fueron realizados en equipos JASCO J-715, partiendo de 10 a 80 o hasta 90 °C en una celda de 0.1 cm, incrementando la temperatura en 1°C por minuto. En todo momento, la solución amortiguadora (KP) pH 8 50 mM con KCl 300 mM fue utilizada. La concentración de proteína por celda fue de 10 µM con LAO\_L117K, y de 7 µM en todos los demás casos. La concentración de ligando fue 100 veces la de la proteína usada.

Termoforésis a microescala (MST por *MicroScale Thermophoresis*). Se comenzó con una solución concentrada (e. g. 5 mM Gln en el caso de LAOBP\_sca) en la solución amortiguadora que ofrece NanoTemper Technologies en el kit para MST, a partir de la cual se realizaron 16 diluciones 1:1 en serie, manteniendo la proteína a una concentración 50 nM en el mismo buffer. Por cada paso, una muestra de ~4 µL fue absorbida por un único capilar del mismo kit. Las muestras fueron medidas usando el equipo NT.LabelFree MST de NanoTemper Technologies y los datos fueron analizados por el software NanoTemper-Analysis (Seidel *et al.*, 2013).

Calorimetría de Titulación Isotérmica (ITC por *Isothermal Titration Calorimetry*). Este ensayo fue realizado en un equipo de TA Instruments modelo nanoITC Low Volume. La concentración de la proteína fue ajustada a 215 µM, y los ligandos a 1500 µM en buffer de diálisis (50 mM Tris pH 8, 150 mM KCl). Las mediciones se llevaron a cabo a 20 °C con una velocidad de agitación de 300 rpm y 300 s entre inyecciones de 2 µL. Los datos fueron analizados con el software NanoAnalyze (TA Instruments). Las constantes de unión fueron determinadas a partir de ajustes sigmoidales en un modelo de unión a un sólo sitio. La línea base del calor de dilución, correspondiente a titulaciones de los ligandos en buffer sin proteína, fue restada de los datos experimentales.

Fluorometría Diferencial de Barrido. En placas Fast Optical de 96 pozos de reacción de 0.1 mL tipo MicroAmp de AppliedBiosystems, un volumen final de 10 µL por pozo, la concentración de proteína fue ajustada a 12 µM. Las soluciones con ligando fueron preparadas para alcanzar 10 mM por pozo. El reportero de desplegamiento fue el Sypro Orange (Sigma Aldricha concentración final 1x. En todo momento, la solución amortiguadora fue (KP) pH 6 50 mM KCl 50 mM, dado que estas eran las condiciones en las que la mutante LAO\_L117K fue más estable bajo esta técnica. Los ensayos fueron realizados en un equipo StepOnePlus Real-Time PCR, y los datos fueron analizados con el software Protein ThermalShift 1.x, disponible en <https://licensn.appliedbiosystems.com/download/index> (Ver Figura 10).

Estructura cristalográfica. La proteína LAOBP\_L117K purificada fue cristalizada en presencia de Gln por el método de “*sitting-drop vapour-diffusion*” a 20 °C. Las gotas de 1 µL fueron preparadas a una razón 1:1 con el “líquido madre” usando un robot Genomic Solutions Honeybee 963. Los cristales aparecieron después de 1 año en la condición H12 (150 mM de bromuro de potasio, 30 % (peso/volumen) PEG 2000 MME, 25 % (volumen/volumen) glicerol. Un set completo de difracción para LAOBP\_L117K en el grupo espacial P212121 fue colectado a 100 K en el equipo de la Sociedad del Anillo de almacenamiento de Electrones de Berlín para el Sincrotrón de Radiación 14.2 (*Berlin Electron Storage Ring Society for Synchrotron Radiation beamline 14.2, BESSY BL 14.2*), el cual fue operado por el Centro Helmholtz de Berlín para Materiales y Energía (Helmholtz-Zentrum Berlin für Materialien und Energie, 2016). Los datos de difracción fueron procesados con el software XDSAPP (Krug *et al.*, 2012) usando la estructura de LAOBP (1laf) para el reemplazo molecular. El modelado “manual” fue realizado con el software Coot (Emsley *et al.*, 2010), y el refinamiento fue realizado con el software Phenix (Adams *et al.*, 2010). Los datos de la cristalización se encuentran en el Apéndice como “Tabla de cristalización”

## Resultados y Discusión

### Comparación entre dos cavidades de unión parecidas pero con diferente especificidad.

Aunque LAOBP y QBP comparten menos del 30% de identidad en su secuencia (Sun *et al.*, 1998), y exhiben diferentes especificidades a ligandos, ambas proteínas se pliegan como típicas PBPs: dos lóbulos conectados por dos bisagras (Oh *et al.*, 1994 y Sun *et al.*, 1998). La comparación de las estructuras cristalográficas en conformación cerrada por la unión de Arg (PDB: 1laf, LAOBP, Oh *et al.*, 1994) y, por otra parte, la unión de Gln (PDB: 1wdn, QBP, Sun *et al.*, 1998) revela una diferencia de menos de 1Å de RMSD (0.74 Å) al superponer los C $\alpha$  del esqueleto (ver Figura 3A).

De acuerdo a la figura 3B, es posible apreciar que los que sitios conservados D11 (D10 en QBP), R77 (R75) y D161 (D157) forman enlaces de hidrógeno o interacciones iónicas con el ligando propio de cada proteína. De hecho, la región variable del ligando presenta una interacción catión- $\pi$ <sup>29</sup> entre los residuos aromáticos Y14 (F13) y F52 (F50). En ambas proteínas, incluso las moléculas de agua están posicionadas en lugares similares (Sun *et al.*, 1998).

### Enfoque de Binding Pocket Grafting

Dado que los residuos cercanos a los respectivos ligandos unidos son muy similares (Figura 3B), una estrategia lógica para lograr cambiar la especificidad de LAOBP hacia L-glutamina, sería trasplantar la cavidad de unión (*Binding Pocket Grafting*, Scheib *et al.*, 2013). De acuerdo a esta técnica, para que LAOBP uniera L-Glutamina, necesitamos las mutaciones S69A, S70G, S72T, L117K, T121G y G160H; todas ellas simultáneamente, dado que las posiciones equivalentes forman enlaces de hidrógeno o interacciones iónicas en QBP con Gln.

Sin embargo, como comentamos desde la introducción, al considerar que pudiese haber residuos más allá de la primera esfera de unión que pudieran afectar las propiedades de las cavidades de unión por diferentes circunstancias, como movimiento interno de las proteínas (Yang & Lai, 2017), extendí el enfoque a residuos que estuvieran hasta 5 Å de cualquier átomo del ligando, ya sea Arg en LAOBP o Gln en QBP. Así fue como las mutaciones Y14F, S120T, Q122S y T195Y también fueron incluidas, dando lugar a la secuencia LAOBP\_graft, la cual presenta 10 mutaciones simultáneas respecto de LAOBP (ver Tabla 2).

---

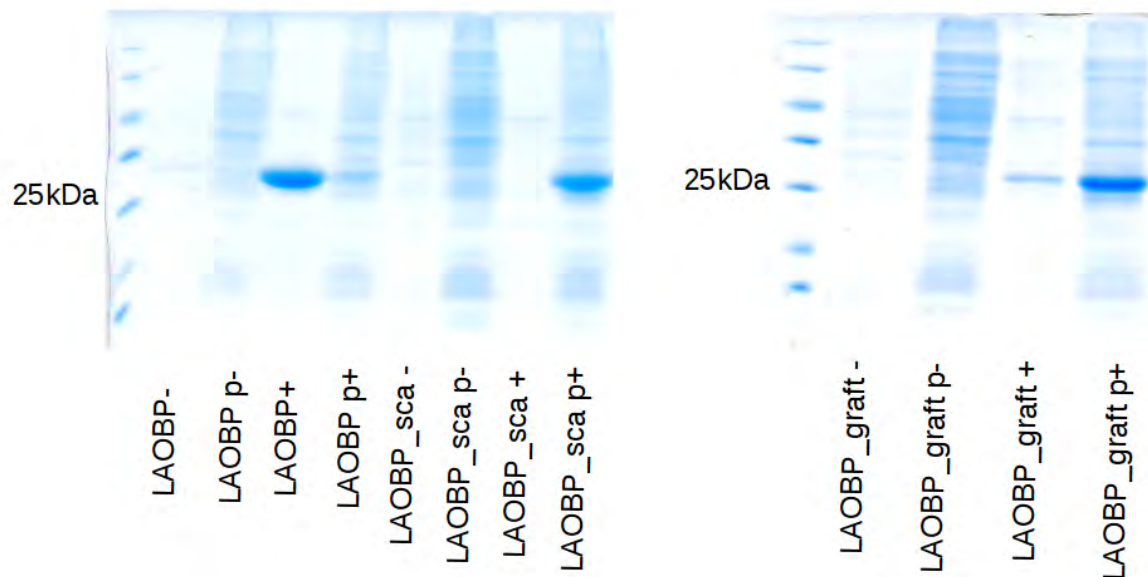
29 La interacción catión-pi (catión- $\pi$ ) consiste en la unión molecular no covalente entre la cara de un sistema  $\pi$  rico en electrones, como un anillo benceno (o el anillo de los aminoácidos aromático)s con un catión adyacente, como un átomo de sodio (Na) cargado. Este tipo de interacción involucra por sí misma diferentes fuerzas intermoleculares, incluyendo electrostáticas e hidrofóbicas. (Mecozzi *et al.*, 1996)

**Tabla 2. Comparación de Cavidades de Unión.** Sitios equivalentes de las cavidades de unión estructuras LAOBP (PDB: 1laf) y QBP (PDB: 1wdn) a partir de superposición estructural. Se muestran en negritas las mutaciones de acuerdo a la secuencia de LAOBP para dar lugar al diseño LAOBP\_graft. Los sitios que involucran mutaciones fisicoquímicamente conservativas están sombreadas en verde, mientras que las no conservativas están sombreadas en rosa. \*Residuos no alterados a una distancia de 5 Å del ligando debido a una falta de información sobre su participación en la unión.

	<i>LAOBP</i>	<i>QBP</i>	<i>LAOBP_graft</i>
<i>Hacen enlaces de hidrógeno con el ligando</i>	D11	D10	D11
	<b>S69</b>	<b>A67</b>	<b>A69</b>
	<b>S70</b>	<b>G68</b>	<b>G70</b>
	<b>S72</b>	<b>T70</b>	<b>T72</b>
	R77	R75	R77
	<b>L117</b>	<b>K115</b>	<b>K117</b>
	<b>T121</b>	<b>G119</b>	<b>G121</b>
	<b>Q160</b>	<b>H156</b>	<b>H160</b>
	D161	D157	D161
	<i>Se encuentran cerca del ligando (<math>\leq 5\text{Å}</math>)</i>	T13	A12
<b>Y14</b>		<b>F13</b>	<b>F14</b>
S18		E17	S18*
F52		F50	F52
L71		I69	L71*
G119		G117	G119
<b>S120</b>		<b>T118</b>	<b>T120</b>
<b>Q122</b>		<b>S120</b>	<b>S122</b>
<b>T195</b>		<b>Y185</b>	<b>Y195</b>

### Expresión de LAOBP\_graft y su comportamiento en presencia de Gln

Alrededor del 40 % de LAOBP\_graft puede expresarse como proteína soluble a 37 °C (ver Figura 5) a juzgar por los contenidos posteriores a la purificación de acuerdo a la intensidad de las bandas en geles SDS-PAGE (ver Metodología). Se registraron espectros de Dicroísmo Circular (CD), los cuales se llevaron a cabo a diferentes longitudes de onda (ver Figura 6) en presencia de diferentes ligandos para observar el contenido de estructura secundaria en ausencia o presencia de ligandos a concentraciones saturantes. Como puede apreciarse en la Figura 6B, los espectros son muy similares a LAOBP, pero parece haber un ligero aumento en el contenido de  $\alpha$ -hélices en presencia de L-glutamina. Sin embargo, con el afán de observar si había un incremento en la temperatura media ( $T_m$ ) aparente de la proteína en presencia de diferentes ligandos (Figura 7), se realizaron mediciones de desnaturalización térmica a 222 nm para seguir la pérdida de estructura de las  $\alpha$ -hélices con el incremento de la temperatura (Kelly & Price, 1997).



**Figura 5. Prueba de expresión a 37°C.** Se transformaron células BL21 DE3 de *E. coli* on plásmidos que contienen los genes ya sea LAOBP, diseño LAOBP\_sca o diseño LAOBP\_graft. Las células fueron (+) inducidas o no (-). El precipitado (p) y las fracciones solubles fueron separadas por centrifugación. El peso molecular esperado siempre fue de 28.3 kDa.



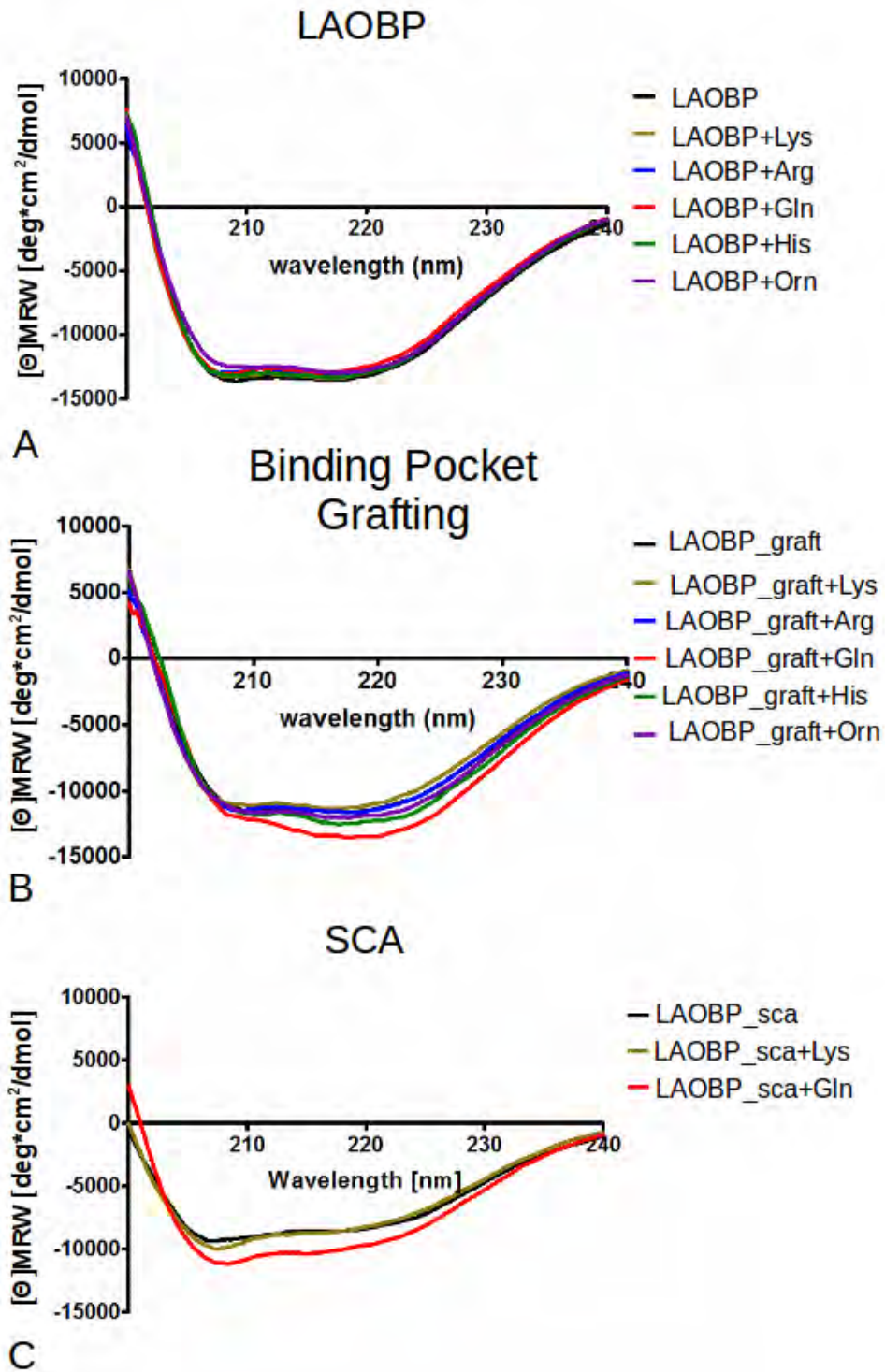


Figura 6. Dicroísmo Circular. Diferentes efectos en proteínas en presencia de Gln.

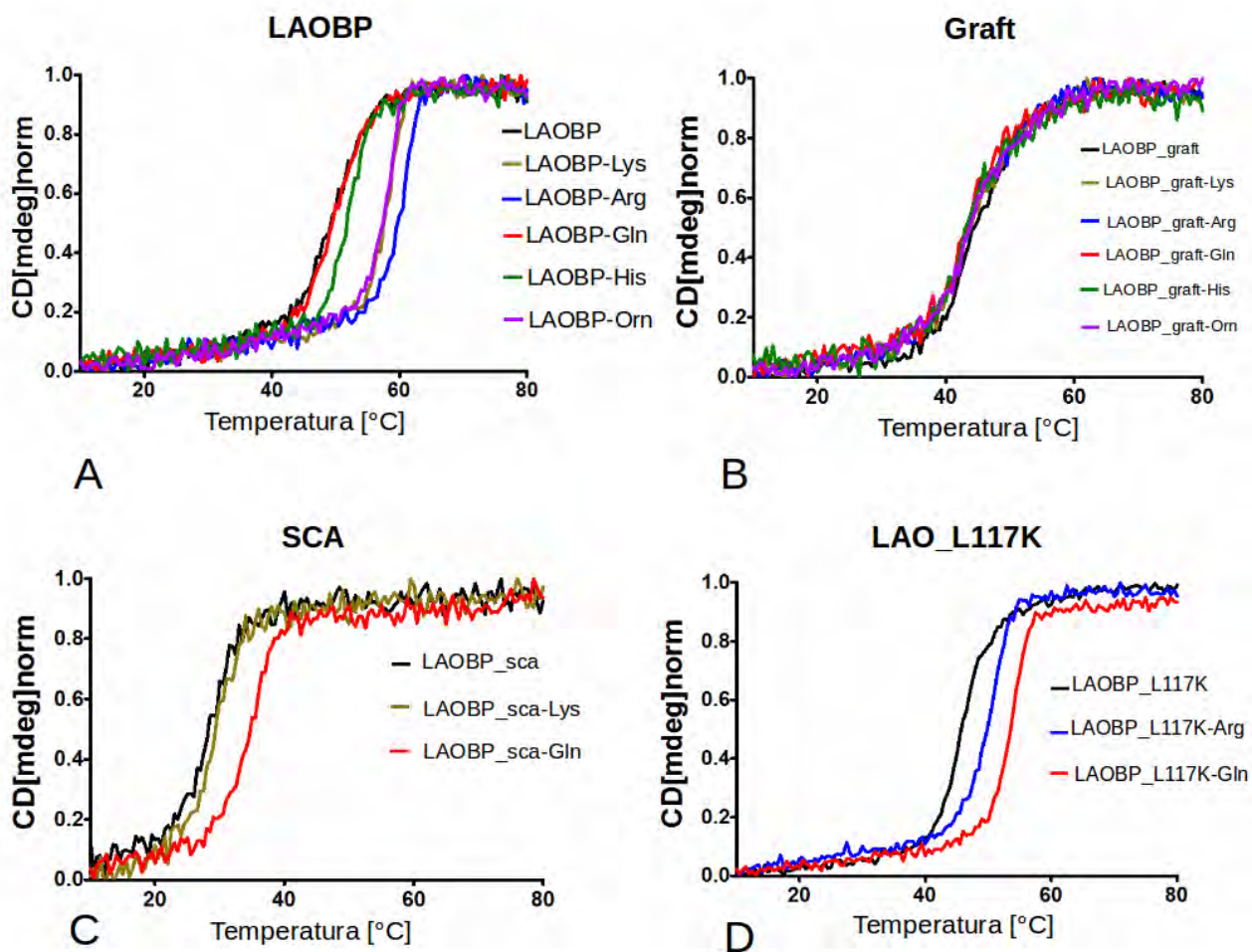


Figura 7. Ensayos de desnaturalización térmica de las principales mutantes generadas en este trabajo. En un principio decidimos realizar el ensayo con todos los aminoácidos de interés (A y B), posteriormente, al observar unión de Gln, sólo utilizamos un aminoácido básico como Lys (C) o Arg (D) para comparar los cambios con respecto a la presencia de Gln.

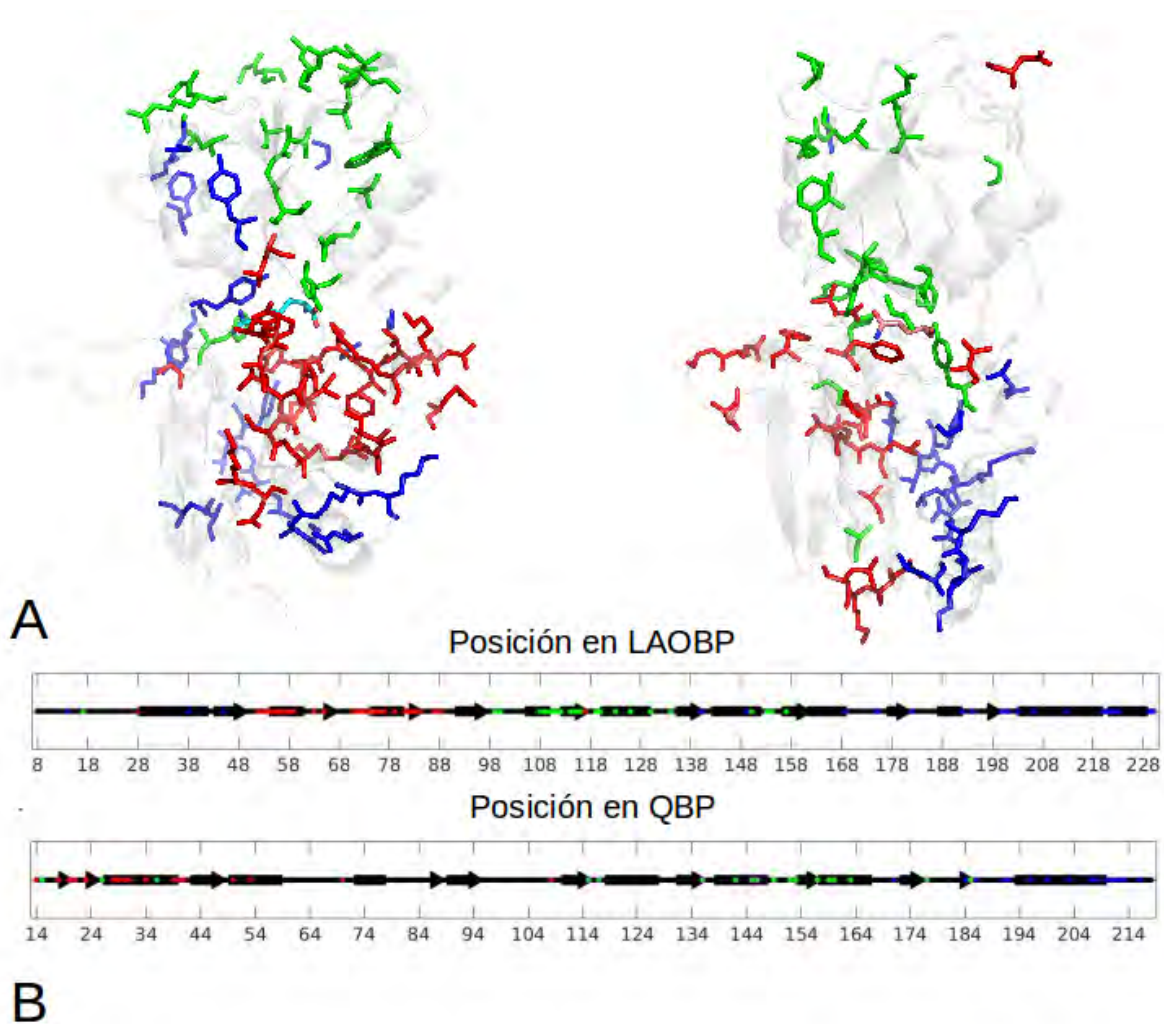
### Enfoque de Acoplamiento Estadístico

Dado que los sitios que pueden ser importantes para un fenotipo particular, como la unión a ligando, no están restringidos únicamente a aquellos en contacto directo con el ligando (Lee & Goodey, 2011; Yang & Lai, 2017), usamos el Análisis de Acoplamiento Estadísticos (SCA, McLaughlin *et al.*, 2012). De acuerdo a SCA, nos fue posible ver qué sitios podrían jugar un rol simultáneo en cuanto a especificidad de unión, así como presentar un cambio concertado con otras posiciones, para así evitar mutaciones no compensadas en el momento del diseño. El razonamiento fue que las posiciones acopladas en sitios equivalentes en dos proteínas (como LAOBP y QBP) identificadas a través de SCA independientes, podrían ser determinantes en el perfil de especificidad y, así, la respuesta estaría en los sitios equivalentes que estaban acoplados y eran diferentes en LAOBP y QBP. Así, se realizaron dos análisis de SCA, uno a partir de 1387 secuencias de proteína similares a LAOBP y, por otro lado, uno con 839 secuencias similares a QBP. Se puso especial cuidado en que estos conjuntos de proteínas no compartieran una sola secuencia.

LAOBP y QBP comparten solamente 28.99% de identidad de secuencia, por lo que las búsquedas con BLASTp para encontrar homólogos que pudieran, confiablemente, ser clasificadas como proteínas tipo LAOBP

pero no QBP y viceversa, fue infructuoso, dado que prácticamente los mismos conjuntos de secuencias fueron hallados usando cualquiera de las secuencias de búsqueda (ver Figura S2). Por tal motivo, utilizamos el enfoque Rd.HMM (Martínez-Castilla & Rodríguez-Sotres 2010), el cual involucra más información que la presente en una única secuencia, puesto que podemos utilizar las restricciones estructurales condensadas en un HMM a partir de un universo de secuencias que son capaces, de acuerdo a Rosetta, de adquirir un mismo esqueleto de proteína, como puede ser una PBP cerrada con su ligando afín. Por medio de los HMM generados, buscamos en la base de datos UniRef90 secuencias que, presumiblemente, pudieran adquirir el mismo esqueleto que LAOBP o QBP en estado cerrado. Después de comparar los resultados encontrados por el HMM para el conjunto putativo LAOBP (LAOBP-like), el universo QBP (QBP-like), y de los encontrados por un HMM para una PBP de unión a cisteína (sólo para tener una referencia de comparación), se segregaron aquellas secuencias exclusivas por conjunto. Al final se recuperaron 1387 secuencias LAO-like y 839 secuencias QBP-like.

Los conjuntos de secuencias fueron alineados de forma independiente, y los SCA fueron realizados también de forma independiente. Tanto con el MSA de LAOBP-like como con el de QBP-like, los sitios estadísticamente acoplados fueron separados en tres sectores: azul, rojo y verde (Figura 8). Al mutar los sitios que están acoplados en LAOBP por aquellos que están acoplados o reportados como involucrados en la unión de Gln en QBP obtuvimos una secuencia a la que llamamos LAOBP\_sca (Tabla 3). La novedad de este enfoque es que utilizamos la información de SCA para hacer alcanzar un fenotipo deseado por mutación dirigida, siendo LAOBP\_sca una prueba directa de que los sitios acoplados influyen decisivamente en aspectos de la función de ciertas PBPs, en contraste con otras metodologías, como son el barrido completo de mutaciones, como se ha realizado antes en otras proteínas (Reynolds *et al.*, 2013) o la utilización de sitios estadísticamente acoplados para generar librerías de mutantes (Liu *et al.*, 2018). Aunque obtuvimos tres sectores en cada caso, decidimos mutar todos los sitios acoplados compartidos entre LAOBP y QBP sin importar el sector de procedencia, con el afán de evitar sesgos, pues no teníamos ninguna certeza de que un sector particular fuera el único responsable de la especificidad. Sospechábamos que tal decisión podía acarrear cambios indeseables junto con el cambio fenotípico deseado, lo que de hecho ocurrió.



**Figura 8. Análisis de Acoplamiento Estadístico.** SCA con las posiciones de sectores identificadas en la conformación cerrada con ligando (A: izquierda para LAOBP y derecha para QBP) o la estructura secundaria (B: arriba para LAOBP, abajo para QBP). Aunque los análisis fueron hechos de manera independiente, con diferentes secuencias en cada caso, los resultados muestran posiciones puntuales que aparecen como acopladas en ambos casos. Es importante mencionar que aquellos residuos en color azul, rojo o verde en LAOBP no necesariamente son azules, rojos o verdes en QBP, respectivamente.

**Tabla 3. Diseño por trasplante de sitios estadísticamente acoplados (SCA-grafting).** La comparación de SCA independientes da lugar a los cambios en negrita de la tercera columna. Las mutaciones conservativas se muestran en verde y las no conservativas en rosa.

<i>LAOBP</i>	<i>QBP</i>	<i>LAOBP_sca</i>
Y14	F13	<b>F14</b>
G28	G26	G28
F52	F50	F52
L55	I53	<b>I55</b>
S72	T70	<b>T72</b>
R77	R75	R77
K110	D108	<b>D110</b>
L117	K115	<b>K117</b>
G119	G117	G119
S120	T118	<b>T120</b>
Y140	F136	<b>F140</b>
L150	L146	L150
G153	N149	<b>N153</b>
A157	A153	A157
G182	G177	G182
K201	K191	K201
L206	L196	L206
A213	A203	A213
D220	N210	<b>N220</b>
G221	G211	G221
T222	T212	T222
Y223	Y213	Y223

### Construcción y caracterización de LAOBP\_sca

En contraste con la proteína LAOBP\_graft, LAOBP\_sca no fue una proteína soluble al expresarse bajo ninguna condición (Figuras 7 y S3). Después del tratamiento para volver a plegar esta proteína (ver Métodos) y su posterior purificación, se realizó su desnaturalización térmica, siguiéndola por CD a 222 nm, en donde LAOBP\_sca muestra un cambio claro en presencia de Gln (ver Figura 7C). De hecho, la  $T_m$  aparente en la que la mitad de la población de  $\alpha$ -hélices pierde estructura es  $\sim 8$  °C superior al de la proteína en ausencia de Gln. Además de eso, observamos que en presencia de Lys, el cambio fue mínimo ( $\sim 1$  °C). Esto difiere del comportamiento de LAOBP, quien se muestra en la figura 7A para fines comparativos.

En ensayos de ITC, LAOBP\_sca siempre fue propensa a agregación, incluso a 15 °C (ver Tabla S2), por lo que la Termoforesis a Micro Escala (MST, por MicroScale Thermophoresis, Seidel *et al.*, 2013) nos permitió estimar una  $K_d=40$   $\mu$ M para Gln.

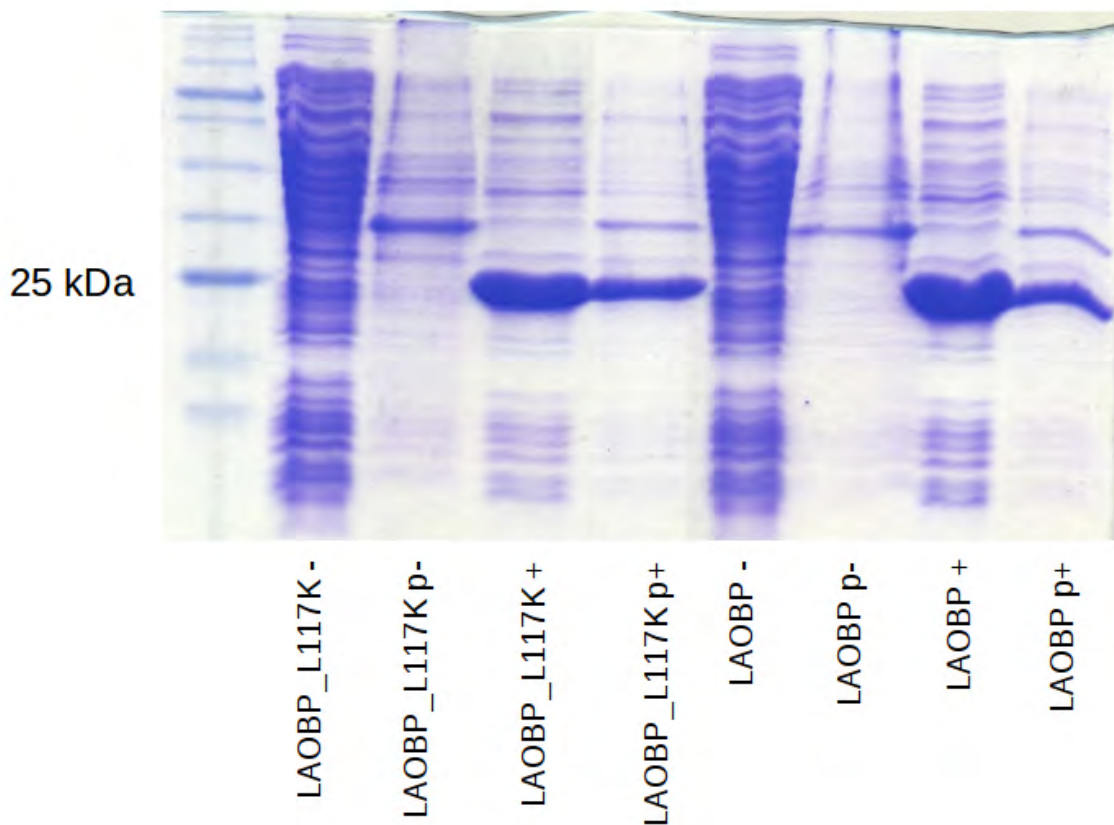
**Un análisis de mutaciones compartidas en ambos enfoques nos revela que un sólo cambio es suficiente para cambiar el perfil de especificidad**

Después de comparar los resultados obtenidos para las secuencias de las proteínas LAOBP\_graft (expresión soluble, pero comportamiento incierto con Gln) y LAOBP\_sca (expresión insoluble, pero evidente unión de Gln), las mutaciones compartidas fueron analizadas. De acuerdo a la Tabla 4, el cambio más importante fue L117K, cambiando un aminoácido hidrofóbico por uno cargado positivamente. Así decidimos inducir la mutación L117K en la secuencia de LAOBP, dando lugar a la mutante LAOBP\_L117K. No es sorprendente que, alrededor del 50% de esta proteína LAOBP\_L117K pueda ser expresada de forma soluble a 37 °C (ver Figura 9).

**Tabla 4. Mutaciones compartidas en las mutantes LAOBP\_graft y LAOBP\_sca.** Se muestran las cuatro mutaciones en ambos enfoques de diseño, incluyendo el cambio bioquímico involucrado en cada caso.

Mutaciones compartidas	Tipo de Mutación
Y14F	Conservativa: aromático por aromático
S72T	Conservativa: polar sin carga por polar sin carga
<b>L117K</b>	No conservativa: hidrofóbico por uno con carga positiva
S120T	Conservativa: polar sin carga por polar sin carga





**Figura 9. Análisis en SDS\_PAGE de Expresión de LAOBP\_L117K.** 0) Marcador de peso molecular, las muestras con LAOBP\_L117K van de los carriles 1 al 4: 1) Fracción soluble sin inducir, 2) Fracción insoluble sin inducir, 3) Fracción soluble inducida, 4) Fracción soluble inducida. Las muestras con LAOBP van de los carriles 5 al 8: 5) Fracción soluble no inducida, 6) Fracción insoluble no inducida, 7) Fracción soluble inducida, 8) Fracción insoluble inducida.

### LAOBP\_L117K une Gln con casi la misma afinidad que QBP

*LAO\_L117K es más estable en presencia de Gln*

El ensayo de desnaturalización térmica de LAO\_L117K fue seguido del cambio en la señal de dicroísmo circular a 222 nm, estimando una temperatura media de pérdida de  $\alpha$ -hélices de  $\sim 46$  °C. Cuando hicimos el mismo ensayo usando Arg a una concentración saturante, la  $T_m$  fue de  $\sim 50$  °C, mientras que con Gln la  $T_m$  fue de  $\sim 54$  °C (Figura 7).

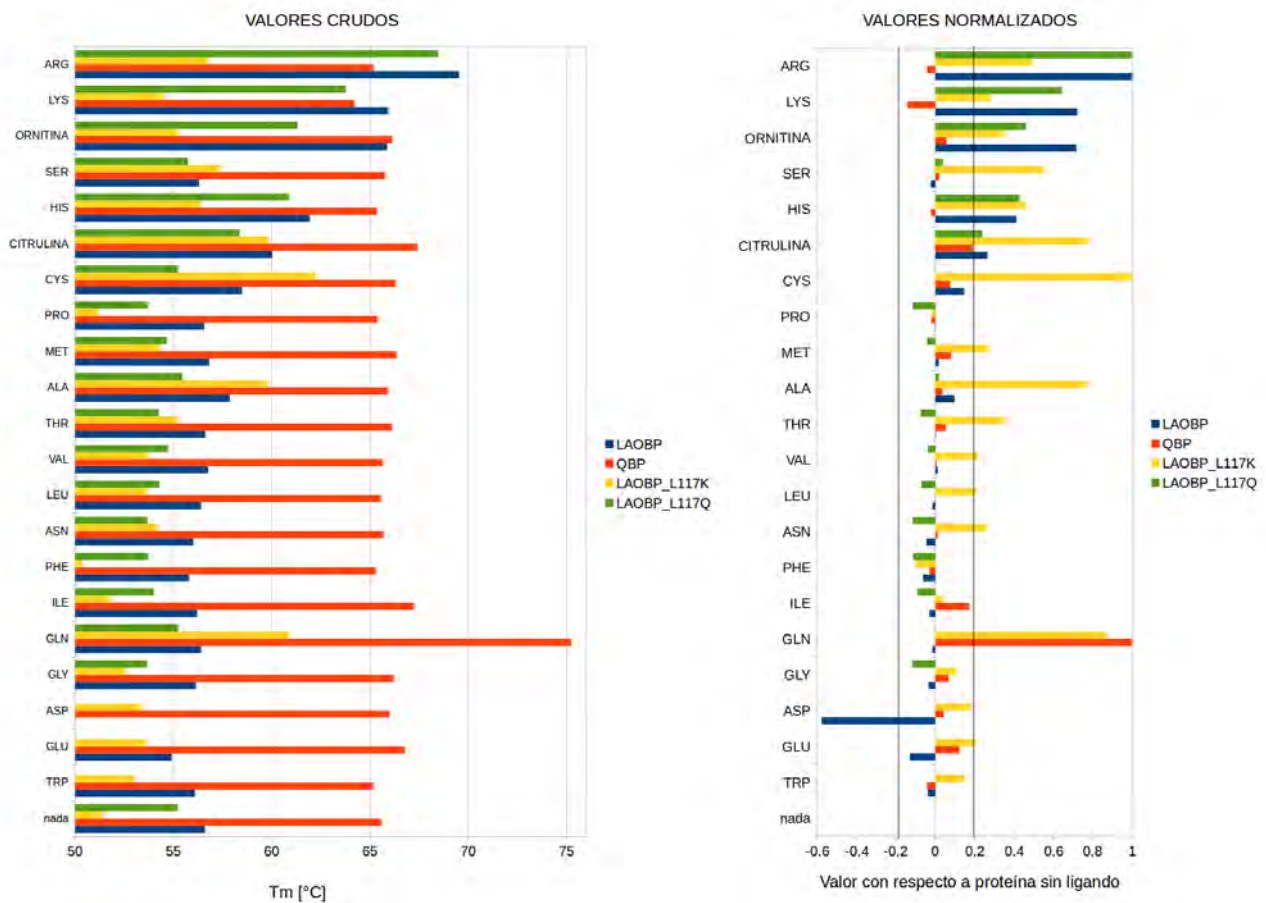
### Fluorimetría Diferencial de Barrido

Con el objetivo de observar la promiscuidad de LAOBP\_L117K en comparación con LAOBP, los ensayos de Fluorimetría Diferencial de Barrido (Vivoli *et al.*, 2014) fueron realizados con una concentración fija de diferentes aminoácidos y con la proteína LAOBP\_L117K a una concentración de 13  $\mu$ M. Por medio de esta técnica, se determinó la  $T_m$  aparente y se normalizaron los datos con respecto al comportamiento de la proteína en ausencia de ligandos. Con fines comparativos, se incluye también a la proteína silvestre QBP, quien une específicamente a Gln, y LAOBP\_L117Q, dado que el ancestro AncQR reportado por Clifton & Jackson en 2016



presenta glutamina en la posición equivalente, y también porque es justo glutamina el aminoácido más común en dicha posición en el alineamiento del conjunto Q.

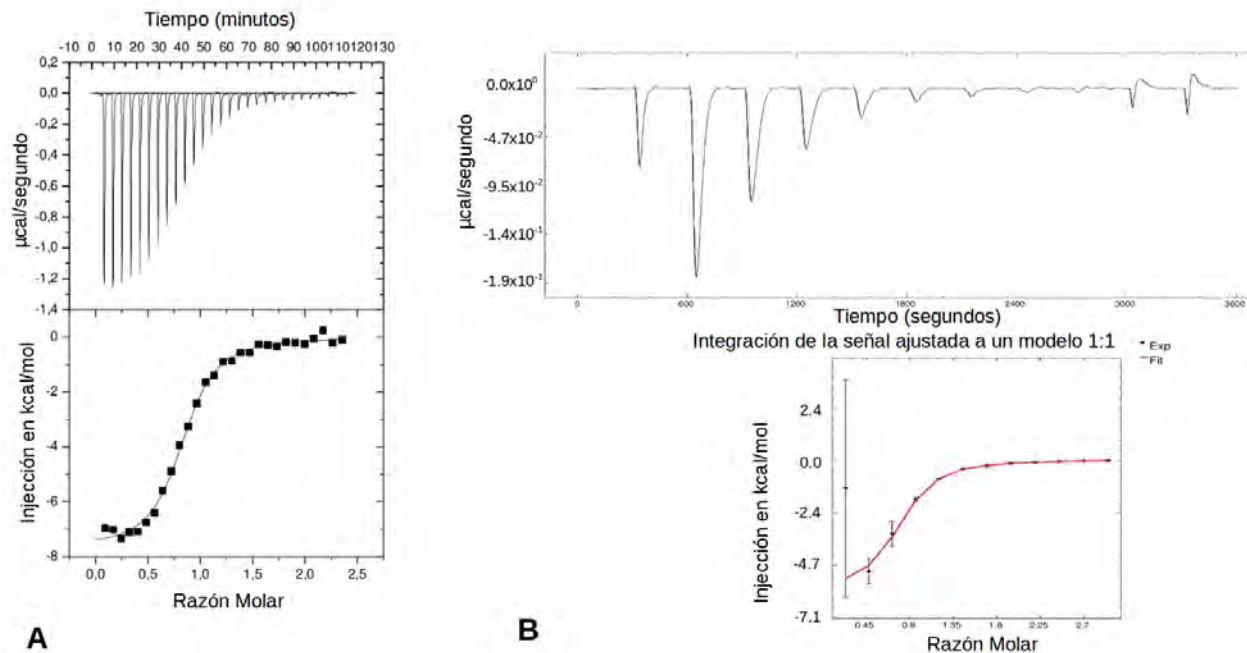
Como puede observarse en las figuras 7 y 10, el cambio en especificidad de LAOBP\_L117K no es completo. De hecho, de acuerdo a la figura 10, LAOBP\_L117K aún puede unir citrulina (otro aminoácido no proteínico e involucrado en el ciclo de la urea), con el cual hubo un cambio marginal de afinidad en comparación con la proteína silvestre. Adicionalmente, pareciera que la cisteína fuera el mejor ligando para LAOBP\_L117K, sin embargo, a juzgar por los ensayos de ITC, en los que se adicionó DTT para evitar la dimerización de cisteína en cistinas (Tabla S3), su afinidad sólo fue mejorada 20 veces en comparación con LAOBP. Por otra parte, Gln demostró ser un ligando obvio para LAOBP\_L117K, pero no es de interacción detectable para LAOBP, representando así el cambio más drástico en afinidad.



**Figura 10. Fluorimetría Diferencial de Barrido.** Comportamiento comparado de unión a diferentes ligandos entre LAOBP, QBP y dos mutantes puntuales de LAOBP en el sitio 117. A) Valores crudos de la Tm en °C para diferentes L-aminoácidos. B) Valores normalizados con respecto a la Tm de cada proteína sola. En los ensayos se incluyeron ornitina y citrulina, el cual presentó indicios de unión en pruebas preliminares (no mostradas), pero no tirosina por problemas de solubilidad e indicios de que no existe unión. Las líneas de tendencia en B indican que en el intervalo (-0.2;0.2) las mediciones odrían ser consideradas ruidosas por efecto de la solubilidad del ligando.

### Ensayos de ITC con LAO\_L117K

Por medio de ITC, es posible determinar una  $K_d = 1.6 \mu\text{M}$  promedio para Gln (ver Figura 11 y Tabla 4), en un proceso dominado por el componente entálpico, como se esperaba de acuerdo a Pulido *et al.*, en 2014 y Clifton & Jackson en 2016, quienes realizaron las mediciones con la versión silvestre de LAOBP y sus ligandos nativos.



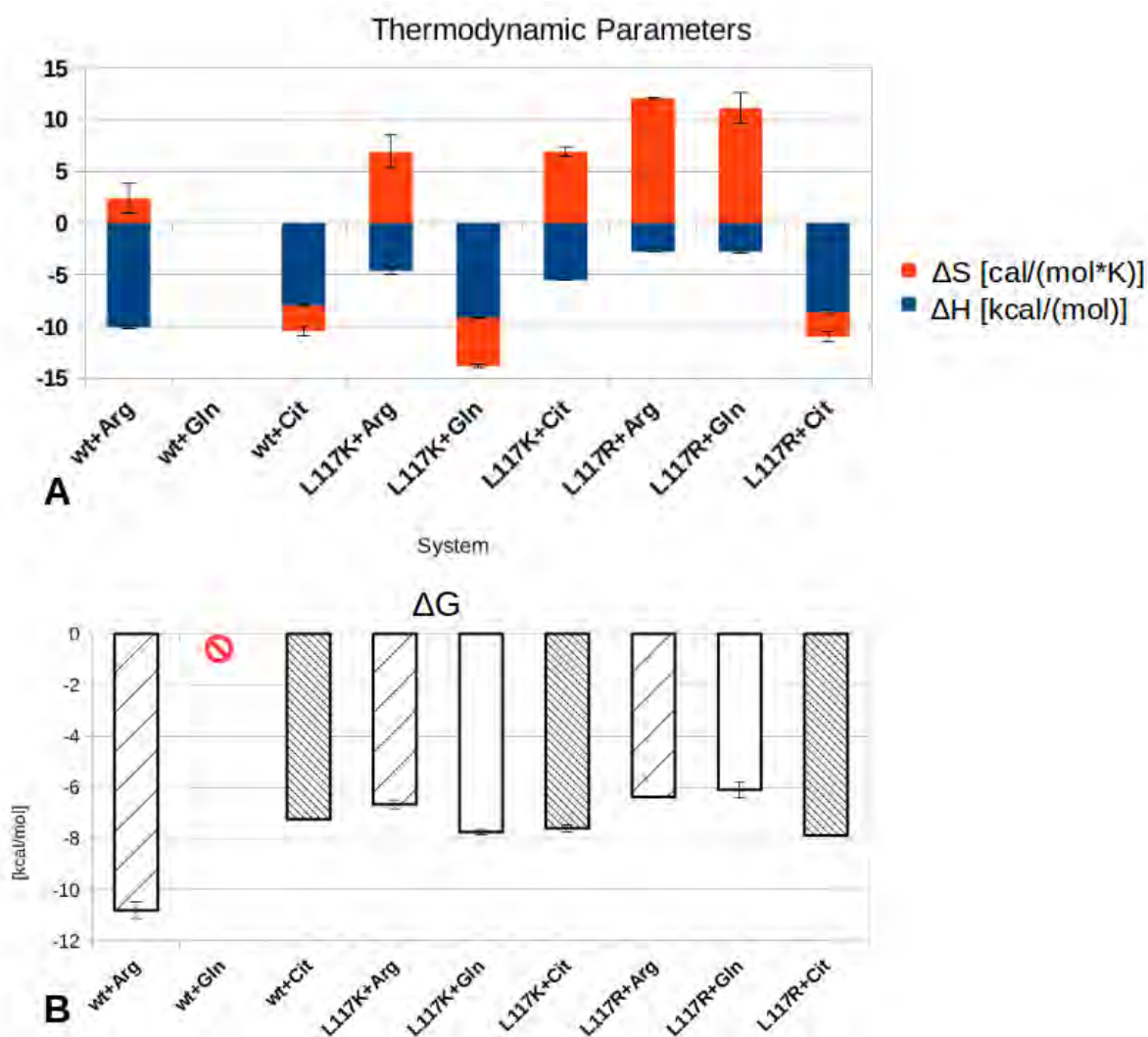
**Figura 11. ITC de LAOBP\_L117K con L-Glutamina.** A) Medición con un equipo nanoITC realizado por nuestros colaboradores en Alemania (grupo de la Dra. Birte Höcker), en donde el ajuste a un modelo de unión 1:1 da lugar a una  $K_d = 1.5 \mu\text{M}$ . B) Medición hecha en nuestro laboratorio con un equipo TA instruments,  $K_d = 2.6 \mu\text{M}$ , en donde un valor de chi cuadrada menor a 1 es sinónimo de significancia estadística.

**Tabla 5. Ensayos de ITC con LAOBP, LAOBP\_L117K y LAOBP\_L117R.** Se muestran los diferentes valores termodinámicos con Gln, Arg y Citrulina.

	$K_d$ [ $\mu\text{M}$ ]	$\Delta H$ [ $\text{kcal} \cdot \text{mol}^{-1}$ ]	$-T\Delta S$ [ $\text{kcal} \cdot \text{mol}^{-1}$ ]	$\Delta G$ [ $\text{kcal} \cdot \text{mol}^{-1}$ ]
<b>L-Glutamina</b>				
LAOBP	N/D	N/D	N/D	N/D
LAOBP_L117K	$1.59 \pm 0.23$	$-9.16 \pm 0.04$	$1.38 \pm 0.05$	$-7.57 \pm 0.08$
LAOBP_L117R	$33.7 \pm 17.0$	$-2.78 \pm 0.13$	$-3.26 \pm 0.44$	$-6.09 \pm 0.31$
<b>L-Citrulina</b>				
LAOBP	$3.84 \pm 0.06$	$-7.98 \pm 0.14$	$0.72 \pm 0.14$	$-7.25 \pm 0.01$
LAOBP_L117K	$2.31 \pm 0.48$	$-5.54 \pm 0.01$	$-2.03 \pm 0.13$	$-7.60 \pm 0.12$
LAOBP_L117R	$1.29 \pm 0.03$	$-8.60 \pm 0.17$	$0.70 \pm 0.01$	$-7.89 \pm 0.01$

L-Arginina				
LAOBP	$0.009 \pm 0.005$	$-10.10 \pm 0.10$	$-0.70 \pm 0.41$	$-10.81 \pm 0.32$
LAOBP_L117K	$11.5 \pm 3.04$	$-4.62 \pm 0.31$	$-2.01 \pm 0.47$	$-6.67 \pm 0.17$
LAOBP_L117R	$19.4 \pm 1.8$	$-2.78 \pm 0.02$	$-3.54 \pm 0.03$	$-6.32 \pm 0.01$

Las  $K_d$  presentadas en la Tabla 5 son consistentes con las observaciones del barrido de fluorimetría diferencial en la figura 10, además de que LAOBP\_L117K presenta una  $K_d$  (Gln) dentro del rango fisiológico de concentración, su afinidad por Arg fue reducida 1000 veces en comparación con la de la proteína silvestre.



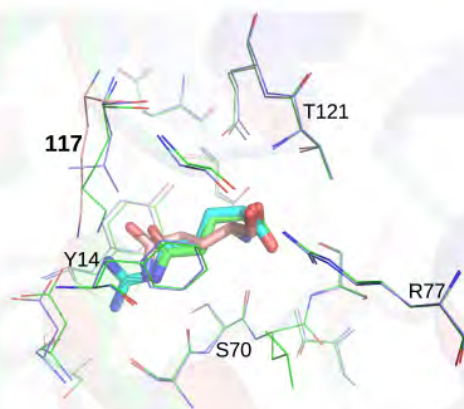
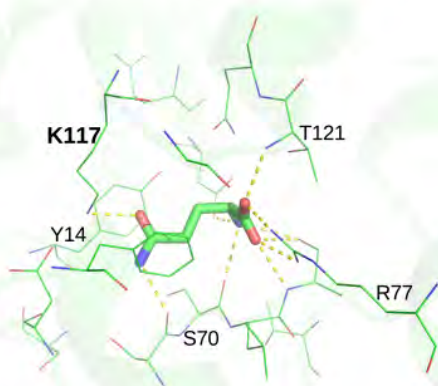
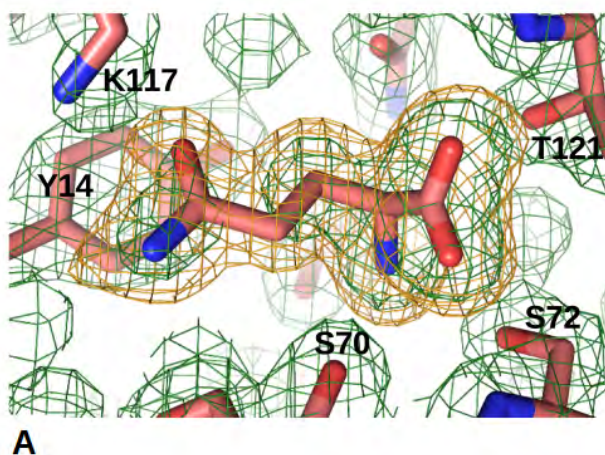
**Figura 12. Parámetros termodinámicos.** Se grafican los valores de la tabla 5 A) Cambios entálpicos ( $\Delta H$ ) y entrópicos ( $\Delta S$ ), en donde éstos últimos están escalados por la temperatura a 293.15 Kelvin (20 °C) y B) Cambio correspondiente en la energía libre de Gibbs ( $\Delta G$ ).

## Estructura Cristalográfica

### *La estructura cristalográfica de LAO\_L117K confirma la unión esperada*

Con el objetivo de observar la localización de Gln unida a LAO\_L117K, el grupo de la Dr. Birte Höcker en Alemania colaboraron con nosotros obteniendo cristales de esta mutante en condiciones saturantes de dicho aminoácido. El mapa de densidad electrónica no es continuo en la cadena lateral de la Gln ligada, aunque sí podemos decir que lo que está unido en la proteína es un aminoácido en la cavidad correcta (ver Figura 13A y 13B). Por medio de un mapa de diferencias, la densidad electrónica de la cadena lateral se vuelve continua. Sin embargo, la forma no descarta la presencia de un aminoácido diferente a la glutamina (ver Figura 13A), de hecho, incluso una metionina podía ajustarse a tal densidad en este punto del análisis (datos no mostrados). Sin embargo, al excluir la densidad del solvente en un versión nueva de mapa de diferencias llamado *polder map* (Liebschner *et al.*, 2017), se obtuvo evidencia de que la densidad sí corresponde a la de una L-glutamina.

Con respecto a la conformación de la proteína, y dado que el único cambio con respecto a la proteína silvestre es en la posición 117, los cambios entre LAOBP y LAOBP\_L117K son mínimos. Con excepción del sitio L117, las cadenas laterales de los distintos residuos tienen conformaciones muy similares. De hecho, como puede observarse en la Figura 13C, el nuevo ligando Gln adopta una conformación más parecida a Arg unida a LAOBP en el PDB 1laf que a la Gln en la cavidad de QBP según el PDB 1wdn. Este fenómeno puede explicarse por el hecho de que el esqueleto de LAOBP\_L117K es aún muy parecido a 1laf, y la Gln debería unirse con una conformación diferente con respecto a lo que se observa en 1wdn.



**Figura 13. Comparación del complejo estructural de LAOBP\_L117K (PDB: 5owf) con los complejos LAOBP (1laf) y QBP (1wdn).** A) Polder map contorneado a  $\pm 3 \sigma$  (naranja) y una densidad de diferencia a  $\pm 1 \sigma$  (verde) con Gln modelada en la cavidad de unión de LAOBP\_L117K. B) Acercamiento de la cavidad de unión de LAOBP\_L117K (verde), 1laf (azul) y 1wdn (rosa) mostrando aquellos residuos cercanos o en contacto con los ligandos. En todos los casos, la representación caricaturizada es translúcida.

### ***Otras mutantes intentadas durante este proyecto***

Durante el desarrollo de este proyecto se realizaron más diseños, de los cuales 10 fueron considerados para realizar análisis informáticos, y de ellos, 7 se llevaron a la experimentación en la mesa de laboratorio comenzando por pruebas de expresión (Figura 14) para determinar la temperatura óptima en que podíamos obtenerlas solubles (además de LAOBP\_sca), posteriormente hicimos ensayos de Fluorescencia (ver Figura 15) para determinar si existía alguna interacción con los ligandos nativos o Gln. El análisis de los resultados obtenidos para estas mutantes alternativas permite profundizar en el significado del sitio 117 como la clave para lograr la unión de Gln en LAOBP y también como la más parsimoniosa. A continuación, se describen los resultados más relevantes con respecto a estas mutantes en cuanto a las mutaciones puntuales compartidas (ver Tabla S2). Las estrategias abordadas para diseñar dichas proteínas y los resultados correspondientes se muestran en mayor detalle en el apéndice (Otras mutantes diseñadas durante este proyecto y Tabla S2) de esta tesis.

En la proteína LAOBP\_sca\_consenso se tomó al residuo consenso según el MSA de las secuencias QBP-like en las posiciones estadísticamente acopladas de LAOBP. La idea era que los sitios más comunes podían servir para cambiar la especificidad de una forma radical sin afectar del todo la estabilidad, pues se sabe que, en general, las proteínas consenso son más estables que las proteínas naturales (Merkl & Sterner, 2016). Sin embargo, sólo observamos proteína soluble al expresar LAOBP\_sca\_consenso a 11 °C, lo cual repercutió en nuestra imposibilidad de realizar ITC con ella, mientras que su comportamiento en presencia de ligandos por las otras técnicas mencionadas en la Tabla S2 sólo mostró un perfil de especificidad esencialmente igual al de LAOBP silvestre. Esta proteína tuvo un comportamiento similar al de LAOBP\_sca\_rosetta, en la cual se le dio libertad al algoritmo *fixbb* de Rosetta de mutar todos los sitios estadísticamente acoplados de LAOBP. Curiosamente, muts presenta mutaciones hacia residuos más pequeños, siendo casi exclusivas en el sector azul, el cual por su distribución en la estructura de LAOBP es sospechoso de ser el más influyente en la estabilidad de la proteína, mientras que en el sector rojo sólo se hicieron dos mutaciones por alanina, una de ellas en el conservado residuo R77, mientras que el sitio 117 fue mutado a una isoleucina. En esta proteína en particular, el hecho de observar ciertos cambios en presencia de ligandos (incluyendo Gln) por medio de fluorescencia y barrido de CD, pero no con Desnaturalización Térmica o MST, nos hizo pensar que las interacciones eran muy débiles, o bien, la proteína era capaz de interactuar con ligandos pero no era capaz de estabilizarlos, posiblemente por problemas en el cerrado de la proteína.

Con la misma idea en que se concibió a la mutante LAOBP\_sca\_rosetta, y en la sospecha de que el sector rojo era el único responsable del cambio en la especificidad, se dio lugar a la proteína LAOBP\_sca\_redsector\_rosetta, en la cual las mutaciones sólo sucedieron en el sector rojo y, con excepción de L117V, las demás mutaciones puntuales fueron las mismas. Dicha proteína era soluble incluso en expresión a 37 °C, y además de unir aminoácidos positivos, el barrido de CD indicó cambio en presencia de Gln. Sin embargo, las demás técnicas, incluyendo ITC, no mostraron indicios de interacción con Gln, por lo que podríamos decir que esta mutante se comporta igual que muts, sólo que es más estable.

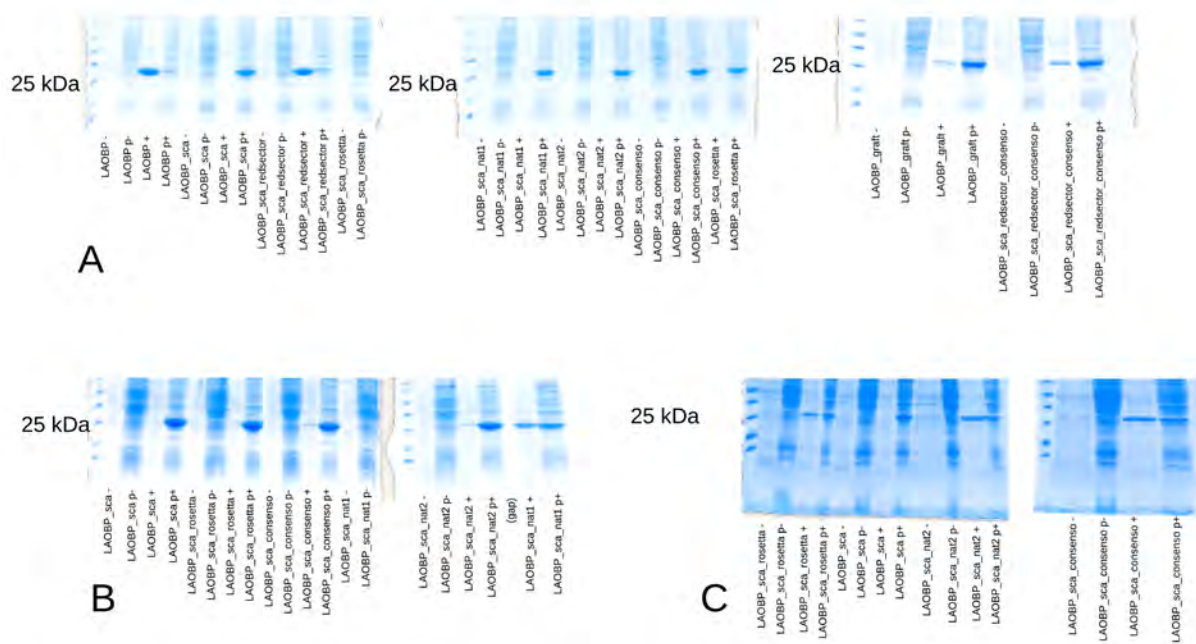
Continuando la exploración del sector rojo de LAOBP exclusivamente, decidí aplicar un consenso en sólo ese sector como en el caso de LAOBP\_sca\_consenso, dando lugar a la proteína LAOBP\_sca\_redsector\_consenso; la

cual se comporta esencialmente igual a LAOBP\_sca\_consenso pero más estable, por lo que pudimos corroborar que, según ensayos de ITC, es incapaz de unir Gln, aunque el barrido de CD indica cierto cambio con tal ligando. De hecho, después de observar los efectos que la mutación L117K provocaba por sí sola, pensamos que esta mutante sería más específica por Gln haciendo un *sca-grafting* de sólo el sector rojo, lo cual podíamos lograr a partir de la secuencia de DNA de LAOBP\_sca\_redsector\_consenso. Aunque la nueva proteína, LAOBP\_sca\_redsector; según ensayos de desnaturalización térmica, era capaz de unir Gln, el cambio no era mejor que con Arg, como tampoco era tan drástico como en el caso LAOBP\_L117K, así que decidimos no hacer más experimentos con LAOBP\_sca\_redsector\_consenso.

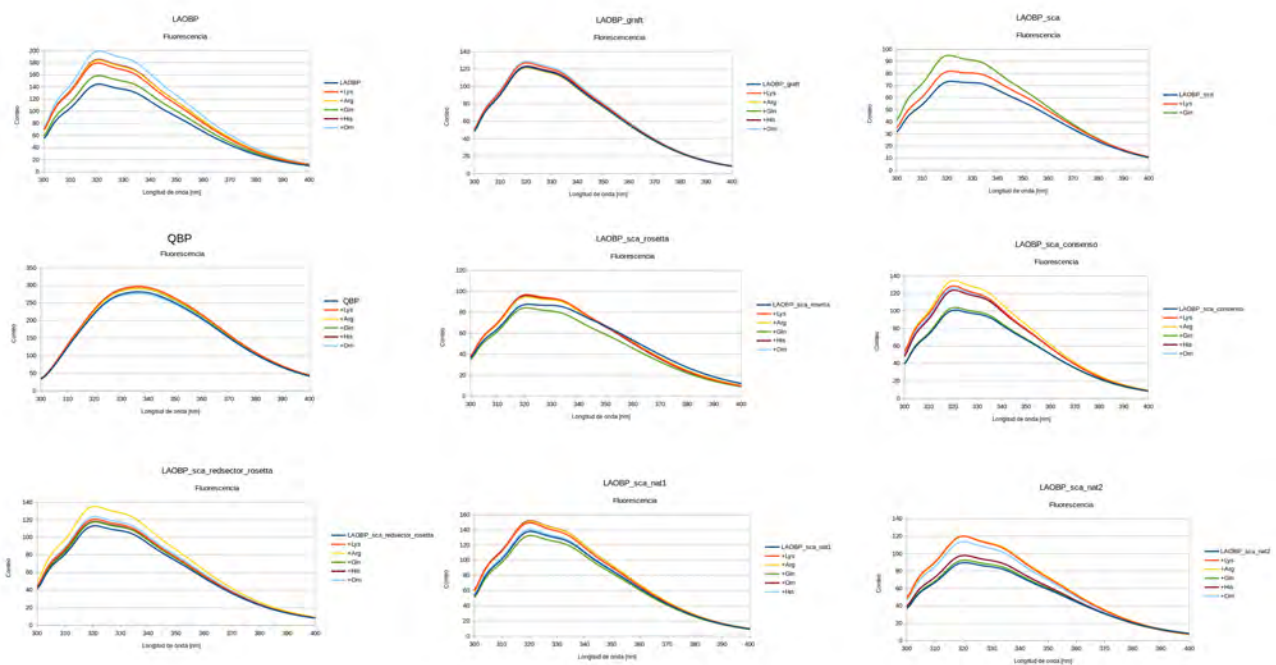
En el caso de las proteínas en que se buscaron combinaciones de aminoácidos acoplados presentes en proteínas naturales, LAOBP\_sca\_nat1 y LOBP\_sca\_nat2, quienes presentan exactamente las mismas mutaciones salvo D220N, es claro que el fenotipo sigue siendo similar al de LAOBP, no hay unión de Gln, y que LAOBP\_sca\_nat2 es menos estable que LAOBP\_sca\_nat1, por lo que es claro que la mutación D220N es desestabilizante, lo cual explica también los problemas que tuvimos con LAOBP\_sca\_consenso, LAOBP\_sca\_rosetta y la misma LAOBP\_sca, quienes presentan la misma mutación puntual; razón por la cual revertir el sitio 220 a un aspártico en LAOBP\_sca (LAOBP\_sca\_220D) da lugar a una proteína más estable, pero con un comportamiento similar al de LAOBP\_sca, siendo casi exclusiva por Gln con respecto a Lys (ver Figura 16).

Es importante aclarar que sometí a todas las mutantes mencionadas en la Tabla S2 a simulaciones de *Docking*, tal como lo describo en la sección de la Metodología, en donde contrasté usando como controles a la misma LAOBP y otras mutantes al azar en la cavidad de unión o sin la Gln en dicha cavidad. Sin embargo, de acuerdo al RMSD y a inspecciones de algunos resultados de *docking* (Figura docking), el ligando nunca salió de la cavidad de unión, puesto que las diferencias se mantuvieron en un rango de entre 0.12 y 3.67 Å, lo cual indicaría, de forma inverosímil, que tanto LAOBP silvestre como los diseños son capaces de unir Gln, por lo cual el *docking* no nos pareció particularmente útil para decidir cuales diseños serían exitosos y cuales no en caso de realizar pruebas experimentales.

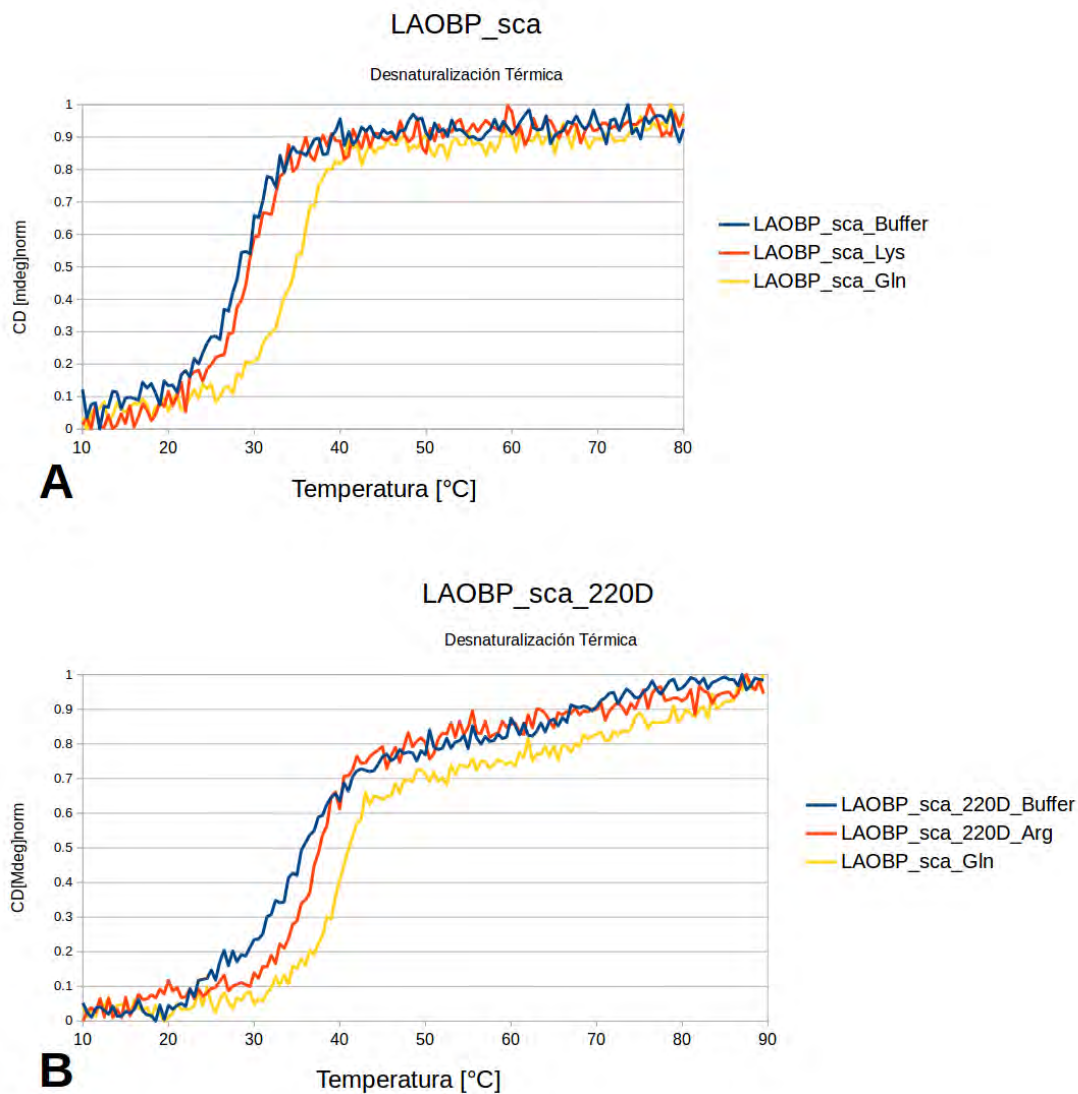




**Figura 14. Pruebas de expresión de las mutantes en la primera fase de diseño.** A) y B) fueron expresadas a 37 °C y 25 °C, respectivamente en células *E. coli* BL21-DE3, mientras que en C) se usaron células Artic Expression (AEX), quienes son unas células BL21-DE3 con un plásmido que expresa una chaperonina para ayudar en la incubación a 11 °C. (-) Extracto procedente de células sin inducción de expresión, (+) Extracto de células inducidas, (p) Precipitado *Pellet*.



**Figura 15. Fluorescencia.** Determinación de posible interacción con los ligandos de interés con las proteínas mutantes de la primera ronda de diseño en contraste con las proteínas LAOBP y QBP silvestre.



**Figura 16. La mutación D220N es desestabilizante.** Ensayos de Tmelt antes y después de revertir la mutación en el sitio 220.

***El sitio 117 siempre fue mutado.***

Dado que el sitio 117 se encuentra en la región considerada como cavidad de unión y está acoplado estadísticamente en ambos análisis, no es sorprendente que en todas las proteínas diseñadas y estudiadas experimentalmente, así como un consenso de la cavidad que no llegó a probarse (LAOBP\_sphere\_consenso), dicho sitio fuese mutado. Lo que sí es de interés es que, cuando se utilizó Rosetta para sugerir mutaciones, la sustitución era hacia otro residuo hidrofóbico, como es el caso de L117I en LAOBP\_sca\_rosetta y L117V en LAOBP\_sca\_redsector\_rosetta, mostrando ambas proteínas indicios de unión por aminoácidos básicos pero nada concluyente con Gln (ver Tabla S2). Por lo que observamos en las mutaciones sugeridas en otros sitios, estos tendieron a ser por aminoácidos pequeños e hidrofóbicos, atribuible a la experiencia de que Rosetta tiende a usar residuos pequeños e hidrofóbicos en cavidades de proteínas (experiencia propia y comunicación personal con desarrolladores del software Rosetta). Por otra parte, la única mutación alternativa hacia un residuo no



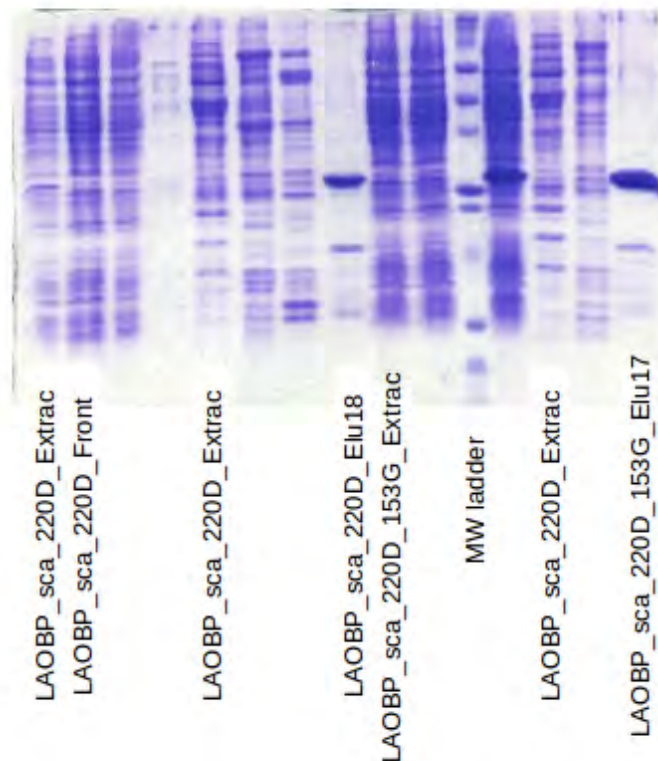
hidrofóbico fue hacia glutamina, debido a su alta frecuencia en el alineamiento del universo QBP-like (ver Figura S4), mostrando todas ellas una especificidad por los aminoácidos básicos de acuerdo a, por lo menos, incrementos en la  $T_m$ , pero nunca con glutamina (Tabla S2).

### ***Haciendo un SCA-grafting de únicamente el sector rojo es posible unir Gln con una afinidad comparable a Arg.***

Las proteínas analizadas experimentalmente en que se sugirió la mutación L117Q, también se sugirieron las mutaciones L55I y S72T, en donde LAOBP\_sca\_redsector\_consensu, quien solo presenta esas tres mutaciones, muestra unión por Lys con un incremento de la  $T_m$  en 5 °C, pero ninguno con Gln (únicos dos ligandos utilizados con esta proteína). Sin embargo, y como comenté anteriormente, a partir de esta mutante pude obtener una versión de *SCA-grafting* de únicamente el sector rojo al inducir el cambio L117K. Esta triple mutante presentó un incremento en la  $T_m$  de 2 °C en presencia de Gln o Arg, lo cual indica un papel importante de tener una lisina en la posición 117 para lograr unión de Gln (ver Tabla S2). Es posible que de haber tenido esta mutante LAOBP\_sca\_redsector desde la primera ronda de diseños, se hubiera llegado a identificar la mutación L117K desde la primera ronda.

### ***Mutar el sitio 220 tiene un impacto en la estabilidad.***

A excepción de LAOBP\_sca, quien siempre se fue a cuerpos de inclusión en todas las condiciones de expresión probadas, en todas las proteínas que presentaron la mutación D220N y analizadas experimentalmente fue necesario expresar a 11 °C para observar proteína soluble (ver Figura 14). Como mencioné al describir a las otras proteínas intentadas en mi proyecto, la mejor comparación para llegar a esta conclusión sobre el sitio 220 fue el comportamiento de LAOBP\_sca\_nat1 y LAOBP\_sca\_nat2, quienes solo difieren en que esta última presenta D220N como mutación adicional. Ambas proteínas, presentan el mismo comportamiento fisicoquímico, excepto en la temperatura máxima para observar expresión, pues LAOBP\_sca\_nat1 puede expresarse a 25 °C y, una vez purificada, no se agrega durante los ensayos de ITC (ver tabla S2). El sitio 220 forma parte del sector azul, quien dada su distribución en la estructura de LAOBP sospechamos que está relacionado con el plegamiento (ver Figura 8). Al observar detenidamente la estructura de LAOBP, el residuo D220 se encuentra en la última  $\alpha$ -hélice y enlaza polarmente a la R40, quien está en la primera  $\alpha$  hélice; funcionando, junto con los residuos adyacentes, como una “grapa” (ver Figura S8). En el caso de LAOBP\_sca, quien presenta la mutación D220N, no se expresa ni a 11 °C. Sin embargo, como ya se mencionó, una vez recuperada de cuerpos de inclusión y purificada, LAOBP\_sca presenta un aumento en la  $T_m$  de 1 °C por Lys y 7 °C por Gln; y al revertir únicamente la mutación D220N, el incremento en  $T_m$  con un ligando positivo, como Arg, es de 1.5 °C, pero con Gln es de 5.5°C; sin embargo, ahora es posible observar expresión soluble incluso a 37 °C (ver Figuras 16 y 17), confirmando no solo afinidad por Gln, sino una mayor especificidad por este ligando, si bien su  $K_d$  pudiera ser mayor (y por lo tanto peor) que la de la mutante LAOBP\_L117K. En este sentido, el *SCA-grafting* ofrece una forma de identificar sitios clave para transformar un fenotipo en otro. Adicionalmente, dado que la mutación G153N sólo se presentaba en LAOBP\_sca, también quisimos hacer una doble revertante LAOBP\_sca\_220D\_153G, deseando obtener una proteína aún más estable (ver Figura 17), lo que parece ser cierto a juzgar por la intensidad de la banda en el gel, sin embargo, no continuamos con más experimentos con esta mutante dado nuestro éxito con LAOBP\_L117K.



**Figura 17. Mutantes de la segunda fase de diseño.** Después de las pruebas de expresión A) LAOBP\_sca\_220D, LAOBP\_sca\_220D\_153G y LAOBP\_L117K fueron purificadas por afinidad a níquel.

***La proteína LAOBP\_graft es soluble pero su comportamiento con L-Glutamina es incierto***

Los resultados de LAOBP\_graft nos revelan dos pistas. La primera es que, de acuerdo a la intensidad de las bandas, las mutaciones que esta mutante presenta no alteran críticamente la estabilidad de la proteína, dado que alrededor del 40 % del péptido es aún soluble expresándose a 37 °C. La segunda pista es, que las mutaciones introducidas, aunque no necesariamente todas ellas, dan lugar a un aparente cambio de contenido de estructura secundaria en presencia de Gln (Figura 6B). Aunque este experimento fue el único en mostrar un cambio en presencia de Gln, podría sugerir la presencia de interacciones favorables que se forman con Gln.

***LAOBP\_sca une L-Glutamina, pero necesita tratamiento para plegarse***

La proteína LAOBP\_sca, cuyo diseño de la secuencia provino de la comparación de dos SCA, es insoluble en todas las condiciones de expresión probadas (Figura 14). Sin embargo, después de solubilizar y permitir el replegamiento de los cuerpos de inclusión, los ensayos de desnaturalización térmica muestran un claro aumento de la T<sub>m</sub>, siendo esta ~33 °C en presencia de L-Glutamina, mientras que la proteína sola tiene una T<sub>m</sub> de ~25 °C, o de ~26 °C en presencia de L-Lisina. Por lo tanto, la combinación de mutaciones llevó a la especificidad deseada, aunque dicha combinación nos dio problemas de estabilidad.

### ***Una comparación de las mutaciones compartidas por ambos enfoques de diseño mutacional nos reveló un cambio clave: La mutación L117K a partir de LAOBP permite la unión de L-Glutamina***

La más radical de las mutaciones compartidas entre las estrategias de diseño (proteína LAOBP\_graft) y SCA (proteína LAOBP\_sca), es decir, L117K, ofreció lo mejor de las dos estrategias abordadas: La capacidad de unir Gln (Figuras 7, 10 y 11) y la solubilidad de la proteína expresada, que se refleja en una  $T_m$  cercana a LAOBP. Como puede verse en la figura 7, LAOBP\_L117K presentó un claro incremento en la tolerancia a la temperatura en presencia de L-Glutamina en comparación con la proteína sola (la  $T_m$  incrementó  $\sim 9$  °C), sugiriendo una reorganización de las interacciones débiles entre los residuos de la proteína, posiblemente debido al cambio conformacional asociado a un fuerte desplazamiento hacia el estado plegado (Sanchez-Ruiz, 2007). De hecho, en la mejora con Gln en esta mutante es incluso mayor que en presencia de Arg, uno de los ligandos preferidos por LAOBP silvestre, demostrando un intercambio de especificidad. Así pues, se concluye que la mutación puntual L117K es una solución suficiente para lograr la unión novedosa a Gln sin comprometer demasiado la estabilidad de la proteína, pues su  $T_m$  en ausencia de ligandos de  $\sim 46$  °C, mientras que la  $T_m$  de la proteína silvestre es de  $\sim 49$  °C en las mismas condiciones (ver Figura 7). Las otras mutaciones presentes en la proteína LAOBP\_graft y/o en LAOBP\_sca necesitarían ser analizadas experimentalmente para determinar el impacto negativo en cuanto a unión a ligando y/o expresión. Por ahora, LAOBP\_L117K es la mejor solución para nuestro propósito. Adicionalmente, es sorprendente que la  $K_d$  promedio ( $1.6 \mu\text{M}$ ) para L-glutamina se encuentre en el mismo orden de magnitud que el reportado para la QBP silvestre ( $K_d = 0.5 \mu\text{M}$ , Sun *et al.*, 1998). A pesar de que fue Sun *et al.*, 1998 quienes sugirieron que la diferencia en los sitios L117 y Q160 en LAOBP con respecto a los equivalentes en QBP (K115 y H156, respectivamente) sean los más sospechosos para explicar la especificidad de unión entre ambas proteínas sin demostrarlo experimentalmente, la combinación de los enfoques tomados en este trabajo, *Binding Pocket Grafting* y *SCA-grafting*, nos permitió encontrar una solución incluso más minimalista. Ambos enfoques ofrecen ventajas y desventajas que, en nuestro caso, no muestran otro sobrelape que algunas pistas de interacción positiva con L-Glutamina. Resulta curioso para mí que LAOBP\_graft presente tanto la mutación L117K como la Q160H sin ofrecer resultados concluyentes de unión de Gln, haciéndome pensar en lo impredecible de las mutaciones antes de los ensayos experimentales, pues antes de concluir que la mutación L117K era el camino más rápido para lograr unión por Gln presenté la figura 3B en cuatro exámenes tutorales, en donde tanto mi tutor principal como mis cotutores tienen preparación académica y práctica en bioquímica (incluso uno de ellos es cristalógrafo), tres seminarios para los cinco grupos de investigación que conforman el laboratorio en donde realicé mi doctorado, un seminario departamental para la comunidad de Bioquímica de la Facultad de Medicina de la UNAM, dos congresos internacionales de bioquímica y diseño de proteínas, tres seminarios en el grupo de la Dra. Birte Höcker en el Instituto Max Planck de Tübingen, Alemania; y nunca, ni una sola vez, hubo alguien que me sugiriera mutar únicamente el sitio L117 por una lisina. En cuanto mostré los resultados de la mutante puntual LAOBP\_L117K, literalmente decenas de personas me han dicho que el resultado era perfectamente predecible. Existe el término *Sesgo de la retrospección* (también llamado *Prejuicio de Retrospectiva* o *Recapitulación*) el cual describe perfectamente este tipo de errores cognitivos en que una vez que los individuos conocen el resultado final, modifican su recuerdo al pensar que siempre supieron lo que iba a ocurrir, lo cual se observa fácilmente cuando se registran las predicciones de los individuos y se demuestra objetivamente que nadie había acertado el resultado antes de conocerlo (Kahneman, 2011).

### ***LAOBP\_L117K presenta un perfil de especificidad diferente comparado con LAOBP o QBP***

Aunque LAOBP\_L117K es una solución para la unión de Gln, tal molécula no es la única, ni la preferida, como ligando a unirse a esta proteína. Como puede verse en la Figura 10, los valores de las  $T_m$  que antes solo cambiaban con aminoácidos positivos (L-arginina, L-lisina, L-ornitina y L-histidina) para LAOBP, ahora lo hacen con ligandos sin carga (L-cisteína, L-glutamina, L-citrulina, alanina y L-serina) para LAOBP\_L117K. En cambio, QBP es específica para únicamente Gln. Se sabe que cierta robustez es necesaria para lograr un cambio de fenotipo, involucrando un cambio gradual de una función previa a una nueva (Stiffler *et al.*, 2015). En este sentido, no es sorprendente que LAOBP\_L117K presente un perfil intermedio en cuanto a su especificidad, la cual no es tan fuerte y delimitada como en LAOBP para aminoácidos positivos, o altamente específica para Gln, como es QBP. Sin embargo, lo que es realmente interesante, es haber encontrado un atajo para lograr la unión de una molécula que anteriormente no era reconocible de forma detectable por LAOBP (Clifton & Jackson 2016, Figuras 7, 10, 11 y Tabla 3). La combinación de estas estrategias nos permitió identificar no sólo un sitio clave para lograr el cambio del perfil de especificidad, sino también el residuo por el cuál necesitábamos mutar para lograr tal cambio, yendo de una leucina a una lisina.

### ***Posibles explicaciones por las que la mutación L117K impacta directamente en la unión de L-Glutamina***

Por la Figura 10, es evidente que una lisina en la posición 117 es un transformador de la especificidad en LAOBP, mientras que otro residuo en la misma posición, como una glutamina, mantiene el fenotipo de especificidad limitado a los aminoácidos positivos, así que podemos concluir que cambiar al residuo hidrofóbico leucina por un residuo sin carga, como es la glutamina, tiene poco o ningún impacto para nuestros objetivos. Esto es interesante, considerando que es precisamente glutamina el residuo más común en la posición equivalente de nuestro conjunto QBP-like de secuencias de proteína alineadas (Figura S4). Sin embargo, glutamina es también el residuo más común en el MSA del conjunto QBP\_like, y es también el residuo encontrado en la posición equivalente (123) del ancestro AncQR de Clifton & Jackson (2016), el cual es capaz de unir L-glutamina con una afinidad milimolar. Tal hecho es lo que podría explicar que en mis construcciones, la mutación L117Q de lugar a un comportamiento esencialmente igual al de LAOBP. Por otra parte, la lisina es un residuo cargado positivamente, el cual, posiblemente, cambia el potencial electrostático de la entrada a la cavidad de unión, haciendo difícil que los ligandos con carga positiva puedan llegar a la cavidad de la PBP (como un fenómeno de selectividad, Najmanovich 2017) e incrementando las oportunidades de que otros aminoácidos, especialmente los que no presentan carga, queden atrapados.

Sin embargo, esta hipótesis no puede sustentarse meramente al usar residuos con cadenas laterales positivas en la posición 117, dado que la mutación L117R no es una mejor solución para unir L-glutamina según ensayos ITC (Tabla 5). Esto es interesante, considerando que la arginina es un residuo más común en la posición equivalente de nuestro universo QBP-like de lo que es en nuestro universo LAOBP-like (Figura S4). Es posible que la carga positiva no esté tan bien localizada debido a los tres grupos amino del grupo guanidinio de la arginina, quienes podrían hacer otras interacciones con los residuos de los alrededores. Tal problema podría evitarse con un único grupo amino en la punta de la cadena lateral, como lo es en el caso de la lisina. Podría también pensarse que el tamaño de la arginina ocupando el sitio 117 pudiera tener algún impacto en el tamaño de los ligandos preferidos, pero recordemos que la L-citrulina, un aminoácido con un volumen similar al de la Arg, puede ser unida en el mismo orden de magnitud en LAOBP\_L117K, LAOBP\_L117R y LAOBP silvestre (ver Figura 10 y Tabla 5).

Por otra parte, con las otras mutantes probadas experimentalmente a lo largo de este estudio (Tabla S2), existen indicios de que el sitio 117 es clave para lograr la modificación de la especificidad, y tal vez hubiésemos llegado más rápidamente a esta conclusión de haber probado a la mutante LAOBP\_sca\_redsector desde la primera ronda

de diseños.

Todo esto refleja el problema de que descripciones químicas más precisas del complejo proteína-ligando no conducen, generalmente, a predicciones más precisas de la afinidad de unión (Ballester *et al.*, 2014) dado que aspectos como la dinámica intrínseca y los grandes cambios conformacionales de las proteínas pueden impactar las propiedades del sitio de unión a ligandos (Yang & Lain, 2017).

En resumen, hemos presentado un caso en el que las observaciones hechas en dos diferentes estrategias de diseño, para intentar lograr la unión de L-glutamina por una PBP, nos llevó a una única mutación puntual, la cual resultó ser suficiente para dar lugar a una proteína capaz de unir L-glutamina casi tan bien como una proteína silvestre de unión a glutamina. La probabilidad de haber logrado esto por azar es igual a uno dividido por el producto del tamaño de la proteína andamio en residuos (LAOBP), multiplicado por los residuos alternativos por posición ( $1/238 \times 19$ ), que es 0.000221141. Tal probabilidad refleja la cantidad de alternativas no racionales (4522) que hubiesen tenido que probarse experimentalmente para obtener el mismo resultado, haciendo que la perspectiva tomada por nosotros en este trabajo fuera una estrategia rápida y barata para alcanzar nuestro objetivo general.

### **Comparación con otros trabajos**

La mutante puntual LAOBP\_L117K, deducida a partir de la comparación entre la estrategia de *binding pocket grafting* y la de *sca-grafting*, y adicionalmente confirmada de forma indirecta por otra proteína en que mutamos el sitio 117, es una solución con efectos positivos y más drásticos que los abordados en otras publicaciones basadas en PBPs. Para comenzar, los resultados experimentales son perfectamente reproducibles, lo cual no sucede con lo afirmado por Looger *et al.*, 2003 con respecto a PBPs supuestamente capaces de unir TNT u otros ligandos, lo cual fue refutado por Schreier *et al.*, 2009. En nuestro caso particular, el grupo de la Dra. Birte Höcker, primero en el Instituto Max Planck de Tübingen y posteriormente en la Universidad de Bayreuth, ambas instituciones en Alemania, pudo observar los mismos fenómenos de Desnaturalización Térmica, fluorescencia y afinidad por ITC que yo en los laboratorios de la Facultad de Medicina e Instituto de Fisiología Celular, ambos en la UNAM, México, a partir de copias del mismo DNA que da lugar a la proteína LAOBP\_L117K, la cual fue confirmada por secuenciación y de forma independiente en ambos países.

Con respecto al trabajo de Scheib *et al.*, 2013, en el que se hizo un *binding pocket grafting* de una PBP de unión a espermidina en una de unión a putrescina, recordemos que la afinidad lograda en la mutante por espermidina es comparable a la ya presente por esta molécula antes del trasplante, por lo que no hubo un intercambio de afinidades pero sí un incremento en la especificidad por espermidina, pues se abatió la afinidad natural por putrescina. Con respecto a las dos formas de hacer injertos abordadas en mi trabajo, la de *SCA-grafting*, dando lugar a LAOBP\_sca, ya es en sí una solución para unir Gln, pues aunque no se provoca un total abatimiento de la unión por aminoácidos positivos como Lys, se obtuvo el fenotipo de unión a Gln, el cual es drástico por pasar de no ser detectable antes de la mutación puntual a ser comparable ( $K_d$  de 40 micromolar) al de la proteína natural de unión a Gln, llamada QBP, además de presentar casi una total especificidad por esta nueva molécula (ver Figura 7C), siendo una solución aún mejor en cuanto a afinidad con respecto al ancestro AncQR deducido y reportado por Clifton & Jackson en 2015, el cual presenta varias mutaciones con respecto a LAOBP y su afinidad por Gln es milimolar. Si bien no obtuvimos resultados concluyentes de unión a Gln al usar la estrategia de *binding pocket grafting*, el comparar las mutaciones de esta estrategia con *SCA-grafting* nos permitió deducir que la mutante puntual LAOBP\_L117K, la cual es más estable que LAOBP\_sca, su  $K_d$  por Gln es aún mejor que en esta última, aunque con una capacidad de unión por los aminoácidos positivos no tan disminuía pero sí mucho más que con la proteína LAOBP silvestre.

Por último, mientras corregía esta tesis durante agosto del 2018, encontré en en la misma época en que mi artículo estaba siendo revisado en *Protein Science*, otro artículo afín fue revisado para su publicación en la revista *Biochemistry* durante el invierno del año pasado. En dicho artículo, Liu *et al.*, 2018 usaron SCA para guiar el diseño de una librería de mutantes que les permitió, en una sólo ronda, descubrir interesantes luciferasas mutantes, las cuales son proteínas capaces de emitir luz. Las justificaciones que estos autores mencionan son prácticamente las mismas que yo menciono en mi artículo y en esta tesis: los problemas que la mutagénesis de proteínas conlleva por el surgimiento de mutaciones neutrales o con efectos negativos para la propiedad de interés, así como la presencia de sinergia entre las mutaciones, las cuales se comportan de forma no aditiva (epistasia). A diferencia de las 13 mutantes que yo tuve que analizar (10 si sólo contamos a las que involucraron SCA, y 8 si de esas sólo consideramos a las que llegaron a fase experimental), Liu y sus colaboradores analizaron experimentalmente 2666 mutantes de  $10^4$  posibles, 330 de las cuales emitieron luz; presentando cada una un promedio 5.8 mutaciones puntuales. Por estos números, los autores concluyen que SCA es una herramienta muy eficaz para diseñar librerías con una alta proporción (12%) de mutantes exitosas, además de que es posible detectar sinergia entre mutaciones, las cuales difieren de las posibles por mera inspección de la estructura de la proteína silvestre al proponer mutaciones que no son intuitivas, así como descartar mutaciones de la cavidad de unión por la presencia de otras mutaciones que reacomodan tales sitios, o incluso el surgimiento de mutaciones que compensen mutaciones útiles para la propiedad de interés pero que por sí solas son desestabilizantes. En lo personal, me siento muy contento de que las justificaciones y conclusiones que escribí en mi artículo coincidan con las de Liu *et al.*, 2018, así como también me siento muy aliviado de no haber hecho librerías de mutantes durante mi proyecto porque así mi trabajo y el de Liu *et al.*, 2018 se complementan.

## Conclusiones

Pudimos alcanzar nuestro objetivo general de cambiar la especificidad de LAOBP para lograr la unión de Gln, el cual es un ligando que la proteína silvestre no es capaz de unir, y nuestra hipótesis sobre los sitios estadísticamente acoplados para cambiar la especificidad resultó ser cierta. Sin embargo, gracias a que combinamos la información de SCA con la de una metodología diferente, como lo fue *Binding Pocket Grafting*, pudimos delimitar la solución para unir Gln a un sólo residuo de la proteína. Esto, sin embargo, no necesariamente es la única solución posible al mismo problema, como tampoco lo es que variar únicamente el sitio 117 de LAOBP, o equivalente en la proteína ancestral, sea la solución que la evolución, por medio de la selección natural, encontró para dar lugar a proteínas con especificidad diferente. Dado que el sitio 117 está acoplado tanto en el conjunto LAOBP-like como en el QBP-like, no es necesario mutar los demás sitios acoplados con él para lograr la unión de Gln debido precisamente a que las perturbaciones fisicoquímicas en el sitio 117 pueden tener efectos mínimos pero suficientes en los sitios acoplados para nuestro objetivo, además de que no todos los demás sitios acoplados con el 117 son diferentes entre LAOBP y QBP. Si bien es cierto que podríamos intentar mutar los sitios acoplados a 117 para lograr que la especificidad por Gln sea exclusiva, ese no fue el objetivo de este proyecto. Sin embargo, lo que he aprendido sobre aplicar SCA a conjuntos de PBPs puede ser aprovechado para lograr intercambios más drásticos y finos entre LAOBP y QBP.

## Perspectivas

-Sería interesante hacer la resonancia magnética nuclear (NMR, por sus siglas en inglés) de la proteína LAOBP\_L117K, pues dado que el grupo del Dr. Carlos Amero en Cuernavaca ya tiene asignación de NMR para LAOBP silvestre, una variante en un único sitio de la proteína con un fenotipo tan distinto seguramente ha modificado su dinámica interna. Ojalá pronto tuviésemos la comparación de ambas proteínas, con y sin ligandos.

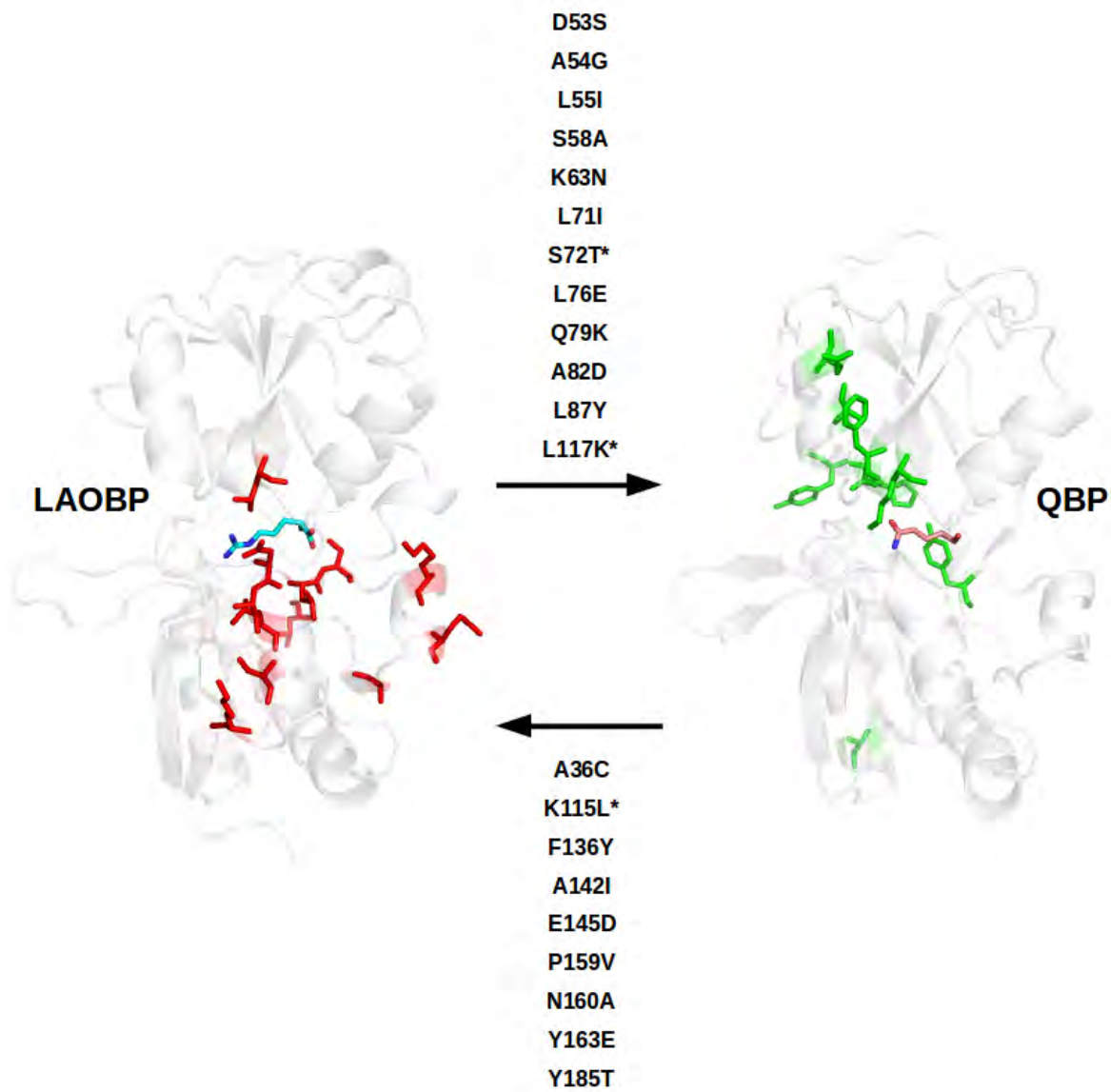
Ahora bien, si como se mencionó en la conclusión, un siguiente objetivo fuera intercambiar completamente la especificidad de LAOBP por la de QBP y viceversa, con el menor número de mutaciones posible, necesito entonces plantear lo siguiente. Pensando en la importancia que los residuos secundarios que tanto he mencionado a lo largo de este texto, la capacidad de SCA para considerar estos e incluso otros sitios aún más alejados de la cavidad de unión, así como la aparente mayor especificidad (pero no tanta afinidad) por Gln de la proteína LAOBP\_sca que LAOBP\_L117K, y que un sólo sector sea el mero responsable de la modificación de la especificidad tanto en LAOBP como en QBP, propongo lo siguiente:

Intercambiar únicamente los sitios que pertenecen a sectores equivalentes en ambas proteínas. En nuestro caso, el sector rojo de LAOBP, como ya comenté, es el más sospechoso para la unión de ligandos, además de ser el único que involucra al sitio L117. Análogamente, y de acuerdo a la tabla 6, el sitio K115 en QBP aparece acoplado, aunque el sector al que pertenece, el verde en este caso, involucra otros sitios. En ambos casos, los sitios estadísticamente acoplados no están enriquecidos en sitios de contacto directo con el ligando en cuestión, pero sí con los residuos secundarios y otros evidentemente alejados de la cavidad de unión (ver Figuras 3 y 18), los cuales pueden influir en reacomodo de los residuos primarios para unir ligandos distintos a los originales. La idea que propongo consiste en intercambiar los sitios del sector rojo de LAOBP por los correspondientes de QBP para alcanzar una especificidad por Gln, mientras que en QBP mutemos los sitios de su sector verde por los correspondientes en LAOBP para provocar una especificidad por aminoácidos positivos en QBP (ver Figura 18).

**Tabla 6. Sitios estadísticamente acoplados.** Después de determinar los sectores de los conjuntos de secuencias LAOBP-like y, por otra parte, de QBP-like; las proteínas LAOBP y QBP fueron alineadas estructuralmente. Se muestra en nomenclatura de tres letras los residuos de LAOBP y sus anotaciones con minúscula, mientras que en los residuos de QBP los residuos están en nomenclatura de una letra las anotaciones en mayúscula. También se etiquetan como “primarios” a los sitios que, según las anotaciones de la base de datos PDB, están involucrados en la unión de ligando.

LAOBP	Anotaciones	QBP	Anotaciones	LAOBP	Anotaciones	QBP	Anotaciones
ASP11	primario	d10	PRIMARIO	LEU117	rojo/primario	K115	VERDE/PRIMARIO
TYR14	azul/primario	f13	PRIMARIO	GLY119	verde	G117	VERDE
PHE17	verde	-		SER120	verde/primario	T118	PRIMARIO
GLU25		L023	ROJO	THR121	primario	G119	PRIMARIO
GLY28	rojo	G026	ROJO/VERDE	GLN122	primario	S120	
ASP30		D028	ROJO	GLU123	verde	V121	
ILE31		V029	ROJO	ALA126	verde	A124	
ASP32		D030	ROJO	TRP130	verde	I128	
LEU33		L031	ROJO	GLY134	verde	-	
GLU36		A034	ROJO	TYR140	azul	F136	VERDE
CYS38	azul	A036	ROJO/VERDE	ILE146		A142	VERDE
MET41		L039	ROJO	ASP149	azul	E145	VERDE
GLN42		K040	ROJO	LEU150	verde	L146	VERDE
VAL43		L041	ROJO	GLY153	verde	N149	VERDE
CYS45	azul	Y043		ARG154	azul/verde	R150	
PHE52	rojo/primario	F050	AZUL/ROJO/PRIMARIO	ASP156		D152	VERDE
ASP53	rojo	S051		ALA157	verde	A153	VERDE
ALA54	rojo	G052		GLN160		H156	PRIMARIO
LEU55	rojo	I053	ROJO	ASP161	primario	D157	VERDE/PRIMARIO
ILE56	rojo	I054		VAL163		P159	VERDE
PRO57	rojo	P055		ALA164		N160	VERDE
SER58	rojo	A056		GLU167		Y163	AZUL/VERDE
LEU59	rojo	L057		PHE169	azul	F164	
LYS63	rojo	N061		GLY175	azul	N170	
ASP65	rojo	D063		LYS176		G171	AZUL
SER69	primario	A067	PRIMARIO	GLY182	azul	G177	AZUL/VERDE
SER70	primario	G068	PRIMARIO	LYS189		Q184	AZUL
SER72	primario	T070	ROJO/PRIMARIO	GLY192	azul	-	
ARG77	primario	R075	PRIMARIO	GLY194	azul	-	
GLN79	rojo	K077		THR195		Y185	VERDE/PRIMARIO
ILE81	rojo	I079		GLY196		G186	AZUL
ALA82	rojo	D080		ARG200	azul	P190	
PHE83	rojo	F081		LYS201	azul	K191	AZUL
SER84	rojo	S082		ASP202		G192	AZUL
LEU87	rojo	Y085		ASP203	azul	S193	
TYR88	rojo	Y086		LEU206	azul	L196	AZUL
GLY99	verde	N097		ALA209		K199	AZUL
SER100	verde	N098		ALA213	azul	A203	AZUL
SER108	verde	D106		LEU217		L207	AZUL
LEU109	verde	L107		ARG218		R208	AZUL
LYS110	verde	D108	AZUL/ROJO	ASP220	azul	N210	AZUL
GLY111	verde	G109		GLY221	azul	G211	AZUL
LYS112	verde	K110		THR222	azul	T212	AZUL
VAL114	verde	V112		TYR223	azul	Y213	AZUL
GLY115	verde	A113		MET226		I216	AZUL
VAL116	verde	V114		LYS229	azul	K219	
				TYR230	azul	W220	

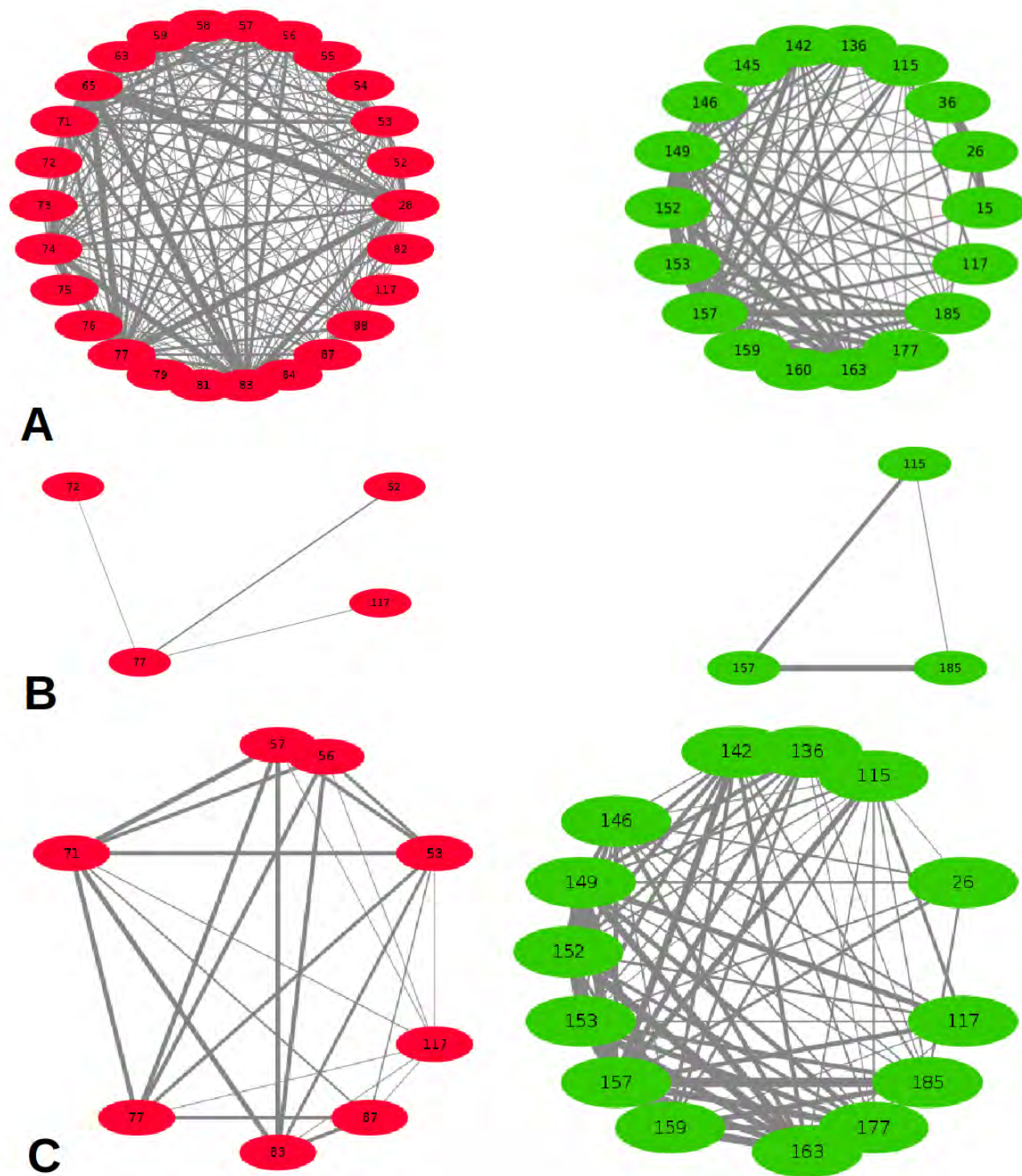




**Figura 18. Interconversión de especificidades por sector.** El sector rojo de LAOBP parece ser el más implicado en su especificidad, pues el residuo L117 se encuentra ahí y sólo ahí, mientras que el residuo equivalente en QBP es K115, quien en el universo QBP-like pertenece al sector verde. Intercambiar los sitios que son diferentes entre las proteínas LAOBP y QBP podrían ser suficientes para intercambiar por completo la especificidad de estas proteínas. Los sitios en cursivas son los únicos que pueden superponerse en el universo de mutaciones entre ambas proteínas. \* Sitios identificados como de unión a ligando según la base de datos del PDB.

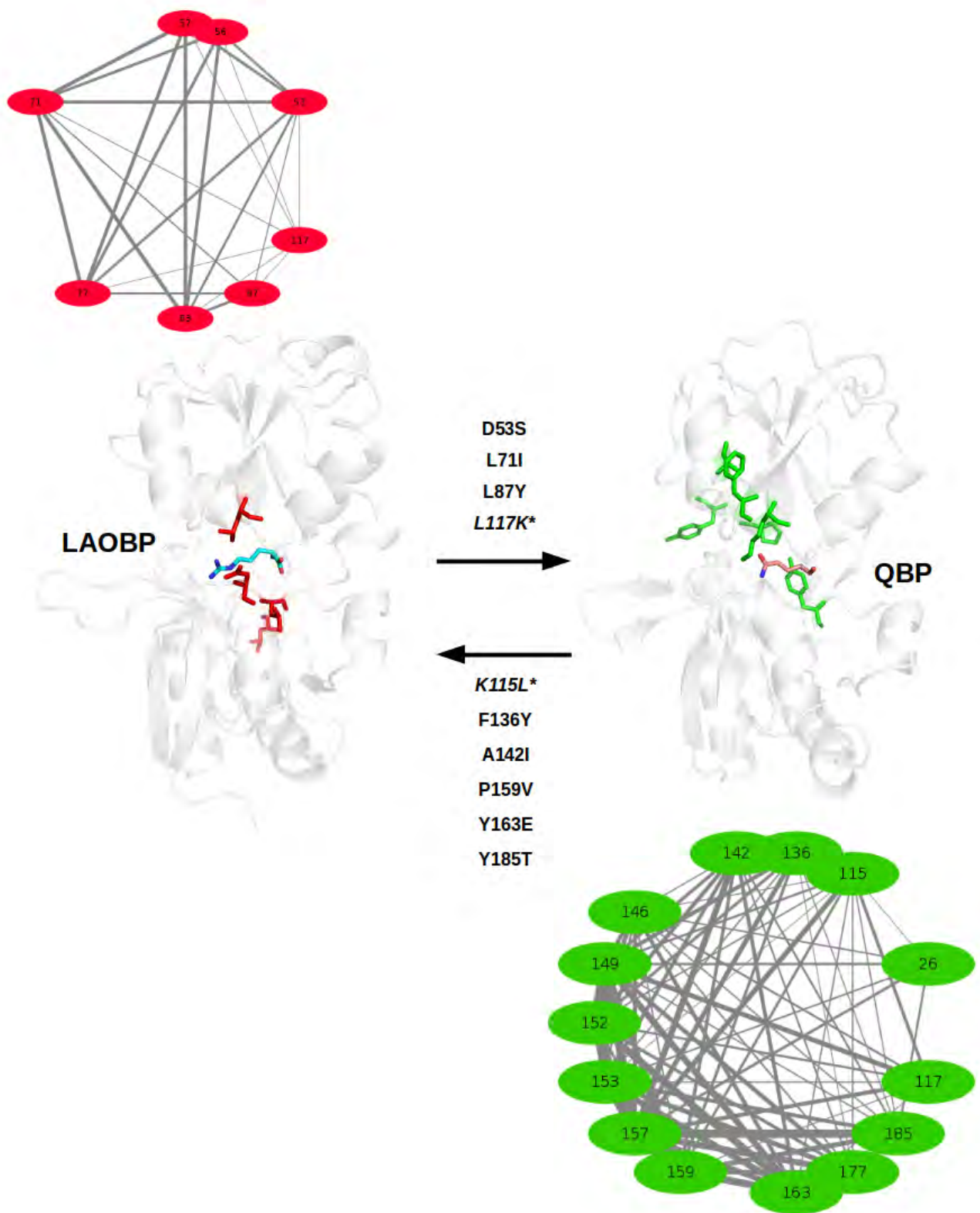
Sin embargo, el número de mutaciones para pasar de una especificidad a otra en la propuesta de la figura 18 es comparable tanto a la estrategia de *Binding pocket grafting* como a la de *SCA-grafting*. Considerando los

problemas para determinar unión de ligandos en LAOBP\_graft y la baja termoestabilidad de LAOBP\_sca, es posible que pudiéramos delimitar la información de los sitios estadísticamente acoplados si los vemos desde una perspectiva de redes (ver Figura 19).



**Figura 19. Redes de sectores de interés.** A) Redes de los sectores equivalentes, LAOBP en rojo y QBP en verde. El grosor de la línea entre posiciones es cualitativamente proporcional a su intensidad de acoplamiento estadístico. B) Subredes de acuerdo a sitios reportados como de contacto con el ligando (sitios primarios). C) Subredes de acuerdo a los vecinos directos del sitio 117 (LAOBP) o su equivalente en QBP (115).

Por la figura 19, es interesante que el sitio 117 de LAOBP (115 en QBP), quien forma parte de los residuos primarios, esté mucho más conectado con sitios secundarios (Figura 19C) que con otros sitios primarios (figura 19B), siendo precisamente las subredes de la figura 19C las que contienen mucho más conexiones internas, además de intensas entre los vecinos inmediatos de 117 (115) que en la figura 19B. Al combinar la información de las subredes de la figura 19C con el enfoque de la figura 18, podemos reducir el universo de mutaciones en la interconversión de especificidad entre LAOBP y QBP.



**Figura 20. Interconversión por residuos estadísticamente acoplados al sitio 117.** Gracias al enfoque de redes, podemos delimitar aún más los sitios que serán mutados en la proteína LAOBP (A) para alcanzar especificidad por glutamina. De manera análoga, en el contexto del sitio 115 de QBP (B), podemos alcanzar una especificidad por residuos positivos.

Curiosamente, el único sitio primario en LAOBP es precisamente el 117, como lo es también el equivalente estructural en QBP (115), mientras que los demás son sitios secundarios. Así, en el contexto del sitio 117, del cual ya sabemos que es primario y tiene por sí sólo un efecto muy grande en la modificación de la especificidad, y lo que he planteado sobre la importancia de los sitios secundarios, me atrevo a hacer la siguiente hipótesis: En el contexto de LAOBP, las mutaciones de los sitios secundarios en la figura 20 ayudarán a acomodar al sitio 117 y a otros sitios primarios (lo cuales no serán mutados) para alcanzar una geometría óptima para una especificidad por Gln; mientras que lo mismo pasará en el contexto de las mutaciones puntuales propuestas en la figura 20 para que la proteína QBP se vuelva específica por aminoácidos positivos.

Esta nueva hipótesis la hago por lo que me ha parecido que he aprendido de la proteína LAOBP y el estudio de su rediseño usando, principalmente, el Análisis de Acoplamiento Estadístico. El tiempo, los recursos y la voluntad de una o más personas que lleven a cabo los experimentos pertinentes con las mutantes que propongo en la figura 20 comprobará o refutará dicha hipótesis.

## **Agradecimientos**

A la Universidad Nacional Autónoma de México, UNAM; de la cual he sido estudiante desde el nivel medio superior. También, mis agradecimientos al Dr. Alejandro Sosa Peinado, por fungir como mi tutor de doctorado y confiar en mi visión para alcanzar el objetivo de cambiar la especificidad de una proteína, así como darme siempre la incondicional oportunidad de resolver todo tipo de problemas que se nos presentaron por mí mismo y apoyarme intelectualmente con entusiasmo. Así mismo, agradezco a la Dra. Georgina Regina Garza-Ramos Martínez, tanto por sus oportunas observaciones a mi trabajo teórico y experimental como a su apoyo moral a lo largo de mi estancia en el laboratorio.

Este trabajo fue parcialmente financiado por el apoyo IN215612-3 de la Dirección General de Apoyos al Personal Académico (DGAPA) de la UNAM, así como por el apoyo 167838 del Consejo Nacional de Ciencia y Tecnología (CONACyT), México, organismo que me dio una beca durante mi doctorado con el número de CVU 388734. Así también, agradezco al Programa de Apoyo a los Estudios de Posgrado (PAEP) por ayudarme económicamente para realizar actividades académicas en el extranjero, las cuales impulsaron en gran medida el desarrollo de este trabajo. Agradezco a la Dra. Birte Höcker y a su técnico Sooruban Shanmugaratnam del Instituto Max Plack y la Universidad de Bayreuth en Alemania, así como ella y nosotros agradecemos el apoyo del Centro Helmholtz de Berlín, y el uso de su sincrotrón. Agradezco también al Dr. Alfredo Torres Larios, quien fue mi cotutor durante mis estudios de doctorado y me permitió realizar experimentos clave para caracterizar la especificidad de proteínas mencionadas en este estudio. Así también, agradezco al Dr. León Martínez Castilla, quien también fue mi cotutor durante el doctorado y siempre me hizo pertinentes observaciones del orden informático y evolutivo, además de ayudarme a comprender el enfoque Rd.HMM.

Agradezco especialmente al Dr. Rogelio Rodríguez Sotres por su visión para que yo pudiera superar la limitante de la búsqueda de secuencias por medio de Rd.HMM y los primeros análisis con este enfoque. Obviamente, he siempre de agradecer a las doctoras Isabel Velázquez y Laura Álvarez por su asistencia técnica en el Laboratorio de Físicoquímica e Ingeniería de Proteínas de la Facultad de Medicina, UNAM, así como a mis



compañeros del laboratorio por sus críticas constructivas y destructivas.

## Bibliografía

1. Allert, M., Rizk, S. S., Looger, L. L., Hellinga, H. W. 2004. Computational design of receptors for an organophosphate surrogate of the nerve agent soman, *Proc Natl Acad Sci U S A.* 101(21):7907-12.
2. Baier, F., Tokuriki, N. 2014. Connectivity between catalytic landscapes of the metallo- $\beta$ -lactamase superfamily. *J. Mol. Biol.* 426: 2442-2456.
3. Ballester, P., Schreyer, A., Blundell, T. 2014. Does a more precise chemical description of protein-ligand complexes lead to more accurate prediction of binding affinity? *J. Chem. Inform. Model.* 54: 944-955
4. Banda-Vázquez, J.A., Shanmugaratnam, S., Rodríguez-Sotres, R., Torres-Larios, A., Höcker, B., Sosa-Peinado, A. 2018. Redesign of LAOBP to bind novel L-amino acid ligands. *Protein Sci.* 27(5):957-968. doi: 10.1002/pro.3403
5. Barelier, S., Sterling, T., O'Meara, M. J., Shichet, B. K. 2015. The Recognition of Identical Ligands by Unrelated Proteins. *ACS. Chem. Biol.* 2015.<http://dx.doi.org/10.1021/acscchembio.5b00683>.
6. Baum, B., Muley, L., Smolinski, M., heine, A., Hangauer, D., Klebe, G. 2010. Non-additivity of Functional Group Contributions in Protein-Ligand Binding: A Comprehensive Study by Crystallography and Isothermal Titration Calorimetry. *J. Mol. Biol.*, 397:1042-1054.
7. Bender, B. J., Cisneros, A. III., Duran, A. M., Fin, J. A., Fu, D., Lokits, A. D., Mueller, B. K., Sangha, A. K., Sauer, M. F., Sevy, A. M., Sliwoski, G., Sheehan, J. H., DiMaio, F., Meiler, J., Moretti, R. 2016. *Protocols for Molecular Modeling with Rosetta3 and RosettaScripts.* *Biochem.* 2016, 55: 4748-4763
8. Clifton, B., Jackson, C. 2016. Ancestral Protein Reconstruction Yields Insights into Adaptive Evolution of Binding Specificity in Solute-Binding Proteins. *Cell Chem. Biol.* 23: 1-10.
9. Coleman, M. L. & Chisholm, S. W. 2010. Ecosystem-specific selection pressures revealed through comparative population genomics. *Proc Natl Acad Sci USA.* 107(43): 18634-18639
10. Combs, S. A., De Luca, S. L., DeLuca, S. H., Lemmon, G. H., Nannemann, D. P., Nguyen, E. D., Willis, J. R., Sheehan, J. H. and Meiler J. (2013). Small-molecule ligand docking into comparative models with Rosetta. *Nature Protocols.* Vol 8, No. 7. doi:10.1038/nprot.2013.074
11. Dwyer, M. A., Hellinga. H. W. 2004. Periplasmic Binding Proteins: A Versatile Superfamily for Protein Engineering. *Curr. Opin. Struc. Biol.* 2004, 14:495-504.
12. Definición de biosensor de acuerdo al diccionario en línea Dictionary.com. Link: <http://www.dictionary.com/browse/biosensor>
13. Diccionario de la Real Academia Española. Asociación de Academias de la Lengua Española. Consultado durante el mes de Abril del 2018 Link: [dle.rae.es](http://dle.rae.es)
14. Diccionario de Oxford en Línea. Oxford University Press. Consultado durante el mes de Abril del 2018 para buscar la palabra "Afinidad". Link: <https://en.oxforddictionaries.com/definition/affinity>
15. Eddy, S. R. 1996. Hidden Markov Models. *Current Opinion in Structural Biology.* 6:362-365.
16. Eddy, S. R. 2011. Accelerated Profile HMM Searches. *PloS Comp. Biol.* 7:e1002195.
17. Genetics Home Reference. NIH: U.S. National Library of Medicine. Consultado durante el mes de Agosto 2018 para la definición de Mutación. Link: <https://ghr.nlm.nih.gov/primer/mutationsanddisorders/genemutation>
18. Glosario del software Rosetta. Consultado durante el mes de Abril del 2018. Link: [https://www.rosettacommons.org/docs/latest/rosetta\\_basics/Glossary/Glossary](https://www.rosettacommons.org/docs/latest/rosetta_basics/Glossary/Glossary)
19. Halabi, N., Rivoire, O. Leibler, S., Ranganathan, R. 2009. Protein sectors: evolutionary units of three-dimensional structure. *Cell.* 138: 774-786.
20. Henning, C. 2015. What are the true clusters?. Elsevier. *Pattern Recongnition Letters.* Vol.64:53-62.

<https://doi.org/10.1016/j.patrec.2015.04.009>

21. Huang, P., Boyken, S.E. & Baker, D. 2016. The coming age of de novo protein design. *Nature*. 537: 320-327
22. Jeffery, C. J. 2011. Engineering periplasmic ligand binding proteins as glucose nanosensors. *Nano Rev.* 2: 5743-5746.
23. Kahneman, Daniel. 2011. *Thinking, Fast and Slow*. Farrar, Straus & Giroux. 499 pp. ISBN: [978-0374275631](https://doi.org/10.1016/j.patrec.2015.04.009)
24. Kelly, S. & Price, N. C. 1997. The application of circular dichroism to studies of protein folding and unfolding. *Biochem. And Biophys. Acta.* 1338 (1997) 161:185.
25. Kufareva, I. & Abagyan, R. 2015. Methods of protein structure comparison. *Methods Mol Biol.* 857: 231-257. doi: 10.1007/978-1-61779-588-6\_10
26. Liu, M. D., Warner, E. A., Morrisey, C. E., Fick, C. W., Wu, T. S., Ornelas, M. Y., Ochoa, G. V., Zhang, B. S., Rathbun, C. M., Porterfield, W. B., Prescher, J. A., Leconte, A. M. 2018. Statistical Coupling Analysis-Guided Library Design for the Discovery of Mutant Luciferases. *Biochemistry.* 57: 663-671
27. Looger, L. L., Dwyer, M. A., Smith, J. J., Hellinga, H. W. 2003. Computational design of receptor and sensor proteins with novel functions. *Nature.* 423(6936):185-90.
28. Martínez-Castilla, L. P., Rodríguez-Sotres, R. 2010. A Score of the Ability of a Three-Dimensional Protein Model to Retrieve Its Own Sequence as a Quantitative Measure of Its Quality and Appropriateness. *PLoS ONE.* 9: e12483. doi:10.1371/journal.pone.0012483
29. McLaughlin Jr, R.N., Poelwijk, F. J., Raman, A., Gosal, W. S., Ranganathan, R. 2012. The spatial architecture of protein function and adaptation. *Nature.* 491: 138-142.
30. Mecozzi, S., West Jr, A. P., Dougherty, D. A. Cation- $\pi$  interactions inn aromatics of biological and medicinal interest: Electrostatic potential surfaces as a useful qualitative guide. *Proc Natl Acad USA.* Vol. 93:10566-10571
31. Merkl, R, Sterner, R. 2016. Ancestral Protein Reconstruction: techniques and applications. *Bio Chem* 397(1):1-21.
32. Miton, C. M., Tokuriki, N. 2016. How Mutational Epistasis Impairs Predictability in Protein Evolution and Design. *Prot. Society.* 25: 1260-1272.
33. Muller, A., Thomas, G. H., Horler, R., Brannigan, J.A., Blagova, E., Levdivkov, V. M., Fogg, M.J., Wilson, K.S., Wilkinson, A.J. 2005. An ATP-binding cassette-type cysteine transporter in *Campylobacter jejuni* inferred from the structure of an extracytoplasmic solute receptor protein. *Mol. Microbiol.* 57. 143-155. doi: 10.1111/j.1365-2958.2005.04691.x
34. Najmanovich, R., Kurbatova, N., Thornton, J. 2008. Detection of 3D atomic similarities and their use in the discrimination of small molecule protein-binding sites. *Bioinformatics,* 24:i105-i111.
35. Najmanovich, R. J. 2017. Evolutionary Studies of Ligand Binding Sites in Proteins. *Curr. Opin. Struct. Biol.* 2017, 45:85-90.
36. Oh, B., Ames, G. F., Kim, S. 1994. Structural Basis for Multiple Ligand Specificity of the Periplasmic Lysine-, Arginine-, Ornithine-binding Protein. *J. Bio. Chem.* 269, 26323-26330.
37. Oh, B., Pandit, J., Kang, C., Nikaido, K., Gokcen, S., Amer, G. F., Kim, S. 1993. Three-dimensional Structures of the Periplasmic Lysine/Arginine/Ornithine-binding Protein with and without a Ligand. *J. Bio. Chem.* 268, 11348-11355.
38. Oliveros, J.C. (2007-2015) Venny. An interactive tool for comparing lists with Venn's diagrams. <http://bioinfogp.cnb.csic.es/tools/venny/index.htm>
39. Pagadala, N. S., Syed, K., Tuszyński, J. 2017. Software for molecular docking: a review. *Biophys. Rev.* 9:91-102

40. Pearson, W. 2014. An Introduction to Sequence Similarity (“Homology”) Searching. *Curr Protoc Bioinformatics*. DOI: 10.1002/0471250953.bi0301s42
41. Planamente, S., Vigouroux, A., Mondy, S., Nicaise, M., Faure, D., Moréra, S. 2010. A Conserved Mechanism of GABA Binding and Antagonism Is Revealed by Structure-Function Analysis of the Periplasmic Binding Protein Atu2422 in *Agrobacterium tumefaciens*. *J. Bio. Chem.* 285, 30294-30303.
42. Pulido, N. O., Silva, D. A., Tellez, L. A., Pérez-Hernández, G., García-Hernández, E., Sosa-Peinado, A., Fernández-Velasco, A. 2014. On the molecular basis of the high affinity binding of basic amino acids to LAOBP, a periplasmic binding protein from *Salmonella typhimurium*. *J. Mol. Recognit.* 28: 108-116.
43. Redelings, B. 2014. Erasing Errors due to Alignment Ambiguity When Estimating Positive Selection. *Mol. Biol. Evol.* 31(8):1979-1993.
44. Reynolds, K. A., Russ, W. P., Socolich, M., Ranganathan, R. 2013. Evolution-Based Design of Proteins. *Methods in Enzymology*, Vol. 523. 213-235
45. Rose, P. W., Prli, A., Bi, C., Bluhm, W. F., Christie, C. H., et al., 2015 The RCSB Protein Data Bank: views of structural biology for basic and applied research and education [Internet]. *Nucleic Acids Res*, 43, D345-D356.
46. Rosetta Commons. Un reservorio de herramientas administrado por David Baker y colaboradores. <https://www.rosettacommons.org/software>
47. Samish, I., Bourne, P. E., Najmanovich, R. J. 2015. Achievements and challenges in structural bioinformatics and computacional biophysics. *Bioinformatics*, 31. 146-150.
48. Sanchez-Ruiz, J. M., 2007. Ligand effects on protein thermodynamic stability. *Biophy. Chem.* 126, 43-49.
49. Scheib, U., Shanmugaratnam, S., Farías-Rico J.A., Höcker B., 2013. Change in protein-ligand specificity through binding pocket grafting. *J. Struct. Biol.* <http://dx.doi.org/10.1016/j.jsb.2013.06.002>
50. Schreier, B., Stumpp, C., Wiesner, S., Höcker, B. 2009. Computational design of ligand binding is not a solved problem. *Proc. Natl. Acad. Sci.* 106: 18491-18496.
51. Skolnick, J., Gao, M., Roy, A., Srinivasan, B., Zhou, H. 2015. Implications of the Small Number of Distinct Ligand Binding Pockets in Proteins for Drug Discovery, Evolution and Biochemical Function. *Bioorg. Med. Lett.* 25:1163-1170.
52. Silva, D. A., Bowman, G. R., Sosa-Peinado, A., Huang, X. 2011. A Role for Both Conformational Selection and Induced Fit in Ligand Binding by the LAO Protein. *PloS Comput. Biol.* May;7(5):e1002054.
53. Software y Licencia para uso de software de Applied Biosystems para realizar Fluorometría Diferencial de Barrido. Obtenida durante el mes de Enero del 2017. Link: <https://licensn.appliedbiosystems.com/download/index>
54. Starr, T. N., Thornton. J. W. 2016. Epistasis in Protein Evolution. *Prot. Society. Prot. Sci.* 25:1204-12089
55. Stiffler, M. A., Hekstra, D. R., Ranganathan, R. 2015. Evolvability as a Function of Purifying Selection in TEM-1  $\beta$ -lactamase. *Cell.* Feb;160(5):882-92.
56. Stoddard, B. L. (ed.), 2016. Computational Design of Ligand Binding Proteins. *Methods in Molecular Biology*, vol. 1414. doi: 10.1007/978-1-4939-3569-7
57. Sun, Y. J., Rose, J., Wang, B. C. Hsiao, C. D. 1998. The structure of glutamine-binding protein complexed with glutamine at 1.98 Å resolution: comparisons with other amino acid binding proteins. *J. Mol. Biol.* 278: 219:229.
58. Suzek, B. E., Huang, H., McGorvey, P. B., Wu, C. H. 2015. The Uniprot Consortium. UniRef clusters: a comprehensive and scalable alternative for improving sequence similarity searches. *Bioinformatics*. Vol. 3, 6:926-9323 <https://doi.org/10.1093/bioinformatics/btu739>

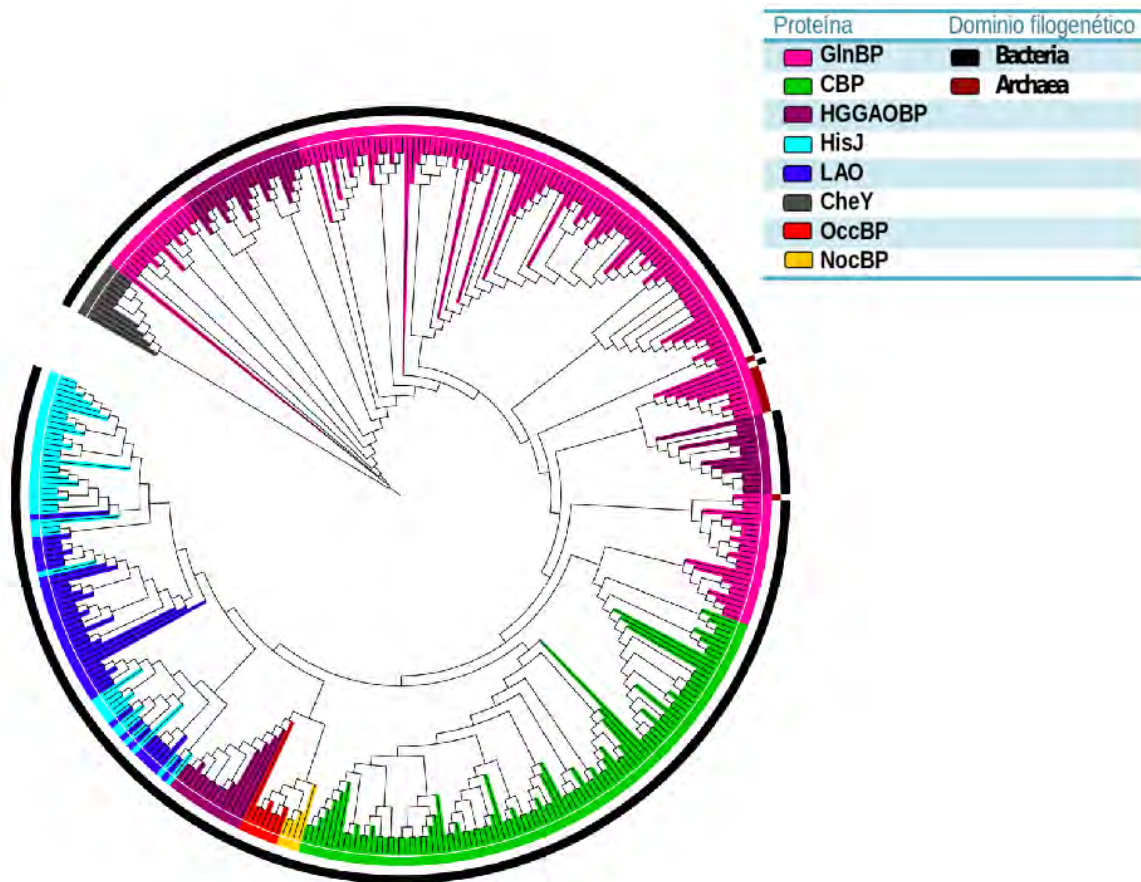


59. Teşileanu, T., Colwell, L. J., Leibler, S. 2015. Protein Sectors: Statistical Coupling Analysis versus Conservation. *PloS Comput. Biol.* 11(2): e1004091. doi:10.1371/journal.pcbi.1004091.
60. Ud-Din, S., Roujeinikova, A. 2017. Methyl-accepting chemotaxis proteins: a core sensing element in prokaryotes and archae. *Cell. Mol. Life Sci.* doi: 10.1007/s00018-017-2514-0.
61. Universal Protein Resources (UniProt). Una base de datos de proteínas, en donde también se pueden buscar algunas definiciones, incluyendo componentes celulares como lo es el Periplasma. Link: <http://www.uniprot.org/locations/SL-0200>
62. Vacic V., Iakoucheva L.M., and Radivojac P. 2006. Two Sample Logo: A Graphical Representation of the Differences between Two Sets of Sequence Alignments. *Bioinformatics*, 22(12): 1536-1537.
63. Vivoli, M., Novak, H. R., Littlechild, J. A., Harmer, N. J. 2014. Determination of Protein-ligand Interactions Using Differential Scanning Fluorimetry. *J. Vis. Exp.* 91: e51809, doi:10.3791/51809 (2014).
64. Whitfield, J. H., Zhang, W. H., Herde, M. K., Clifton, B. E., Radziejewski, J., Janovjak, H., Henneberger, C., Jackson, C. 2015. Construction of a robust and sensitive arginine biosensor through ancestral protein reconstruction. *Prot. Sci.* 25: 1412-1422.
65. Yang, G., Hong, N., Baier, F., Jackson, C. J., Tokuriki, N. 2016. Conformational Tinkering Drives Evolution of Promiscuous Activity through Indirect Mutational Effects. *ACS Biochem.* 55, 4583-4593
66. Yang, W., Lai, L. 2017. Computational Design of Ligand-Binding Proteins. *Curr. Opin. Struct. Biol.* 45:67-73

# APÉNDICE

## Primera motivación para la realización de este proyecto

Poco antes de la llegada del autor de esta tesis al laboratorio de Físicoquímica e Ingeniería de Proteínas (LFQIP) de la Facultad de Medicina de la UNAM, el Dr. Alejandro Sosa contaba con la entonces estudiante de la carrera en Biología Marisol Maya, cuyo trabajo consistió, en parte, en determinar la filogenia de PBPs que unieran ciertos aminoácidos o moléculas similares (ver Figura S1), obteniendo una separación en tres grandes grupos: PBPs de unión a aminoácidos positivos, PBPs de unión a Gln y PBPs de unión a Cys.



**Figura S1. Árbol filogenético de PBPs.** Las PBPs analizadas por la Biól. Marisol Maya. LAOBP corresponde a LAO y QBP a GlnBP. *Figura adaptada de la Tesis de Licenciatura de Marisol Maya, disponible en el repositorio en línea de la UNAM.*

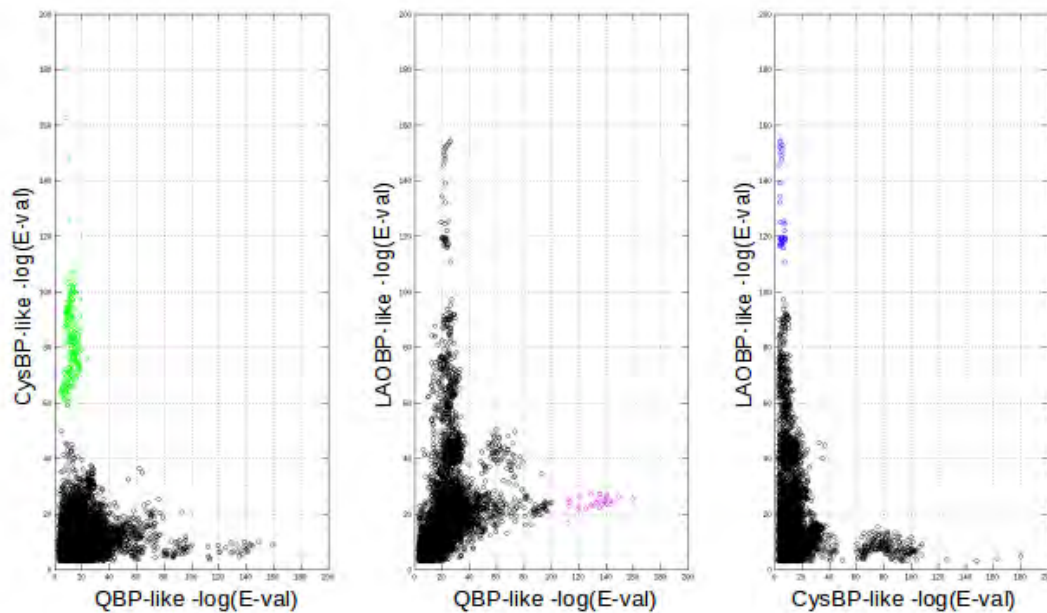
Este trabajo previo nos motivó con la idea de que la información contenida en las secuencias de aminoácidos de proteínas naturales, desde una perspectiva estadística y posteriormente aplicada a casos particulares, podría ser utilizada para delatar sitios en una proteína cuya mutación pertinente lograría la capacidad de reconocer un nuevo ligando, como es el caso de

LAOBP por Gln.

Durante la realización de mi trabajo de doctorado no concluí nada significativo al intentar reproducir el trabajo de Marisol con nuevas búsquedas de secuencias y alineamientos, pero tales resultados no me detuvieron para inferir los sitios para la mutante LAOBP\_sca. Los detalles del trabajo de Marisol se encuentran en su tesis de licenciatura, disponible en el repositorio en línea de la UNAM.

### Obstáculos para la construcción de las bases de datos

La construcción de bases de datos (DB por sus siglas en inglés), o en otras palabras, conjuntos de secuencias de aminoácidos de proteínas naturales como LAOBP, QBP o una PBP de unión a cisteína fue clave para realizar los alineamientos con los cuales diseñar la secuencia LAOBP\_sca. Sin embargo, esta empresa tuvo la limitante de que cada una de las 3 DB presentaba prácticamente el mismo contenido al utilizar la herramienta BLASTp. Es decir, independientemente de la secuencia de búsqueda utilizada los resultados eran los mismos, y eso se convirtió en un problema para aprender de las diferencias entre las distintas DB, pues necesitábamos que estas fuesen exclusivas de un universo LAOBP-like pero no QBP-like o CBP-like. Aunque dependiendo de la secuencia de búsqueda los resultados se obtenían con un E-valor<sup>30</sup> diferente, determinar puntos de corte para clasificar las secuencias homólogas en categorías diferentes (ver figura S2) nos dejaba con muy pocas secuencias (de 30 o menos) para realizar estadísticas confiables con SCA, pues lo recomendado suele ser un mínimo de 100 secuencias por alineamiento (Halabi *et al.*, 2009). Este problema se presentó aunque la identidad de secuencia entre los tres “anzuelos” siempre fue menor a 30% (ver Tabla S1).



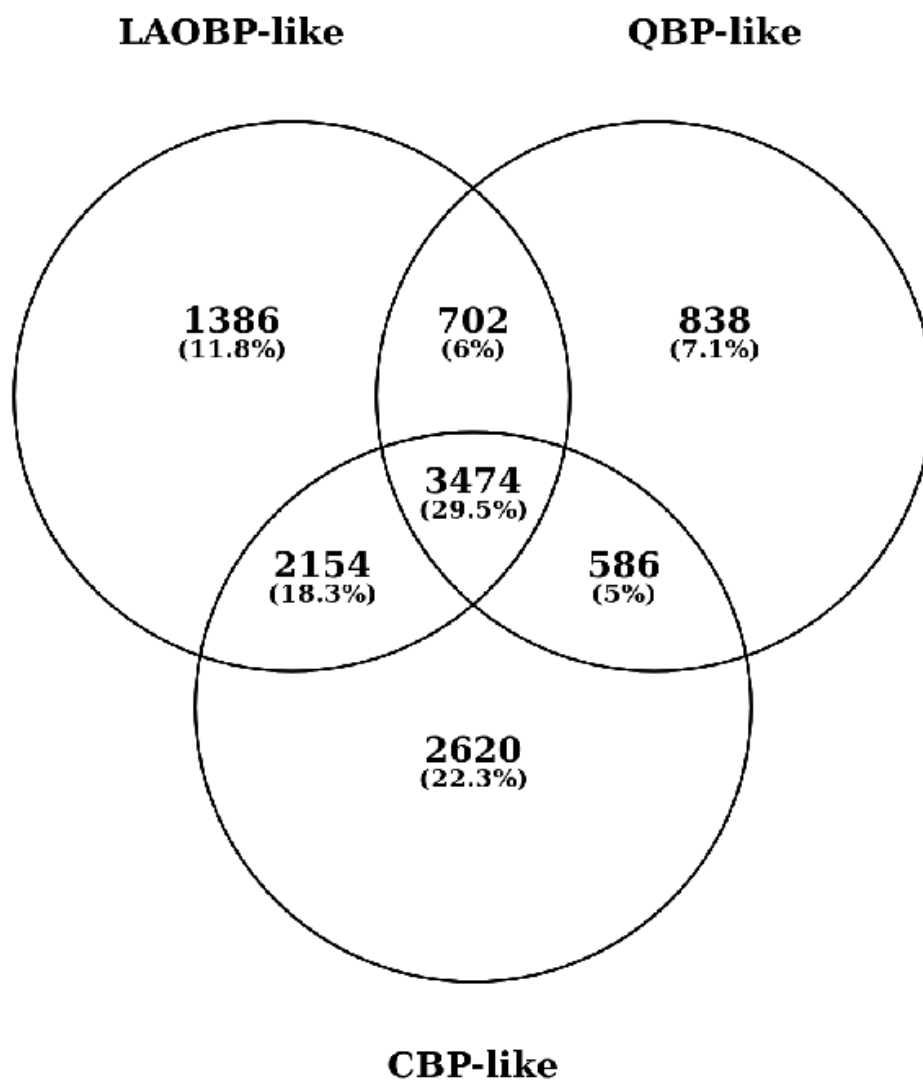
**Figura S2. E-values ganadores las búsquedas con BLASTp.** Independientemente de la secuencia anzuelo utilizada: LAOBP, QBP y CBP, los resultados son los mostrados en los 3 gráficos de esta figura. Se muestran en color aquellos resultados que, por un punto de corte arbitrario, pertenecen a un tipo particular de PBP y no a las otras dos. Para fines de visualización, los E-values están transformados a su valor logarítmico en base 10.

30 El E-valor, también llamado Valor Esperado (E) es un parámetro que describe el número de resultados que alguien espera observar al buscar en una base de datos de un tamaño particular, describiendo así el ruido debido al azar. Por ejemplo, un E-valor de 1 en un resultado particular puede interpretarse como que, en la base de datos usada, uno esperaría ver un resultado o coincidencia con una calificación (score) similar simplemente por azar. Por lo tanto, entre más cercano a cero sea el E-valor, el resultado coincidente es más significativo. Fuente [https://blast.ncbi.nlm.nih.gov/Blast.cgi?CMD=Web&PAGE\\_TYPE=BlastDocs&DOC\\_TYPE=FAQ#expect](https://blast.ncbi.nlm.nih.gov/Blast.cgi?CMD=Web&PAGE_TYPE=BlastDocs&DOC_TYPE=FAQ#expect)

**Tabla S1. Identidad y similitud entre secuencias semilla.** Cálculos de acuerdo a nuestros alineamientos entre LAOBP de *S. typhimurium*, QBP de *E. coli* y la PBP de unión a cisteína (CBP) de *Campylobacter jejuni*.

<b>Identidad / similitud</b>	<b>LAOBP</b>	<b>QBP</b>	<b>CBP</b>
<b>LAOBP</b>	-	28.99% / 42.43%	20.19% / 30.79%
<b>QBP</b>	28.99% / 42.43%	-	21.92% / 31.22%
<b>CBP</b>	20.19% / 30.79%	21.92% / 31.22%	-

Afortunadamente, durante una conversación, el Dr. Rogelio Rodríguez-Sotres nos sugirió hacer una adaptación del enfoque Rd.HMM (Martínez-Castilla & Rodríguez-Sotres) para incorporar información estructural en la búsqueda de secuencias homólogas a una proteína cuyas coordenadas cristalográficas se conocieran. En nuestro caso, LAOBP, QBP y CBP presentan estructura reportada en complejo con sus ligandos respectivos, y los resultados con Rd.HMM fueron satisfactorios, en el sentido de que ahora contábamos con cientos de secuencias exclusivas por búsqueda (ver Figura S3).



**Figura S3.**  
**Resultados de Rd.HMM en UniRef90.** Nos quedamos con los resultados exclusivos en cada caso para hacer los SCA.

### Otras mutantes diseñadas durante este proyecto

Antes de realizar experimento alguno en el laboratorio, y previendo la posibilidad de que ni la mutante LAOBP\_graft ni LAOBP\_sca mostraran unión por Gln, exploramos otras ideas *in silico*, como mutar con ayuda de Rosetta sitios de la cavidad de unión de LAOBP o de los sitios declarados como mutables en nuestro enfoque SCA-grafting, así como elegir consensos (ver Tabla S2). Al final, de acuerdo a estos resultados, decidimos sintetizar las secuencias de DNA correspondientes sólo a ciertas mutantes.

**Tabla S2. Otras mutantes diseñadas y sus análisis. Algunas mutantes no fueron analizadas experimentalmente. (-) Ensayo o medición no realizada. Los ensayos que indican “Precipitación” fueron realizados a 20 y 15 °C.**

Diseño	Fase Computacional				¿Cambios observados en Fase Experimental con Gln?				
	Mutaciones con respecto a LAOBP	Filosofía de creación	Herramientas de evaluación	Observaciones	Fluorescencia	CD	Tmelt	MST	ITC
LAOBP_sca <i>Descrito más detalladamente en el texto principal</i>	Sector azul: Y14F, Y140F, D220N  Sector rojo: L55I, S72T, L117K  Sector green: K110D, S120T, G153N	Trasplantar todos los sitios mutables distintos entre LAOBP y GlnBP identificados por medio de SCA.	- Estimación de la estabilidad con Rosetta fixbb con alrededor de 10,000 modelos estructurales. -Docking con Gln usando el protocolo RosettaLigand.	El total_score de Rosetta (fixbb) siempre fue negativo y comparable al de LAOBP. El docking siempre mostró score negativo y la comparación estructural pareada (~5x10 <sup>7</sup> ) de ligandos siempre mantuvo un RMSD entre 0.12Å y 3.67Å.	Sí	Sí	Sí	Sí, K <sub>d</sub> =40 µM	Precipitación
LAOBP_sca_consenso	Sector azul: Y140F, D220N  Sector rojo: L55I, S72T, L117Q  Sector verde: K110C, S120T	Usar el residuo más común por sitio mutable según el alineamiento del universo GlnBP-like.	- Estimación de la estabilidad con Rosetta fixbb con alrededor de 10,000 modelos estructurales. -Docking con Gln usando el protocolo RosettaLigand.	El total_score de Rosetta siempre fue negativo y comparable al de LAOBP. El docking siempre mostró score negativo y la comparación estructural pareada (~5x10 <sup>7</sup> ) de ligandos siempre mantuvo un RMSD entre 0.12Å y 3.67Å.	No	No	No	No	Precipitación
LAOBP_sca_rosetta	Sector azul: Y14N, Y140F, K201L, L206S, D220N, T222G	Permitir a Rosetta fixbb decidir qué residuo	- Estimación de la estabilidad con Rosetta fixbb con alrededor	El total_score de Rosetta (fixbb) siempre fue negativo y comparable al de LAOBP. El docking siempre mostró score	No	No	No	No	Precipitación

	Sector rojo: S72A, R77A, L117I  Sector verde: Ninguno	o usar en cada uno de los sitios mutables.	de 10,000 modelos estructurales. -Docking con Gln usando el protocolo RosettaLigand.	negativo y la comparación estructural pareada (~5x10 <sup>7</sup> ) de ligandos siempre mantuvo un RMSD entre 0.12Å y 3.67Å.					
LAOBP_bindingpocket_rosetta (no se construyó)	Y14N, S70G, S72A, R77A, T121A, Q122A	Permitir a Rosetta decida qué residuos usar dentro en los sitios que presentan por lo menos un átomo a 5Å de cualquier átomo de la Arg en la estructura 1laf. de LAOBP.	- Estimación de la estabilidad con Rosetta fixbb con alrededor de 10,000 modelos estructurales. -Docking con Gln usando el protocolo RosettaLigand.	El total_score de Rosetta (fixbb) siempre fue positivo (energéticamente desfavorable); aunque el docking siempre mostró score negativo y la comparación estructural pareada (~5x10 <sup>7</sup> ) de ligandos siempre mantuvo un RMSD entre 0.12Å y 3.67Å. DECIDIMOS NO LLEVAR ESTE DISEÑO A LA FASE EXPERIMENTAL	No aplica	No aplica	No aplica	No aplica	No aplica
LAOBP_sphere_consenso (No se construyó)	D11Y, T13P, Y14F, S70G, L71I, S72T, L117Q, S120T, Q122S	Trasplantar única mente los sitios que presentan por lo menos un	- Estimación de la estabilidad con Rosetta fixbb con alrededor de 10,000 modelos estructurales.	El total_score de Rosetta siempre fue positivo (energéticamente desfavorable); aunque el docking siempre mostró score negativo y la comparación estructural pareada (~5x10 <sup>7</sup> ) de ligandos	No aplica	No aplica	No aplica	No aplica	No aplica

		átomo a 5Å de cualquier átomo de la Gln en la estructura 1wdn en la posición equivalente de LAOBP.	-Docking con Gln usando el protocolo RosettaLigand.	siempre mantuvo un RMSD entre 0.12Å y 3.67Å. DECIDIMOS NO LLEVAR ESTE DISEÑO A LA FASE EXPERIMENTAL					
LAOBP_sca_re dsector_rosetta	S72A, R77A, L117V	Permitir a Rosetta decidir qué residuo usar en cada sitio mutable de única mente el sector rojo del SCA de LAOBP.	- Estimación de la estabilidad con Rosetta fixbb con alrededor de 10,000 modelos estructurales. -Docking con Gln usando el protocolo RosettaLigand.	El total_score de Rosetta (fixbb) siempre fue negativo y comparable al de LAOBP. El docking siempre mostró score negativo y la comparación estructural pareada (~5x10 <sup>7</sup> ) de ligandos siempre mantuvo un RMSD entre 0.12Å y 3.67Å.	No	Sí	No	No	No
LAOBP_sca_re dsector_consens o	L55I, S72T, L117Q	Usar el residuo más común por cada sitio que se presente a, por lo	- Estimación de la estabilidad con Rosetta fixbb con alrededor de 10,000 modelos estructurales.	El total_score de Rosetta siempre fue negativo y comparable al de LAOBP. El docking siempre mostró score negativo y la comparación estructural pareada (~5x10 <sup>7</sup> ) de ligandos	No	Sí	No	No	No



		menos , un átomo a 5Å de cualquier átomo de la Gln en la estructura 1wdn en la posición equivalente de LAOBP.	-Docking con Gln usando el protocolo RosettaLigand.	siempre mantuvo un RMSD entre 0.12Å y 3.67Å.					
LAOBP_sca_re dsector	L55I, S72T, L117K	Trasplantar única mente los sitios mutables distintos entre LAOBP y GlnBP identificados en el sector rojo del SCA de LAOBP.	No se realizó análisis computacional	No se realizó análisis computacional	No se realizó	No se realizó	Sí	No se realizó	No se realizó
LAOBP_sca_na t1	Sector azul: Y14F, Y140F  Sector Rojo: L55I, S72T,	Combinación de mutaciones acorde s con lo observ	-Scripts propios para búsqueda de combinaciones naturales, modelado	Una de las 72 combinaciones naturales de las 2 <sup>22</sup> exploradas, y, consistentemente , el mejor score entre 1000 relajaciones.	No	No	No	No	No

	L117Q Sector Verde: K110C, S120T	ado en secuencias naturales del universo GlnB P-like. (Detalles en Métodos)	por Rosetta Fixbb y evaluación por medio de Rosetta Relax.						
LAOBP_sca_nant2	Sector azul: Y14F, Y140F, D220N  Sector Rojo: L55I, S72T, L117Q  Sector Verde: K110C, S120T	Combinación de mutaciones acordadas con lo observado en secuencias naturales del universo GlnB P-like. (Detalles en Métodos)	-Scripts propios para búsqueda de combinaciones naturales, modelado por Rosetta fixbb y evaluación por medio de Rosetta Relax.	Una de las 72 combinaciones naturales de las 2 <sup>22</sup> exploradas, y, consistentemente, el segundo mejor score entre 1000 relajaciones.	No	No	No	No	Precipitación
LAOBOP_sca_220D	Sector azul: Y14F, Y140F  Sector rojo: L55I, S72T, L117K  Sector green: K110D, S120T, G153N	Reversión del sitio 220 hacia un aspártico en el contexto de LAOBOP_sca	No aplicable	No aplicable	No	Sí	Sí	No	No



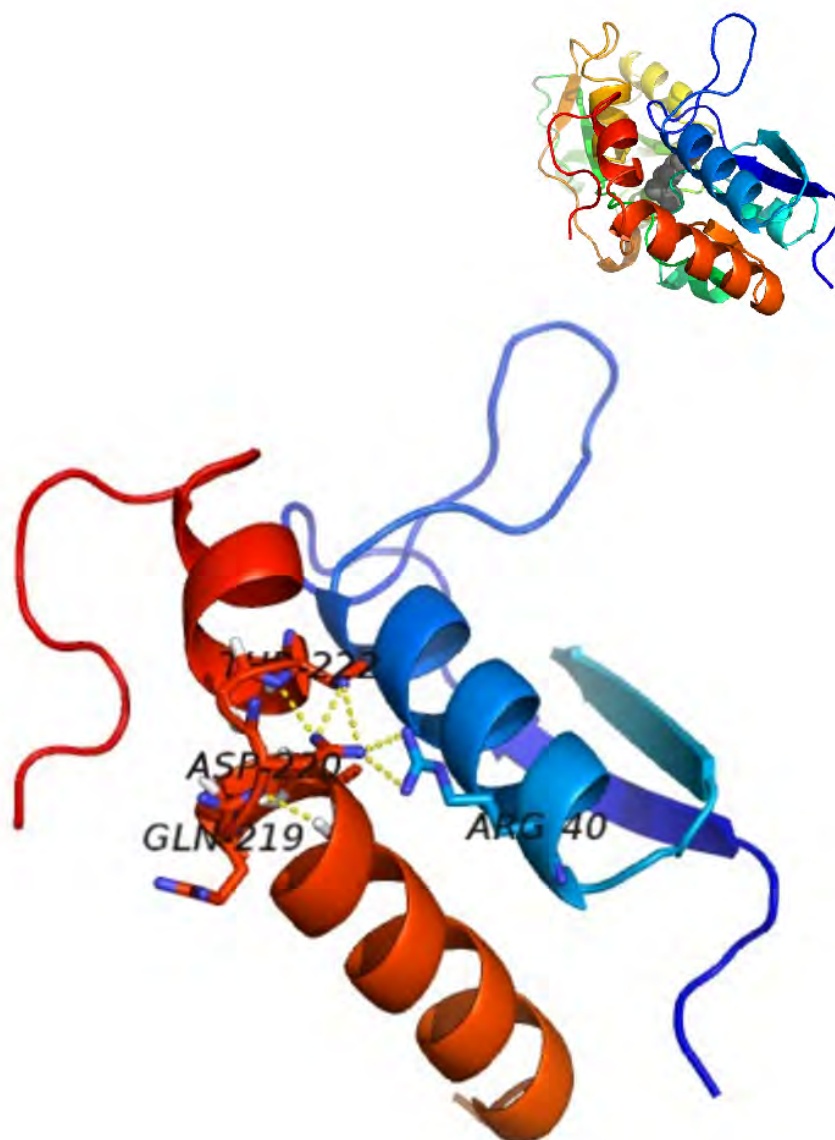


Figura S5. Acercamiento del sitio 220 en LAOBP. Interacciones polares del sitio 220.

Tabla S3. Constantes de disociación y valores termodinámicos con ligandos de interés.

Protein	Ligand	Kd ( $\mu\text{M}$ )	$\Delta\text{H}$ (kcal/mol)	$-\text{T}\Delta\text{S}$ (kcal/mol)	$\Delta\text{G}$ (kcal/mol)
LAOBP	Glutamine	No binding	No value	No value	No value
	Cysteine	$51.62 \pm 2.69$	$-4.81 \pm 0.08$	$-1.04 \pm 0.08$	$-5.85 \pm 0.03$
	Citrulline	$3.84 \pm 0.06$	$-7.98 \pm 0.13$	$0.72 \pm 0.14$	$-7.25 \pm 0.01$
	Arginine	$(9.5 \pm 4) \times 10^{-3}$	$-10.10 \pm 0.10$	$-0.69 \pm 0.41$	$-10.81 \pm 0.32$
	Glutamine	$1.59 \pm 0.23$	$-9.16 \pm 0.04$	$1.38 \pm 0.05$	$-7.57 \pm 0.08$
LAOBP_L117K	Cysteine	$2.61 \pm 0.14$	$-8.16 \pm 0.17$	$0.54 \pm 0.17$	$-7.62 \pm 0.03$
	Citrulline	$2.31 \pm 0.48$	$-5.54 \pm 0.01$	$-2.03 \pm 0.13$	$-7.60 \pm 0.12$
	Arginine	$11.5 \pm 3.04$	$-4.62 \pm 0.31$	$-2.01 \pm 0.47$	$-6.67 \pm 0.16$

**Tabla de cristalización.** Parámetros obtenidos durante el experimento de difracción de rayos X en los cristales de la proteína LAOBP\_L117K con Gln como ligando.

Identificador en PDB	5OWF
Longitud de onda	0.918400Å
Rango de resolución	41.14 - 1.91 (1.978 - 1.91)
Grupo espacial	P 21 21 21
Celda Unitaria	36.972 58.73 115.312 90 90 90
Total de reflexiones	194942 (14635)
Reflexiones únicas	20169 (1946)
Multiplicidad	9.7 (7.5)
Porcentaje completado	99.54 (97.49)
Media I/sigma(I)	8.67 (1.12)
Wilson B-factor	29.05
R-merge	0.1741 (1.954)
<i>R-meas</i>	0.1838 (2.093)
<i>R-pim</i>	0.05782 (0.7327)
CC1/2	0.995 (0.375)
CC*	0.999 (0.738)
Reflexiones usadas en el refinamiento	20129 (1946)
Reflexiones usadas para el <i>R-free</i>	1007 (97)
<i>R-work</i>	0.1896 (0.3122)
<i>R-free</i>	0.2314 (0.3286)
CC(work)	0.958 (0.659)
CC(free)	0.935 (0.532)
Número de átomos que no son hidrógeno	2073
macromoléculas	1827
ligandos	10
solvente	145
Residuos de proteína	235
RMS(enlaces)	0.006
RMS(ángulos)	0.76
Favorable a un gráfico de Ramachandran (%)	97.85
Dentro del gráfico de Ramachandran (%)	2.15
Fuera del gráfico de Ramachandran (%)	0
Rotámeros no naturales (%)	0
<i>Clashscore</i>	6.7
<i>B-factor</i> promedio	35.63
macromoléculas	33.85
ligandos	52.69
solvente	45.34
Número de grupos TLS	6