



UNIVERSIDAD NACIONAL AUTÓNOMA
DE MÉXICO

FACULTAD DE CIENCIAS

“Filogenia molecular de proteínas periplásmicas de unión a
aminoácidos como herramienta de estudio de la interacción
proteína – ligando”

T E S I S

QUE PARA OBTENER EL TÍTULO DE:

BIÓLOGA

P R E S E N T A :

KARINA MARISOL MAYA RAMÍREZ



FACULTAD DE CIENCIAS
UNAM

Director de tesis:

DR. ALEJANDRO SOSA PEINADO

Á

Á

México, D.F., a 15 de mayo del 2014



Universidad Nacional
Autónoma de México



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.

1. Datos del alumno
Maya
Ramírez
Karina Marisol
50 27 56 84
Universidad Nacional Autónoma de México
Facultad de Ciencias
Biología
306290503
2. Datos del tutor
Dr.
Alejandro
Sosa
Peinado
3. Datos del sinodal 1
Dr.
Bernardo
Pérez
Zamorano
4. Datos del sinodal 2
Dr.
Arturo Carlos II
Becerra
Bracho
5. Datos del sinodal 3
Dr.
Daniel Alejandro
Fernández
Velasco
6. Datos del sinodal 4
M en C.
Enrique
Scheinvar
Gottdiener
7. Datos del trabajo escrito
Filogenia molecular de proteínas periplásmicas de unión a aminoácidos como herramienta de estudio de la interacción proteína – ligando.
66 p
2014

A mí madre que me enseñó el arte
de nadar contra corriente
y a mí padre que me enseñó el valor
de la honestidad.

«Hay quien cruza el bosque y sólo ve leña para el fuego»
Lev Nikoláyevich Tolstói

Este trabajo se llevo a cabo en el Laboratorio de Fisicoquímica e Ingeniería de Proteínas del Departamento de Bioquímica de la Facultad de Medicina, UNAM.

Durante el desarrollo del mismo conté con la beca del proyecto PAPIIT IN215612-3, así como con el apoyo por parte de CONACYT (167838).

AGRADECIMIENTOS

«Los grandes espíritus siempre han encontrado una violenta oposición de parte de mentes mediocres»

Agradezco por tradición a la Universidad Nacional Autónoma de México y a la Facultad de Ciencias por brindarme la oportunidad de terminar una licenciatura en ésta institución de talla internacional, esperando algún día los alumnos tengan más voz que la burocracia que en ésta existe.

A mi asesor de tesis, el Dr. Alejandro Sosa Peinado por abrirme las puertas del laboratorio y brindarme la oportunidad de realizar un proyecto de tesis. Al Dr. Bernardo Pérez Zamorano por su apoyo durante este proyecto. Y sobre todo a mis compañeros de laboratorio que hicieron de mi estancia algo divertido y enriquecedor, gracias Tania y Renan por sus consejos, gracias Andrés por siempre mantener el buen humor, gracias Julio por tu amistad incondicional, gracias Aarón por tu amistad y orientación académica, y en general al Laboratorio de Físicoquímica e Ingeniería de Proteínas por el excelente equipo de trabajo que forman.

A Gerardo y Artemisa, personas sumamente importantes en mi desarrollo en esta etapa, fueron mis amigos y con todo y nuestros bemoles siempre estuvieron ahí. Sus regaños, sus bromas, sus consejos, esos cafés, esos desayunos, son cosas que no olvidaré y que aunque con el tiempo todo cambie, me encargaré de guardar solo aquellos momentos que nos unieron como amigos y que marcaron mi vida en este ciclo tan importante.

Agradezco sobre todo a mi familia, sabemos lo insoportables que somos cada uno de sus miembros y aún así jamás nos hemos dejado de apoyar; a mi mamá y mejor amiga que si no hubiese sido por ella ninguno de mis logros en la vida hubiesen sido posibles, tengo mucho que aprender de ti; a mi papá por ser pilar de la familia; a mi hermano Daniel por nuestros debates filosóficos que nunca llegan a ningún lugar pero que me han ayudado mucho a abrir mi perspectiva del mundo; a Carlos mi hermano que me ha enseñado que no importa la basura que el mundo crea que eres, la lucha es con uno mismo; y a mi amiga y hermana Eva que le ha aportado a mi vida ese lado dulce, noble y divertido. Gracias familia, por sus enseñanzas, por preocuparse por mí, por sembrar en mi valores que en la sociedad contemporánea hacen falta, gracias por no dejarme caer.

Y por supuesto, a ti, Eduardo, porque por el simple hecho de existir en mi vida le diste a ésta una motivación, esas ganas de ser mejor persona; gracias por haberme dado sueños, momentos felices, momentos tristes, nunca había aprendido tanto de mi misma como lo hice contigo. Gracias por haber formado parte de mi vida y sobre todo porque al tenerte en mi camino entendí y aprendí muchas cosas que no conocía. Caminar de tu mano ha sido lo más especial que me ha pasado jamás. Tu presencia ha sido la clave para que todo fuera real.

ÍNDICE

ÍNDICE DE FIGURAS	2
ÍNDICE DE TABLAS	3
ABREVIATURAS	4
RESUMEN	5
1. INTRODUCCIÓN	6
1.1 <i>¿Qué son las proteínas?</i>	6
1.2 <i>Proteínas periplásmicas de unión</i>	7
1.3 <i>Proteína de unión a Lisina-Arginina-Ornitina (LAO)</i>	12
1.4 <i>La evolución biológica y su alcance a nivel molecular</i>	14
1.5 <i>El estudio de la evolución molecular y su implicación en la ingeniería de proteínas</i>	15
1.6 <i>La filogenia como herramienta de análisis evolutivo</i>	16
1.7 <i>Evolución de proteínas periplásmicas de unión</i>	17
1.8 <i>Principales métodos utilizados en análisis filogenético</i>	20
1.9 <i>Reconstrucción de secuencias ancestrales</i>	22
2. JUSTIFICACIÓN	24
3. OBJETIVOS	24
3.1 <i>General</i>	24
3.2 <i>Particulares</i>	24
4. MATERIALES Y MÉTODOS	25
4.1 <i>Obtención de secuencias homólogas a LAO</i>	25
4.2 <i>Alineamiento</i>	26
4.3 <i>Construcción del árbol filogenético</i>	27
4.4 <i>Reconstrucción y análisis estructural de las secuencias ancestrales</i>	28
5. RESULTADOS	29
5.1 <i>Número de secuencias obtenidas homólogas a la proteína LAO y diversidad</i>	29
5.2 <i>Alineamiento de secuencias</i>	32
5.3 <i>Árbol filogenético de las proteínas periplásmicas homólogas a LAO</i>	33
5.4 <i>Secuencias ancestrales</i>	44
6. DISCUSIÓN	53
7. CONCLUSIÓN Y PERSPECTIVAS	59

REFERENCIA	61
APÉNDICE I: Estructuras hipotéticas de los nodos ancestrales.....	66

ÍNDICE DE FIGURAS

Figura 1. Esquematización del transporte de membrana por proteínas periplásmicas de unión.....	8
Figura 2. Clasificación de las PBP's basada en el número y disposición de las láminas β presentes en cada lóbulo	10
Figura 3. Clasificación de las PBP's basada en el número de segmentos polipeptídicos que unen ambos dominios.....	10
Figura 4. Proteína periplásmica de unión a L - Lisina, L - Arginina y L – Ornitina.....	12
Figura 5. Esquematización de los principales residuos de aminoácidos involucrados en la unión a ligandos de la proteína LAO	13
Figura 6. Representación genealógica de las 15 PBP's.....	18
Figura 7. Esquema propuesto para el origen evolutivo de las proteínas periplásmicas de unión tipo I y tipo II, a partir de la proteína CheY	19
Figura 8. Representación gráfica del número de secuencias por proteína y por clase taxonómica. Mostrando la diversidad taxonómica para cada una	31
Figura 9. Clases taxonómicas obtenidas para ambos dominios filogenético y el porcentaje de secuencias pertenecientes a cada clase.....	31
Figura 10. Visualización general del Alineamiento utilizando el programa Jalview	32
Figura 11. Vista circular del árbol filogenético de las proteínas periplásmicas de unión homólogas a LAO, utilizando como grupo externo a la proteína involucrada en quimiotaxis CheY	35
Figura 12. Árbol filogenético de la proteína periplásmica LAO (vista rectangular).....	36
Figura 13. Clado perteneciente al árbol original donde se encuentran secuencias de LAO e HisJ	37
Figura 14. Clado perteneciente al árbol original donde se encuentran secuencias de NocBP, OccBO y HGGAOBP	38
Figura 15. Clado perteneciente al árbol original donde se encuentran secuencias de CBP	39
Figura 16. Clado perteneciente al árbol original donde se encuentran secuencias de GlnBP y CBP	40

Figura 17. Clado perteneciente al árbol original donde se encuentran secuencias de GlnBP	41
Figura 18. Clado perteneciente al árbol original donde se encuentran secuencias de GlnBP y HGGAOPB.....	42
Figura 19. Árbol colapsado para mostrar las ramas más cercanas a la raíz donde se encuentran secuencias de GlnBP, HGGAOPB y CheY.....	43
Figura 20. Árbol filogenético (Vista convexa) donde están marcados los ocho nodos ancestrales de interés.....	46
Figura 21. Estructura de LAO	47
Figura 22. Alineamiento de la secuencia LAO de <i>Salmonella typhimurium</i> y de los nodos ancestrales.	48
Figura 23. Alineamiento estructural de LAO con cada uno de los nodos ancestrales	49
Figura 24. Alineamiento estructural de la región de unión a ligando de LAO con cada uno de los ancestros hipotéticos.....	51

ÍNDICE DE TABLAS

Tabla 1. Número de secuencias obtenidas para cada proteína.....	30
Tabla 2. Número de secuencias obtenidas para cada proteína por clase taxonómica en cada dominio	30
Tabla 3. Propiedades de los alineamientos sin y con las secuencias del grupo externo ..	32

ABREVIATURAS

aLRT : Approximate Likelihood-Ratio Test

BLAST Basic Local: Alignment Search Tool

CBP: Proteína periplásmica de unión a Cisteína

CheY: Proteína involucrada en quimiotaxis

GlnBP: Proteína periplásmica de unión a Glutamina

HGGAOBP: Proteína periplásmica de unión a Histidina, Glutamina, Glutamato, Arginina y Opina

HisJ: Proteína periplásmica de unión a Histidina

iTOL : Interactive Tree of Life

LAO: Proteína periplásmica de unión a lisina, arginina y ornitina

LG: Matriz de sustitución de aminoácidos mejorada a partir del modelo WAG nombra así por sus autores Le y Gascuel

MAFFT: Multiple Aligment – fast Fourier transform

NCBI: Nacional Center for Biotechnology Information

NocBP: Proteína periplásmica de unión a Nopalina

OccBP: Proteína periplásmica de unión a Octopina

PBP: Proteína periplásmica de unión a ligando

PhyML: Phylogenetic Maximum Likelihood

SPR: Subtree Pruning and Regrafting

WAG: Matriz de sustitución de amminoácidos propuesto por Whelan y Goldman

RESUMEN

En el presente trabajo se realizó una filogenia a partir de la proteína periplásmica LAO utilizando como método la Máxima Verosimilitud. LAO es una proteína que forman parte de los sistemas de transporte bacteriano, implicada en la unión de aminoácidos básicos: L-histidina, L-lisina, L-arginina y L-ornitina. Se reconstruyeron las secuencias ancestrales hipotéticas a partir de dicha filogenia, con el fin de dilucidar los residuos de aminoácidos implicados en la unión a ligando. Además de realizar una comparación a nivel estructural de los ancestros contra una proteína periplásmica actual, con el fin de explorar las bases moleculares y posibles rediseños de esta interacción proteína-ligando

La ingeniería de proteínas es una herramienta de gran interés actualmente, puesto que permite comprender y adaptar los aspectos moleculares de una proteína, modificando y manipulando su secuencia, que es el código en el cuál subyace la información precisa del cómo es que ocurre cada proceso en ella, desde el plegamiento hasta la función. El objetivo particular de modificar una proteína es obtener una función deseada para poder posteriormente darle una aplicación con gran potencial industrial. Se han desarrollado muchos métodos para obtener este fin, sin embargo, gran parte de ellos han fallado, puesto que su enfoque ha sido reducido a pensar que la única parte que es preciso modificar para obtener una función deseada son residuos de aminoácidos específicos que generalmente son los que interactúan con el ligando; se ha dejado de lado que la proteína es una unidad donde sus aminoácidos interactúan en conjunto para poder hacer posible su existencia. De esta manera, el estudio de la evolución de una proteína resulta de gran apoyo para poder entender el “experimento” que se ha llevado a cabo a lo largo de miles o millones de años en el que dicha proteína tuvo las modificaciones precisas para su función actual.

La inferencia filogenética en conjunto con la reconstrucción de secuencias ancestrales, son métodos que nos ayudan a entender a una proteína actual, asomándonos de manera hipotética en su pasado, observando los cambios en secuencia que tuvo, o bien aquellos aminoácidos que se conservan. La obtención de ancestros hipotéticos puede presentar la clave que de paso a la modificación de una proteína actual exitosa, sirviendo como guía para identificar los cambios de residuos de aminoácidos de los cuales se derivó a la proteína actual.

1. INTRODUCCIÓN

1.1 *¿Qué son las proteínas?*

Desde un punto de vista químico, las proteínas son polímeros de aminoácidos unidos covalentemente por enlaces peptídicos que pueden adquirir diferentes niveles conformacionales, lo que las hace estructural y funcionalmente más complejas; esto no resulta extraño pues la estructura y las propiedades químicas de cada proteína se han ido afinando a lo largo de miles de años de la historia evolutiva (Berg, L., & Stryer, 2002).

Por otro lado, a pesar de que en el DNA se encuentra almacenada una cantidad increíble de información, no es capaz de llevar acción alguna por sí solo. Las proteínas son los “caballos de batalla moleculares” de la célula, encargadas de construir componentes celulares y de llevar a cabo las reacciones bioquímicas necesarias. Cada célula contiene un juego específico de proteínas cuyas actividades determinan la forma, los movimientos, la función y la capacidad de reproducción de la célula (Audesirk, Audesirk, & Byers, 2004).

Las proteínas constituyen la mayor parte del peso seco de una célula y llevan a cabo prácticamente todas las funciones celulares. El ejemplo más común son las enzimas, que aceleran enormemente la velocidad de las reacciones químicas de una célula, por ejemplo, la degradación de azúcares a CO_2 tardaría más de 1000 años en un ambiente sin enzimas, pero gracias ellas, el proceso toma minutos. Ejemplo de otras funciones son las proteínas embebidas en la membrana plasmática que forman canales y bombas que controlan la entrada y salida de pequeñas moléculas. Otras proteínas transportan mensajes de una célula a otra, mientras que algunas actúan como integradoras de señales, capaces de transmitir la información desde la membrana plasmática hasta el núcleo de la célula. Existen también proteínas que llevan a cabo funciones de máquinas moleculares, como elementos móviles, donde uno de los ejemplos más conocidos es la topoisomerasa que modifica la topología de ácidos nucleicos. También podemos encontrar proteínas más especializadas que actúan como anticuerpos, toxinas, hormonas, moléculas anticongelantes, fibras elásticas, etc (Alberts et al., 2003).

Las proteínas son conocidas también como polipéptidos debido a su naturaleza química. Cada tipo de proteína tiene una secuencia única de aminoácidos. La secuencia repetida de átomos a lo largo de la cadena se conoce como esqueleto polipeptídico. Unidas a esta cadena repetitiva están aquellas partes de los aminoácidos que no participan en la

formación del enlace peptídico, conocidas como cadenas laterales, y que confieren a cada uno de los aminoácidos sus propiedades características (Lehninger, Nelson, & M.Cox, 2005).

Las fuerzas de van der Waals las cuales no permiten el solapamiento de átomos, restringen el número posible de ángulos de un enlace en una cadena polipeptídica. Estas fuerzas y otras limitaciones estéricas reducen severamente el número de disposiciones conformacionales posibles; no obstante, una proteína puede plegarse en un número enorme de posibilidades. El plegamiento de una cadena proteica también está condicionado por muchos tipos diferentes de enlaces no covalentes débiles, como son los enlaces de hidrógeno y los enlaces iónicos, por lo tanto la estabilidad de cada forma plegada se verá afectada por la fuerza combinada de un gran número de estos enlaces. Otro factor sumamente importante que condiciona el plegamiento de cualquier proteína es la distribución de sus aminoácidos polares y apolares (Alberts et al., 2003).

Todas las moléculas proteicas están formadas por miles de átomos unidos entre sí por enlaces covalentes y no covalentes, los cuales están orientados de una manera precisa. Existen cuatro niveles de estructuración en las proteínas: la estructura primaria se define como la secuencia de aminoácidos sin especificar su disposición espacial; la estructura secundaria se refiere a los segmentos de la cadena polipeptídica que forman motivos conocidos como hélices α y láminas β ; y la estructura terciaria se refiere a la conformación tridimensional que forma la cadena polipeptídica. Si una proteína está constituida por un complejo de más de un polipéptido, la relación espacial entre los polipéptidos se refiere entonces a la estructura cuaternaria (Lehninger et al., 2005).

1.2 Proteínas periplásmicas de unión

Las bacterias Gram negativas presentan un espacio llamado periplasma, que se encuentra entre la membrana externa y la membrana citoplasmática. En el espacio periplásmico se encuentran gran diversidad de proteínas, donde muchas funcionan como sistema de transporte y se denominan *proteínas periplásmicas de unión*, también conocidas como PBPs por sus siglas en inglés (Periplasmic Binding Proteins) (Gilson et al., 1988).

Los sistemas de transporte que emplean PBP's, en conjunto con un transportador de membrana y proteínas que hidrolizan ATP (Figura 1), son nombrados sistemas de transporte ABC por sus siglas en inglés ATP-binding cassette. En procariontes se han

encontrado más de 200 sistemas diferentes que transportan compuestos orgánicos e inorgánicos (Madigan, Martinko, Dunlap, & Clark, 2009).

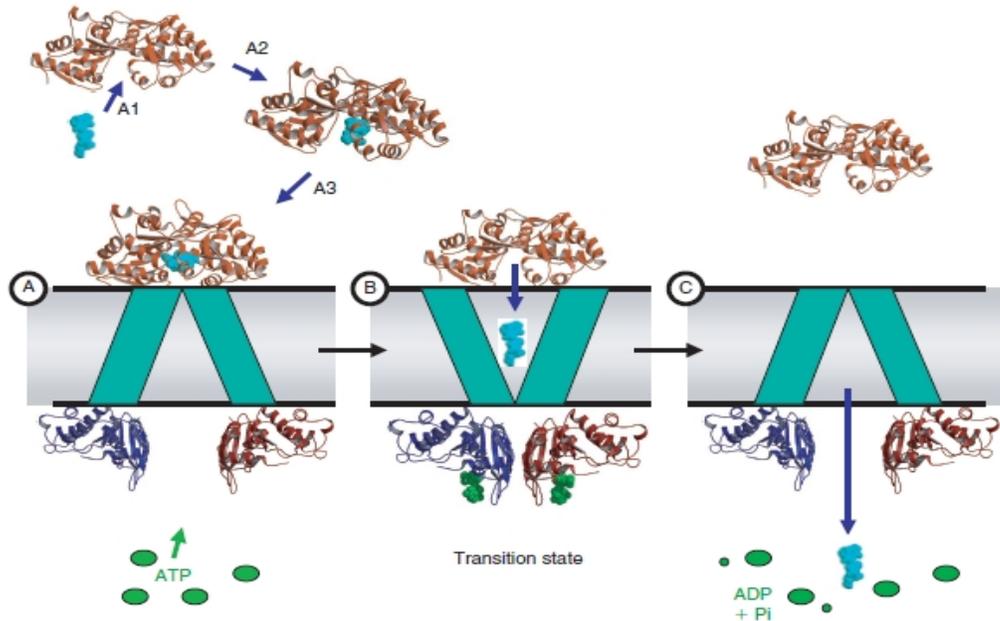


Figura 1. Esquematación del transporte de membrana por proteínas periplásmicas de unión. (A) La PBP se une al ligando ocasionando un cambio conformacional de su forma abierta (A1) a su forma cerrada (A2). A continuación la proteína periplásmica se une al componente de membrana (representado con barras azules A3). Esta interacción ocasiona un cambio conformacional en el componente de membrana (B). Hidrólisis de ATP; este paso se asocia a la apertura de la PBP liberando el sustrato. La liberación de ADP y fosfato a partir de las subunidades de la ATPasa (C) da como resultado el paso del sustrato al citoplasma, así como también la liberación de la PBP del complejo de membrana. Holland, I. (2003). ABC Proteins: From Bacteria To Man [Imagen].

Se pueden encontrar transportadores homólogos en bacterias Gram positivas, que a diferencia de las Gram negativas no presentan membrana externa, por lo que la proteína receptora se ancla en la superficie externa de la membrana citoplasmática. Las PBP's se caracterizan por tener alta afinidad a su sustrato en un rango desde nanomolar a milimolar (Gilson et al., 1988). Se ha observado que la internalización de ciertas moléculas es mucho más lenta e ineficiente en mutantes sin PBP's (Davidson, Shuman, & Nikaido, 1992).

En el estudio del genoma de *Escherichia coli* se ha propuesto que existen entre 40 y 50 proteínas periplásmicas de unión. Las PBPs son monómeros que varían en tamaño, teniendo afinidad a diversos sustratos que incluyen oxianiones, aminoácidos, azúcares, péptidos, poliaminas, vitaminas, iones metálicos o quelatos (Gilson et al., 1988).

Aunque existe una amplia diversidad de unión a ligandos, la mayoría de las PBPs tienen una topología (arreglo tridimensional) común en su estructura, que en general comprende dos dominios lobulares de topología similar. Cada dominio consta de láminas β plegadas en la parte central, flanqueadas por hélices α (Nikaido & Ames, 1992). Ambos dominios están unidos por segmentos polipeptídicos, los cuales pueden ser uno, dos y (algunas veces) tres, que generalmente están en conformación extendida (Holland, Kuchler, Higgins, & Cole, 2003).

La unión del ligando a la proteína está acompañada por el movimiento de ambos lóbulos. Se han determinado estructuras cristalográficas que muestran una conformación “abierta” y otra “cerrada” para la proteína con y sin ligando respectivamente (Gilson et al., 1988). Se especula que el ligando inicialmente interactúa con sólo uno de los dos dominios, ya que en la mayoría de las estructuras cristalográficas de PBPs con su sustrato uno de los dominios contribuye con una proporción significativamente mayor de la superficie de unión que el otro dominio. Los dominios posteriormente se acercan y el ligando interactúa con ambos dominios (Holland et al., 2003).

Se consideran dos clasificaciones para las PBP's de acuerdo a su estructura. La primera de ellas está basada en el número y disposición de las láminas β (Figura 2) que conforman la región central de cada lóbulo, agrupando a las PBP's en proteínas de tipo I, las cuales presentan seis láminas β ; y las tipo II, que presentan solo cinco (Dwyer & Hellinga, 2004). La segunda clasificación se basa en el número de segmentos polipeptídicos que unen ambos dominios (Figura 3), reuniéndolas en tres diferentes grupos: el grupo I presentan tres segmentos polipeptídicos uniendo ambos lóbulos, las del grupo II presentan dos segmentos, y las del grupo III solo un segmento (Shi et al., 2009).

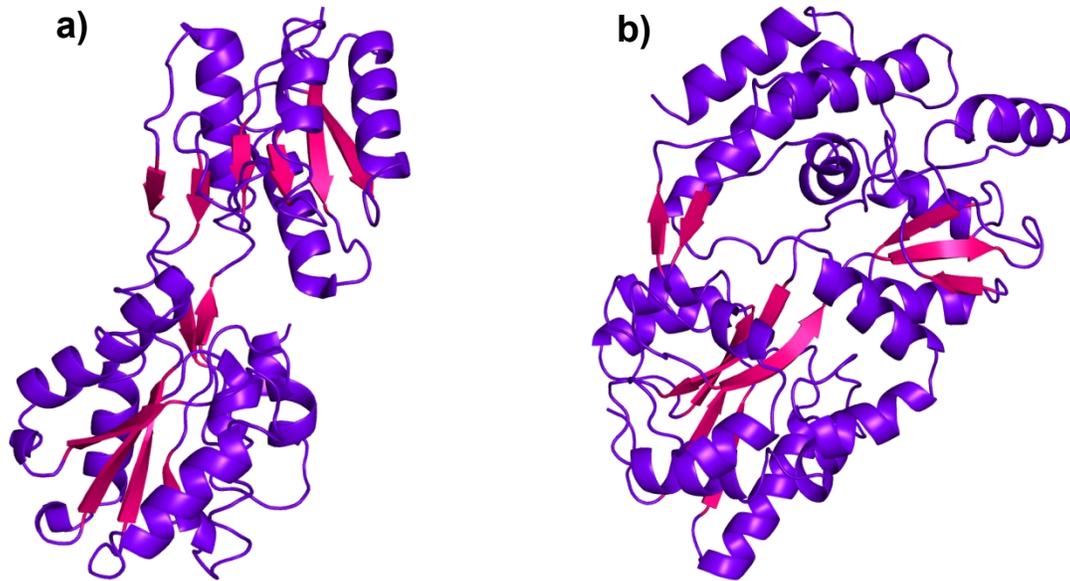


Figura 2. Clasificación de las PBPs basada en el número y disposición de las lámina β (en color rosa) presentes en cada lóbulo (Fukami-Kobayashi, Tateno, & Nishikawa, 1999). a) Proteína de tipo I: PBP de unión a arabinosa (PDB: 2WRZ); b) Proteína de tipo II: PBP de unión a maltosa (1MPC).

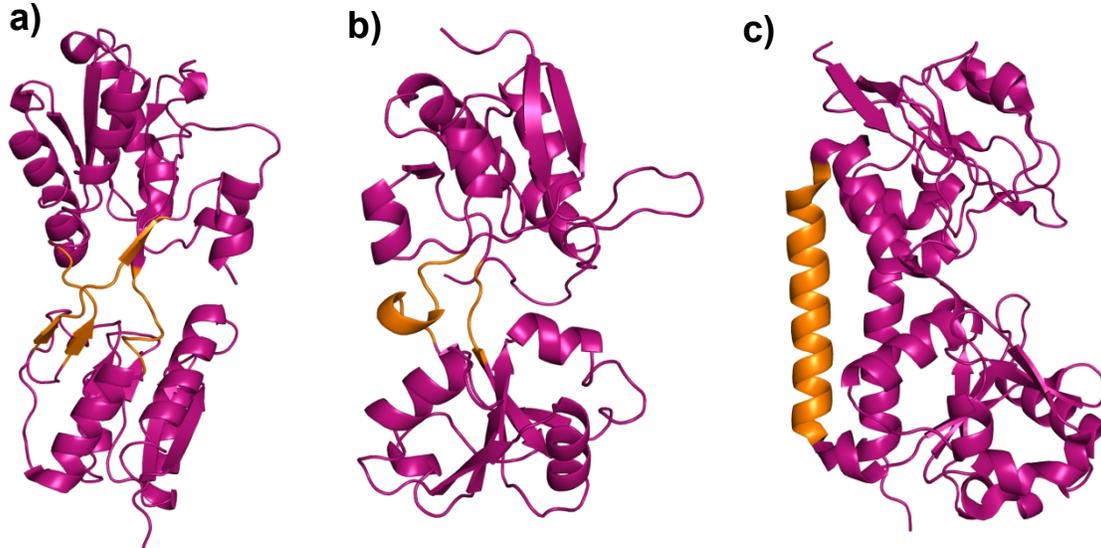


Figura 3. Clasificación de las PBPs basada en el número de segmentos polipeptídicos que unen ambos dominios (en color naranja) (Shi et al., 2009). a) Proteína de tipo I: PBP de unión a arabinosa (PDB: 2WRZ); b) Proteína de tipo II: PBP de unión a Lisina, Arginina y Ornitina (PDB: 2LAO); y c) Proteína de tipo III: PBP de unión a hierro (PDB: 3BE5).

Existe otra clasificación más reciente, donde se incluyen las proteínas de unión a sustrato de los tres dominios de la vida, entre ellas las PBPs; utilizaron los PDBs de 107 proteínas y se consideran seis grupos basados en las características de sus estructuras tridimensionales (Berntsson, Smits, Schmitt, Slotboom, & Poolman, 2010):

Grupo A: las proteínas de éste grupo presentan una hélice α que sirve como la bisagra entre los dos dominios. Esta hélice asegura una estructura rígida que se refleja en el pequeño movimiento de los dos dominios en la unión del sustrato. Todas las proteínas de unión a sustrato de éste grupo juegan un papel importante en la unión de metales.

Grupo B: las proteínas de éste grupo son de unión a carbohidratos principalmente, tales como ribosa, glucosa y arabinosa. En este grupo están contenidas aquellas proteínas con tres segmentos polipeptídicos entre ambos dominios. Además de que el N- y C- terminal no están localizados dentro del mismo dominio.

Grupo C: éstas proteínas unen diferentes ligandos, tales como di- y oligopéptidos, arginina, iones níquel y celobiosa. Las características principales para éste grupo es que son de tamaño grande y tienen un dominio adicional en comparación a las demás proteínas de unión a sustrato.

Grupo D: la característica discernible de estas proteínas es que su región de bisagra está compuesta por dos cadenas polipeptídicas cortas, de entre 4-5 aminoácidos de largo. Este grupo de proteínas de unión se une a una gran variedad de sustratos tales como carbohidratos, putrescina, tiamina, oxianiones tetraédricos como hierro férrico o ferroso.

Grupo E: las características sobresalientes de este grupo de proteínas son una gran única lámina β que forma parte de cada uno de los dos conjuntos de cinco láminas β de cada dominio; además de contener una lámina β que conecta ambos dominios. Una segunda característica distintiva es una hélice larga que se extiende por ambos dominios.

Grupo F: las proteínas de unión a sustrato de éste grupo tienen una región de bisagra compuesta por dos segmentos polipeptídicos que unen ambos dominios, éstos segmentos son significativamente largos de entre 8 a 10 aminoácidos, lo cual probablemente confiera mayor flexibilidad. Une a una gran variedad de sustratos que van desde aniones plana trigonal (nitrato y bicarbonato) a los aminoácidos y solutos compatibles, tales como la glicina betaína. LAO se encuentra dentro de éste grupo.

1.3 Proteína de unión a Lisina-Arginina-Ornitina (LAO)

La Proteína de unión a Lisina-Arginina-Ornitina (LAO) participa en el transporte de los aminoácidos básicos L-histidina, L-lisina, L-arginina y L-ornitina, del periplasma al citoplasma. Esta proteína está constituida por 238 aminoácidos, y la conforman dos lóbulos: el lóbulo I (o dominio B), el cual es discontinuo y está formado por los residuos del 1 al 88 y del 195 al 238; y el lóbulo II (o dominio A), que es continuo formado por los residuos 93 al 185. Ambos lóbulos conectados por dos segmentos cortos de cadenas polipéptidicas, formando la región de bisagra del residuo 88 al 91 y 186 al 194 (Ames et al., 1993). Cada lóbulo está constituido a su vez por cinco láminas β , cuatro paralelas y una antiparalela, rodeadas por cinco y cuatro hélices α (Figura 4); tiene dos triptófanos en las posiciones 47 y 130, y un puente disulfuro entre los residuos 38 y 45. La afinidad por sus ligandos es de 15 nM para lisina, 14 nM para arginina, 29 nM para ornitina y 500 nM para histidina (Nikaido & Ames, 1992).

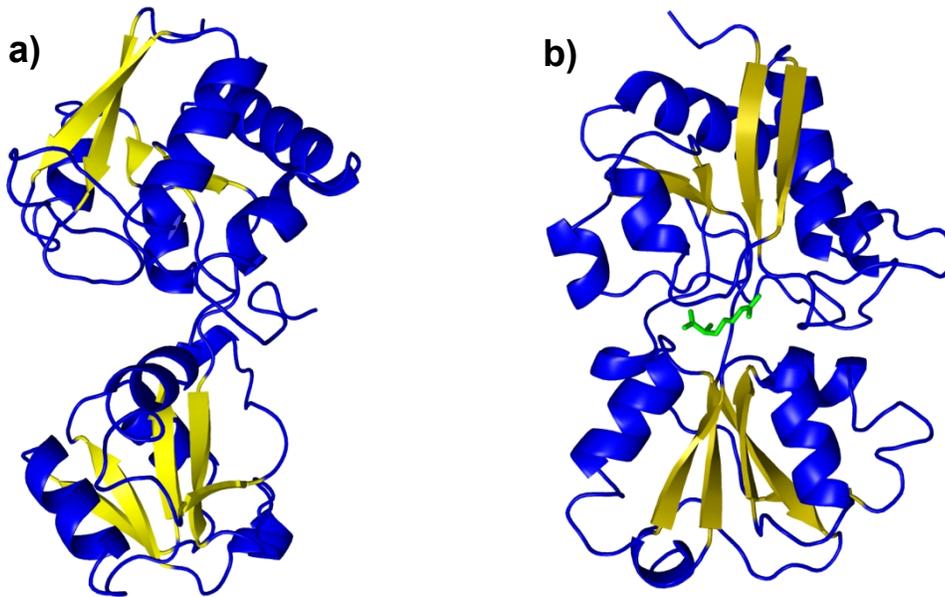


Figura 4. Proteína periplásmica de unión a L - Lisina, L - Arginina y L - Ornitina. a) Proteína LAO sin ligando (forma abierta, PDB: 2LAO); b) Proteína LAO con Arginina como ligando (forma cerrada, PDB: 1LAF); en azul se presentan las hélices α , en amarillo las láminas β y en verde la Arginina como ligando.

La unión al ligando da lugar a movimientos conformacionales de los cuales resultan los estados de la proteína abierta y cerrada; este movimiento tiene una rotación de 52° alrededor de un eje imaginario ubicado en la zona de bisagra. Los residuos involucrados en la unión de los ligandos son Tyr14, Phe52, Asp30, Ser69, Ser70, Ser72, Arg77, Leu117, Thr121, Asp161 y Asp11 (Ames, Oh, & Kim, 1994). Los anillos aromáticos Tyr14 y la Phe52 mantienen a la cadena lateral del ligando (lisina) en una estructura tipo “sándwich” (Figura 5).

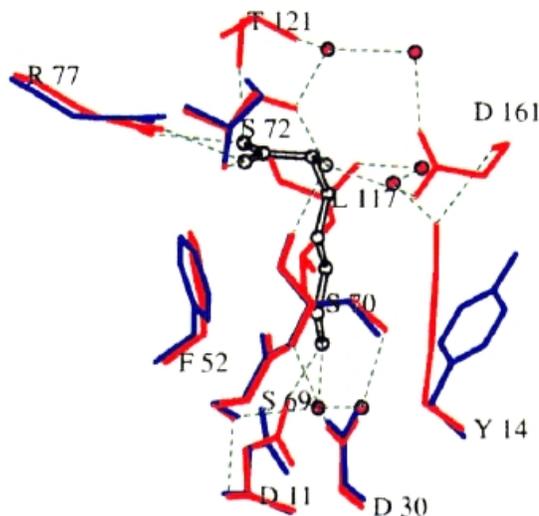


Figura 5. Esquematación de los principales residuos de aminoácidos involucrados en la unión a ligandos de la proteína LAO. En azul se muestran los residuos de aminoácidos de LAO sin ligando, y en rojo se muestran las interacciones entre los residuos de aminoácidos de LAO unido a Lisina; el ligando se muestra en color blanco(Ames et al., 1993). Ames, G. (1993). Three-dimensional structures of the periplasmic lysine/arginine/ornithine-binding protein with and without a ligand [Imagen].

En la estructura cristalográfica con lisina no se observa que el ligando interactúe con la región de bisagra, indicando que el ligando no provoca directamente el cambio conformacional. Por el contrario se observa que presenta mayor interacción con el lóbulo discontinuo (Ames et al., 1993).

1.4 La evolución biológica y su alcance a nivel molecular

La evolución biológica es una de las teorías más fundamentada de la ciencia, unificando todos los conceptos biológicos, Theodosius Dobzhansky en 1973 mencionó lo siguiente: "nada en biología tiene sentido excepto a la luz de la evolución". De manera general el termino evolución se refiere a los cambios acumulativos que ocurren en los organismos de generación en generación dentro de un mismo linaje, que pueden ir desde cambios macroscópicos hasta microscópicos, desde el cambio de una estructura y su función hasta cambios en su información genética que se almacena en el DNA, y que a su vez codifica para la información de la secuencia de proteínas a través del dogma central de la biología (Ridley, 2004).

En 1859 Charles Darwin con su publicación "El origen de las especies" mostró al mundo los fundamentos que aún hoy en día son la base de los mecanismos evolutivos; en la lucha por la existencia, sólo aquellos organismos mejor adaptados al medio sobrevivirán y dejarán descendencia, este proceso es mejor conocido como *selección natural*. Para que haya evolución es imprescindible que exista variación de las características de los organismos, y las principales fuentes de variación conocidas son la mutación y la recombinación. Actualmente sabemos que estas características están "contenidas" en los genes de los seres vivos, y que puede haber diferentes formas de los genes en una misma población llamadas alelos.

La evolución, a nivel molecular, es observable en los cambios de nucleótidos en el DNA y algunas veces da como resultado cambios de aminoácidos en las proteínas. Las evidencias bioquímicas que más ejemplifican el proceso evolutivo, provienen de la comparación de la estructura primaria de proteínas homólogas en diferentes especies, junto con el análisis de secuencias de DNA correspondientes a las mismas. La homología de las secuencias de aminoácidos encontradas en proteínas de diferentes especies constituye una poderosa prueba del origen común de éstas y por lo tanto del origen común de los seres vivos (Cardella & Hernández, 2002).

Algunas modificaciones a nivel molecular que pueden dar origen a cambios con posible significado evolutivo son las mutaciones, la duplicación génica, translocación de material genético, adquisición de genes (trasferencia horizontal), pérdida de genoma, y recombinación génica, principalmente (Futuyma, 2005).

1.5 El estudio de la evolución molecular y su implicación en la ingeniería de proteínas.

Gran parte de los estudios de la biología molecular están interesados en la caracterización e interpretación funcional de la información codificada en las secuencias de ácidos nucleicos y proteínas. El campo del desarrollo de la evolución molecular ha provisto una gran variedad de análisis evolutivos, los cuales recuperan información acerca de la historia filogenética de las secuencias y de los mecanismos por los cuales se han desarrollado (Easteal, 1998).

El gran interés en el estudio de las proteínas y ácidos nucleicos se debe a que la información contenida en el fenotipo podría ser deducida y obtenida de una manera más directa que a partir de caracteres morfológicos (de Chadarevian, Rheinberger, & Suárez-Díaz, 2009). Se ha visto que los aminoácidos esenciales para la actividad de una proteína determinada se conserva comúnmente a lo largo del tiempo evolutivo, por otro lado, los residuos de aminoácidos que no participan directamente con dicha actividad, presentan mayor variación. Sin embargo, las sustituciones de aminoácidos no siempre se dan de manera aleatoria, existen ciertas regiones en la estructura primaria que sólo toleran ciertas sustituciones de aminoácidos, con el fin de conservar la función (Lehninger et al., 2005).

La reciente acumulación en bases de datos de miles de estructuras tridimensionales de proteínas y el desarrollo de herramientas sensibles para la búsqueda de similitud de secuencias están empezando a arrojar nuevos resultados sobre la evolución de las estructuras de las mismas, en relación con sus funciones. La aparición de un nuevo paradigma donde las estructuras de las proteínas son evolutivamente plásticas y cambiantes, tiene aplicaciones importantes para el diseño de proteínas y abre nuevas puertas en la ingeniería proteica dándole como fin la obtención de las propiedades funcionales deseadas. Una de las muchas preguntas aún sin clarificar es la existencia de proteínas con secuencias similares pero diferentes estructuras, lo cual dificulta métodos de modelado con respecto a la homología (Kinch & Grishin, 2002).

El diseño computacional de proteínas ha cambiado fundamentalmente la forma en que se pueden manipular estas biomacromoléculas, sin embargo, no ha sustituido a la evolución dirigida como el método de elección para la ingeniería de proteínas (Holliger, Jaeger, Fenical, Hill, & Lutz, 2010). En cambio, los investigadores deben reconocer la

complementariedad de las dos estrategias e integrar ambas para una manipulación más eficaz.

1.6 La filogenia como herramienta de análisis evolutivo

Una filogenia es la representación de la historia evolutiva de un linaje, que muestra su agrupación de acuerdo a la relación ancestro – descendiente. Es una hipótesis estadística de relaciones evolutivas que nos habla no sólo de los parentescos entre los organismos, genes, etc., sino también del proceso que se llevó a cabo en el transcurso de la evolución (Ridley, 2004).

El análisis filogenético de las secuencias de DNA o proteínas se ha convertido en una importante herramienta para el estudio de la historia evolutiva de los organismos desde las bacterias hasta los seres humanos. Se puede estudiar las relaciones evolutivas de prácticamente todos los niveles de organización, reinos, phyla, clases taxonómicas, familias, géneros, especies y poblaciones intraespecíficas (Nei, 1996). El análisis filogenético también es importante para aclarar el patrón evolutivo de familias de genes, así como para la comprensión de la evolución adaptativa a nivel molecular. La reconstrucción de árboles filogenéticos inició mediante el uso de métodos estadísticos; en un principio se desarrollaron principalmente para caracteres morfológicos, pero en los últimos años se han desarrollado nuevos métodos para el análisis de datos moleculares (Walsh & Sharma, 2009).

La lógica de la inferencia filogenética es idéntica para los caracteres moleculares y morfológicos, pero los dos tienen distintas propiedades, los métodos y conceptos que se utilizan para cada uno pueden parecer muy diferentes. Debido a que hay muchos métodos de construcción de árboles filogenéticos, es importante conocer las ventajas y desventajas de cada uno de ellos (Blair & Murphy, 2010). Existen diferentes criterios con respecto a la comparación de los métodos para la construcción de un árbol filogenético. Los más importantes son: (a) velocidad de cálculo, (b) consistencia en el estimador de la topología, (c) prueba estadística para la construcción de los árboles filogenéticos, (d) probabilidad de la obtención de la topología correcta, y (e) fiabilidad de las estimaciones para la longitud de cada rama (Nei, 1996).

Las filogenias son importantes para abordar diversas cuestiones biológicas como las relaciones entre las especies o genes; el avance de las tecnologías de secuenciación ha dado al análisis filogenético un nuevo nivel, las filogenias han alcanzado casi todas las

ramas de la biología (Yang & Rannala, 2012). La fuente primaria de obtención de datos moleculares son los ácidos nucleicos, nucleares y extranucleares, en cuyas secuencias ha quedado registrada la historia evolutiva (Olivares, 2007).

1.7 Evolución de proteínas periplásmicas de unión

Se ha considerado que las PBPs provienen de un ancestro en común dada la similitud que existe entre sus mecanismo de unión a ligando, el orden de los genes en el operón de cada una de ellas, y su estructura tridimensional, principalmente. Las PBPs que provienen de organismos evolutivamente divergentes pero que unen al mismo substrato están más relacionadas entre sí, que con otras proteínas específicas para diferentes ligandos provenientes del mismo organismo (Tam & Saier, 1993).

En *E. coli* las PBPs son codificadas por el mismo operón junto con proteínas asociadas al sistema de transporte, tales como las proteínas ABC y permeasas. La organización de los genes en el operón se encuentra generalmente en el siguiente orden: PBPs, permeasas, y proteínas ABC. Los patrones de agrupación de genes para diferentes ligandos son consistentes para estos tres tipos de transportadores, lo que sugiere que los operones involucrados en el sistema de transporte probablemente se originaron de un operón ancestral por un evento de duplicación de genes (Fukami-Kobayashi et al., 1999).

De acuerdo a la existencia de los dos grupos principales de PBPs, con base a la clasificación de la disposición y número de láminas β en cada lóbulo, se puede inferir que dichas diferencias debieron haberse generado a partir de un ancestro común. Se propone que estos cambios estructurales surgieron de la pérdida e intercambio entre ambos lóbulos de láminas β dando como resultado el reordenamiento espacial de los elementos de la estructura secundaria, proceso conocido como dislocación de dominios (Fukami-Kobayashi et al., 1999).

En 1999, Fukami Kobayashi y colaboradores realizaron un análisis de secuencia de 15 PBPs, ocho de ellas eran del tipo I y las otras siete del tipo II. Se obtuvieron dos árboles filogenéticos, el primero de ellos basado en la comparación de aminoácidos en la secuencia y el segundo basado en la estructura tridimensional. Los resultados en ambos árboles fueron muy similares, ya que en ambos hubo una separación completa de los grupos de proteínas de tipo I y tipo II, indicando que en efecto, ambos grupos de proteínas comparten un ancestro común y que la dislocación de dominios se llevó a cabo una sola vez en el proceso evolutivo (Figura 6).

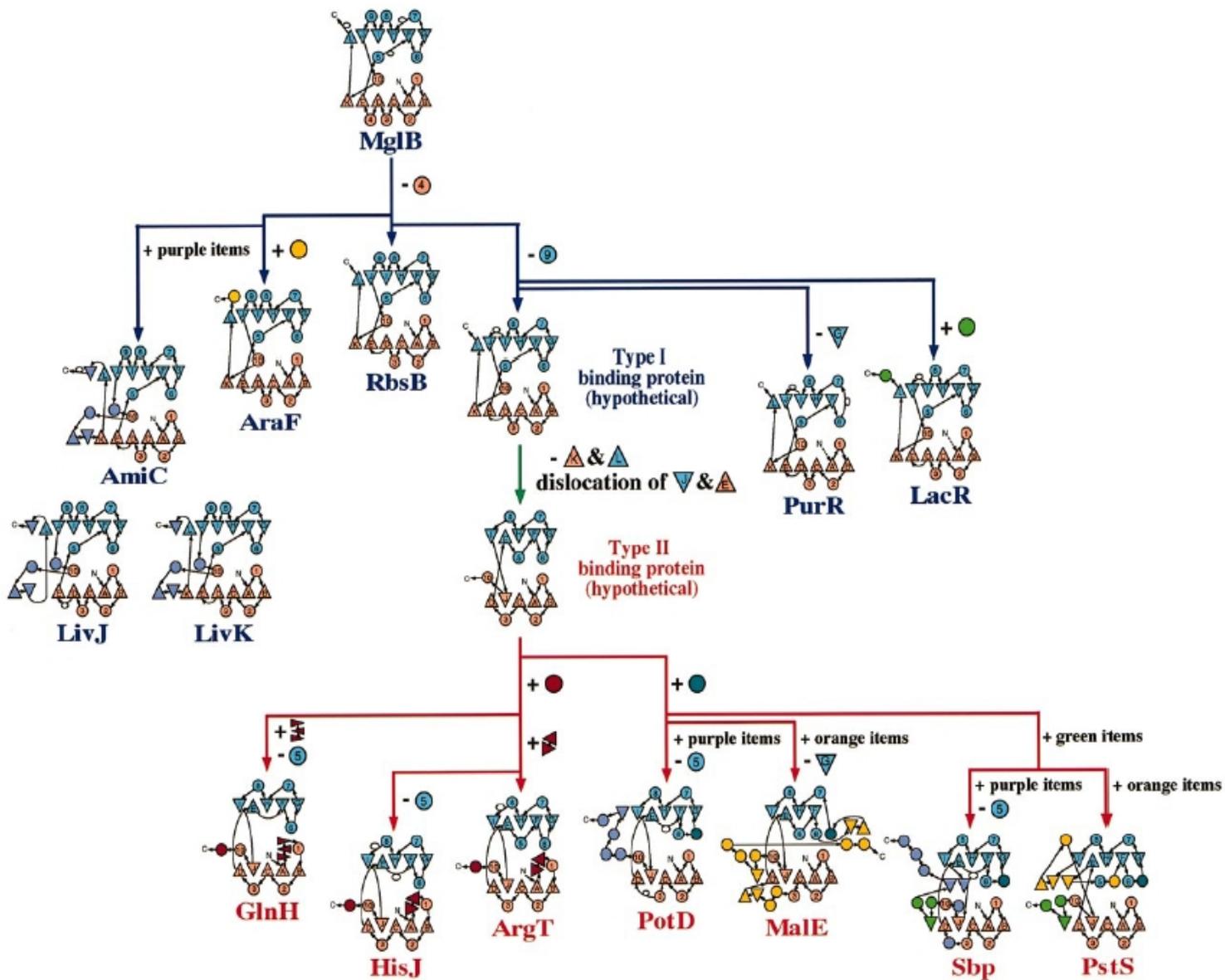


Figura 6. Representación genealógica de las 15 PBPs, donde se muestra los cambios estructurales que pudieron haber dado lugar a las proteínas de tipo I y tipo II a partir de la proteína de unión a galactosa MglB. Los dominios se encuentran coloreados en azul y rojo. Las hélices α están representadas con círculos y las láminas β con triángulos. Fukami-Kobayashi, K. (1999). Domain dislocation: a change of core structure in periplasmic binding proteins in their evolutionary history [Imagen].

Se sugiere que las PBPs tipo I dieron origen a las tipo II debido a que la repetición de elementos en la estructura secundaria es más regular en las proteínas tipo I que en las proteínas tipo II, así como también la mayor complejidad en la topología de las hojas β en las proteínas tipo II. La topología del ancestro común podría ser la misma de la proteína MglB (de unión a galactosa) ya que presenta características topológicas que ninguna otra PBP tiene. Se propone que las PBPs se originaron a partir de la duplicación y fusión del gen de la proteína CheY (Figura 7), involucrada en la quimiotaxis bacteriana (Fukami-Kobayashi et al., 1999).

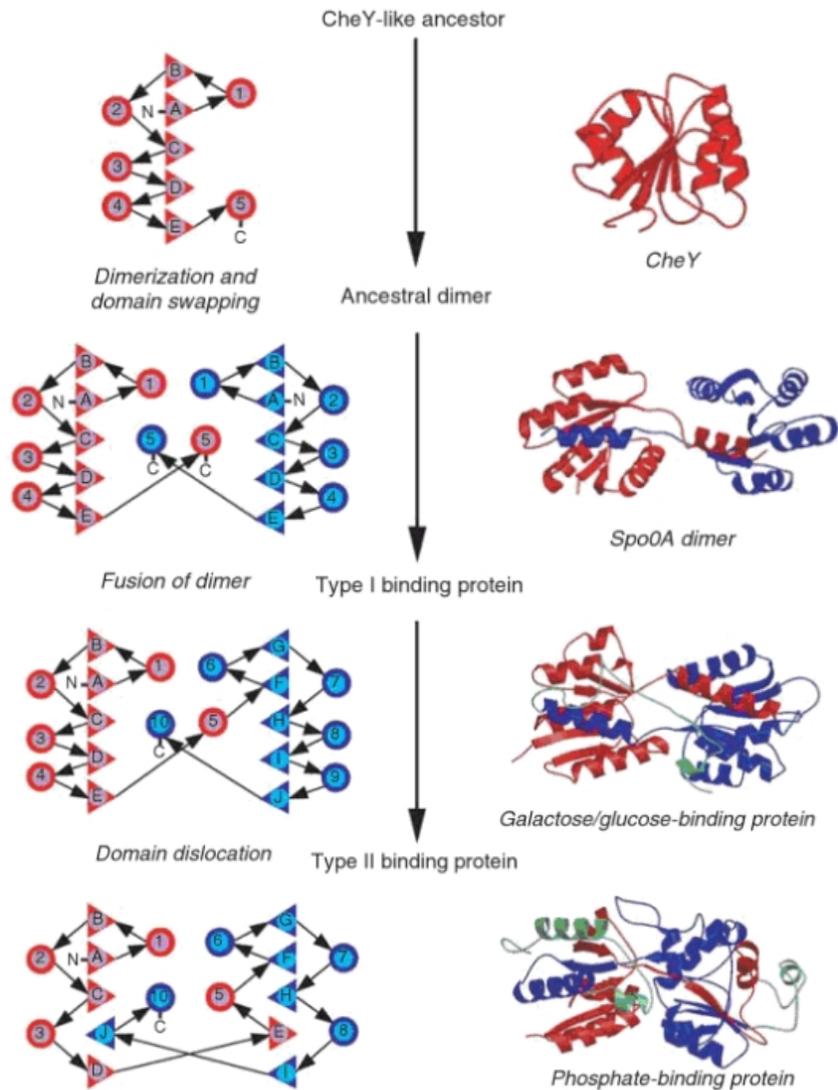


Figura 7. Esquema propuesto para el origen evolutivo de las proteínas periplásmicas de unión tipo I y tipo II, a partir de la proteína CheY (adaptación del esquema de Fukami-Kobayashi et al., 1999). De la proteína CheY se generaron las PBP tipo I, la cual formó un dímero ancestral cuyos monómeros se fusionaron. Las PBP tipo II se originaron a partir de una “dislocación de dominio” de una PBP tipo I. Los triángulos representan láminas β y los círculos las α hélices. Holland, I. (2003). ABC Proteins: From Bacteria To Man [Imagen].

1.8 Principales métodos utilizados en análisis filogenético

El objetivo del análisis filogenético es construir un árbol que refleje las relaciones evolutivas a partir de un conjunto de datos, los cuales pueden ser de muchos tipos, secuencias, genomas o características, entre otros. Los avances en las tecnologías para la secuenciación de DNA han dado al análisis filogenético un peso considerable, alcanzando hoy en día casi todas las ramas de la biología. En la inferencia de la evolución de proteínas, la filogenia se ha utilizado principalmente para rastrear las sustituciones sinónimas y no sinónimas a lo largo de las ramas, así como para identificar las tasas de cambio de aminoácidos (Blair & Murphy, 2010).

Existen distintas aproximaciones para la construcción de árboles filogenéticos, los más conocidos son los métodos de distancia, los cuales consisten en calcular las distancias de las secuencias pareadas. Este método trabaja con una matriz de distancia, el ejemplo más usado y de implementación más eficiente es Neighbourjoining (vecino más cercano) (Saitou & Nei, 1987).

Máxima parsimonia es otro método utilizado para la construcción de árboles filogenéticos, éste método intenta minimizar la cantidad de cambios evolutivos implicados por el árbol, es decir, busca el árbol que con el menor número de cambios evolutivos. El algoritmo para encontrar el número mínimo de cambios en un árbol fue desarrollado por Fitch y Hartigan. Los programas de parsimonia utilizados más comunes son PAUP, MEGA y TNT (Yang & Rannala, 2012).

La inferencia bayesiana es una metodología general de inferencia estadística, los parámetros en este modelo son considerados variables aleatorias con distribuciones estadísticas. Antes del análisis de los datos, se le asigna a los parámetros una distribución *a priori*, que se combina con los datos para generar la distribución posterior. Por lo tanto, todas las inferencias a los parámetros se basan entonces en la distribución posterior. El lanzamiento del programa MrBayes hizo de éste un método popular entre los sistemáticos moleculares (Huelsenbeck & Ronquist, 2001). Un programa de más reciente de implementación bayesiana es BEAST.

Máxima verosimilitud (MV) es un método desarrollado por R. A. Fisher en la década de 1920, como una metodología estadística para la estimación de parámetros desconocidos en un modelo (Yang & Rannala, 2012). La Máxima verosimilitud es una estimación de inferencia estadística, la cual supone la obtención del árbol filogenético con la mayor

verosimilitud dados los datos observados (Felsenstein, 1981), las inferencias por verosimilitud se basan solamente en los datos y el modelo (Evans & Rosenthal, 2005). El método de MV representa toda la información de los datos sobre los parámetros maximizando la verosimilitud. Las estimaciones de máxima verosimilitud tienen usualmente muestras grandes y propiedades deseables: son imparciales, coherentes (que se acercan a los valores reales) y eficientes (tienen menor varianza entre las estimaciones) (Felsenstein, 1981).

El primer algoritmo para el análisis de datos con el método de MV para secuencias de DNA fue desarrollado por Felsenstein. Dos pasos involucrados en la estimación de árboles con éste método son la optimización de la longitud de las ramas para calcular la puntuación de cada árbol candidato, y una búsqueda de la mejor topología del árbol en el espacio. Desde un punto de vista estadístico, la topología del árbol es un modelo en lugar de un parámetro, mientras que longitudes de las ramas en el árbol y las tasas de sustitución son los parámetros en el modelo (Felsenstein, 1981).

Las primeras implementaciones del método de MV fueron PHYLIP, MOLPHY y PAUP 4.0. Entre las más modernas encontramos a PhyML, RAxML, Garli y MEGA 5, las cuales, no sólo son computacionalmente mucho más rápidas sino también son más eficaces en la búsqueda de árboles con altas puntuaciones de verosimilitud (Yang & Rannala, 2012). La velocidad de los métodos para reconstrucciones filogenéticas es un atributo importante, así como también la precisión de los resultados. El enfoque del método de MV es especialmente preciso para la construcción de filogenias moleculares. No obstante, los cálculos para la construcción de un árbol éste método son difíciles, ya que la verosimilitud del árbol no sólo depende de la topología del mismo sino también de otros parámetros numéricos incluyendo la longitud de las ramas, optimizada para cada posición de inserción y el reordenamiento de las mismas (Guindon & Gascuel, 2003). El algoritmo implementado en el programa PHYML es uno de los más rápidos que existe en comparación con otros de máxima verosimilitud, siendo equivalente en precisión topológica y la maximización en la verosimilitud; su velocidad permite la inferencia de árboles muy grandes. La velocidad de PHYML se explica en parte por el hecho de que inicia con un árbol basado en el método de distancia (Guindon & Gascuel, 2003).

Las tasas de sustitución reflejan aspectos biológicos, químicos y propiedades físicas de los aminoácidos. Se sabe que los sitios de la secuencia de una proteína no evolucionan al mismo tiempo ni a la misma velocidad, algunos sitios son lentos, o simplemente se

conservan debido a fuertes restricciones funcionales o estructurales, mientras que otros evolucionan rápidamente. En los métodos de MV se utilizan las probabilidades de cambio a lo largo de las ramas del árbol para obtener la verosimilitud de los datos. Las matrices de sustitución de aminoácidos son esenciales para las filogenias de proteínas. Para aminoácidos PHYML utiliza la matriz LG (por default), la cual es una versión mejorada de la matriz de WAG propuesta por Whelan y Goldman en el 2001, a la cual se le incorporaron parámetros como variabilidad en la tasa de evolución a través de los sitios en la estimación de la matriz y el uso de una base de datos mucho más grande y diversa (Le & Gascuel, 2008).

Existe también, una variante del modelo LG, basada en la estructura secundaria y accesibilidad al solvente de la proteína. En este modelo, se clasifican los sitios en internos o expuestos, lo que involucra matrices de remplazo que difieren dependiendo de las propiedades estructurales de los sitios. Se estimaron tres matrices diferentes: 1) EX tiene dos categorías correspondientes a sitios expuestos e internos. 2) EHO tiene tres categorías correspondientes a sitios extendidos, alfa-hélice y otros sitios. Y por último 3) EX – EHO tiene seis categorías, los sitios son clasificados en: expuestos y extendidos, internos y extendidos, expuestos y alfa-hélice, internos y alfa-hélice, expuestos y otros, o internos y otros. Siendo esta última la que arroja mejores resultados (Le & Gascuel, 2010).

1.9 Reconstrucción de secuencias ancestrales

El propósito de la reconstrucción filogenética es tratar de inferir las relaciones de parentesco o de ancestría-descendencia de un conjunto de datos representativos de entidades biológicas, además de ayudarnos a inferir las secuencias proteicas de los nodos internos del árbol filogenético para de esta manera reconstruir la “proteína ancestral” del nodo de interés. La reconstrucción de secuencias ancestrales nos ayuda a dilucidar cuales son los residuos de aminoácidos de una proteína que se encuentran bajo presiones de selección, a detectar puntos de recombinación, identificar mutaciones, obtener información sobre la estructura y función de las proteínas, entre otros aspectos (Eguiarte, Souza, & Aguirre, 2007).

Existen tres métodos para la reconstrucción de secuencias ancestrales, los métodos de máxima parsimonia, los métodos de máxima verosimilitud y los bayesianos. Los métodos de máxima parsimonia minimizan la cantidad de cambios evolutivos a lo largo de las

ramas del árbol filogenético, asumiendo tasas de cambio lentas y consistentes, sin tomar en cuenta los patrones de sustitución entre aminoácidos o las diferentes longitudes de las ramas del árbol. Los métodos de máxima verosimilitud no tienen estas limitaciones, y por lo general arrojan resultados más fiables pues incluyen tasas de sustitución variables de nucleótidos y aminoácidos (Cai, Pei, & Grishin, 2004).

Benner y colaboradores demostraron que para intervalos evolutivos cortos las mutaciones pueden ser entendidas usando las tasas de sustitución de ácidos nucleicos sin embargo, para intervalos más largos, hay requerimientos biológicos con respecto a la estructura y función proteica que limitan el camino evolutivo (Bennet, Cohen, & Gonnet, 1994).

Al tener en cuenta la ruta evolutiva a nivel de aminoácidos es posible tomar ventaja de estas limitaciones con el fin de modelar correctamente el proceso de la evolución molecular a escala. Este enfoque también nos permite incluir directamente la información sobre la estructura local en cada sitio en la reconstrucción mediante el uso de probabilidades a priori y matrices de mutaciones específicas para cada tipo de estructura (Koshi & Goldstein, 1996). La reconstrucción de secuencias ancestrales nos permite esclarecer la pérdida de una actividad que de acuerdo a una hipótesis la proteína ancestral presentaba, así como especialización en proteínas.

2. JUSTIFICACIÓN

La proteína de unión a Lisina, Arginina y Ornitina (proteína LAO) perteneciente a la superfamilia de las PBPS (proteínas periplásmicas de unión), representa un buen modelo para el estudio de las bases moleculares de la interacción proteína – ligando, ya que ésta proteína tiene la capacidad para unirse a múltiples ligandos. La base molecular de la diversidad en las interacciones entre estas proteínas y sus ligandos es aún un problema no resuelto. Por lo que en este trabajo reconstruiremos la historia evolutiva de LAO a partir de su filogenia con el fin de aportar información que contribuya en la solución de éste problema, así como con la ayuda del estudio estructural a partir de la reconstrucción de secuencias ancestrales para identificar los residuos de aminoácidos implicados en la unión de diversos ligandos en las PBPs.

3. OBJETIVOS

3.1 General

Construir la filogenia de la proteína LAO para obtener un patrón de ancestría – descendencia, que aporte información de los cambios en secuencia que originaron la diversidad de afinidad a sus ligandos en las PBPs.

3.2 Particulares

- Correlacionar la divergencia evolutiva de las PBP's de unión a aminoácidos homólogos a la proteína LAO con el nicho ecológico de las especies en las que se encuentran a nivel de clase taxonómica, así como la relación que existe con el ligando al que unen.
- Obtener las secuencias ancestrales de los nodos principales donde coincida más de una clase de PBP a partir de la filogenia generada.
- Obtener y analizar las estructuras teóricas de los posibles ancestros de las proteínas homólogas a LAO identificando los posibles residuos de aminoácidos que están directamente implicados con la diversidad a la unión a diferentes ligandos.

4. MATERIALES Y MÉTODOS

Para la obtención de la filogenia de la proteína LAO se realizaron una serie de pasos divididos en distintos bloques, que se indican a continuación.

4.1 Obtención de secuencias homólogas a LAO

Para realizar las búsquedas de las secuencias homólogas a la proteína LAO se utilizó la secuencia con número de registro GI:157835300, obtenida de la base de datos NCBI (National Center for Biotechnology Information). La cual consta de 238 aminoácidos sin su secuencia de exportación al periplasma, perteneciente al organismo *Salmonella typhimurium* puesto que dicha secuencia tiene estructura cristalográfica reportada:

```
1   ALPQTVRIGT DTTYAPFSSK DAKGEFIGFD IDLGNECKR MQVKCTWVAS DFDALIPSLK
61  AKKIDAIISS LSITDKRQQE IAFSDKLYAA DSRLIAAKGS PIQPTLESLE GKHVGVLQGS
121 TQEAYANDNW RTKGVVVVAY ANQDLIYSDL TAGRLDAALQ DEVAASEGFL KQPAGKEYAF
181 AGPSVKDKKY FGDGTGVGLR KDDTELKAAF DKALTELRQD GTYDKMAKKY FDFNVYGD
```

A partir de ésta secuencia, se realizó un censo en la base de datos NCBI con ayuda de la herramienta BLAST (Basic Local Alignment Search Tool) (Altschul, Gish, Miller, Myers, & Lipman, 1990). Esta herramienta calcula la similitud en los resultados para alineamientos locales (las regiones más similares entre dos secuencias), para ello utiliza matrices de puntuación específica, generando una tabla con las mejores coincidencias de la base de datos (Altschul et al., 1990). Los parámetros que delimitaron el censo correspondieron a una búsqueda en secuencias de proteínas no redundantes, en un principio se restringió el resultado a sólo secuencias pertenecientes a organismos del dominio Bacteria y se realizó una sondeo original que arrojó como límite máximo 5 000 secuencias, de las cuales se localizaron las diversas PBPs obtenidas. Posteriormente, se realizó un BLAST utilizando la misma secuencia de la proteína LAO pero esta vez con la restricción de que solo se obtuvieran aquellas secuencias pertenecientes a cada una de las PBPs obtenidas en el censo. Para cada BLAST se obtuvieron 500 secuencias por cada PBP especificada en la búsqueda.

Una vez obtenidas las secuencias para cada proteína, se seleccionaron algunas de éstas con criterios específicos, con el fin de elegir solo aquellas que aportaran información no sesgada o reiterativa. Los criterios de depuración fueron los siguientes: 1) porcentaje de empalme (query coverage) superior a 60 %, este valor indica qué porcentaje de la

secuencia de búsqueda está alineada con cada secuencia obtenida; 2) un valor de expectancia estadístico (E-value) menor a 1×10^{-12} , lo que indica el número de resultados que se puede esperar encontrar únicamente por azar cuando se hace una búsqueda en una base de datos de tamaño particular, esencialmente describe el ruido de fondo de la búsqueda por azar, cuanto menor sea éste valor o sea más cercano a cero resulta más significativo que la secuencia encontrada sea homóloga a la secuencia de búsqueda; por último 3) se eliminaron secuencias hipotéticas, putativas, inferidas de gen, secuencias sin nombre, así como secuencias que no tenían registrada la especie a la que pertenecen o su etiqueta de registro no contenía la información del ligando al cual unen.

Una vez descartadas las secuencias no confiables, el conjunto de datos resultante disminuyó notablemente, se prosiguió a eliminar secuencias idénticas o redundantes de una misma especie, tomando en cuenta cuál de las secuencias con organismos repetidos tenían mejores parámetros, tanto los ya mencionados así como también el 4) valor de puntuación, este valor es una puntuación normalizada del alineamiento, matriz de sustitución y penalización de gaps; y 5) máxima identidad, refiriéndose al porcentaje de similitud entre la secuencia de búsqueda y la secuencia resultante sobre la longitud del área que converge; y la 6) puntuación total, que es la suma de puntuación de todas las secuencias alineadas. Del conjunto de datos obtenido a partir de la depuración, se descargaron las secuencias seleccionadas.

Se eligió como grupo externo la proteína involucrada en quimiotaxis CheY, la cual es una candidata para ser ancestro de las PBPs, ya que se piensa que las tipo I podrían ser un estado dimérico de la proteína CheY (Fukami-Kobayashi et al., 1999); se realizó un BLAST con la secuencias de la proteína CheY de *Salmonella typhimurium* con número de acceso AAA27037.1, se utilizaron los mismos criterios de depuración y se eligieron las primeras diez secuencias.

Se especificó la clase taxonómica de cada secuencia de acuerdo al organismo con ayuda de la base de datos de “List of Prokaryotic names with Standing in Nomenclature” (Euzéby, 1998).

4.2 Alineamiento

El alineamiento de la secuencias se realizó con el programa MAFFT (Multiple Aligment – Fast Fourier Transform), este programa está especializado para alinear múltiples secuencias de residuos de aminoácidos, y está basado en la transformación de Fourier; la

ventaja es que reduce el tiempo de cómputo drásticamente comparado con otros métodos existentes. MAFFT incluye una matriz de sustitución optimizada para aminoácidos, donde las regiones homólogas son rápidamente identificadas por medio de una transformación de Fourier, en el cual la secuencia de aminoácidos es convertida a una secuencia compuesta de valores de volumen y polaridad para cada residuo (Kato, Misawa, Kuma, & Miyata, 2002). Se utilizó el servidor MAFFT versión 7, con los parámetros asignados por default.

4.3 Construcción del árbol filogenético

Para la construcción del árbol filogenético de la proteína LAO, se empleó el programa PhyML (Phylogenetic Maximum Likelihood) que está basado en el método de máxima verosimilitud, que se caracteriza por ser rápido y preciso para la estimación filogenética de ML a partir de secuencias de residuos de aminoácidos. Este programa proporciona la estimación de varios parámetros evolutivos, con el fin de realizar análisis filogenéticos integrales en grandes conjuntos de datos en tiempo de cálculo razonable (Guindon, Lethiec, Duroux, & Gascuel, 2005).

Se utilizó el modelo EX-EHO, éste modelo involucra matrices de remplazo que difieren dependiendo de las propiedades estructurales de los sitios, expuestos / internos. EX – EHO tiene seis categorías: expuestos y extendidos, internos y extendidos, expuestos y alfa-hélice, internos y alfa-hélice, expuestos y otros, o internos y otros (Le & Gascuel, 2010); para el rearrreglo topológico del árbol se utilizó el método SPR (subtree pruning and regrafting), que se refiere a los reordenamientos topológicos del árbol, este método corta cada subárbol y los coloca en posiciones distintas permitiendo una búsqueda exhaustiva de la topología (Hordijk & Gascuel, 2005); para la estimación del soporte de las ramas se utilizó aLRT SH-like (approximate likelihood-rate test Shimodaira-Hasegawa-like procedure) la cual es una rápida prueba aproximada de verosimilitud para cada rama. El aLRT SH-like permite identificar y mejorar en la verosimilitud comparando la topología de las ramas contra nuevos re-arreglos locales (Anisimova & Gascuel, 2006)

Para la edición y visualización del árbol se utilizaron: 1) un servidor en línea llamado iTOL (Interactive Tree of Life), utilizado principalmente para representar la distribución de las PBPs por su afinidad a cierto ligando y el grupo externo, así como los dominios de la vida a los cuales correspondían a cada secuencia (Letunic & Bork, 2011); y 2) el programa Archaeopteryx, el cual se utilizó para asignarle como raíz al árbol el grupo de las

proteínas CheY, para observar las distancias evolutivas entre cada grupo y para asignar especie y clase taxonómica a cada secuencia (Han & Zmasek, 2009).

4.4 Reconstrucción y análisis estructural de las secuencias ancestrales

Se utilizó la herramienta de reconstrucción de secuencias ancestrales del servidor de Datamonkey (Delport, Poon, Frost, & Kosakovsky Pond, 2010) usando la forma de reconstrucción *joint* la cual toma en cuenta las probabilidades de sustitución de todos los nodos, ya que es la más adecuada cuando se quieren contar los cambios de cada sitio, ésta forma de reconstrucción consiste en encontrar la ruta de cambios de estado de carácter más verosímil, tomando en cuenta las probabilidades de sustitución de todos los nodos (Pupko, Pe'er, Shamir, & Graur, 2000).

Se eligieron las secuencias de los nodos ancestrales con posibles afinidades promiscuas. Para predecir las estructuras tridimensionales (3D) de cada una de ellas se utilizó el servidor de predicción de estructura de proteínas RaptorX (Källberg et al., 2012). Este servidor predice la estructura secundaria y terciaria de la secuencia del nodo ancestral, así como las regiones desordenadas y la accesibilidad al solvente.

Una vez obtenidas las estructuras en 3D de las secuencias de cada uno de los nodos seleccionados, con la ayuda del programa PyMOL (Schrödinger, 2010) se realizó la visualización y comparación estructural de los PDBs de las secuencias de los nodos ancestrales contra el PDB de LAO (2LAO), con el propósito de identificar las diferencias entre la estructura de los ancestros hipotéticos de cada nodo seleccionado y la estructura actual de una PBP.

5. RESULTADOS

5.1 Número de secuencias obtenidas homólogas a la proteína LAO y diversidad taxonómica

Del censo realizado para elegir a las proteínas homologas a LAO, del dominio Bacteria se obtuvieron siete tipos de PBPs que unen diferentes ligandos: la proteína de unión a Lisina, Arginina y Ornitina (LAO), la proteína de unión a Histidina (HisJ), la proteína de unión a Glutamina (GlnBP), la proteína de unión a Cisteína (CBP), la proteína de unión a Nopalina (NocBP), la proteína de unión a Octopina (OccBP) y la proteína de unión a Histidina, Glutamato, Glutamina, Arginina y Opina (HGGAOBP). Siendo las proteínas tipo GlnBP y CBP las más numerosas (Tabla 1).

Después de la depuración se obtuvieron un total de 389 secuencias, de las cuales 369 fueron PBPs pertenecientes al dominio Bacteria y sólo 10 secuencias fueron representadas en el dominio Archaea pertenecientes a proteínas tipo GlnBP; el grupo externo fue constituido por 10 secuencias de la proteína involucrada en la quimiotaxis CheY correspondientes al dominio Bacteria (Tabla 1).

Las clases taxonómicas representadas para el dominio Bacteria fueron diez, donde Gammaproteobacteria, Bacilli y Betaproteobacteria fueron las clases con mayor número de secuencias. En el caso del dominio Archaea se representaron tres clases taxonómicas con Halobacteria como la clase más numerosa (Tabla 2).

Se observa que la PBP con mayor diversidad taxonómica es la GlnBP, seguida de la CBP; por otro lado LAO e HisJ presentan una gran similitud en cuanto a diversidad, CheY se muestra como una de las proteínas con menor diversidad, lo cual puede deberse al número de secuencias seleccionadas que componen a este grupo. En la proteína tipo NocBP, también se observa con una diversidad taxonómica mínima, representada por una clase Alphaproteobacteria. En cuanto a la proteína de tipo HGGAOBP, presenta tres clases taxonómicas Bacilli, Alphaproteobacteria y Clostridia (Figura 8).

La clase taxonómica con mayor número de secuencias representadas con un 40% del total de secuencias, es Gammaproteobacteria, seguida de la clase Bacilli con un 16% del total y Alphaproteobacteria con un 7% (Figura 9).

Tabla 1. Número de secuencias obtenidas para cada proteína

Proteína	Número de secuencias
Proteína de unión a Glutamina (GlnBP)	144
Proteína de unión a Cisteína (CBP)	95
Proteína de unión a Histidina, Glutamato, Glutamina, Arginina y Opina (HGGAOBP)	50
Proteína de unión a Histidina (HisJ)	40
Proteína de unión a Lisina, Arginina y Ornitina (LAO)	39
Proteína involucrada en la quimiotaxis (CheY)	10
Proteína de unión a Octopina (OccBP)	7
Proteína de unión a Nopalina (NocBP)	4
Total	389

Tabla 2. Número de secuencias obtenidas para cada proteína por clase taxonómica en cada dominio.

Proteína	GlnBP	CBP	HGGAOBP	HisJ	LAO	OccBP	NocBP	CheY
<i>Clases taxonómicas del dominio Bacteria</i>								
Gammaproteobacteria	46	32	-	34	32	2	-	10
Bacilli	49	14	35	-	-	-	-	-
Betaproteobacteria	11	40	-	6	7	-	-	-
Alphaproteobacteria	2	1	14	-	-	5	4	-
Negativicutes	16	-	-	-	-	-	-	-
Clostridia	7	-	1	-	-	-	-	-
Deinococci	-	7	-	-	-	-	-	-
Actinobacteria	2	-	-	-	-	-	-	-
Dictyoglomia	1	-	-	-	-	-	-	-
Cyanobacteria		1		-	-	-	-	-
<i>Clases taxonómicas del dominio Archaea</i>								
Halobacteria	8	-	-	-	-	-	-	-
Archaeoglobi	1	-	-	-	-	-	-	-
Metanococci	1	-	-	-	-	-	-	-

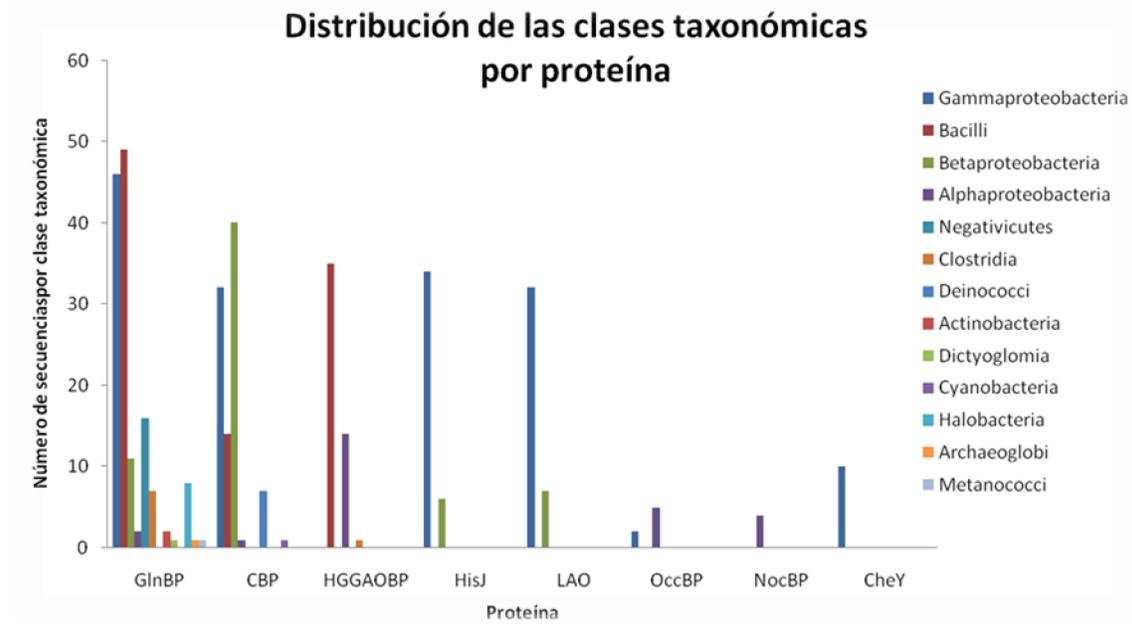


Figura 8. Representación gráfica del número de secuencias por proteína y por clase taxonómica. Mostrando la diversidad taxonómica para cada una.

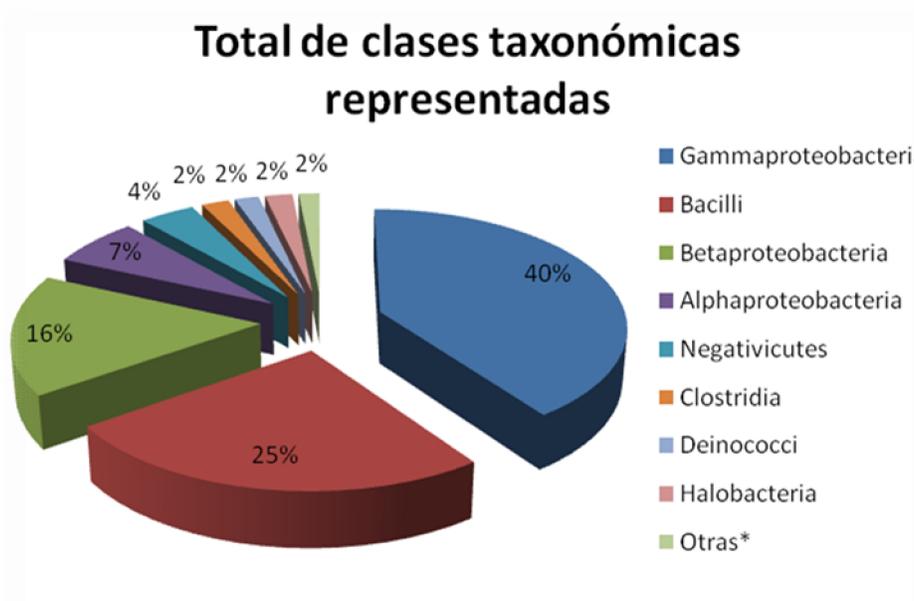


Figura 9. Clases taxonómicas obtenidas para ambos dominios filogenéticos y el porcentaje de secuencias pertenecientes a cada clase. Se muestran las ocho clases más numerosas. En la categoría de "Otras" se representan aquellas clases con un número menor a 5 secuencias.

5.2 Alineamiento de secuencias

El alineamiento con las secuencias de bacterias y arqueas muestra gaps consistentes, teniendo una continuidad en los bloques de los sitios; el gap más grande corresponde a las secuencias de las proteínas NocBP y OccBP. En el caso del alineamiento de las secuencias de bacterias y arqueas más el grupo externo los bloques mantienen su distribución, infiriendo así, que las secuencias de la proteína CheY no perturban el alineamiento para la consecuente construcción del árbol filogenético (Figura 10).

En cuanto a las propiedades de ambos alineamientos el valor que difiere es el de la longitud mínima de secuencia, la cual es de la proteína CheY, que está constituida por 129 residuos; en cuanto al número de posiciones se mantiene, así como la longitud máxima. (Tabla 3).

Tabla 3. Propiedades de los alineamientos sin y con las secuencias del grupo externo.

Bacteria y Archaea	Bacteria, Archaea y CheY
Longitud min de la secuencia: 159	Longitud min de la secuencia: 129
Longitud máx de la secuencia: 260	Longitud máx de la secuencia: 260
Promedio: 222	Promedio: 220
Posiciones : 321	Posiciones: 321

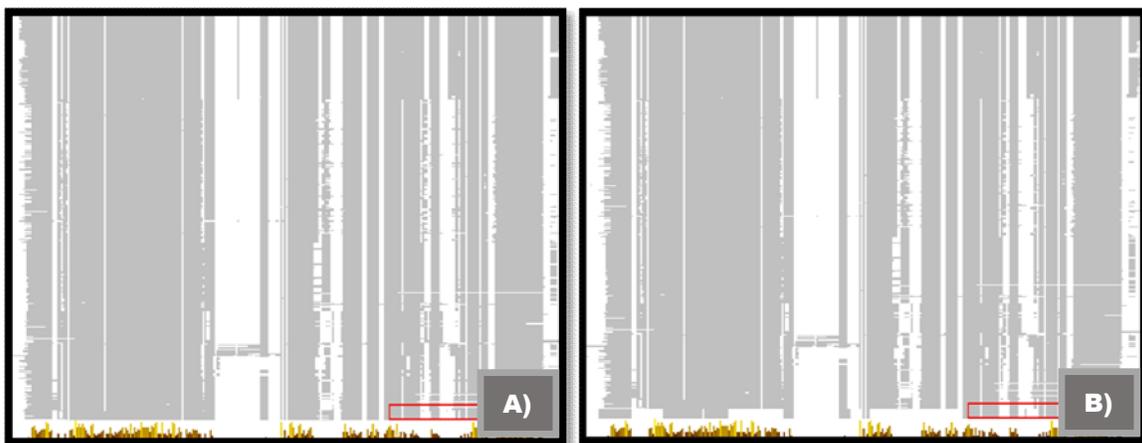


Figura 10. Visualización general del alineamiento utilizando el programa Jalview A) sin las secuencias del grupo externo y B) con las secuencias del grupo externo (CheY); mostrando consistencia en los bloques y gaps.

5.3 Árbol filogenético de las proteínas periplásmicas homólogas a LAO

El árbol filogenético final se muestra en la Figura 11, donde se presentan las relaciones entre las secuencia de cada PBP con respecto al ligando al que unen, así como las relaciones taxonómicas entre ellas. El soporte probabilístico de las ramas principales del árbol (Figura 12) varía dependiendo de cada rama, presentando valores de aLRT cercanos a 1 en las ramas principales; el valor más bajo de alguna de éstas corresponde al nodo perteneciente a las secuencias del grupo externo teniendo un aLRT de 0.784, el nodo donde se separan la mayoría de las GlnPBs de las otras PBPs tiene un valor de aLRT de 0.827, mientras que la rama donde se encuentran algunas GlnPBs y HGGAOBPs tiene un valor de aLRT de 0.938; dentro de esta misma rama se encuentran nueve de las secuencias correspondientes al dominio Archaea; el nodo que agrupa a estas últimas presenta un valor de aLRT de 0.995. Por otro lado la rama que agrupa a la mayoría de las CBP's también presenta un valor alto de 0.891, así como también la rama que agrupa a todas las secuencias de NocBP y OccBP, y algunas de HGGAOBP siendo este último valor de aLRT de 0.964. Se obtuvieron 18 valores de aLRT iguales a cero localizados en las hojas del árbol.

Se puede observar que las HGGAOBPs formaron 3 agrupaciones alrededor de todo el árbol (Figura 11), dos de estas, intercaladas entre las secuencias de GlnBP; las secuencias de arqueas también forman un grupo, solo una de ellas aparece en una rama distinta, la cual pertenece al organismo *Archaeoglobus fulgidus*, localizada también junto a secuencias de GlnBP. Las CBPs aparecen agrupadas en dos ramas principales, una de ellas corresponde sólo secuencias de CBP, y la otra rama es compartida con un pequeño grupo de GlnBPs. Las secuencias de todas las NocBPs y OccBPs se encuentran en una misma rama junto con uno de los tres grupos que forman las HGGAOBPs. Por otro lado, las secuencias de LAO e HisJ se agrupan en una rama principal, siendo éstas las secuencias menos resueltas, así como las más recientes con respecto al grupo externo.

En las Figuras 13 - 19, se muestran los cladogramas a partir de cada rama principal; en primera instancia en el clado que agrupa a HisJ y a LAO se puede observar que la mayoría de las secuencias de ambas proteínas corresponden a la clase taxonómica de Gammaproteobacteria, y sólo unas cuantas de la proteína LAO pertenecen a la clase Betaproteobacteria, las cuales aparecen juntas (Figura 13).

Por otro lado, en el clado de la Figura 14, se encuentran las secuencias de NocBP las cuales pertenecen a la clase taxonómica Alphaproteobacteria: tres del género *Rizobium* y una del género *Agrobacterium*; también encontramos las secuencias de HGGAOBP donde sin excepción todas pertenecen a la clase Alphaproteobacteria; y las secuencias de OccBP, donde seis de ellas son grupo hermano de NocBP: cuatro de la clase Alphaproteobacteria y dos de la clase Gammaproteobacteria (*Erwinia billingiae* y *Pectobacterium astrosepticum*); la séptima secuencias de OccBP perteneciente a *Azospirillum brasilense* se encuentra en la base de la rama que agrupa a todas las HGGAOBP, siendo todas las secuencias de esta rama de la clase Alphaproteobacteria.

Uno de los clados donde se encuentran las secuencias de CBP (Figura 15) está representado por las clases taxonómicas Betaproteobacteria, Bacilli, Gammaproteobacteria y Deinococci en este orden de abundancia de mayor a menor, y por una secuencia del phylum Cyanobacteria representada por el género *Synechococcus*. En el otro clado donde se encuentran la otra parte de las secuencias de CBP (Figura 16), la gran mayoría son de la clase taxonómica Gammaproteobacteria, muy pocas Betaproteobacterias y sólo una secuencia de la clase Alphaproteobacteria, Éste grupo de CBPs son grupo hermano de nueve secuencias correspondientes a GlnBP, representadas estas últimas por clase taxonómica Bacilli.

En el clado de la Figura 17, se observan algunas de las secuencias pertenecientes a GlnBP, las clases que representan esta rama son Bacilli, Betaproteobacteria, Gammaproteobacteria, Clostridia, Dyctioglomia y una secuencia del dominio Archaea clase Archaeoglobi del organismo *Archaeoglobus fulgidus*. Mientras que en el clado de la Figura 18, donde se encuentran el resto de las secuencias de arqueas, se observa que en el grupo de GlnBP aparece una secuencia de bacteria representada por *Halanaerobium saccharolyticum* de la clase taxonómica Clostridia, las demás secuencias de arqueas con las que comparte la rama son en su mayoría de la clase Halobacteria y una de la clase Methanococci localizada en la base del grupo como precedente de estas. Su grupo hermano pertenece a uno de los tres grupos que forman las secuencias de HGGAOBP, representadas en su mayoría por la clase Bacilli y una secuencia de la clase Clostridia.

En la Figura 19 se muestra el árbol colapsado, con el fin de observar las secuencias más cercanas a la raíz, en su mayoría son GlnBPs, de las clases taxonómicas Gammaproteobacteria y Bacilli, así como también Negativicutes, Betaproteobacteria y Clostridia en menor número. Dentro de estas secuencias se puede observar que se

intercaló uno de los grupos de las secuencias de HGGAOBP que por su localización en el árbol precede a los otros dos, éste se compone en su totalidad por secuencias de la clase Bacilli. Por último, en el grupo externo representado en gris, las diez secuencias que lo componen pertenecen a la clase Gammaproteobacteria, donde la secuencia de *Citrobacter rodentium* se encuentra en la base de ésta rama. En general, la distribución de las secuencias no sólo muestra homogeneidad con respecto al ligando al que unen, sino también a su clasificación taxonómica.

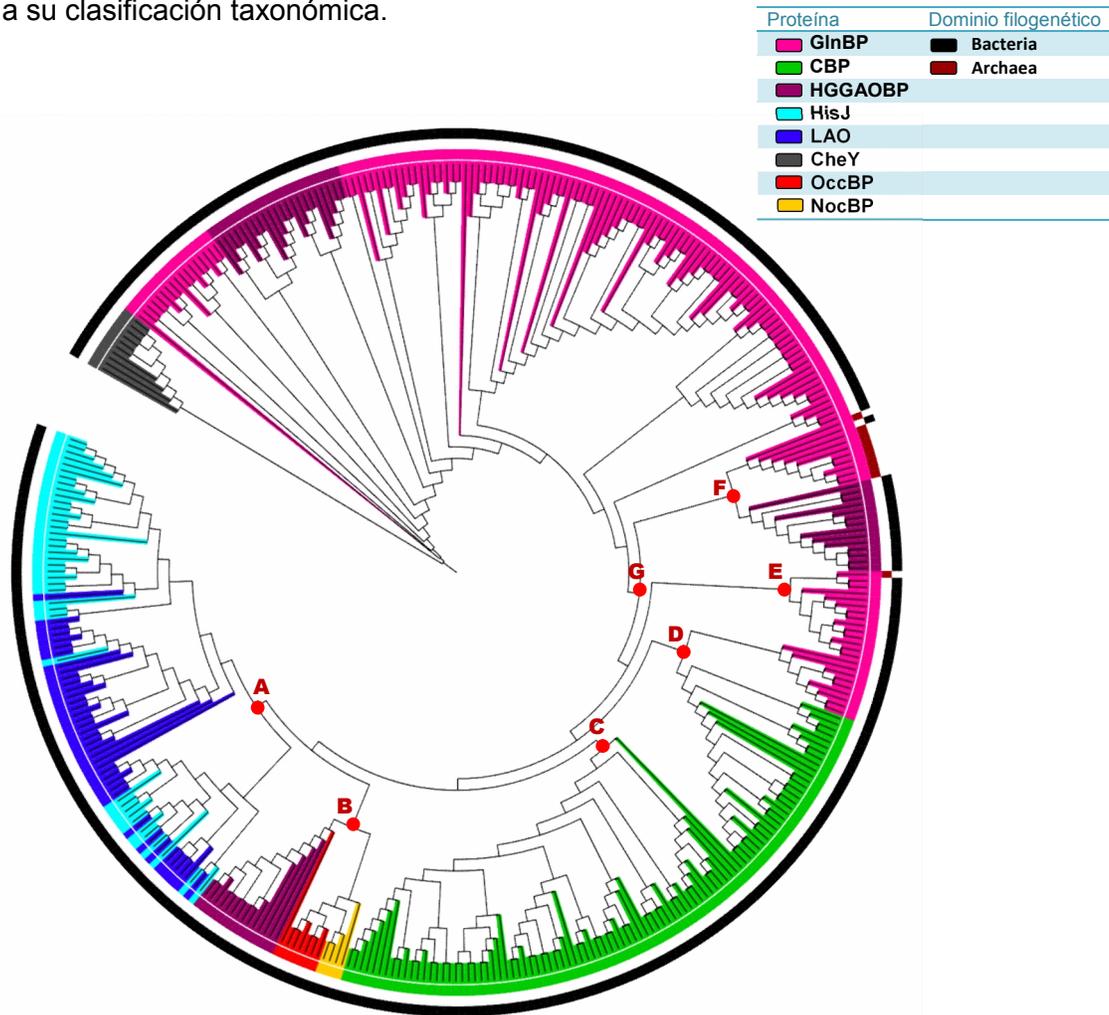


Figura 11. Vista circular del árbol filogenético de las proteínas periplásmicas de unión homólogas a LAO, utilizando como grupo externo a la proteína involucrada en quimiotaxis CheY. Se puede apreciar la distribución de las secuencias de acuerdo a su carga del aminoácido que une cada proteína y al dominio de la vida al cual pertenecen. Los puntos rojos señalan las ramas principales del árbol. Figura realizada en iTOL (Letunic & Bork, 2007, 2011).

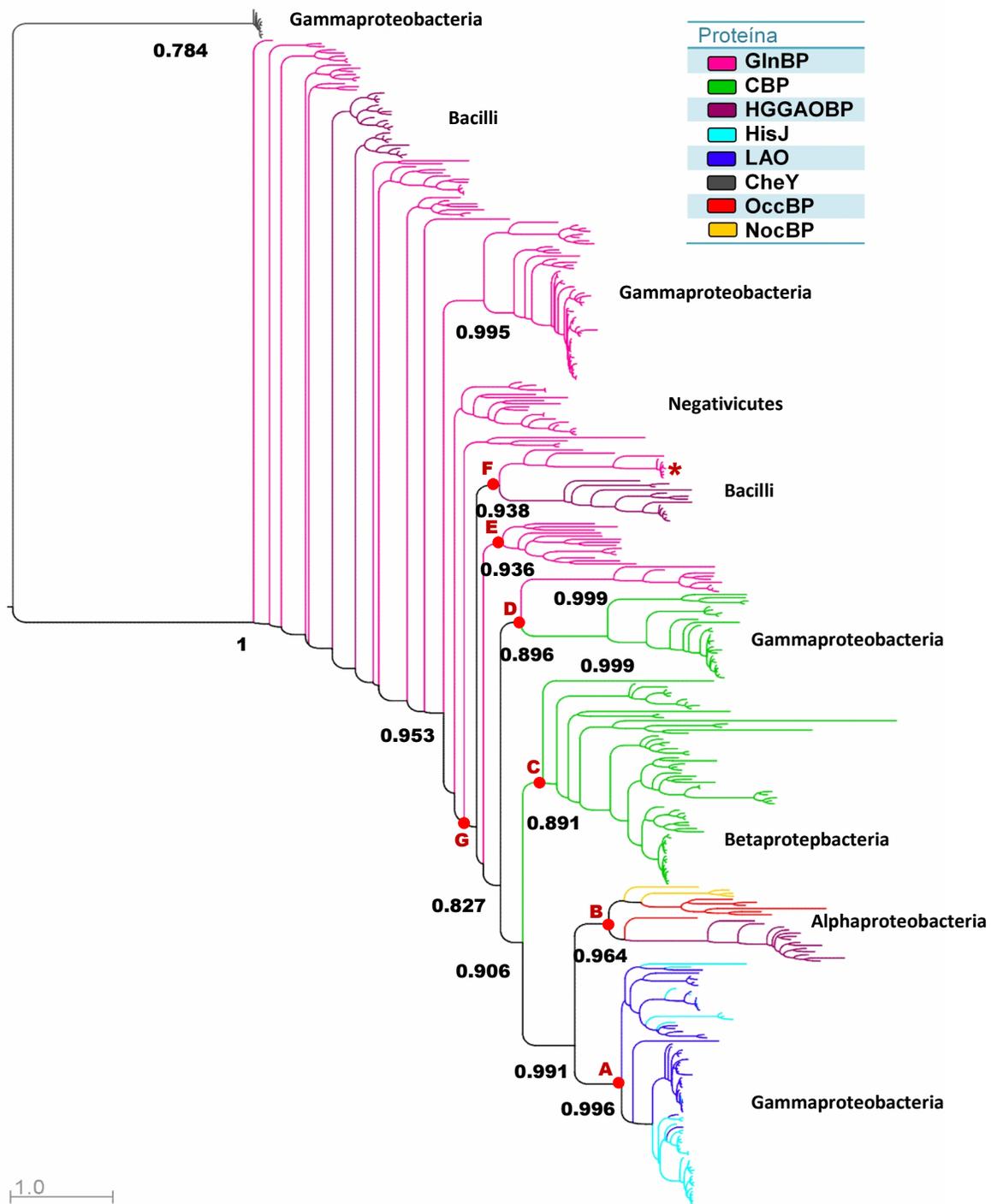


Figura 12. Árbol filogenético de la proteína periplásmica LAO. Vista rectangular del árbol donde se presentan los principales valores de aLRT y las clases taxonómicas más numerosas en cada rama principal, así como las distancias evolutivas de cada secuencia. El asterisco rojo indica la rama donde se encuentran las secuencias del dominio Archaea. Los puntos rojos señalan las ramas principales del árbol. Figura realizada con Archaeopteryx (Han & Zmasek, 2009).

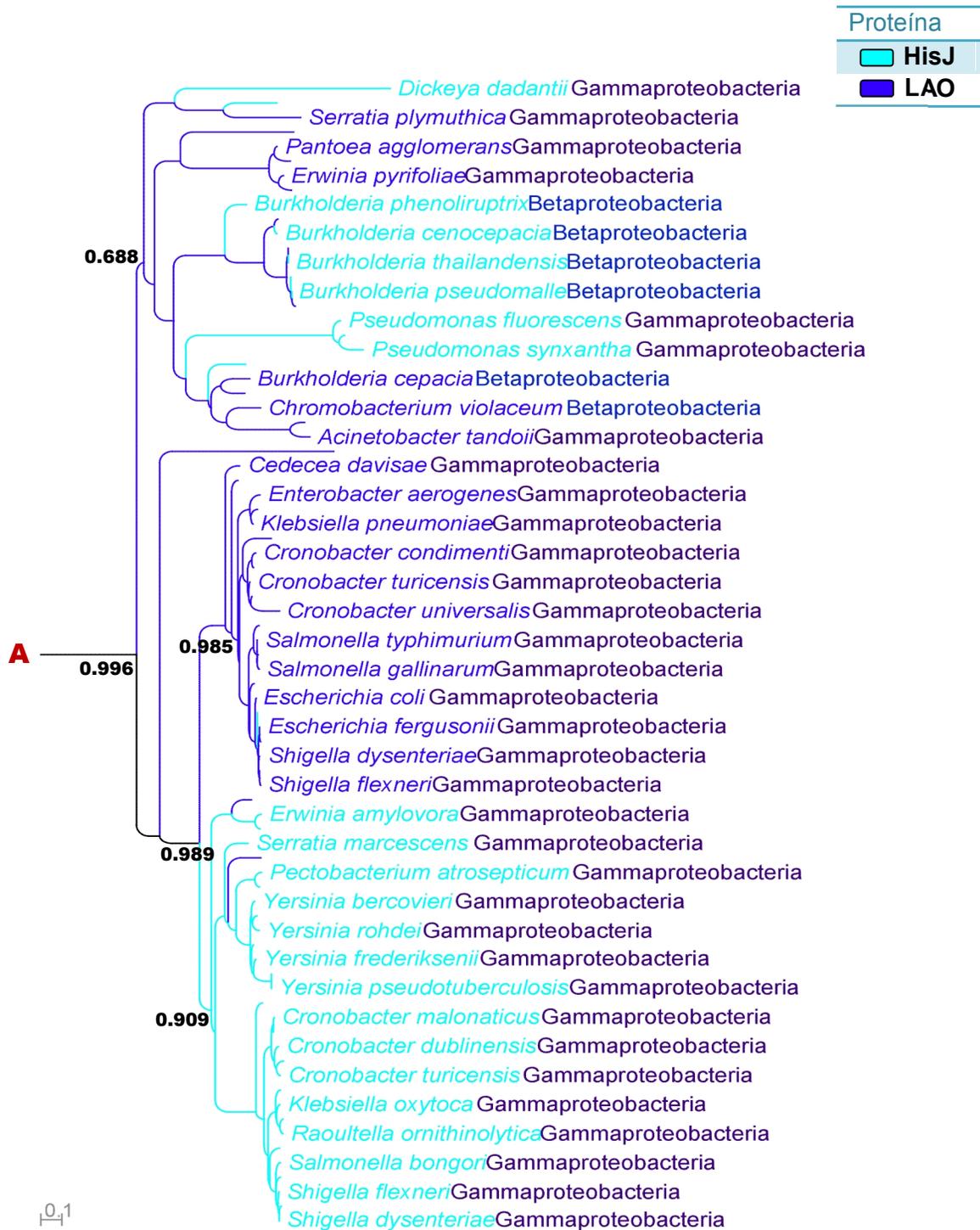


Figura 13. Clado perteneciente al árbol original donde se encuentran secuencias de LAO e HisJ, mostrándose los nombres de algunas especies y clases taxonómicas, así como los principales valores de aLRT. Figura realizada en Archaeopteryx (Han & Zmasek, 2009).

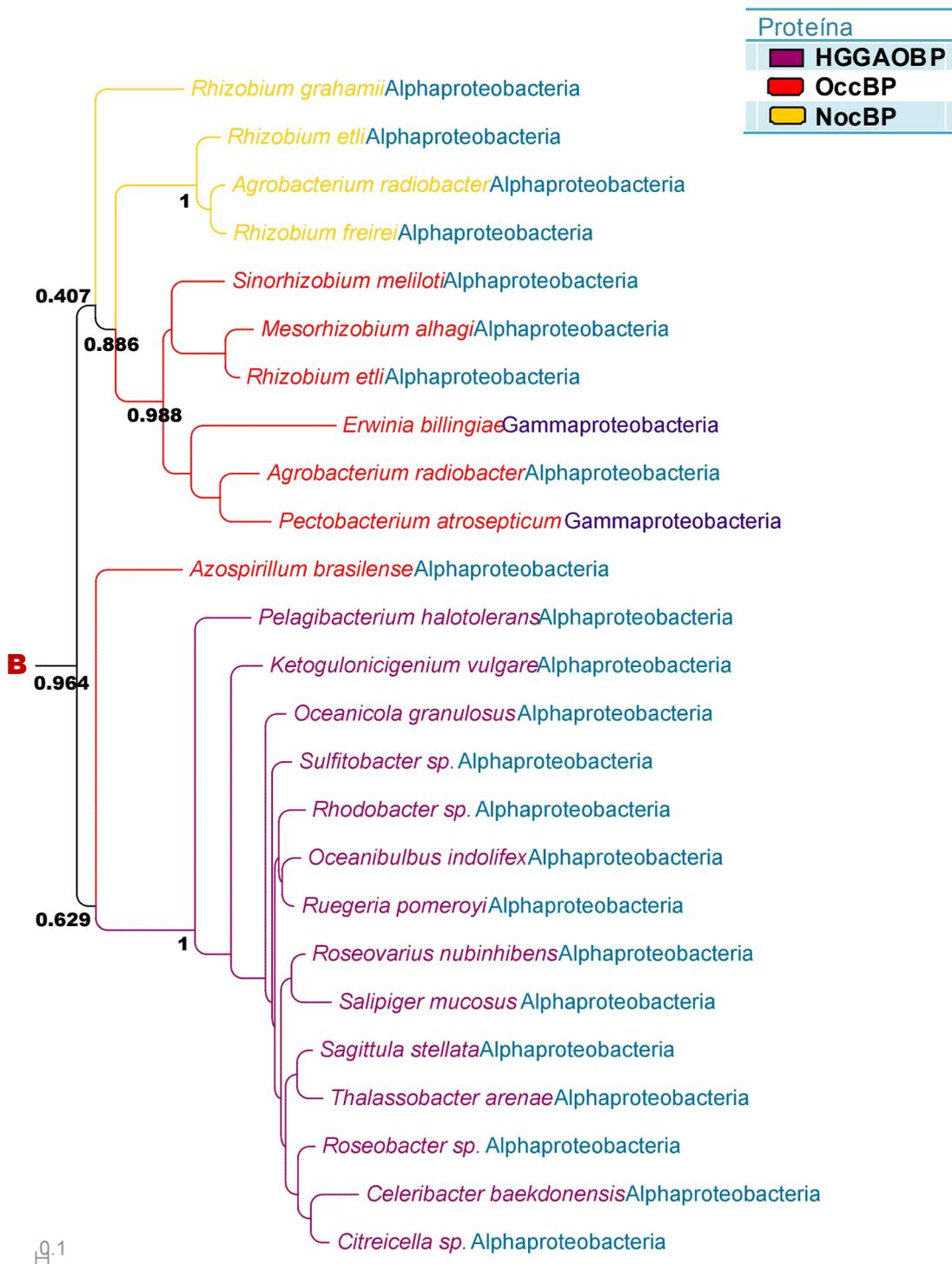


Figura 14. Clado perteneciente al árbol original donde se encuentran secuencias de NocBP, OccBO y HGGAOBP, mostrándose los nombres de algunas especies y clases taxonómicas, así como los principales valores de aLRT. Figura realizada en Archaeopteryx (Han & Zmasek, 2009).



Figura 15. Clado perteneciente al árbol original donde se encuentran secuencias de CBP, mostrándose los nombres de algunas especies y clases taxonómicas, así como los principales valores de aLRT. Figura realizada en Archaeopteryx (Han & Zmasek, 2009).

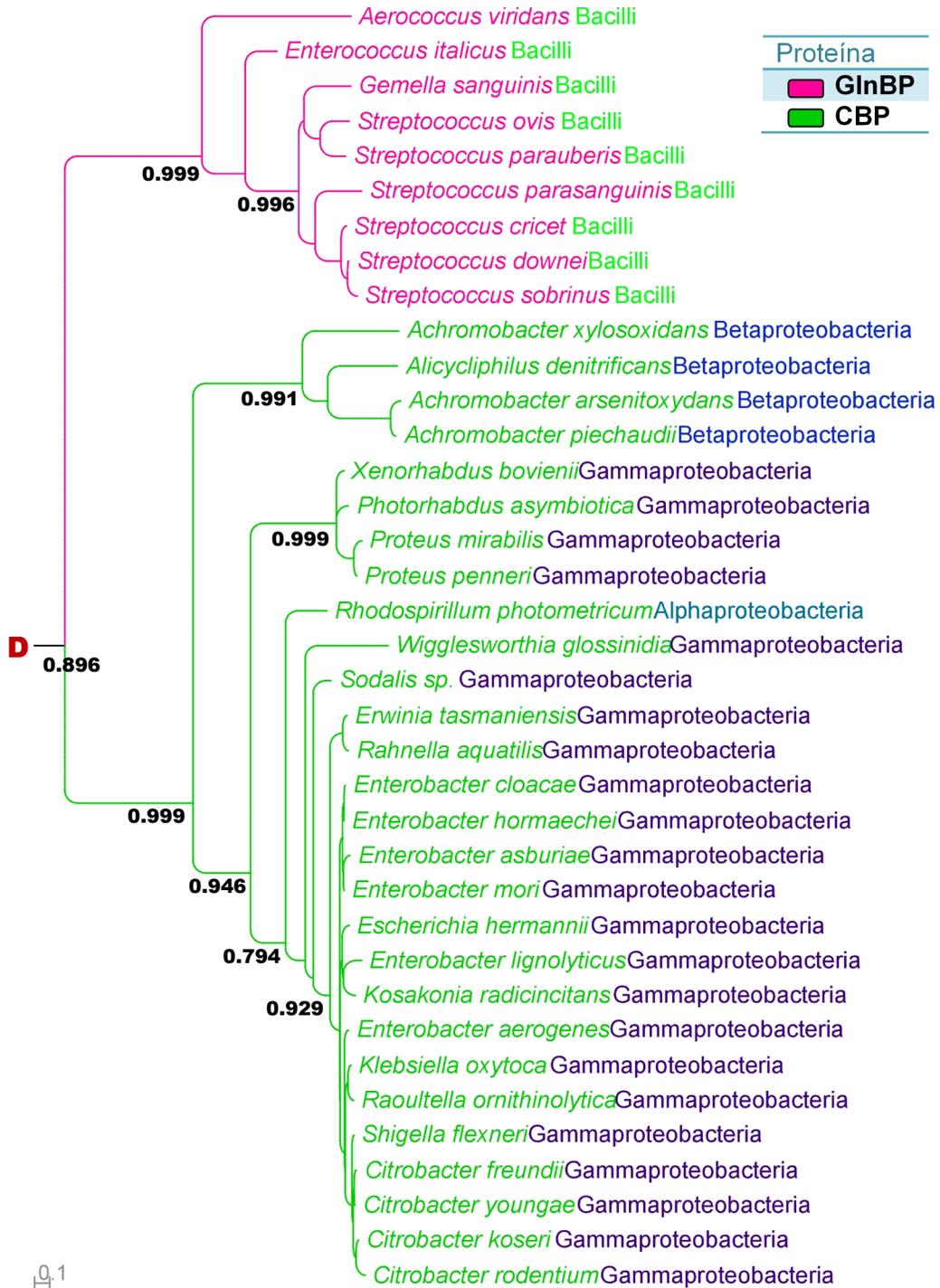


Figura 16. Clado perteneciente al árbol original donde se encuentran secuencias de GlnBP y CBP, mostrándose los nombres de algunas especies y clases taxonómicas, así como los principales valores de aLRT. Figura realizada en Archaeopteryx (Han & Zmasek, 2009).

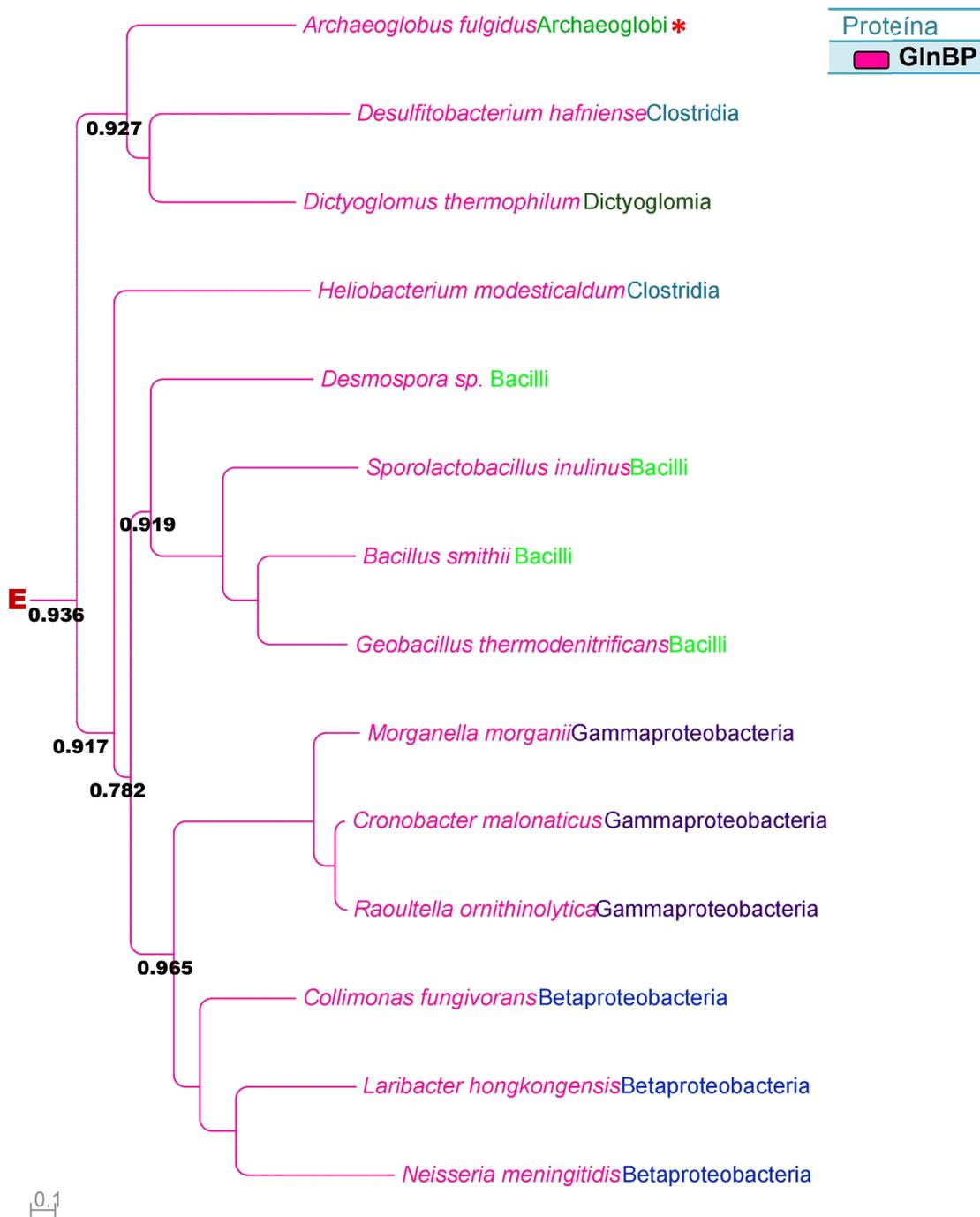


Figura 17. Clado perteneciente al árbol original donde se encuentran secuencias de GlnBP, mostrándose los nombres de algunas especies y clases taxonómicas, así como los principales valores de aLRT. Se marca con un asterisco rojo la secuencia perteneciente al dominio Archaea. Figura realizada en Archaeopteryx (Han & Zmasek, 2009).

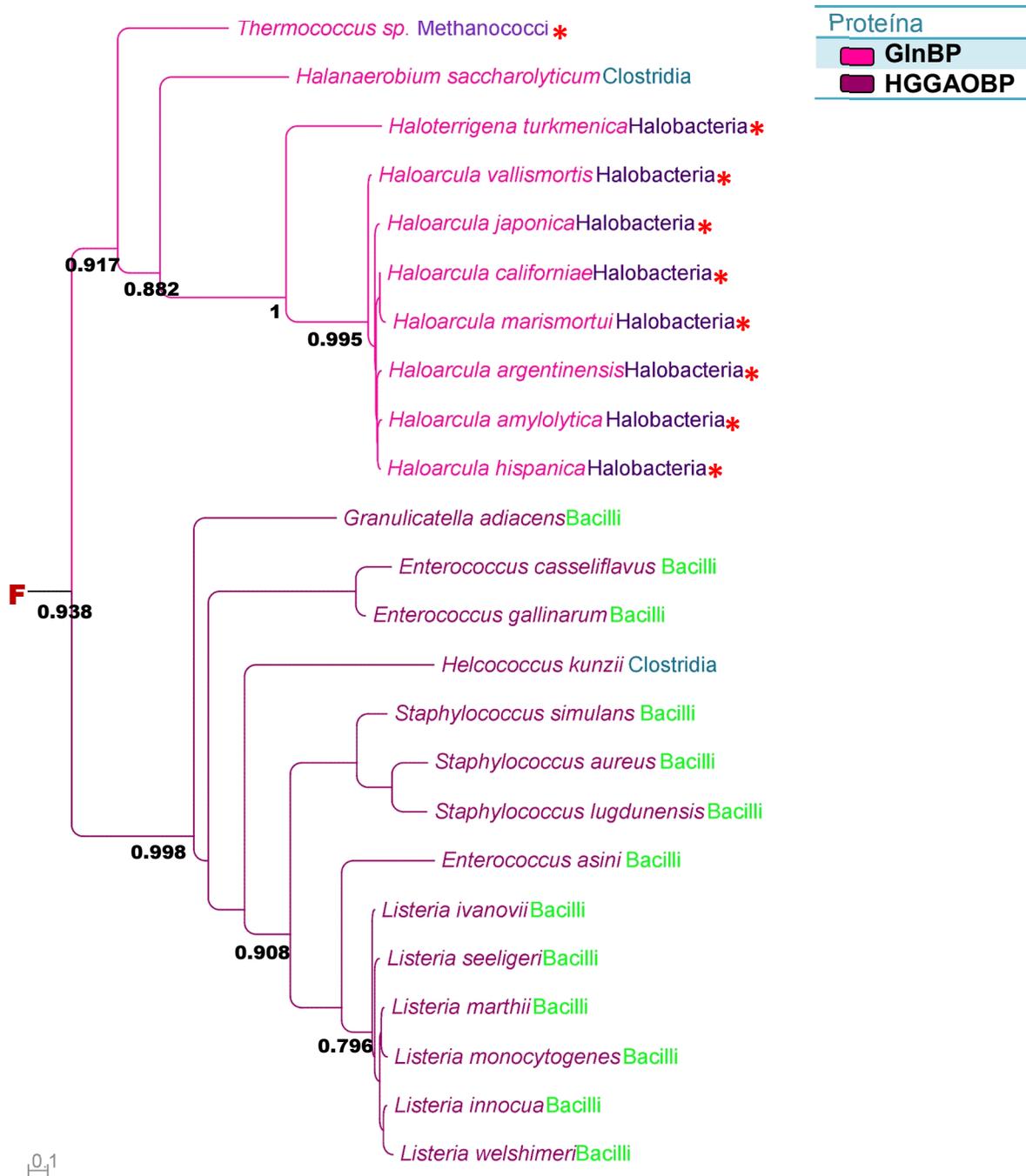


Figura 18. Clado perteneciente al árbol original donde se encuentran secuencias de GlnBP y HGGAOBP, mostrándose los nombres de algunas especies y clases taxonómicas, así como los principales valores de aLRT. Se marca con un asterisco rojo las secuencias pertenecientes al dominio Archaea. Figura realizada en Archaeopteryx (Han & Zmasek, 2009).

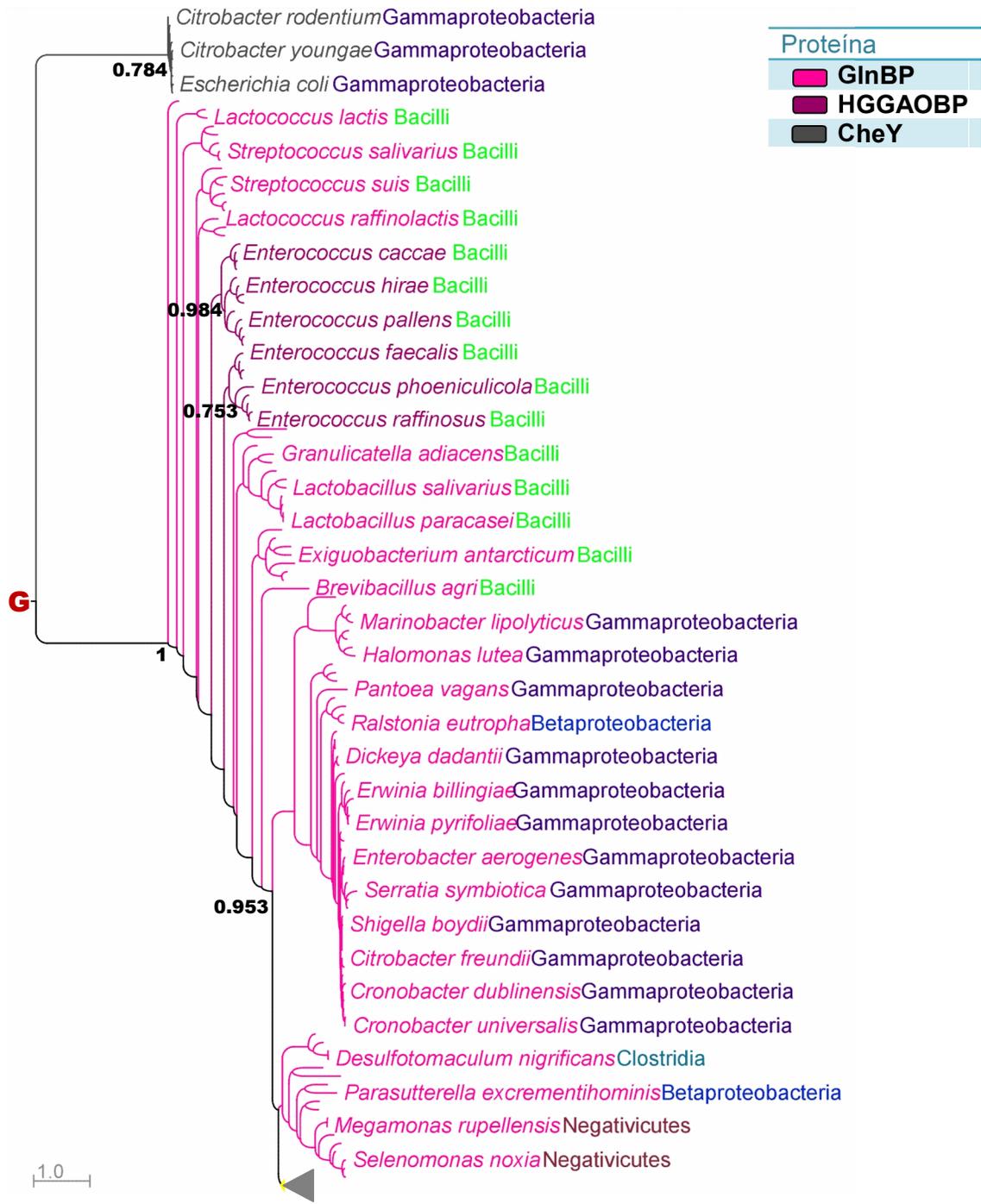


Figura 19. Árbol colapsado para mostrar las ramas más cercanas a la raíz. Se encuentran secuencias de GlnBP, HGGAOPB y CheY, mostrándose los nombres de algunas especies y clases taxonómicas, así como los principales valores de aLRT. Figura realizada en Archaeopteryx (Han & Zmasek, 2009).

5.4 Secuencias ancestrales

Dado que se utilizaron todas las secuencias del árbol para generar los ancestros hipotéticos, se seleccionaron los nodos de interés tomando en cuenta la posición de cada uno de ellos en el árbol de manera que pudiera indicar una posible actividad primitiva de las PBPs agrupadas en esa rama. Como resultado se seleccionaron ocho nodos (Figura 20): **N316** representante del ancestro hipotético de NocBP, OccBP y HGGAOBP; **N32** ancestro hipotético de GlnBP y HGGAOBP (ésta agrupación de secuencias de GlnBP's está conformado por las secuencias de aqueas); **N108** representante de GlnBP y CBP; **N365** ancestro hipotético de HisJ y LAO; **N315** es el ancestro hipotético de NocBP, OccBP, HGGAOBP, LAO e HisJ; **N181** representante de CBP, NocBP, OccBP, HGGAOBP, LAO e HisJ; **N107** ancestro hipotético de GlnBP, CBP, NocBP, OccBP, HGGAOBP, LAO e HisJ; y por último el **N15** representante del ancestro hipotético de todas las PBPs del árbol sin tomar en cuenta el grupo externo.

La longitud de las secuencias seleccionadas aumento en algunos casos. El N15 tuvo la longitud mayor con un número de residuos de 321, el nodo con menor longitud fue N365 con 239 residuos de aminoácidos; el promedio de la longitud de aminoácidos de las ocho secuencias ancestrales es de 281 residuos. Tomando en cuenta que la secuencia de LAO del organismo *Salmonella typhimurium* presenta 238 aminoácidos de longitud podemos decir que la longitud de los ancestros no difiere en extremo con la longitud de las PBPs extantes.

El alineamiento múltiple (Figura 22) de la secuencia de LAO de *Salmonella typhimurium* junto con los ocho nodos ancestrales muestra que las ancestros hipotéticos presentaron una mayor parte de los residuos conservados en las regiones cercanas a las láminas β y hélices α , además de que es claro que en la región de bisagra que corresponde a la secuencia de LAO la mitad de los ancestros presenta láminas β en esa región. Siete de las secuencias ancestrales no presentan las láminas β_4 y β_{10} . La secuencia con mayor número de residuos corresponde al nodo más primitivo y muestra consistencia con algunas secuencias que forman el gap más marcado enseguida de la primera cadena polipeptídica de la región de bisagra. También se observa que las hélices α son consistentes en todos la mayoría de los nodos, sólo en N181 presenta una hélice α extra al inicio de la secuencia, en N15 una hélice α al final de su secuencia, en N32 y N107 no se presenta la cuarta hélice α .

Con respecto a los residuos que participan directamente con la unión al ligando, en todos los ancestros se conservan cuatro residuos, que corresponden a Tyr14 siendo éste un residuo aromático, Asp30 ubicado en la hélice α A, Arg77 ubicado en la hélice α C y Asp161 ubicado en la lámina β 8 (las posiciones de los residuos corresponden a la secuencia de LAO de *Salmonella typhimurium* como modo de comparación). Aunque los demás residuos involucrados con la unión no se encuentran estrictamente conservados, cabe señalar, que en la mayoría de los casos solo una o dos secuencias difieren de residuo con respecto a las demás (Figura 22).

Los residuos de aminoácidos (que interactúan directamente en la unión con el ligando) no conservados en las secuencias son: Asp (ubicado en la posición 11 de LAO), no se encuentra presente en los nodos N316 y N32, éstos presentan Glu y Ser respectivamente. El segundo residuo aromático Phe52 se conserva en N365, N32 y N15, y es reemplazado por Trp en los nodos N108, N107, N181, N316 y N315. Para el residuo Ser69 difiere únicamente en el nodo N15 sustituido por Ala. El residuo Ser70 se encuentra conservado sólo en el nodo N365, en las siete secuencias restantes el residuo de aminoácido que lo reemplaza corresponde a Gly (Figura 22).

La Ser72 se mantiene conservada para el nodo N316, en las secuencias restantes este residuo es suplido por Thr. Por otro lado, el residuo Leu117, ubicado en la lámina β 6 de la secuencia de LAO, es reemplazado por Lys para el nodo N15 y por Gln para los demás nodos. Y por último el residuo Thr121 ubicado en la hélice α E, fue cambiado por Ala para el nodo N15. Los únicos dos Trp presentes en la secuencia de LAO se conservan sólo en la secuencia del ancestro hipotético N108 (Figura 22).

En los alineamientos estructurales (Figura 23) de los ancestros hipotéticos en comparación con la estructura de LAO de *Salmonella typhimurium* (Figura 21), se puede observar que el motivo estructural se conserva. Como generalidad, se distinguen ambos lóbulos, así como una región de bisagra, algunas de las estructuras presentan regiones de loops considerablemente sobresalientes, la posición del amino y carboxilo terminal se mantiene, en otras se logran distinguir dos láminas β en la región que enlaza ambos lóbulos ubicadas en cada uno de los segmentos polipéptidicos que los conecta, y en otros casos se pueden observar dos láminas β perpendiculares a las láminas β del lóbulo I. El lóbulo II presenta menos diferencias estructurales para todos los casos.

Los residuos de los ancestros hipotéticos (Figura 24) que coinciden con los aminoácidos que interactúan con el ligando en LAO, se encuentran ubicados en la misma región. Presentan gran similitud, se puede observar notoriamente que los ancestros conservan la estructura de “sándwich” y que a simple vista la estructura de esta región no parece variar excesivamente.

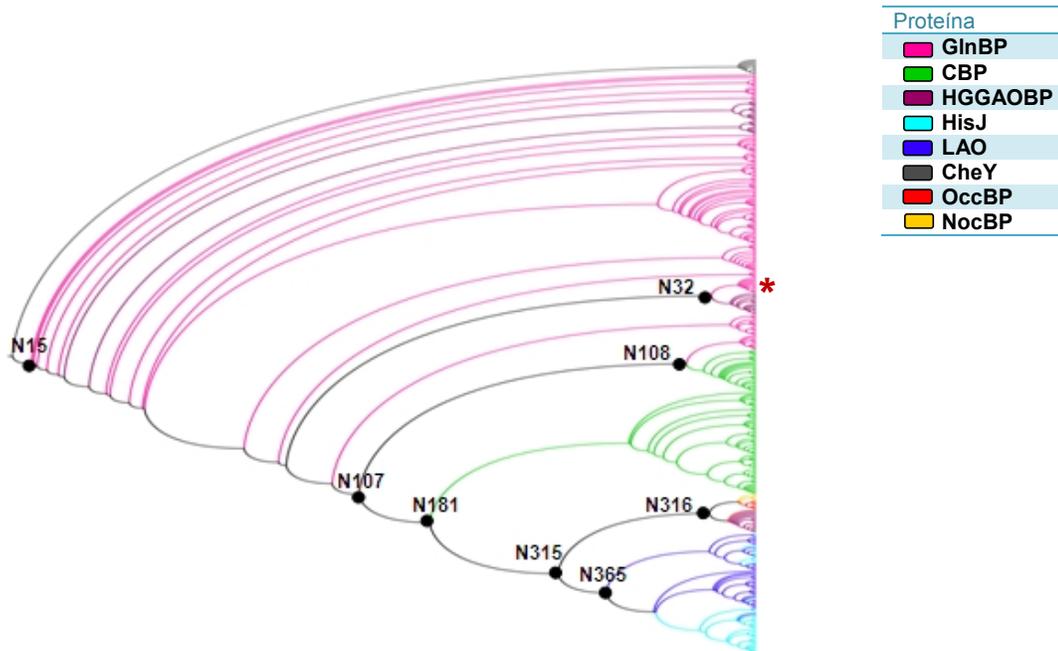


Figura 20. Árbol filogenético (Vista convexa) donde están marcados los ocho nodos ancestrales de interés (la rama donde están ubicadas las secuencias de arqueas está marcada con asterisco rojo). Figura realizada en Archaeopteryx (Han & Zmasek, 2009).

A partir de las secuencias obtenidas para los nodos, se generaron modelos estructurales utilizando el programa RaptorX. Las principales diferencias entre las estructuras hipotéticas contra la estructura de LAO son:

- N316 (Figura 23, a): no presenta las láminas $\beta 4$ y $\beta 10$, y muestra dos láminas β yuxtapuestas al lóbulo I.
- N32 (Figura 23, b): no presenta las láminas $\beta 4$ y $\beta 10$, muestra dos láminas β yuxtapuestas al lóbulo I, además contiene otras dos láminas β ubicadas en la región de bisagra y no se observa la presencia de la hélice αD .
- N108 (Figura 23, c): no presenta las láminas $\beta 4$ y $\beta 10$, muestra dos hélices β yuxtapuestas al lóbulo I y presenta otras dos láminas β en la región de bisagra en cada una de las cadenas polipeptídicas que conforman esta zona.

- N365 (Figura 23, d): no presenta las láminas $\beta 4$ y $\beta 10$, y muestra dos láminas β yuxtapuestas al lóbulo I.
- N315 (Figura 23, e): no presenta las láminas $\beta 3$, $\beta 4$ y $\beta 10$, además muestra dos láminas β yuxtapuestas al lóbulo I y un loop que destaca en la región del lóbulo II.
- N181 (Figura 23, f): presenta una hélice α en los residuos 2-5, no contiene las láminas $\beta 4$ y $\beta 10$, muestra otras dos láminas β en la región de bisagra, y tiene un loop sobresaliente en la región del lóbulo II
- N107 (Figura 23, g): no presenta las láminas $\beta 4$ y $\beta 10$, presenta un loop sobresaliente en la región del lóbulo II, y no presenta la hélice αD .
- N15 (Figura 23, h): no presenta la lámina $\beta 3$, muestra dos láminas β ubicadas en la región de bisagra, se observa una hélice α en los residuos 316 – 320, y presenta un loop sobresaliente en la región del lóbulo II.

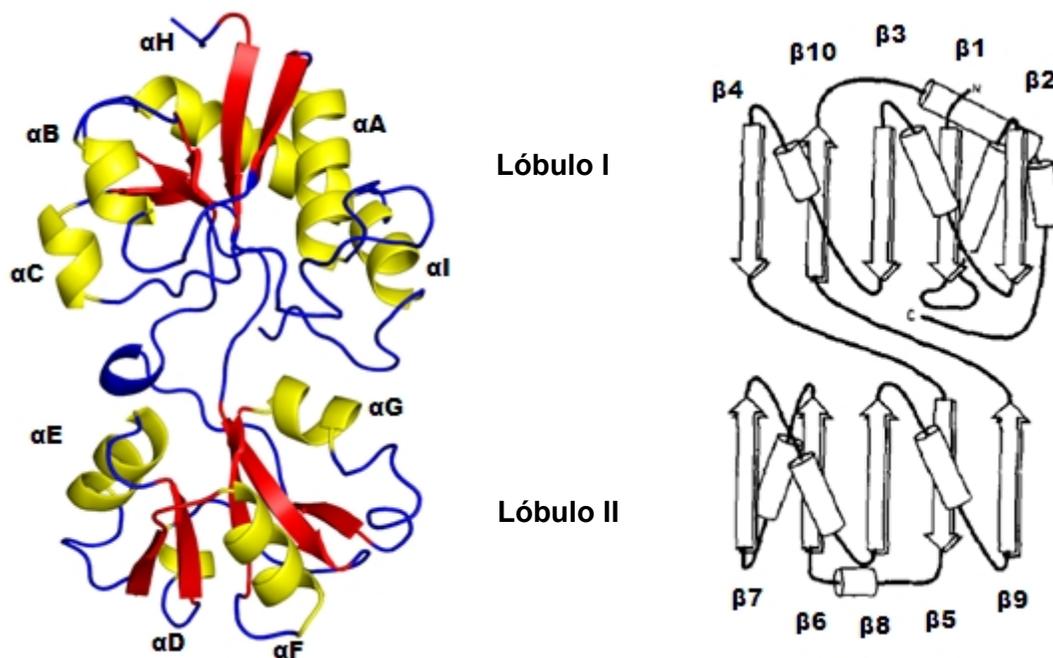


Figura 21. Estructura de LAO, indicando la posición de las 5 láminas β para cada lóbulo (en rojo) y 5 y 4 hélices α para el lóbulo I y II respectivamente (en amarillo), enunciadas de acuerdo a su localización en la secuencia de aminoácidos.

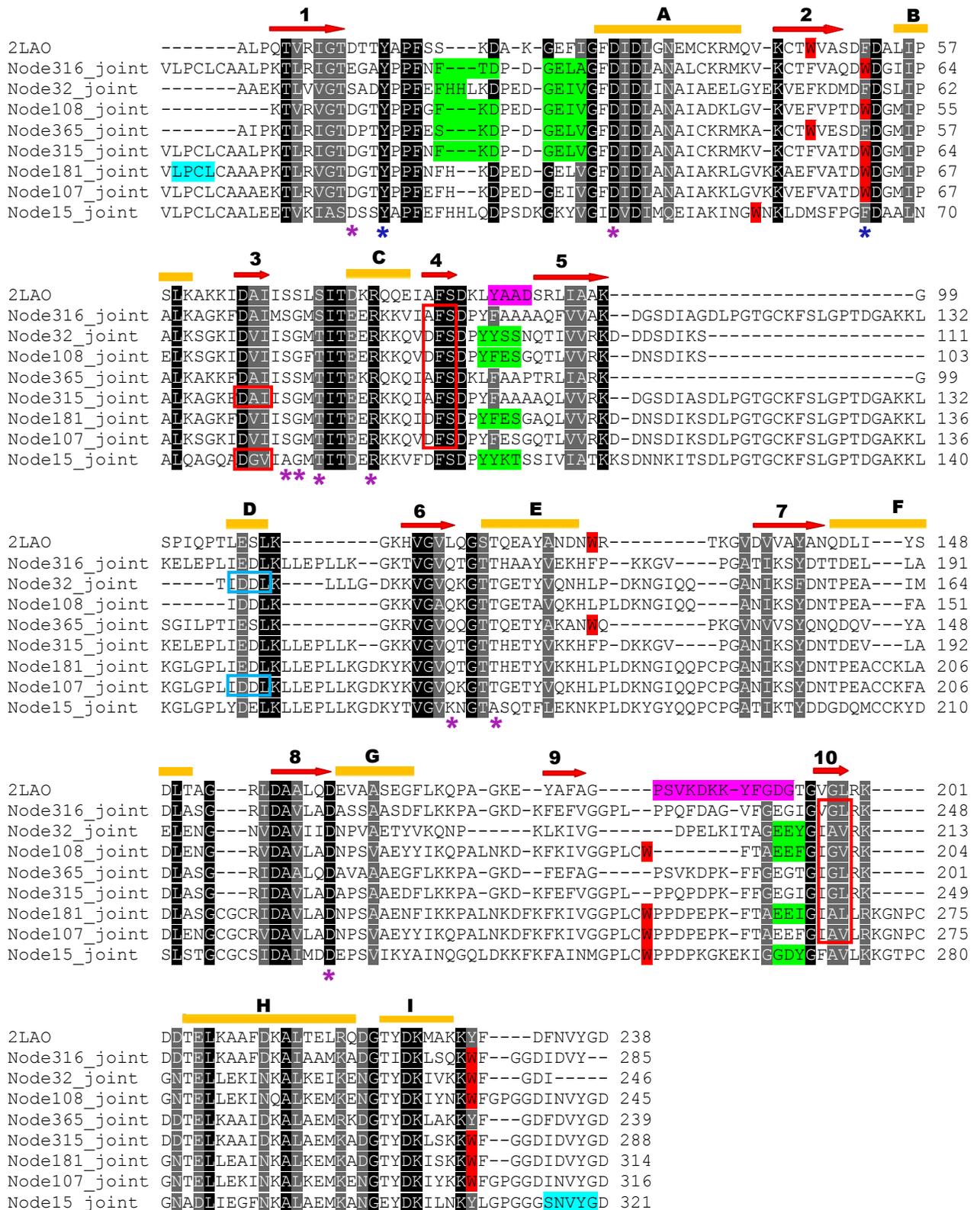


Figura 22. Alineamiento de la secuencia LAO de *Salmonella typhimurium* y de los nodos ancestrales. La estructura secundaria está representada por flechas en rojo (láminas β) y barras en amarillo (hélices α). Las regiones sombreadas en verde representan las láminas β extras presentes en los ancestros hipotéticos en comparación con la secuencia de LAO. Los residuos de aminoácidos sombreados en color cian representan las hélices α extras en los ancestros. Las regiones sombreadas en rosa resaltan la región de bisagra en LAO, y en naranja se resalta los Trp. Los asteriscos representan los residuos involucrados en la interacción con el ligando (asteriscos azules representan los residuos aromáticos). Las secciones contorneadas muestran las láminas β (rojo) y hélices α (azul) que no se presentan en los ancestros. Figura realizada con Color Align Conservation (Stothard, 2000).

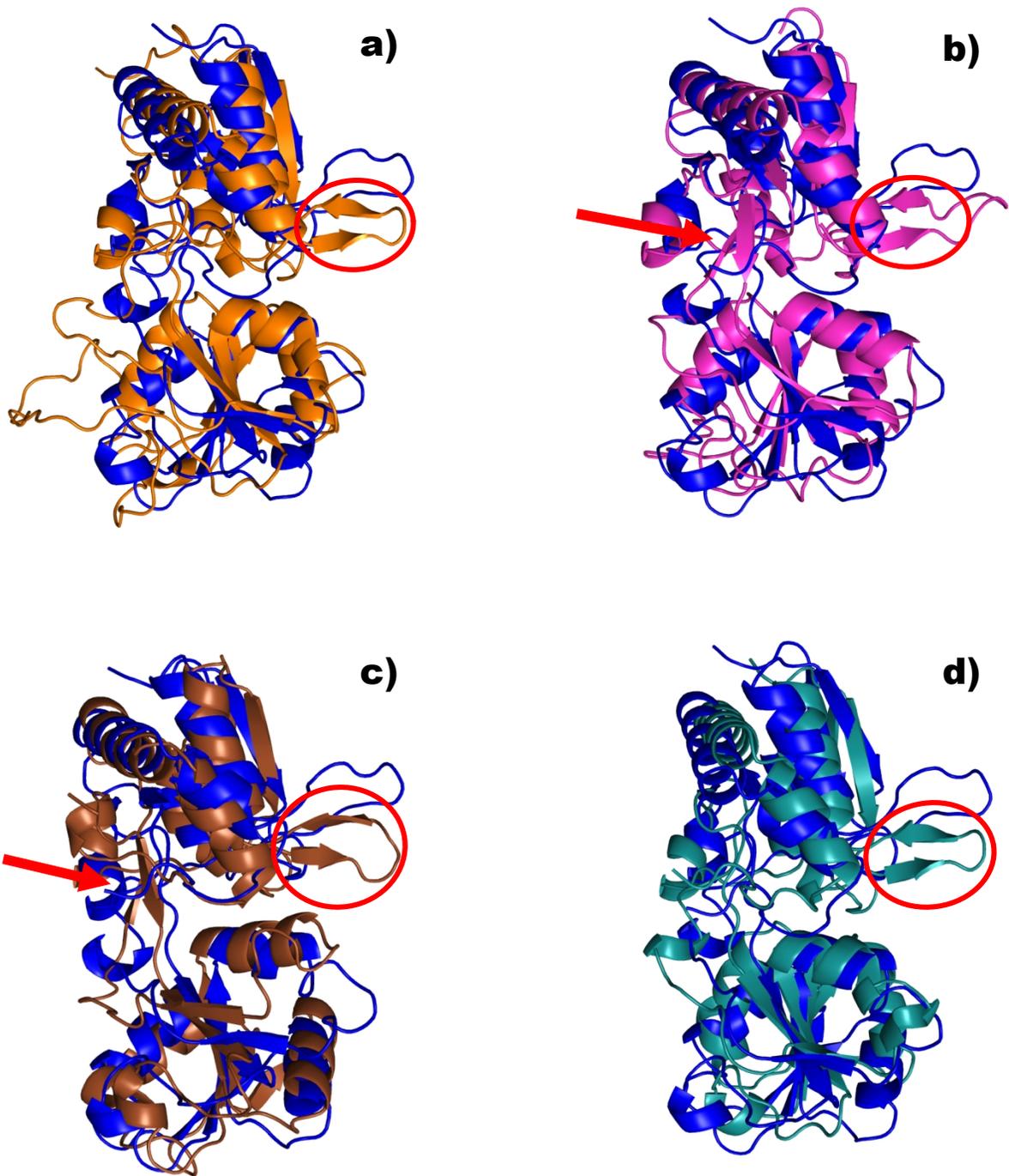


Figura 23. Alineamiento estructural de LAO con los modelos obtenidos para cada uno de los nodos ancestrales. a) **N316**: ancestro hipotético de NocBP, OccBP y HGGAOBP. b) **N32**: ancestro hipotético de HGGAOBP y GlnBP. c) **N108**: ancestro hipotético de CBP y GlnBP. d) **N365**: ancestro hipotético de HisJ y LAO. Se señala con círculos y flechas rojas las hélices α y láminas β extras que presentan los modelos estructurales de los ancestros hipotéticos. Figuras realizadas con PyMOL (Schrödinger, 2010).

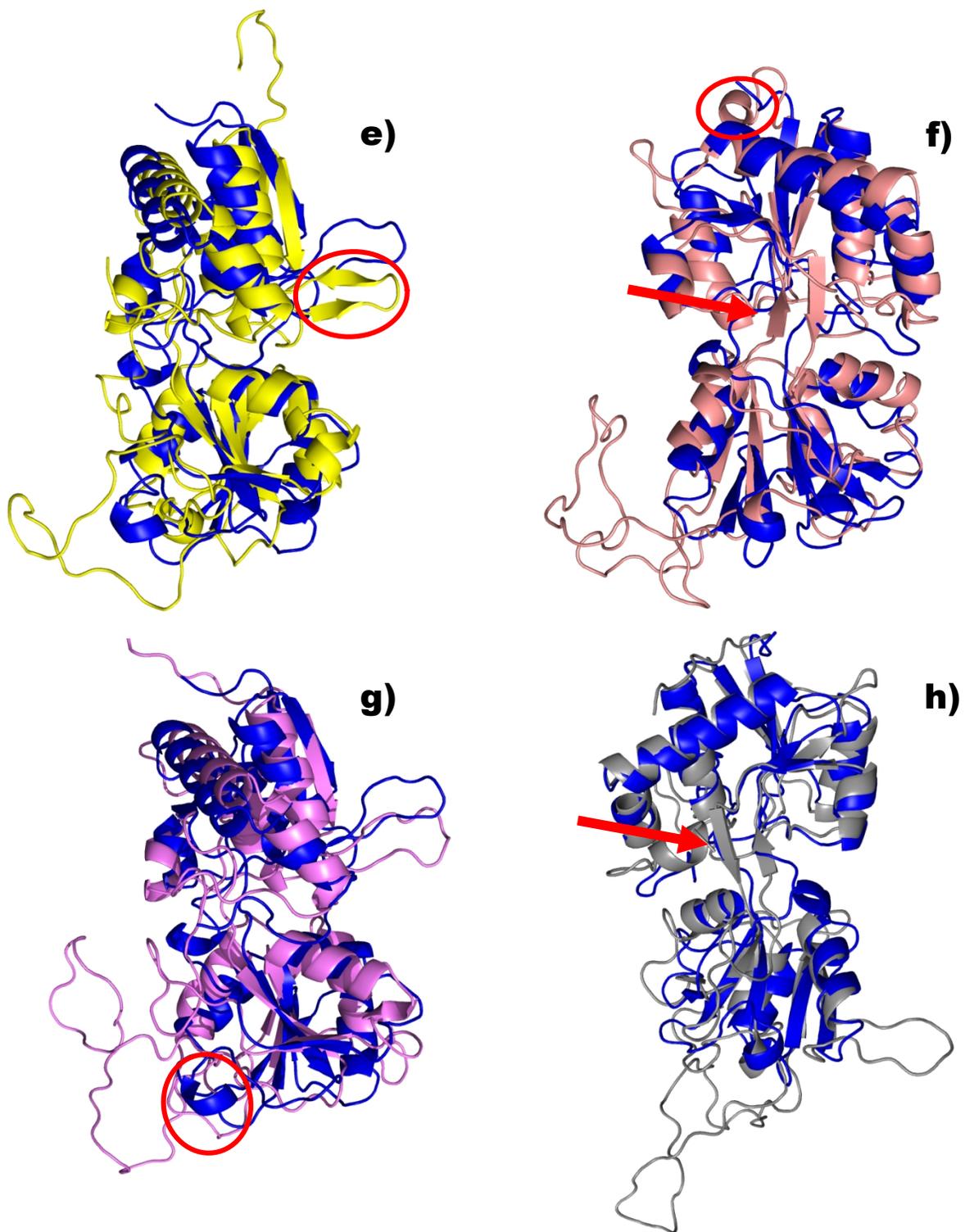


Figura 23. Alineamiento estructural de LAO con los modelos obtenidos para cada uno de los nodos ancestrales. e) N315: ancestro hipotético de NocBP, OccBP, HGGAOBP, LAO e HisJ. f) N181: ancestro hipotético de CBP, NocBP, OccBP, HGGAOBP, LAO e HisJ. g) N107: ancestro hipotético GlnBP, HGGAOBP, CBP, NocBP, OccBP, LAO e HisJ. h) N15: ancestro hipotético de todas las PBPs, sin tomar en cuenta a CheY. Se señala con círculos y flechas rojas las hélices α y láminas β extras que presentan los modelos estructurales de los ancestros hipotéticos. Figuras realizadas con PyMOL (Schrödinger, 2010).

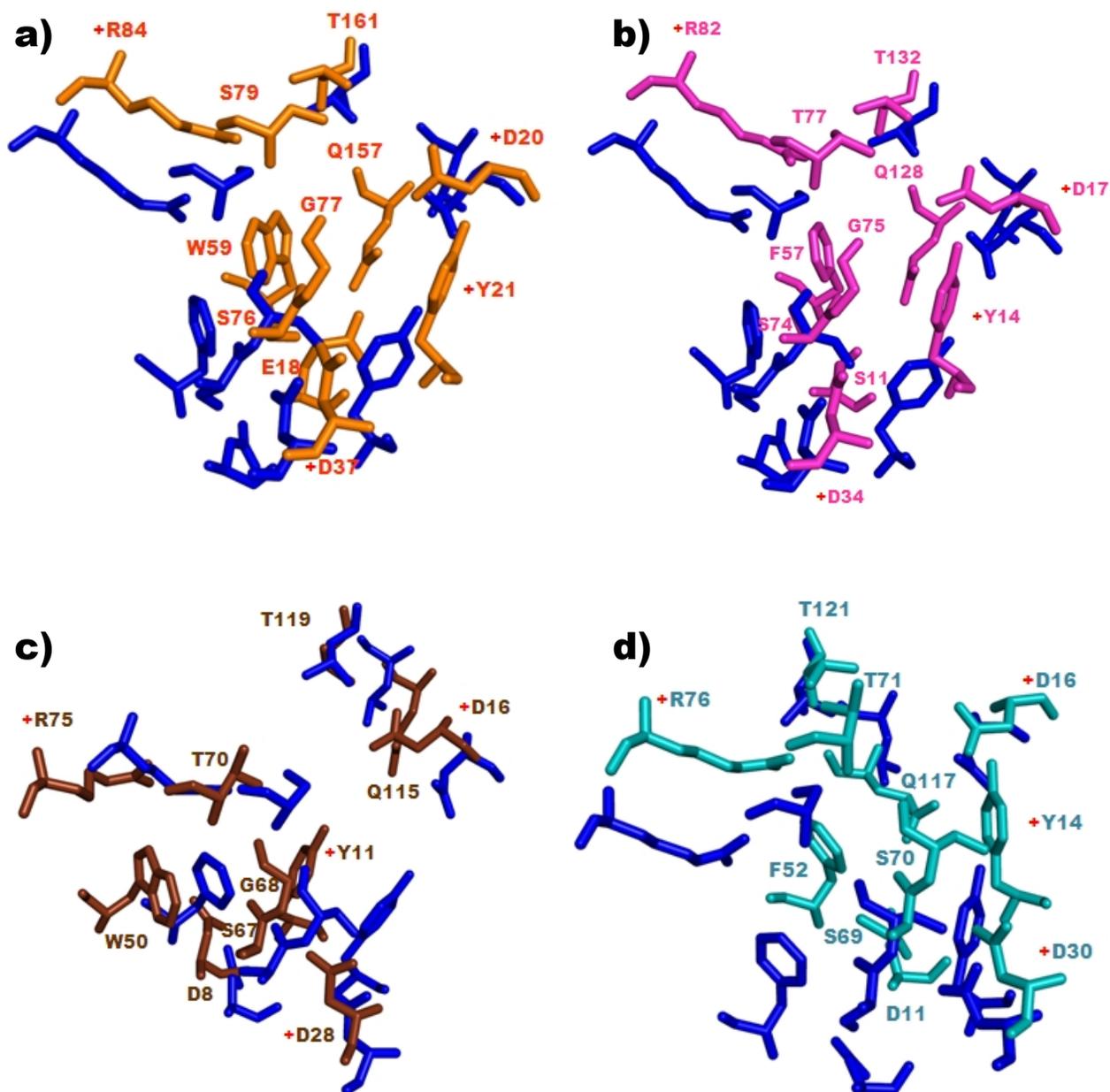


Figura 24. Alineamiento estructural de la región de unión a ligando de LAO con cada uno de los ancestros hipotéticos. Los residuos de los ancestros marcados con una cruz (+) representan aquellos que están conservados en todas las secuencias. a) N316: ancestro hipotético de NocBP, OccBP y HGGAOBP. b) N32: ancestro hipotético de HGGAOBP y GlnBP. c) N108: ancestro hipotético de CBP y GlnBP. d) N365: ancestro hipotético de HisJ y LAO. Figuras realizadas con PyMOL (Schrödinger, 2010).

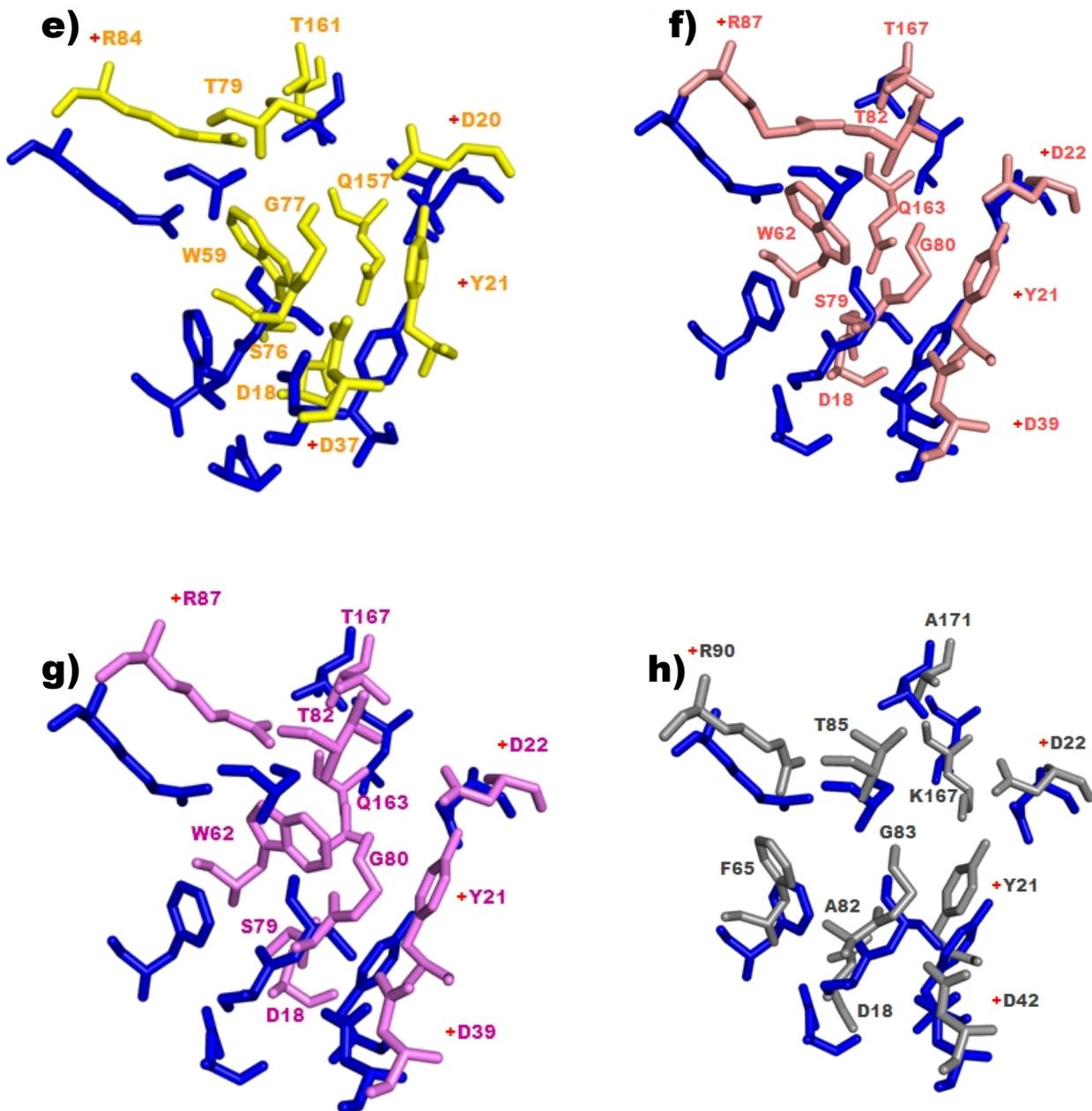


Figura 24. Alineamiento estructural de la región de unión a ligando de LAO con cada uno de los ancestros hipotéticos. Los residuos de los ancestros marcados con una cruz (+) representan aquellos que están conservados en todas las secuencias. e) N315: ancestro hipotético de NocBP, OccBP, HGGAOBP, LAO e HisJ. f) N181: ancestro hipotético de CBP, NocBP, OccBP, HGGAOBP, LAO e HisJ. g) N107: ancestro hipotético GlnBP, HGGAOBP, CBP, NocBP, OccBP, LAO e HisJ. h) N15: ancestro hipotético de todas las PBPs, sin tomar en cuenta a CheY. Figuras realizadas con PyMOL (Schrödinger, 2010).

6. DISCUSIÓN

Es evidente que el número de secuencias y la diversidad taxonómica presente para cada tipo de PBP se ve afectado por el sesgo de información en la base de datos, es decir la anotación es siempre una limitante, ya que la información bioquímica de la funcionalidad de las proteínas es escasa. Un claro ejemplo de esto es el número de secuencias para GlnBP, esta proteína es uno de los sistemas de transporte más caracterizado, ya que la glutamina (aminoácido esencial) es un sustrato de energía para la mayoría de las células. Las función general de las GlnBPs es recoger y detectar la glutamina en el medio ambiente y transportarla al citoplasma (Sun, Rose, Wang, & Hsiao, 1998). Dada su importancia, es por lo cual ha sido una de las PBPs de unión a aminoácidos más estudiada en diversos organismos, y que constituye en el presente trabajo el 37% de las secuencias totales.

En este trabajo las PBPs relacionadas con la proteína LAO son miembros específicamente implicados con la unión y transporte de aminoácidos con carga positiva en su grupo R. En el árbol filogenético propuesto podemos observar que en primera instancia las PBPs del árbol se agrupan en éste por el tipo de aminoácido al que unen, es decir se agrupan por la carga neta del aminoácido, por un lado tenemos a las proteínas que unen aminoácidos positivos como Lisina, Arginina, Ornitina e Histidina, las que unen aminoácidos sin carga neta como Glutamina y Cisteína, y por otro lado encontramos a las proteínas que unen aminoácidos poco comunes como Nopalina y Octopina. En el caso específico de la PBP que se encuentra anotada en la base de datos con unión a cinco aminoácidos, HGGAOBP no cuenta con referencia experimental. Por otro lado, en el caso de la GlnBP, sólo esta reportado como único sustrato de unión la glutamina, por lo tanto se pueden derivar dos interpretaciones a esto: las anotaciones de las HGGAOBPs no son correctas pues no se les ha medido la afinidad por su(s) ligando(s) y probablemente unan solo glutamina y no han podido ser nombradas correctamente, o bien, HGGAOBP si une todos esos aminoácidos y por eso aparece no solo junto a las GlnBPs.

En el caso de que las HGGAOBPs unieran esos cinco aminoácidos, podríamos explicar que sus tres apariciones en el árbol se deben a que una proteína ancestral muy parecida a ésta es predecesora de las demás (siendo un carácter plesiomórfico), con actividad promiscua, la cual se duplicó y por presiones de selección en algunos organismos se comenzó a especializar, no obstante, siempre quedaban proteínas promiscuas, puesto que resultaría evidente que unir cinco aminoácidos probablemente mejora la adecuación de la

población que presenta esta proteína, por ejemplo, un organismo que presenta un transportador específico para glutamina, siendo esta molécula su única fuente de energía tiene menor probabilidad de sobrevivir en ambientes desfavorables (escasez de glutamina) que un organismo que no solo dependa de un aminoácido como recurso primario de energía.

Dos de los tres grupos de las HGGAOBPs que se encuentran en ramas cercanas a las GlnBPs son en su mayoría de la clase taxonómica *Bacilli*, mientras que las que se encuentran como grupo hermano de las GlnBPs de arqueas pertenecen a bacterias patógenas (Figura 18). Este primer aspecto nos hace notar que las arqueas halófilas en su mayoría no comparten el mismo nicho ecológico con las HGGAOBPs de organismos patógenos, deduciendo que hubo una duplicación de genes. Por otro lado encontramos al grupo de NocBP y OccBP como grupo hermano del tercer grupo de HGGAOBP, donde se observa que ocurre algo similar, las HGGAOBPs presentes en esta rama pertenecen a organismos que fueron aislados de ambientes marinos, y por otro lado los organismos que representan a NocBP y OccBP están asociados a nichos terrestres. Estos dos puntos nos ayudan a descartar transferencia horizontal entre estos organismos puesto que difieren totalmente en sus nichos ecológicos, y podría de alguna manera apoyar la hipótesis de que algunos ellos conservan una PBB promiscua tal como HGGAOBP de donde derivaron las demás.

Los organismos que presentan a NocBP y OccBP, están relacionados directamente con plantas, tales géneros como *Rhizobium*, estos organismos viven en simbiosis con determinadas plantas en su raíz y su función es fijar nitrógeno, o bien, patógenos de plantas como el género *Erwinia* y *Pectobacterium* que también presentan transportadores de éstos aminoácidos poco comunes estrechamente relacionados a su nicho ecológico donde se encuentran, ya que las plantas les proveen fuentes de nitrógeno y energía. Los sistemas de transporte para los aminoácidos que tienen cadenas laterales polares o con carga: lisina, histidina, ornitina, arginina, glutamina, glutamato, cistina y ácido diaminopimélico, opinas como octopina (N^2 -(1,3-D-carboxietil)-L-arginina) y nopalina (N^2 -(1,3-D-dicarboxypropyl) - L-arginina) (Zanker, von Lintig, & Schroder, 1992) son transportados en agrobacterias por los transportadores de la familia PAO. Con base a esto podemos explicar que NocBP y OccBP comparten una rama principal con LAO e HisJ debido a que transportan ligandos derivados de la Arginina.

Parece probable que los sistemas de transporte para octopina y nopalina funcionan de la misma manera que los sistemas de transporte para arginina-lisina-ornitina, e histidina en *S. typhimurium*, aunque esta hipótesis debe ser comprobada en detalle (Zanker et al., 1992). Tanto la octopina como la nopalina son derivados de L-arginina, lo que podría ser una explicación probable para la similitud con los sistemas de transporte para los aminoácidos básicos en otras bacterias. LAO, HisJ, NocBP, OccBP y HGGAOBP (Figura 11) comparten un ancestro en común y son el grupo más reciente de PBPs del el árbol.

El caso específico grupo de LAO e HisJ, las secuencias no se encuentran completamente separadas (clado no resuelto) resulta evidente por su alto porcentaje de identidad y similitud, 70% y 83% respectivamente (Kang et al., 1991). Las PBPs con unión a lisina, arginina, ornitina e histidina, siendo estas las proteínas con funciones más “recientes” podrían estar en proceso de especializarse para la unión de un solo ligando. En general, se puede observar que las actividades más derivadas se encuentran más lejos de la raíz. En la figura 12 se observa que LAO, HisJ, NocBP y OccBP forman un grupo separado, siendo estas proteínas las más derivadas y con mayor número de cambios. Es probable que de la duplicación de una CBP, haya derivado éste grupo.

La divergencia de las PBPs además del enfoque actual del diseño racional de proteínas, requiere un análisis vertical que tome en cuenta las diferencias y similitudes de las secuencias actuales con las ancestrales. Dados los ancestros hipotéticos obtenidos encontramos que 4 de los aminoácidos que interactúan directamente con la unión a ligando en la proteína LAO se conservan en todas las secuencias, en algunos casos los aminoácidos no son substituidos por otros de la misma clase, pudiendo esa variación estar estrechamente relacionada con el ligando ancestral, el cual pudo ser también un aminoácido.

Para todos los ancestros se predice la estructura común de una proteína periplásmica de unión: compartiendo una topología de la estructura secundaria α/β , con dos lóbulos conectados por dos segmentos polipeptídicos. Hay evidencias para creer que existe un número limitado de plegamientos en la naturaleza, probablemente debido a las limitaciones fisicoquímicas del plegamiento proteico.

El transporte de L-glutamina es uno de los más activos de los sistemas de transporte de aminoácidos en bacterias, así como de los más específicos (Weiner & Heppel, 1971). En la PBP que une glutamina el grupo α -amino del ligando es estabilizado por la interacción con

los residuos Gly68, Thr70 y Asp157 (Hsiao, Sun, Rose, & Wang, 1996). Una interacción iónica puede también encontrarse entre la cadena lateral del Asp157 y NH_3^+ del ligando. El grupo α -carboxilo del ligando interactúa con los residuos Thr70, Arg75 y Gly119. La cadena principal del Thr70 y Gly119 forman un puente de hidrógeno mientras que la cadena lateral de Arg75 neutraliza las cargas opuestas. La cadena lateral del ligando (glutamina) es anclada y estabilizada por la interacción con los residuos Asp10, Ala67, Lys115 e His156. La porción alifática del ligando Glutamina está insertada en un hueco hidrofóbico formado entre Phe13 y Phe50 (Sun et al., 1998). Es importante destacar que las secuencias ancestrales obtenidas de los nodos donde se encontraban secuencias de GlnBP conservan en su mayoría los aminoácidos o la clase de aminoácidos que interactúan en una GlnBP actual con su ligando, infiriendo así que el mecanismo de unión a ligando es muy parecido para este ancestro hipotético (Figura 24).

La secuencia de la proteína GlnBP muestra que contiene dos triptófanos localizados en las posiciones 32 y 220, los cuales no están localizados en la región de unión de ligando, Weiner y Heppel en 1971 concluyeron que al menos uno de los residuos de Trp en GlnBP está involucrado en el transporte de L-Gln a través de la membrana citoplasmática (Shen et al., 1989). Todos los ancestros contienen triptófanos en secuencia los cuales podrían estar de igual forma relacionados con el transporte de su ligando (Figura 22).

La función principal de las cadenas polipépticas que unen ambos lóbulos de una PBP además de conectar ambos dominios, es conferir flexibilidad a la proteína. En GlnBP los segmentos que forman la región de bisagra (láminas β) no están directamente relacionados con la interacción de unión de Glutamina. Una red de enlaces de hidrógeno que conecta el dominio grande a la bisagra, y un segundo que conecta el pequeño dominio a la bisagra, interactúan con la glutamina. Así, esta red de enlaces de hidrógeno pueden jugar un importante papel en la cascada conformacional que acompaña a la unión del ligando en GlnBP (Sun et al., 1998); éste podría ser el caso de N181, N15 y N32. Estos dos últimos nodos pertenecen a secuencias relacionadas con GlnBP; en el caso de N181 esto podría deberse a que la secuencia de la PBP actual encargada de transportar cisteína se asemeja mucho a la proteína de unión a glutamina de *Bacillus subtilis* (Müller et al., 2005) y la rama que corresponde a este nodo contiene las secuencias de las CBPs pertenecientes a la clase *Bacilli* (Figura 15).

En LAO los dos anillos aromáticos de Tyr14 y Phe52 que forman una interacción catión-pi con el sustrato, y Leu117 muestran una interacción hidrofóbica con el grupo metileno de la

lisina (ligando), con esta interacción la lisina es completamente atrapada dentro de la proteína (Ames et al., 1993), lo cual refleja la importancia de la aparición de estos anillos aromáticos en todas las estructuras hipotéticas, localizados en similar región que en la proteína LAO (Figura 24).

LAO aprovecha las cargas de las cadenas laterales (Arg77, Asp161) y los grupos hidroxilos (Ser70, Ser72, Thr121) para estabilizar los grupos α -amino y α -carboxilo. Aunque estos grupos son comunes para todos los aminoácidos, parece que el modo de interacción entre ellos y la proteína puede variar con la proteína de unión. Una característica notable de la unión a ligando es que el lóbulo I provee más residuos iónicos o hidrofílicos (Asp11, Ser70, Ser72, Arg77, directamente al ligando) e interacciones hidrofóbicas (Tyr14, Phe52) con la lisina del lóbulo II (Thr121, Asp 161, Leu117, todos directamente con el ligando) (Ames et al., 1993). Los ancestros además de conservar en todos los casos los aminoácidos Tyr14, Arg77 y Asp 161 del alineamiento con LAO, los aminoácidos Asp11, Phe52 Ser70, Ser72, Thr121 de la secuencia de LAO corresponden también para la mayoría de las secuencias de los ancestros hipotéticos (Figura 22).

Las dos posiciones que presentaron más de un cambio en secuencia fueron en el Asp (11 en LAO) y Leu (117 en LAO), los demás solo presentan un cambio. Los otros residuos conservados en las secuencias de los ancestros, no tienen reportada ninguna función específica en la literatura, por lo que podrían estar implicados en la estabilidad de la proteína. Probablemente el sustrato que unía cada uno de los ancestros era muy parecido al de una PBP actual.

En las estructuras reportadas para HGGAOBP (3I6V) y GlnBP (1WDN) se pueden observar dos láminas β en la región que conecta ambos lóbulos, como los presentan los ancestros N108, N181, N32 y N15. A pesar de que los residuos de aminoácidos relacionados con la unión no están completamente conservados, la posición estructural de estos se mantiene (Figura 24). En GlnBP y la mayoría de las PBPs el lóbulo más grande proporciona la mayoría de las interacciones con el ligando (Schrödinger, 2010). En todos los casos los dos aminoácidos aromáticos se mantienen, uno de ellos conservado y otro reemplazado por otro de la misma clase de aminoácido.

En una filogenia realizada acerca de sistemas de transporte en los tres dominios de la vida se observó que sistemas homólogos de los tres reinos están presentes en la punta de las ramas del árbol. Esto sugiere que los sistemas ABC comenzaron a especializarse muy

temprano, probablemente antes de la separación de los tres reinos de los seres vivos. Otra explicación sería la ocurrencia masiva de la transferencia horizontal de genes entre los tres reinos. Se han descrito en algunos estudios proteínas de eucariontes homologas a proteínas periplásmicas, así como se han identificado sistemas de eucariontes que son análogos a sistemas de transporte bacteriano (O'Hara et al., 1993).

Las posiciones de las láminas β y hélices α no variaron significativamente de ancestro a ancestro, hay mayor número de residuos de aminoácidos conservados para todos los ancestros (Figura 22) en la primera mitad de la estructura secundaria, lo cual podría deberse a que en esta es donde se encuentran la mayoría de los residuos que interactúan con el ligando.

7. CONCLUSIÓN Y PERSPECTIVAS

La reconstrucción filogenética es una herramienta útil que permite realizar trabajos de ingeniería de proteínas mejor orientados, con bases evolutivas. Dada la importancia y la complejidad de ésta herramienta, es preciso tomar en cuenta consideraciones específicas para la proteína de la cual se desea trabajar. En principio se deben estudiar las generalidades de la proteína, para poder así elegir el método óptimo en conjunto con el programa con que se realizará la construcción.

Las tres clasificaciones propuestas para las proteínas periplásmicas de unión se consideran principalmente aspectos estructurales, no fue objetivo de este trabajo elaborar una clasificación natural, sin embargo como resultado de la filogenia de las proteínas homologas a LAO, se pudo observar claramente que existe una agrupación consistente con respecto al ligando que une cada una de estas.

De acuerdo a los resultados observados en el árbol filogenéticos se puede concluir que los genes de las PBP de unión a aminoácidos son parálogos, el principal mecanismo implicado en la generación de variabilidad de las proteínas periplásmicas de unión fue la duplicación génica que se dio en etapas tempranas de la diversificación de los dominios filogenéticos de la vida, no descartando la transferencia horizontal entre grupos, que probablemente no se vea reflejada en los resultados por la falta de información de secuencias en algunas PBP en las bases de datos.

Es probable que exista cierta promiscuidad entre los miembros de la PBP que unen a aminoácidos, y que HGGAOBP sea un claro ejemplo de plesiomorfía, la cual brinda esta característica.

A pesar de que no se identificaron los residuos de aminoácidos particulares que determinen la especificidad de cada proteína se puede concluir que existe una muy fuerte correlación entre la secuencia total y la función de la PBP. Puesto que los aminoácidos que interactúan con el ligando en su mayoría se conservan en las PBP, se puede descartar la idea de que únicamente los residuos que interactúan directamente con el ligando son los que determinan la especificidad. De esta manera se concluye la importancia del estudio de la secuencia completa de la proteína y la correlación entre cada uno de sus residuos.

Las perspectivas de este proyecto son:

- Realizar experimentos computacionales (Rosetta y docking molecular) para constatar que el plegamiento de los ancestros sea posible, y además conocer de manera hipotética su ligando.
- Estudiar la co-evolución de los residuos (Xdet – Mtreedet) que no están directamente implicados con la unión del ligando.
- Sintetizar los genes ancestrales a partir de los nodos propuestos y caracterizar bioquímicamente su unión.
- A partir de la identificación de residuos utilizando la reconstrucción ancestral, realizar las mutaciones puntuales en LAO con la finalidad de obtener una proteína de unión a un ligando diferente al ligando por proteína silvestre.

REFERENCIA

- Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., & Walter, P. (2003). *Biología Molecular de la Célula* (Cuarta edi.). Ediciones Omega.
- Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of molecular biology*, 215(3), 403–10. doi:10.1016/S0022-2836(05)80360-2
- Ames, G. F., Oh, B. H., & Kim, S. H. (1994). Structural basis for multiple ligand specificity of the periplasmic lysine-, arginine-, ornithine-binding protein. *The Journal of biological chemistry*, 269(42), 26323–30. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/7929349>
- Ames, G. F., Oh, B. H., Pandit, J., Kang, C. H., Nikaido, K., Gokcen, S., & Kim, S. H. (1993). Three-dimensional structures of the periplasmic lysine/arginine/ornithine-binding protein with and without a ligand. *The Journal of biological chemistry*, 268(15), 11348–55. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/8496186>
- Anisimova, M., & Gascuel, O. (2006). Approximate likelihood-ratio test for branches: A fast, accurate, and powerful alternative. *Systematic biology*, 55(4), 539–52. doi:10.1080/10635150600755453
- Audesirk, T., Audesirk, G., & Byers, B. E. (2004). *Biología. Ciencia y Naturaleza*. México: Pearson.
- Bennet, S. A., Cohen, M. A., & Gonnet, G. H. (1994). Amino acid substitution during functionally constrained divergent evolution of protein sequences. *Protein Engineering, Design and Selection*, 7(11), 1323–1332. doi:10.1093/protein/7.11.1323
- Berg, J. M., L., J. T., & Stryer, L. (2002). *Bioquímica* (Quinta edi.). W. H. Freeman and Company.
- Berntsson, R. P.-A., Smits, S. H. J., Schmitt, L., Slotboom, D.-J., & Poolman, B. (2010). A structural classification of substrate-binding proteins. *FEBS letters*, 584(12), 2606–17. doi:10.1016/j.febslet.2010.04.043
- Blair, C., & Murphy, R. W. (2010). Recent trends in molecular phylogenetic analysis: where to next? *The Journal of heredity*, 102(1), 130–8. doi:10.1093/jhered/esq092
- Cai, W., Pei, J., & Grishin, N. V. (2004). Reconstruction of ancestral protein sequences and its applications. *BMC evolutionary biology*, 4(1), 33. doi:10.1186/1471-2148-4-33
- Cardella, L., & Hernández, R. (2002). *Bioquímica Médica Tomo IV (Bioquímica Especializada)*.
- Dassa, E., & Bouige, P. (2001). The ABC of ABCs: a phylogenetic and functional classification of ABC systems in living organisms. *Research in Microbiology*, 152(3),

211–229. Retrieved from
<http://www.sciencedirect.com/science/article/pii/S0923250801011949>

Davidson, A. L., Shuman, H. A., & Nikaido, H. (1992). Mechanism of maltose transport in *Escherichia coli*: transmembrane signaling by periplasmic binding proteins. *Proceedings of the National Academy of Sciences of the United States of America*, 89(6), 2360–4. Retrieved from
<http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=48657&tool=pmcentrez&rendertype=abstract>

De Chadarevian, S., Rheinberger, H.-J., & Suárez-Díaz, E. (2009). Molecular evolution: concepts and the origin of disciplines. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, 40(1), 43–53. Retrieved from
<http://www.sciencedirect.com/science/article/pii/S1369848608000940>

Delport, W., Poon, A. F. Y., Frost, S. D. W., & Kosakovsky Pond, S. L. (2010). Datamonkey 2010: a suite of phylogenetic analysis tools for evolutionary biology. *Bioinformatics (Oxford, England)*, 26(19), 2455–7. doi:10.1093/bioinformatics/btq429

Dwyer, M. A., & Hellinga, H. W. (2004). Periplasmic binding proteins: a versatile superfamily for protein engineering. *Current opinion in structural biology*, 14(4), 495–504. doi:10.1016/j.sbi.2004.07.004

Easteal, S. (1998). Disciplining molecular evolution. *Trends in Ecology & Evolution*, 13(8). Retrieved from <http://www.sciencedirect.com/science/article/pii/S0169534798014049>

Eguiarte, L., Souza, V., & Aguirre, X. (2007). *Ecología molecular*. México.

Euzeby, J. P. (1998). List of Bacterial Names with Standing in Nomenclature. Retrieved from <http://www.bacterio.cict.fr/>

Evans, M. J., & Rosenthal, J. S. (2005). *Probabilidad y estadística* (p. 758). Reverte. Retrieved from <http://books.google.com/books?id=ZU3MEKZFgsMC&pgis=1>

Felsenstein, J. (1981). Evolutionary trees from DNA sequences: a maximum likelihood approach. *Journal of molecular evolution*, 17(6), 368–76. Retrieved from
<http://www.ncbi.nlm.nih.gov/pubmed/7288891>

Fukami-Kobayashi, K., Tateno, Y., & Nishikawa, K. (1999). Domain dislocation: a change of core structure in periplasmic binding proteins in their evolutionary history. *Journal of molecular biology*, 286(1), 279–90. doi:10.1006/jmbi.1998.2454

Futuyma, D. J. (2005). *Evolution*. Sinauer Associates, INC.

Gilson, E., Alloing, G., Schmidt, T., Claverys, J. P., Dudler, R., & Hofnung, M. (1988). Evidence for high affinity binding-protein dependent transport systems in gram-positive bacteria and in *Mycoplasma*. *The EMBO journal*, 7(12), 3971–4. Retrieved from
<http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=454997&tool=pmcentrez&rendertype=abstract>

- Guindon, S., & Gascuel, O. (2003). A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Systematic biology*, 52(5), 696–704. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/14530136>
- Guindon, S., Lethiec, F., Duroux, P., & Gascuel, O. (2005). PHYML Online--a web server for fast maximum likelihood-based phylogenetic inference. *Nucleic acids research*, 33(Web Server issue), W557–9. doi:10.1093/nar/gki352
- Han, M. V., & Zmasek, C. M. (2009). phyloXML: XML for evolutionary biology and comparative genomics. *BMC bioinformatics*, 10, 356. doi:10.1186/1471-2105-10-356
- Holland, I. B., Kuchler, K., Higgins, C., & Cole, S. (2003). *ABC Proteins: From Bacteria To Man*. Academic Press Inc.
- Holliger, P., Jaeger, K. E., Fenical, W., Hill, R., & Lutz, S. (2010). Beyond directed evolution—semi-rational protein engineering and design. *Current Opinion in Biotechnology*, 21(6), 734–743. Retrieved from <http://www.sciencedirect.com/science/article/pii/S0958166910001540>
- Hordijk, W., & Gascuel, O. (2005). Improving the efficiency of SPR moves in phylogenetic tree search methods based on maximum likelihood. *Bioinformatics (Oxford, England)*, 21(24), 4338–47. doi:10.1093/bioinformatics/bti713
- Hsiao, C. D., Sun, Y. J., Rose, J., & Wang, B. C. (1996). The crystal structure of glutamine-binding protein from Escherichia coli. *Journal of molecular biology*, 262(2), 225–42. doi:10.1006/jmbi.1996.0509
- Huelsenbeck, J. P., & Ronquist, F. (2001). MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics (Oxford, England)*, 17(8), 754–5. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/11524383>
- Källberg, M., Wang, H., Wang, S., Peng, J., Wang, Z., Lu, H., & Xu, J. (2012). Template-based protein structure modeling using the RaptorX web server. *Nature protocols*, 7(8), 1511–22. doi:10.1038/nprot.2012.085
- Kang, C. H., Shin, W. C., Yamagata, Y., Gokcen, S., Ames, G. F., & Kim, S. H. (1991). Crystal structure of the lysine-, arginine-, ornithine-binding protein (LAO) from Salmonella typhimurium at 2.7-Å resolution. *The Journal of biological chemistry*, 266(35), 23893–9. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/1748660>
- Katoh, K., Misawa, K., Kuma, K., & Miyata, T. (2002). MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic acids research*, 30(14), 3059–66. Retrieved from <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=135756&tool=pmcentrez&rendertype=abstract>
- Kinch, L. N., & Grishin, N. V. (2002). Evolution of protein structures and functions. *Current Opinion in Structural Biology*, 12(3), 400–408. doi:10.1016/S0959-440X(02)00338-X

- Koshi, J. M., & Goldstein, R. A. (1996). Probabilistic reconstruction of ancestral protein sequences. *Journal of Molecular Evolution*, 42(2), 313–320. doi:10.1007/BF02198858
- Le, S. Q., & Gascuel, O. (2008). An improved general amino acid replacement matrix. *Molecular biology and evolution*, 25(7), 1307–20. doi:10.1093/molbev/msn067
- Le, S. Q., & Gascuel, O. (2010). Accounting for solvent accessibility and secondary structure in protein phylogenetics is clearly beneficial. *Systematic biology*, 59(3), 277–87. doi:10.1093/sysbio/syq002
- Lehninger, A. L., Nelson, D. L., & M.Cox, M. (2005). *Principios de Bioquímica* (Cuarta edi.).
- Letunic, I., & Bork, P. (2011). Interactive Tree Of Life v2: online annotation and display of phylogenetic trees made easy. *Nucleic acids research*, 39(Web Server issue), W475–8. doi:10.1093/nar/gkr201
- Madigan, M. T., Martinko, J. M., Dunlap, P. V., & Clark, D. P. (2009). *Biología de los microorganismos* (12^o edició.). Pearson.
- Müller, A., Thomas, G. H., Horler, R., Brannigan, J. A., Blagova, E., Levnikov, V. M., ... Wilkinson, A. J. (2005). An ATP-binding cassette-type cysteine transporter in *Campylobacter jejuni* inferred from the structure of an extracytoplasmic solute receptor protein. *Molecular microbiology*, 57(1), 143–55. doi:10.1111/j.1365-2958.2005.04691.x
- Nei, M. (1996). Phylogenetic analysis in molecular evolutionary genetics. *Annual review of genetics*, 30, 371–403. doi:10.1146/annurev.genet.30.1.371
- Nikaido, K., & Ames, G. F. (1992). Purification and characterization of the periplasmic lysine-, arginine-, ornithine-binding protein (LAO) from *Salmonella typhimurium*. *The Journal of biological chemistry*, 267(29), 20706–12. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/1400387>
- O'Hara, P. J., Sheppard, P. O., Thøgersen, H., Venezia, D., Haldeman, B. A., McGrane, V., ... Mulvihill, E. R. (1993). The ligand-binding domain in metabotropic glutamate receptors is related to bacterial periplasmic binding proteins. *Neuron*, 11(1), 41–52. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/8338667>
- Olivares, A. N. (2007). *Sistemática molecular del género Fissurella en el Pacífico Sudoriental*. UNIVERSIDAD DE SANTIAGO DE COMPOSTELA FACULTAD DE BIOLOGÍA Departamento de Bioquímica y Biología Molecular.
- Pupko, T., Pe'er, I., Shamir, R., & Graur, D. (2000). A fast algorithm for joint reconstruction of ancestral amino acid sequences. *Molecular biology and evolution*, 17(6), 890–6. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/10833195>
- Ridley, M. (2004). *Evolution* (Tervera ed.). Blackwell.
- Saitou, N., & Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.*, 4(4), 406–425. Retrieved from <http://mbe.oxfordjournals.org/content/4/4/406.short>

- Schrödinger, L. (2010). The PyMOL Molecular Graphics System, Version 1.5.
- Shen, Q., Simplaceanu, V., Cottam, P. F., Wu, J.-L., Hong, J.-S., & Ho, C. (1989). Molecular genetic, biochemical and nuclear magnetic resonance studies on the role of the tryptophan residues of glutamine-binding protein from *Escherichia coli*. *Journal of Molecular Biology*, *210*(4), 859–867. Retrieved from <http://www.sciencedirect.com/science/article/pii/0022283689901137>
- Shi, R., Proteau, A., Wagner, J., Cui, Q., Purisima, E. O., Matte, A., & Cygler, M. (2009). Trapping open and closed forms of FitE: a group III periplasmic binding protein. *Proteins*, *75*(3), 598–609. doi:10.1002/prot.22272
- Stothard, P. (2000). The sequence manipulation suite: JavaScript programs for analyzing and formatting protein and DNA sequences. *BioTechniques*, *28*(6), 1102, 1104. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/10868275>
- Sun, Y. J., Rose, J., Wang, B. C., & Hsiao, C. D. (1998). The structure of glutamine-binding protein complexed with glutamine at 1.94 Å resolution: comparisons with other amino acid binding proteins. *Journal of molecular biology*, *278*(1), 219–29. doi:10.1006/jmbi.1998.1675
- Tam, R., & Saier, M. H. (1993). Structural, functional, and evolutionary relationships among extracellular solute-binding receptors of bacteria. *Microbiol. Mol. Biol. Rev.*, *57*(2), 320–346. Retrieved from <http://mmbr.asm.org/content/57/2/320.short>
- Walsh, D. A., & Sharma, A. K. (2009). Molecular phylogenetics: testing evolutionary hypotheses. *Methods in molecular biology (Clifton, N.J.)*, *502*, 131–68. doi:10.1007/978-1-60327-565-1_9
- Weiner, J. H., & Heppel, L. A. (1971). A Binding Protein for Glutamine and Its Relation to Active Transport in *Escherichia coli*. *J. Biol. Chem.*, *246*(22), 6933–6941. Retrieved from <http://www.jbc.org/content/246/22/6933>
- Yang, Z., & Rannala, B. (2012). Molecular phylogenetics: principles and practice. *Nature reviews. Genetics*, *13*(5), 303–14. doi:10.1038/nrg3186
- Zanker, H., von Lintig, J., & Schroder, J. (1992). Opine transport genes in the octopine (occ) and nopaline (noc) catabolic regions in Ti plasmids of *Agrobacterium tumefaciens*. *J. Bacteriol.*, *174*(3), 841–849. Retrieved from <http://jb.asm.org/content/174/3/841.abstract?sid=22251a15-80bd-441b-9c3c-29c8b6ed486e>

APÉNDICE I: Estructuras hipotéticas de los nodos ancestrales.

RaptorX arroja como resultado los mejores PDBs que sirvieron como templatado para realizar el alineamiento con la secuencia del nodo ancestral, así como los valores de p-value y uGDT para calificar la calidad de la estructura resultante.

El valor de uGDT representa un puntaje no normalizado de GDT (Global Distance Test) que se define como $1*N(1)+0.75*N(2)+0.5*(4)+0.25*N(8)$, donde $N(x)$ representa el número de residuos con el RMSD local más pequeño que x . El valor GDT es calculado dividiendo uGDT entre la longitud del dominio, multiplicado por 100. Graficando los valores de $-\log(p\text{-value})$ contra los valores de uGDT, un modelo estructural se considera razonablemente bueno si el puntaje GDT es mayor a 50 (Källberg et al., 2012)

SECUENCIA	RESULTADOS DE RAPTORX
N316 1 - 285	p-value: 3.23e-11 uGDT (GDT): 204(72) uSeqId (SeqId): 116(41) Score: 243 Templates: 1l1sA 3vv5A 2y7iA 2ylnA
N32 1 - 246	p-value: 5.93e-10 uGDT (GDT): 191 (78) uSeqId (SeqId): 88 (36) Score: 239 Templates: 4h5fA 4i62A 4g4pA 1l1sA 4f3pA
N108 1 - 245	p-value: 1.88e-10 uGDT (GDT): 196 (80) uSeqId (SeqId): 85 (35) Score: 245 Templates: 3hv1:A
N365 1 - 239	p-value: 2.18e-10 uGDT (GDT): 229 (96) uSeqId (SeqId): 172 (72) Score: 239 Templates: 1l1s:A

N315 1 - 288	p-value: 2.36e-11 uGDT (GDT): 205 (71) uSeqId (SeqId): 128 (44) Score: 244 Templates: 1l1sA 2y7iA 3vv5A 4g4pA
N181 1 - 314	p-value: 3.27e-11 uGDT (GDT): 178 (57) uSeqId (SeqId): 96 (31) Score: 196 Templates: 1l1sA 3vv5A 2y7iA 4f3pA
N107 1 - 316	p-value: 9.55e-11 uGDT (GDT): 173(55) uSeqId (SeqId): 87 (28) Score: 191 Templates: 1l1sA 2pvuA 4g4pA 4f3pA
N15 1 - 321	p-value: 5.93e-11 uGDT (GDT): 181(56) uSeqId (SeqId): 101(32) Score: 211 Templates: 4g4pA 4kqpA 4la9A 2pvuA