



INSTITUTO DE INVESTIGACIONES BIOMÉDICAS

**Análisis Estructural y Evolutivo de la glicil-tRNA
sintetasa bacteriana**

T E S I S

PARA OBTENER EL TÍTULO EN

**LICENCIADO EN INVESTIGACIÓN BIOMÉDICA BÁSICA
PRESENTA**

Adriana Hernández González

Tutor: Dr. Marcelino Arciniega Castro

CIUDAD UNIVERSITARIA, CD. MX., 2022



Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

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

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

Agradecimientos

Al Dr. Alfredo Torres Larios por permitirme ser su alumna y trabajar con él durante mi desarrollo académico.

Al Dr. Marcelino Arciniega Castro por su tutoría en la escritura de esta tesis, así como a los miembros del Jurado por su revisión y sugerencias.

A mi familia y amigos que formaron parte en estos años de mi vida.

Índice

Resumen	1
Introducción	2
Estructura de la subunidad beta	8
Distribución	11
Establecimiento de homología	12
Búsqueda de homólogos lejanos por secuencia	12
Otras aproximaciones	13
Búsqueda de homólogos por estructura	15
Hipótesis	17
Objetivo general	18
Objetivos particulares	18
Métodos	19
Resultados	21
Distribución de las dos versiones de glyRS	21
Análisis de la transferencia horizontal	22
Búsqueda de homólogos por secuencia	24
Predicción de estructura con HHPRED	26
Búsqueda de homólogos por estructura	27
Discusión	29
Ancestralidad y distribución de la glicil tRNA Sintetasa Bacteriana	29

Dos versiones de la misma enzima	31
Similitudes estructurales de la bac-glyRS con otras proteínas	31
Barajeo de dominios y permutación de secuencias	36
Conclusiones	38
Perspectivas	39
Referencias	40
Información Suplementaria	44

Resumen

Existen dos glicil tRNA sintetetasas (glyRSs). Ambas pertenecen a la familia monofilética conocida como Clase II. Una glyRS es dimérica (α_2) y está presente en los tres dominios de la vida. La otra es tetramérica ($(\alpha_2\beta_2)$) y se encuentra presente sólo en bacteria. Aunque estas dos descienden del mismo ancestro común estudios de la subunidad α de glyRS establecieron que la versión bacteriana está más cercanamente relacionada a la alanil tRNA sintetasa. Esto parece indicar que la función de esterificación de la glicina surgió dos veces dentro de la misma familia. En este trabajo de tesis se exponen los estudios realizados a nivel de secuencia y estructura de la subunidad β de la glicil tRNA sintetasa.

Gracias a la comparación de estructuras y a métodos de predicción de secuencia se establecieron similitudes de plegamiento con otras proteínas. Entre ellas se encuentran el dominio fosfodiesterasa, la arginil tRNA sintetasa, la enzima de adición de CCA Clase I y otras enzimas de la Superfamilia Polimerasas β .

También con el objetivo de esclarecer la antigüedad de la glyRS se realizó el mapeo de ausencia y presencia en árboles de referencia. El mapeo en el árbol de la vida publicado por Banfield y colaboradores [25] mostró indicios de eventos de reemplazo no homólogo entre ambas versiones de esta enzima. Se usó un modelo para obtener la probabilidad de estos eventos en árboles generados aleatoriamente. Estos mostraron menos eventos significativos de ganancia de la versión heterotetramérica en contraste con múltiples eventos para la ganancia de la versión dimérica. Ya que los grupos cercanos a la raíz entre archaea y bacteria en este árbol han sido considerados un artefacto del método se incluyó también la distribución en estudios recientes de la raíz del último ancestro común de la bacteria. En el árbol publicado por Coleman y colaboradores [8] el grupo Fusobacteria sería cercano a la raíz del último ancestro común de la bacteria. La homogeneidad de presencia de la versión bacteriana en Fusobacteria podría sugerir que la versión bacteriana de la glyRS es antigua.

Introducción

Las aminoacil-tRNA sintetasas (aaRSs) son una familia de enzimas que catalizan la esterificación de aminoácidos con su tRNA correspondiente. Estas enzimas son el punto clave en la fidelidad del código genético ya que son las encargadas de verificar la correspondencia triplete-aminoácido. Al ser el único punto de control se reconoce que la actividad de las aaRSs es tan antigua como el código genético [37, 39].

Las aaRSs pueden ser clasificadas en dos familias monofiléticas diferentes conocidas como Clase I y Clase II. Esto implica que existieron dos ancestros diferentes que dieron origen a los ejemplares conocidos de estas enzimas. Actualmente se conocen 24 diferentes aminoacil tRNA sintetasas, 20 correspondientes a cada aminoácido común en los seres vivos, una para pirrolisina, una para fosfoserina y las últimas dos corresponden a una segunda versión para los aminoácidos lisina y glicina respectivamente. Estas enzimas suelen nombrarse con el código de tres letras del aminoácido sustrato seguido por RS. El código de tres letras y clasificación de las aaRS correspondiente puede observarse en la **Tabla 1**.

Clase I	Met Leu Ile Val	Cys Gln Glu Tyr	Trp Arg Lys I
Clase II	Gly (I / II) Pro Thr His	Ser Ala Asp Asn	Pyl Phe Sep Lys II

Tabla 1. Las dos clases de aminoacil tRNA sintetasas. Se muestra la clasificación de las aminoacil tRNA sintetasas. Ilustradas por el código de 3 letras de los aminoácidos que esterifican. Se muestra Lys I en la Clase I y Lys II en la Clase II, mientras que para glicina en la Clase II están presentes ambas versiones.

Las dos clases difieren en plegamiento y sitios conservados. La Clase I tiene un plegamiento tipo Rossmann con 5 hebras β paralelas conectadas por α hélices. Sus sitios conservados son conocidos como KMSKS y HIGH, nombrados así por los aminoácidos que los conforman. En contraste la Clase II tiene 7 hebras β antiparalelas flanqueadas por α hélices con tres sitios conservados conocidos como motivos 1, 2, 3 respectivamente. Estos plegamientos se pueden observar en la **Figura 1**.

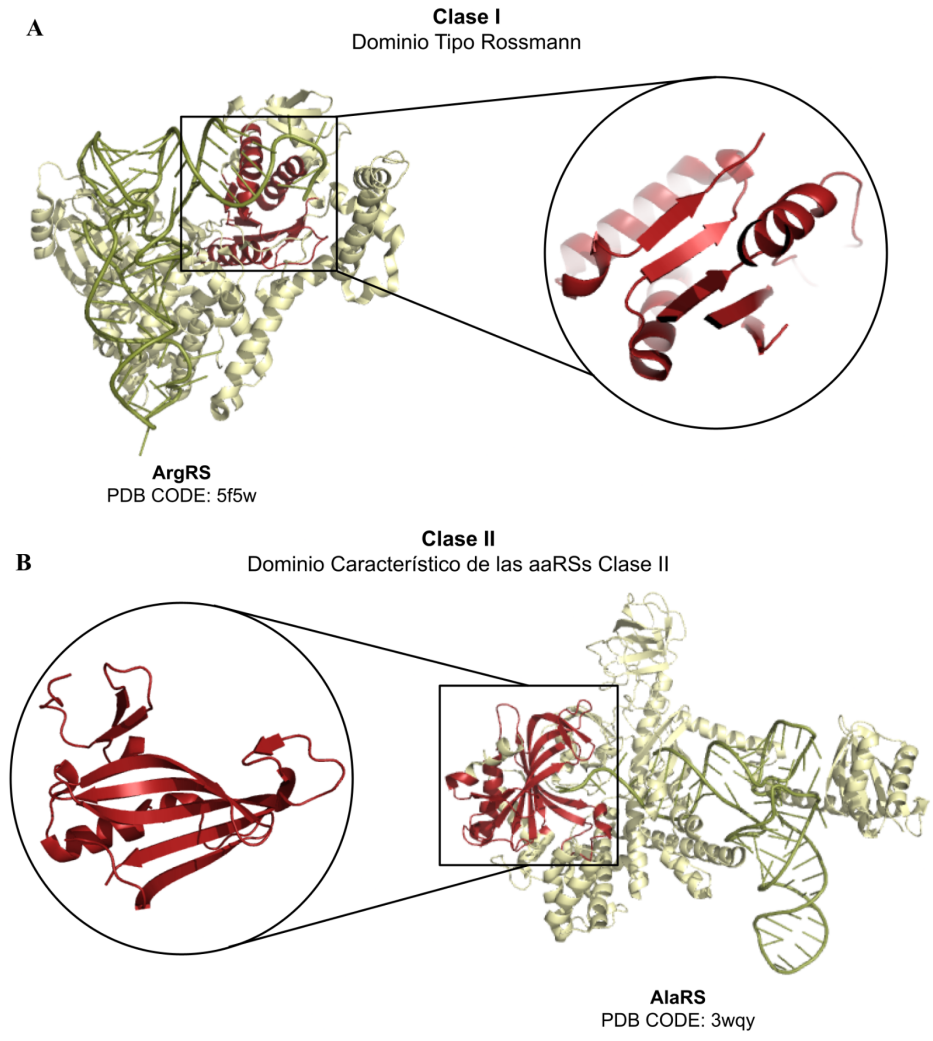


Figura 1. Diferencias en el plegamiento de la Clase I y Clase II de aaRSs. A. Se muestra la representación de la estructura de la arginil tRNA sintetasa (ArgRS) en complejo al tRNA. Ésta pertenece a la Clase I. Acercamiento al plegamiento tipo Rossmann (en rojo). B. El modelo de la alanil tRNA sintetasa (AlaRS) en complejo con el tRNA de alanina. El plegamiento conservado del sitio catalítico de las aminoacil tRNA sintetetas de Clase II (en rojo).

A nivel cuaternario también podemos observar diferencias entre las dos clases. Las aaRSs de Clase I generalmente se encuentran en estado monomérico. Las enzimas de la Clase II se encuentran como dímeros y en el caso de la glyRS y pheRS una subunidad accesoria β .

En cuanto a la correspondencia de una aminoacil tRNA sintetasa por cada aminoácido las dos excepciones son la lysRS y la glyRS. En el caso de la lysRS ambas versiones pertenecen a distintas clases por lo que la existencia de éstas se considera un evento de convergencia en función [40]. En el caso de las dos glyRS ambas pertenecen a la Clase II pero varían en su estado oligomérico, una es dimérica (α_2) y otra es heterotetramérica ($\alpha_2\beta_2$) como se muestra en la **Tabla 2**. Debido a que la versión α_2 es la única en organismos eucariotas y en la mayoría de arqueas la referiremos como eu-glyRS. La versión heterotetramérica será referida como bac-glyRS.

Clase I		Clase II	
aaRS	Oligomerización	aaRS	Oligomerización
Met	α, α_2	eu-Gly	α_2
Leu	α	Pro	α_2
Ile	α	Thr	α_2
Val	α	His	α_2
Cys	α, α_2	Ser	α_2
Gln	α	Asp	α_2
Glu	α	Asn	α_2
Tyr	α_2	Lys II	α_2
Trp	α	Pyl	α_2
Arg	α	Phe	$(\alpha\beta)_2$
Lys I	α	Sep	α_4
		Ala	α_2
		bac-Gly	$\alpha_2\beta_2$

Tabla 2. El estado oligomérico de las aaRSs. Para cada aminoacil tRNA sintetasa se muestra el estado oligomérico, se ilustra con el código de tres letras del aminoácido que esterifican. Mientras que la mayoría de las aaRSs de Clase II se encuentran como monómeros. En la Clase II son dímeros. Entre las excepciones más notorias se encuentra bac-glyRS al contar también con una subunidad β .

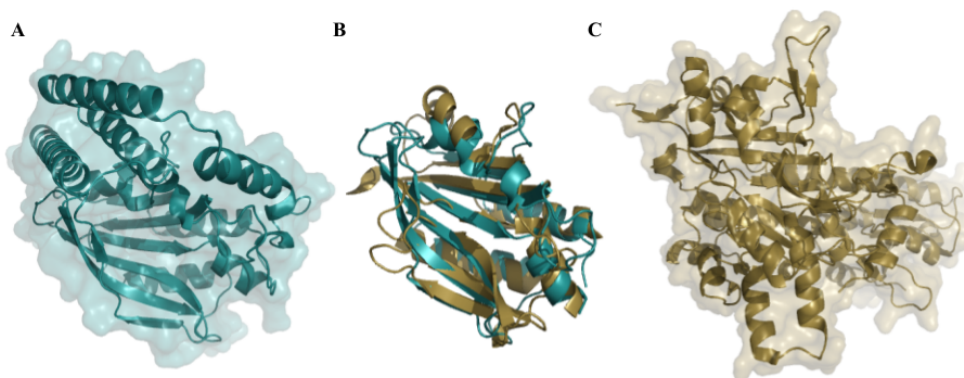


Figura 2. Las dos versiones de glicil tRNA sintetasa pertenecen a la Clase II. A. Estructura de la subunidad α de la bac-glyRS (PDB: 7LU4 cadena A 1-172). B. Alineamiento estructural recortado de subunidad α de bac-glyRS (PDB: 7LU4) y eu-glyRS (PDB: 2zt8). C. Estructura de la eu-glyRS (PDB: 2zt8)

La clasificación de ambas glyRSs fue hecha a nivel de secuencia al comparar la subunidad α de la bac-glyRS con otras aminoacil tRNA sintetatas de la Clase II. Como se puede ver en la **Figura 2** éstas tienen el plegamiento de las aaRSs de Clase II y la eu-glyRS puede ser sobrepuesta con la subunidad α de la bac-glyRS con un RMSD de 2.7 Å [46]. Aunque el plegamiento se encuentra conservado en los centros catalíticos muestran una estrategia diferente de reconocimiento del sustrato. La versión bacteriana interactúa con la glicina mediante aminoácidos como la glutamina, treonina y un ácido glutámico. En contraste, su contraparte eucarionte reconoce a la glicina en parte mediante 3 ácidos glutámicos en distintas posiciones como se muestra en la **Figura 3**.

Comparaciones a nivel de estructura de la subunidad α demostraron que la bac-glyRS α es más parecida a la alaRS que a su contraparte eucarionte. A pesar de que la alaRS y la glyRS unen distintos aminoácidos ambas tienen la misma estrategia para reconocerlos [45]. Esto podría implicar que la actividad de aminoacilación de glicina surgió dos veces en una misma familia de enzimas.

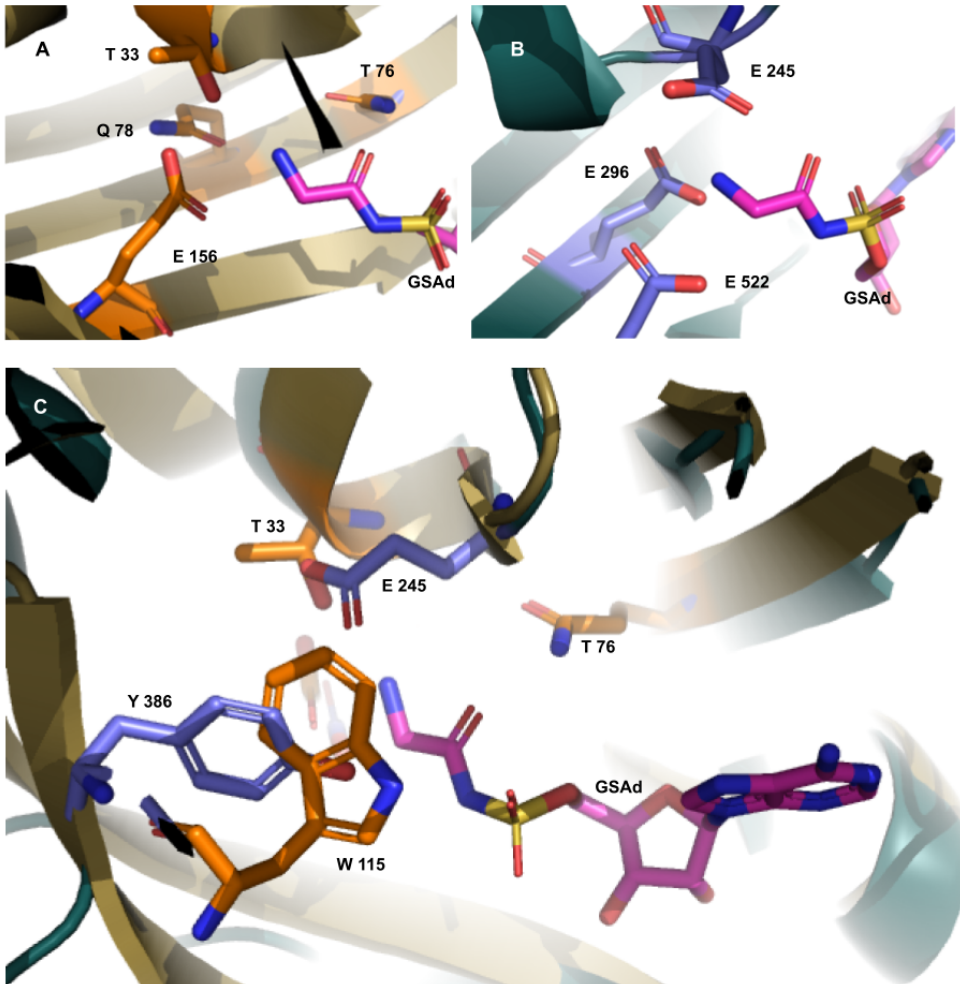


Figura 3. La forma de reconocer a la glicina es diferente en ambas glyRSs. A. Centro catalítico de la bac-glyRS en unión con GSAAd (5'-O-[N-(l-glicil)sulfamoil] adenosin) un análogo del adenilato de glicina. Se muestran los aminoácidos involucrados en el reconocimiento de la glicina (PDB: 1j5w). B. Centro catalítico de eu-glyRS en unión con GSAAd. Se muestra la representación de la cadena lateral de los ácidos glutámicos 245, 295 y 522 (PDB: 2zt8). C. Superposición del sitio de unión a GSAAd en ambas versiones de la glicil tRNA sintetasa.

En cuanto a la estructura completa de la bac-glyRS biológicamente se encuentra como un heterotetrámero $\alpha_2\beta_2$ con una masa total de 100 kDa. La subunidad β tiene una masa aproximada de 36.5 kDa. Ésta flanquea a las subunidades α que se unen en el centro del complejo como se muestra en la **Figura 4**. La catálisis de la aminoacilación se lleva a cabo en la subunidad α por lo que se hipotetiza que la subunidad β podría tener un papel en la unión con el tRNA ya que su presencia tiene efecto en la velocidad de reacción *in vitro* [46].

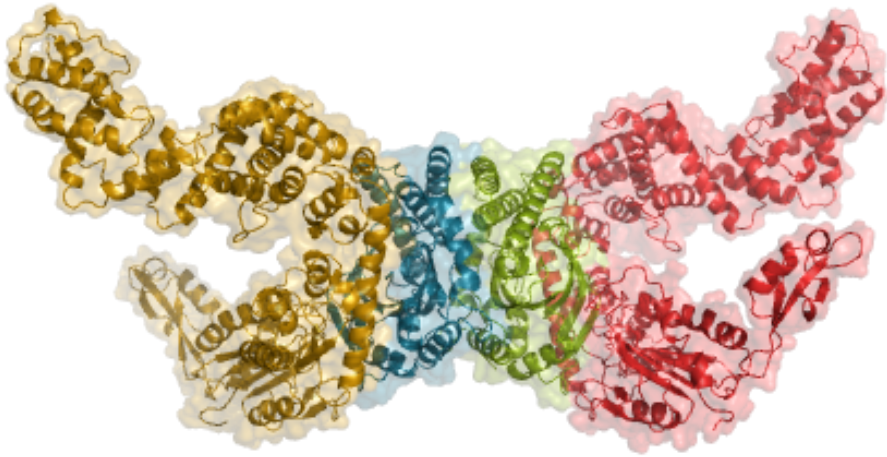


Figura 4. Estructura de la bac-glyRS. La versión heterotetramérica de la glicil-tRNA sintetasa. Las dos subunidades α en el centro (en colores azul y verde). Las subunidades beta flanquean a las subunidades alfa (en colores amarillo y rojo) PDB:7LU4.

Estructura de la subunidad β

La subunidad β es única en su tipo, dentro de las aminoacil tRNA sintetasa la fenilalanil tRNA sintetasa (pheRS) y la glicil tRNA sintetasa bacterianas son las únicas que tienen otra subunidad. En el caso de la pheRS las dos subunidades tienen un tamaño similar y su interfaz de contacto se da a lo largo de toda la proteína [1]. La representación de la estructura de la pheRS se puede observar en la **Figura 5**.

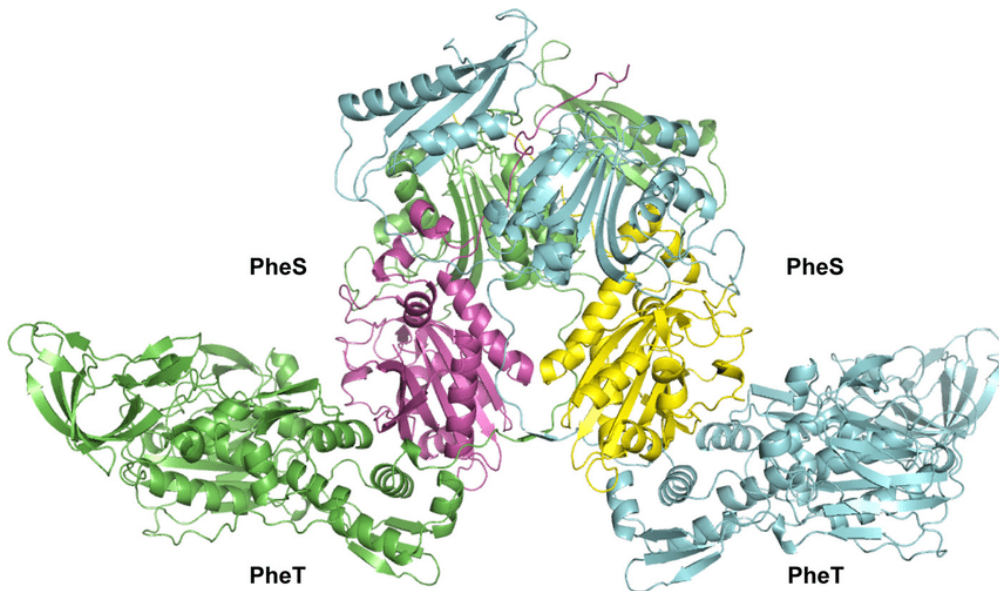


Figura 5. La estructura de la PheRS muestra una oligomerización tipo $(\alpha\beta)_2$. Se muestra el modelo de la estructura de fenil-tRNA sintetasa. Las dos subunidades PheS o α (morado y amarillo) se encuentran rodeadas por una subunidad PheT o β cada una (verde y azul respectivamente). La disposición se considera $(\alpha\beta)_2$ ya que es la unión de dos dímeros $\alpha\beta$ en contraste a la glicil tRNA sintetasa bacteriana que tiene interfaces entre cada subunidad. En este caso dos subunidades PheT cristalizaron junto con el $(\alpha\beta)_2$.

La subunidad β de la bac-glyRS puede ser dividida en 5 dominios distintos como se muestra en la **Figura 6**. Estos podrían tener un papel en el acomodo del tRNA de glicina (gly-tRNA).

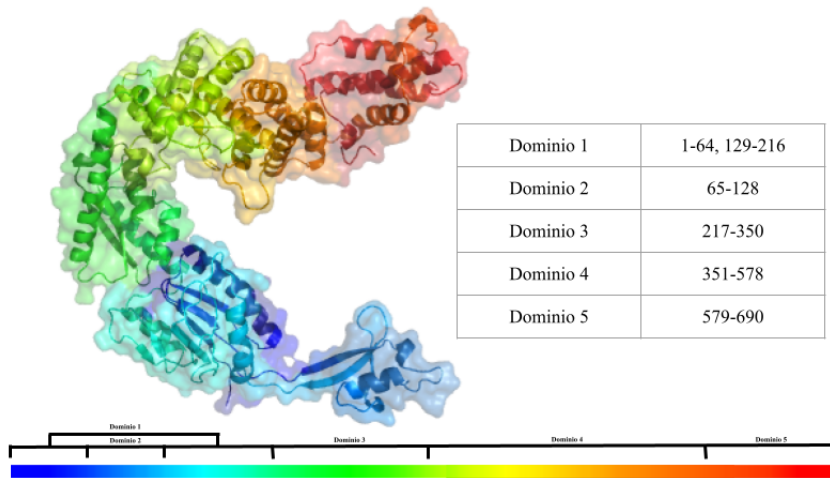


Figura 6. La subunidad β de la bac-glyRS en dominios. Se muestran la estructura de la subunidad β de la bac-glyRS con los colores del arcoiris de N-terminal (azul) a C-terminal (rojo). Se muestran en la tabla los aminoácidos correspondientes a los 5 dominios con base en el programa vast [27].

El grado de conservación de la secuencia varía en los dominios de la subunidad β . El dominio que muestra mayor conservación es el 3, posiblemente por su papel de interfaz entre la subunidad α y β de esta enzima. En la **Figura 7** se muestran los diferentes grados de conservación en la superficie de la proteína.

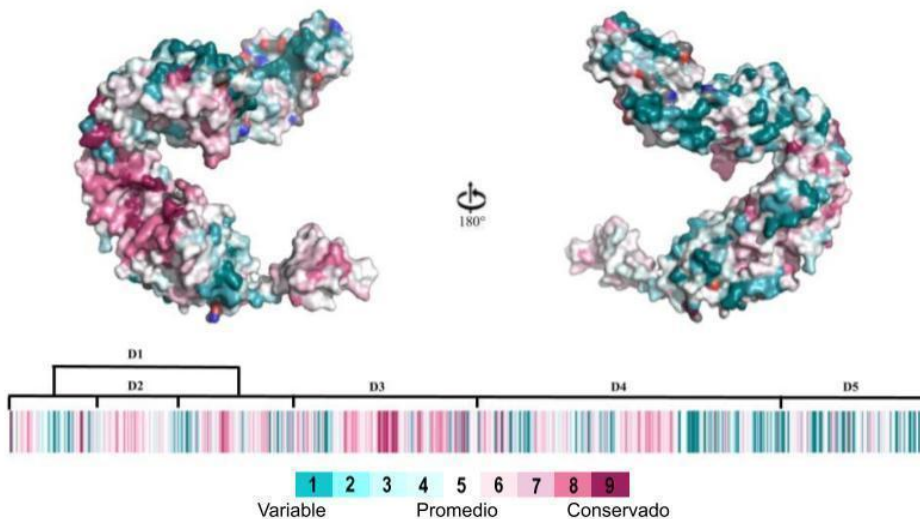


Figura 7. Conservación de secuencia en la subunidad β de la bac-glyRs. Se observa desde dos ángulos opuestos la estructura de la subunidad β , la superficie se muestra conforme a un gradiente de conservación de secuencias generado con MUSCLE e imagen realizada con el plugin de Consurf para PyMol [10].

Distribución

La antigüedad de una proteína se puede inferir a partir de realizar un análisis de cuántos y cuáles organismos cuentan con ella. Una proteína universalmente conservada puede ser considerada ancestral. La presencia esporádica de las dos formas de glyRS indica eventos de transferencia horizontal por lo que no es sencillo concluir que tan antiguas son éstas.

Un mapeo de presencia-ausencia de la bac-glyRS puede ser usado para inferir eventos de transferencia horizontal de genes (LGT por sus siglas en inglés). A los eventos de transferencia horizontal de este tipo se les conoce como reemplazo no homólogo, involucrando la pérdida de una enzima y la ganancia de otra que tiene la misma función pero no comparte un ancestro común. El modelo de la probabilidad de LGT propuesto por Stern y colaboradores [43] cuenta con 4 parámetros que representan las velocidades de ganancia y pérdida de los dos genes. Estos parámetros se calcularon usando técnicas de máxima verosimilitud. El modelo de Markov consiste en representar la probabilidad de que un evento estocástico suceda, en este caso la transferencia horizontal del gen. Éste simula la evolución sobre un alfabeto de 3 estados describiendo en este caso la presencia de bac-glyRS, eu-glyRS o las dos. Se asume que los eventos de ganancia y pérdida simultáneos son imposibles.

El modelo probabilístico fue inicialmente propuesto para las enzimas timidilato sintetasas (TS). También se considera válido para las glyRSs ya que ambas proteínas comparten:

- 1) La existencia de dos versiones de la enzima las cuales no están cercanamente relacionadas evolutivamente
- 2) Una distribución esporádica en los organismos bacterianos.
- 3) Ambas son proteínas fundamentales, lo que significa que la pérdida de éstas compromete la supervivencia del organismo. Lo cual, a su vez, implica que el reemplazo no homólogo pasa por 3 estados: (1) inicial donde ocurre la ganancia de la segunda versión, (2) intermedio con la

presencia de ambas versiones y (3) estado final cuando ocurre la pérdida de la versión inicial de la proteína.

- 4) Su papel enzimático se considera primordial y antiguo, posiblemente encontrado en el Último Ancestro Común (LUCA por sus siglas en inglés). Parte de la replicación del DNA en el caso de las timidilato sintetasas y fidelidad de la traducción proteica en el caso de las glicil tRNA sintetasas.

El cálculo de las probabilidades de transición depende de la topología del árbol generado con marcadores universales. Dicho árbol debe ser generado a partir de otros marcadores independientes de la proteína de interés. Aplicar el modelo evolutivo nos permite calcular la probabilidad posterior de los eventos de LGT.

Establecimiento de homología

La homología es una hipótesis de origen en común, en este caso, entre dos proteínas. Aunque no existe una regla general para declarar dos enzimas homólogas, la comparación de distintos aspectos entre éstas es necesaria.

Búsqueda de homólogos lejanos por secuencia

Para la comparación a nivel secuencia es necesario realizar un alineamiento. La identidad entre las secuencias se refiere al porcentaje de aminoácidos en común. Mientras que la similitud entre dos proteínas es el porcentaje de identidad, más el porcentaje de aminoácidos que al estar alineados son parecidos en propiedades fisicoquímicas. Porcentajes altos de identidad y similitud pueden sustentar una hipótesis de ancestro común .

Entre más antiguo sea el ancestro común habrá mayor divergencia de las secuencias, esto se ve reflejado en porcentajes bajos de identidad (30-50%) [36]. La mayoría de los algoritmos enfrentan problemas para identificar relaciones filogenéticas entre secuencias con niveles bajos de similitud. Para reconocer homólogos remotos es necesario aumentar la flexibilidad de las comparaciones mediante varios métodos.

PSI-BLAST (Position-Specific Iterative Basic Local Alignment Search Tool) es una versión de BLAST que crea su propia matriz de puntuación específica por posición (PSSM por sus siglas en inglés) con el alineamiento múltiple generado con las secuencias encontradas en cada búsqueda. La nueva PSSM es usada para la siguiente búsqueda. De esta manera con cada iteración los valores en el PSSM cambian, conteniendo información de secuencias cada vez más alejadas a la ingresada en el inicio. Está reportado que PSI-BLAST tiene una sensibilidad más grande para encontrar homólogos lejanos comparado con el protocolo de BLASTp [4].

Otra aproximación es HMMER, un algoritmo que implementa los modelos ocultos de Markov. Con estos se crea una matriz análoga a las PSSMs para crear un perfil que describe la variabilidad de homólogos reportados de una enzima. Posteriormente este perfil se usa para realizar una búsqueda contra una base de datos de proteínas y así encontrar secuencias lejanas.

Los resultados son reportados con un valor E asociado. El valor E puede ser interpretado como una medida de la probabilidad de encontrar falsos positivos. Por esta razón entre más bajo sea el valor E se considera mejor. El manual de HMMER sugiere considerar los resultados con valores menores a 1×10^{-3} [37].

Otras aproximaciones

El software HHPRED sirve para predecir estructura a partir de secuencias. Éste tiene los siguientes pasos:

1. Construye un alineamiento múltiple de la secuencia de interés: Usando PSI-BLAST se realizan hasta 8 búsquedas en la base de datos de proteínas no redundantes de NCBI. Enriqueciendo cada búsqueda con los resultados de la búsqueda anterior.
2. Búsqueda de modelos homólogos: Se calculan perfiles de cadenas ocultas de Markov con los alineamientos del paso anterior y se compara con modelos disponibles de otras proteínas en la base de datos PDB70 para ser incluidos. La PDB es construida a su vez con PSI-BLAST siendo la muestra de aquellas proteínas que tienen un porcentaje de identidad mayor a 70%.
3. Se reorganizan los modelos considerados usando una red neuronal: HHsearch asigna una puntuación y organiza a todos los modelos encontrados. La puntuación se considera como la probabilidad de que los perfiles sean homólogos.
4. Se generan alineamientos múltiples disminuyendo la diversidad de secuencias para la secuencia de interés y los modelos encontrados: Las probabilidades de HHPRED en este punto no reflejan directamente el parentesco de los modelos encontrados con la secuencia de interés. Por

esto hay que decidir cuáles tienen una relación con la secuencia inicial descartando varios modelos.

5. Asignar puntuaciones a los modelos no descartados: Ya que todos los modelos fueron hechos a partir de PDB70, todos tienen una estructura disponible en el Protein Data Bank. Para cada uno de los alineamientos se predice un puntaje de similitud topológica.
6. Elegir modelos que serán usados para el siguiente paso: HHPRED elige un sólo modelo por dominio. El alineamiento entre el modelo y la secuencia de interés debe cubrir mínimo 40 residuos.
7. Crear Modelos Tridimensionales: En este punto se recurre a Modeller, un software para creación de modelos de proteínas.

Aunque HHPRED está diseñado para predecir la estructura tridimensional de una proteína a partir de su secuencia, es posible usarlo para encontrar candidatos homólogos.

El uso de este software no sólo implica una alternativa a otros métodos a nivel de secuencia, sino que también al considerar la estructura tridimensional predicha brinda una idea de las proteínas que podrían alinear estructuralmente con nuestra secuencia de interés [22].

Búsqueda de homólogos por estructura

La estructura está más conservada que la secuencia, por esto comparaciones a nivel estructural prometen el poder reconocer relaciones evolutivas más lejanas. Protocolos como DALI SEARCH permiten reconocer entradas del PDB que se alinean estructuralmente con la estructura de interés. Este programa toma las coordenadas cartesianas de las posiciones atómicas de una proteína y retorna una lista de estructuras similares disponibles en el PDB así como un alineamiento estructural entre la proteína de interés y la encontrada. Este análisis es independiente de la secuencia de la proteína.

El algoritmo de búsqueda empleado por DALI es heurístico, por lo cual no se garantiza encontrar la solución más óptima. Éste sacrifica exactitud y precisión a favor del desempeño computacional. Para llevar a cabo el alineamiento emplea matrices de distancia. Las matrices de distancia contienen una

descripción de la estructura tridimensional de la proteína al registrar las distancias entre todos los carbonos α . Las comparaciones estructurales se realizan a partir de calcular las similitudes geométricas definidas como la suma con pesos de las distancias intramoleculares registradas en las matrices de ambas proteínas. Regiones que sean equivalentes estructuralmente serán filtradas y pueden ser separadas para presentar el alineamiento estructural de sólo ese segmento de las proteínas.

Los resultados encontrados son organizados de mejor a peor con base en varias métricas como la raíz cuadrada media de errores cuadrados (RMSD por sus siglas en inglés) y el puntaje Z. El RMSD muestra el promedio de distancias entre las coordenadas atómicas de las dos estructuras, una estructura alineada consigo mismo tendría un RMSD de 0, ya que no hay variabilidad entre las posiciones de los átomos. Entre más bajo sea el RMSD mejor será el alineamiento estructural entre ambas proteínas. Aunque no hay una regla general para usar el RMSD para soportar una hipótesis de homología en modelado molecular, un valor menor a 2 se considera aceptable. En cuanto a la puntuación Z, un puntaje mayor a 20 significa que las dos estructuras son homólogas, entre 8 y 20 existe una probabilidad que lo sean. Los valores más bajos de 8 son un área gris en cuanto a su interpretación y los menores a 2 no se consideran significativos. El tamaño de las proteínas tiene un efecto en el Z-score, estructuras pequeñas tendrán valores más bajos y estructuras más grandes tendrán puntajes más altos. También estructuras con un plegamiento α/β tienen valores más altos que proteínas con sólo plegamientos β [22, 23].

Aunque ninguno de los métodos por sí mismo puede establecer relaciones evolutivas, un consenso entre éstos puede dar soporte a una hipótesis de homología. Además describir relaciones evolutivas entre la bac-glyRS con otras proteínas y definir si ésta es antigua es un comienzo para plantear hipótesis sobre el origen de esta proteína.

Hipótesis

La estructura de la subunidad β de la bac-glyRS puede darnos indicios del pasado evolutivo de la enzima. La presencia de la bac-glyRS en grupos basales del árbol de la vida indicaría si es una proteína ancestral.

Objetivo general

Describir las posibles relaciones evolutivas de la subunidad β de la bac-glyRS con otras proteínas.

Objetivos particulares

- Mapear la presencia o ausencia de la bac-glyRS en organismos de árboles previamente reportados.
- Evaluar la existencia de eventos de transferencia horizontal entre las dos versiones de la glyRS.
- Buscar homólogos remotos de la subunidad β de la bac-glyRS a nivel de secuencia y estructura.
- Describir las posibles relaciones evolutivas entre los candidatos encontrados y la bac-glyRS.

Métodos

1) Distribución

Para el árbol de distribución de las glyRSs se usaron como referencia los datos disponibles del árbol de la vida publicado por Banfield y colaboradores en 2016 [25]. En dicho estudio se construyó un árbol con una concatenación de las 16 proteínas ribosomales conservadas universales (RpL2, 3, 4, 5, 6, 14, 15, 16, 18, 22, and 24, and RpS3, 8, 10, 17, and 19). El árbol contiene 1871 organismos por lo que inicialmente se hizo una selección de 90 organismos considerados como representativos. Se eligieron ciertos organismos con el criterio de ilustrar en cada grupo ejemplos de ambas versiones si es que las había. Finalmente se añadieron otros como *Thermoanaerobacter daxensis*. Para la figura 9 se usaron 98 organismos.

El árbol se construyó como está descrito en [24] integrando datos previos del árbol de la vida construido con el mismo método [25]. MUSCLE v3.8.155 se usó para obtener los alineamientos. La filogenia de máxima verosimilitud se realizó con PhyML v3.1 [10, 18, 21]. La asignación del tipo de glyRS fue hecho al descargar automáticamente el primer resultado del NCBI los resultados de la búsqueda: “Nombre del organismo” + “glycyl tRNA synthetase beta” y “Nombre del organismo” + “glycyl tRNA synthetase beta” y posteriormente alinearlas. Las asignaciones no contundentes fueron verificadas mediante BLAST [3]. El árbol fue anotado usando iTOL [29, 30].

2) Análisis de transferencia horizontal de genes

Para hacer posible el análisis de LGT y probar la robustez de este método se generaron aleatoriamente 10 árboles. La selección de organismos se dio posteriormente de haber clasificado a los organismos en grupo. La selección final consistió de 146 organismo ya que esto mantendría la proporción del árbol original publicado por Banfield en 2016 [24].

Se usó el código proporcionado en los materiales suplementarios del artículo de Stern y colaboradores (2010) [40].

3) Búsqueda de homólogos lejanos con PSI-BLAST

La programa de PSI-BLAST se alimentó con la secuencia de la subunidad β de la bac-glyRS. Se realizaron 10 iteraciones con el umbral conservador de 1×10^{-4} [4].

4) Análisis de secuencia usando HMMER

Para crear el perfil de HMMER de la subunidad beta se recuperaron 2575 secuencias de la subunidad beta de la base de datos del Centro Nacional de Información Biotecnológica (NCBI por sus siglas en inglés). Para encontrar todas las coincidencias se prosiguió a hacer una búsqueda con la base de datos uniprot_sprot, de acuerdo con las instrucciones del manual de HMMER en su versión 3.3.2. [39].

5) Predicción de Estructura HHPRED

Para el uso de HHPRED se recurrió al servidor de Herramientas de Bioinformática del Instituto Max Plank [49, 39]. Se usó la secuencia de la subunidad β en la glyRS de *Thermanaerothrix daxensis*.

6) Búsqueda por medio de alineamiento de estructuras.

Se realizó el protocolo PDB Search con DaliLite en su versión 4.

Se hicieron búsquedas separadas de la subunidad beta y de cada uno de los 5 dominios [23, 24].

Resultados

Distribución de las dos versiones de glyRS

Se realizó el mapeo de presencia-ausencia en el árbol de Banfield (2016) [25] el cual se puede ver en la **Figura 8**. Este árbol consiste en 1872 hojas por lo que para la mejor representación de los datos se muestra en la **Figura 9** un árbol derivado con 98 organismos bajo el criterio de que fueran representativos de la presencia o ausencia de la bac-glyRS del grupo al que pertenecen.

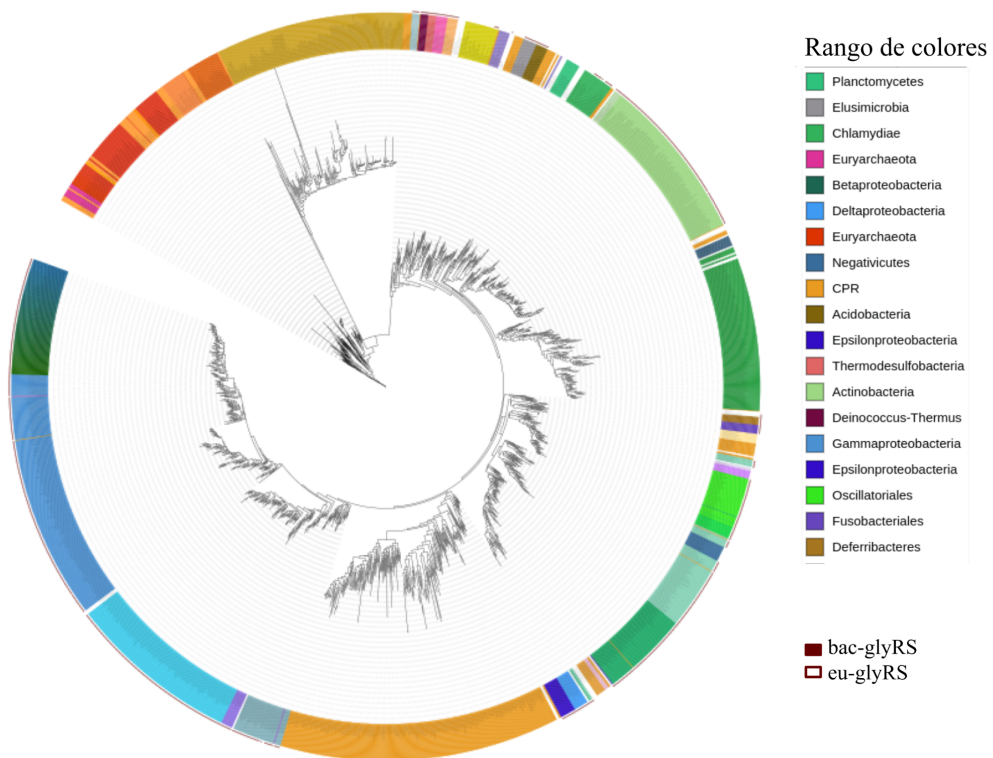


Figura 8. Presencia de las glicil-tRNA sintetisas en el Árbol de la Vida. Se muestra el árbol de Banfield [24] anotado, los organismos fueron clasificados por grupos.

Como se observa en la **Figura 8 y 9** en *Proteobacteria* existe una dominancia de la bac-glyRS. Grupos como *Chloroflexota*, *Deinococcus-Thermus* y *Streptomyces* cuentan con algunos organismos con bac-glyRS y otros eu-glyRS. Su distribución suele ser mutuamente excluyente, con excepción de los organismos que cuentan con la versión bacteriana en el cloroplasto. El grupo *Patescibacteria* o CPR (*Candidate phyla radiation* por sus siglas en inglés) destaca por sólo contar con la versión eucariote.

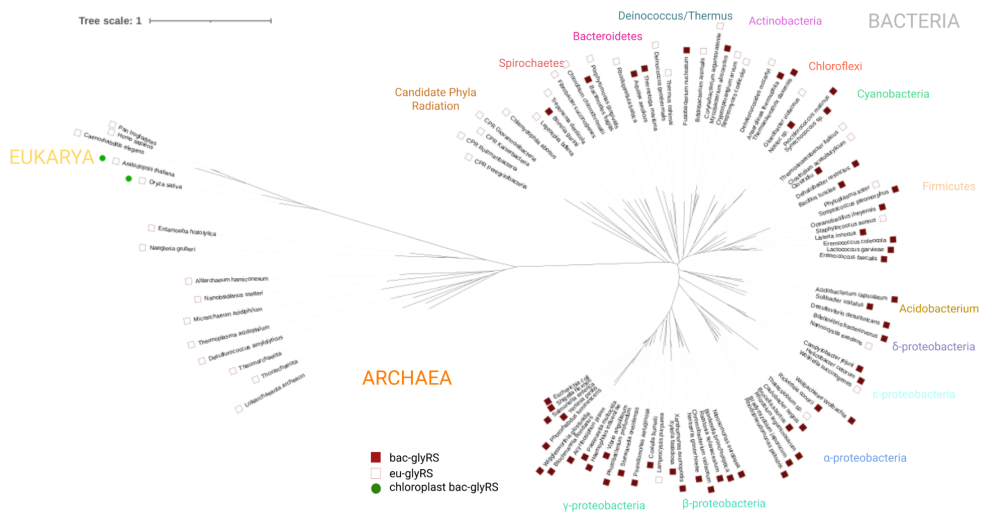


Figura 9. Distribución de las dos versiones de glyRS en los tres dominios de la vida. Árbol sin raíz con organismos seleccionados del árbol de Banfield (2016) [25], además de ciertos organismos modelo añadidos posteriormente. Los marcadores usados son las 16 proteínas ribosomales universales (L2, L3, L4, L5, L6, L14, L16, L18, L22, L24, S3, S8, S10, S17, S19). En un rectángulo rojo se indica la presencia de bac-glyRS, un rectángulo blanco la de su contraparte eucariote y en un círculo verde se indica la presencia de la bac-glyRS en el cloroplasto. En la **Figura Suplementaria 3** se encuentra esta figura en un tamaño más grande.

No hay reportes de la existencia de las dos versiones de la glyRS a excepción de organismos que tienen la versión bacteriana en el cloroplasto y la eucariota en el núcleo.

Análisis de transferencia horizontal

Para los distintos árboles se reportaron diferentes eventos de transferencia horizontal de genes, los eventos significativos de ganancia de eu-glyRS superan en número a los de ganancia de la bac-glyRS (**Tabla 3**). El único evento significativo de ganancia de la bac-glyRS que se dio en todos los árboles fue en el nodo interno que representa el ancestro común entre los grupos CPR y otras bacterias. En este caso es el nodo 6 como se ilustra en la **Figura 10**.

Tabla 3. Probabilidad de eventos de ganancia de la glicil tRNA sintetasa de eucariontes.

Nombre del organismo	Probabilidad de la ganancia eu-glyRS
<i>Thermoanaerobacter italicus</i>	0.761
<i>Staphylococcus aureus</i>	0.856
<i>Dehalococcoides mccartyi</i>	0.866
<i>Gloeobacter violaceus</i>	0.805
<i>Thermus oshimai</i>	0.671
<i>Leptospira biflexa</i>	0.511
<i>Treponema denticola</i>	0.524
<i>Rhodopirellula báltica</i>	0.672
<i>Wolinella succinogenes</i>	0.603
<i>Thalassiosira sp.</i>	0.859

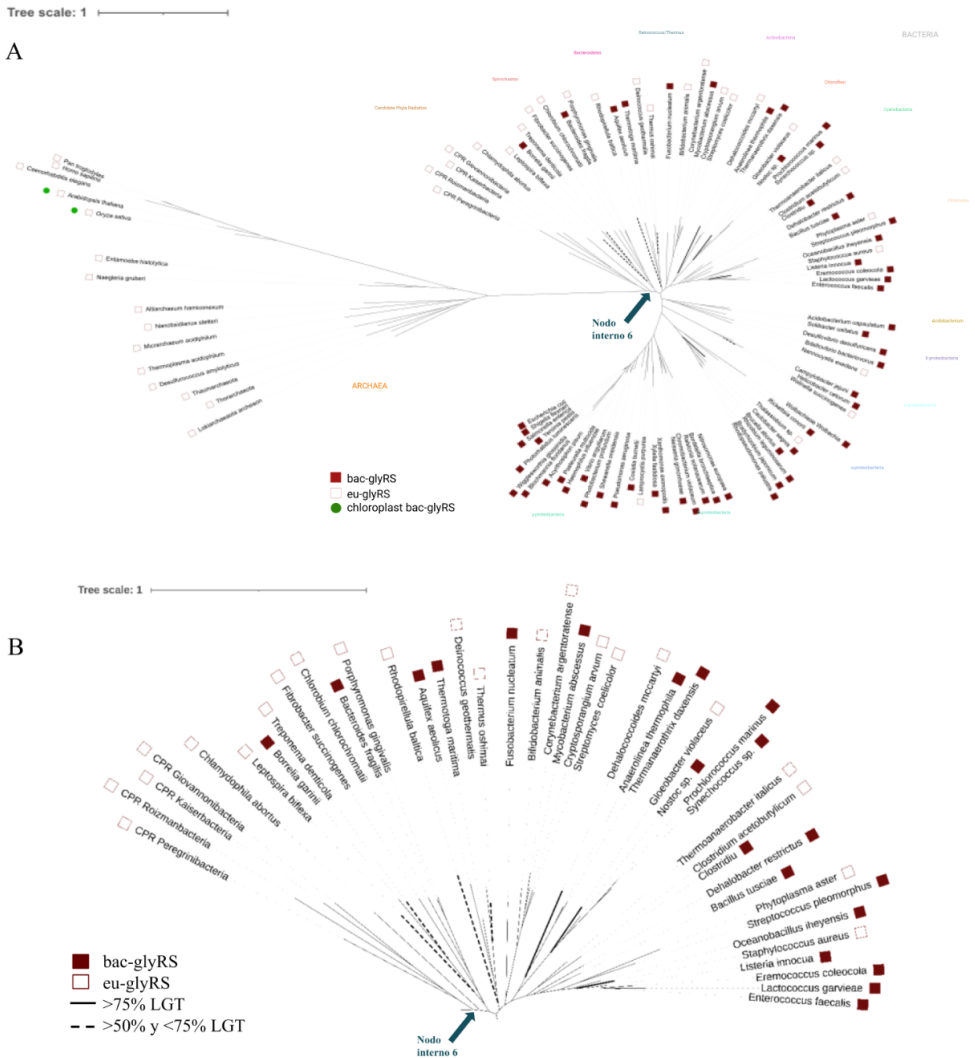


Figura 10. Transferencia horizontal de las glicil tRNA sintetetas. A. Se muestra árbol sin raíz de los 98 organismos modelo seleccionados. Ramas con resultados significativos LGT resaltadas en negro ramas con una probabilidad mayor a 0.75 de ganancia de eu-glyRS, ramas con línea punteada una con una probabilidad entre 0.5 y 0.75. También se indica con una flecha azul el nodo interno 6. B. Acercamiento de una zona con varias ramas resaltadas. Todas éstos eventos se encuentran en la **Tabla 3** con su respectiva probabilidad de ganancia de eu-glyRS.

Búsqueda de homólogos por secuencia

Con el fin de establecer la ancestralidad de la bac-glyRS por otros métodos se prosiguió a buscar su relación evolutiva con otras proteínas. Para hacer esto se siguió la rutina de PSI-BLAST. Se encontraron secuencias de la bac-glyRS con valores de identidad bajos, como lo sería el caso de 31.7% que se muestra en la **Tabla 4**.

Entre los resultados diferentes a otras glyRSs se encuentra la arginil tRNA sintetasa (argRS) y el dominio DARL de unión al anticodón. Los dominios DARL son conocidos así por la secuencia de los aminoácidos aspartato - alanina - arginina - lisina (D-A-R-L) conservada. Este dominio de hélices alfa está presente en los sitios de reconocimiento del anticodón de las sintetastas de arginina, glicina y cisteína [25].

Tabla 4. Algunos resultados de la iteración 10 con PSI-BLAST.

Nombre	Código Uniprot	Puntaje	Valor E	Identidad
Glycyl-tRNA synthetase	Q8L785	907	0	47.75%
Glycyl-tRNA synthetase	P77891	326	3e-103	31.70%
Arginyl-tRNA synthetase	Q32H69	179	3e-46	21.52%
DARL anticodon binding domain	Q6PJN8	50	8e-05	15.09%
Debajo del umbral				
Ribonuclease Y	Q01444	42.3	0.017	19.85%
Alanyl-tRNA ligase	A6US09	42.7	0.019	28.28%

Como alternativa a PSI-BLAST se decidió realizar un perfil HMMER de la β -bac-glyRS y realizar una búsqueda de posibles homólogos. Se usaron las 2575 secuencias disponibles de la subunidad β para crear el perfil HMMER y con éste se realizó una búsqueda contra la base de datos uniprot. Una muestra de los resultados de HMMER se pueden observar en la **Tabla 5**. Una comparación de las proteínas en estos resultados con la bac-glyRS a nivel de estructura y sitios conservados no dio soporte a que pudiesen estar relacionadas.

Tabla 5. Algunos resultados de HMMER search con el perfil de bac-glyRS.

Valor E	ID de la secuencia	Nombre
2.2e-64	Q4WV19	Adenylyltransferase and sulfurtransferase Uba4
6.9e-12	P12282	Molybdopterin-synthase adenylyltransferase
Debajo del umbral		
0.19	Q46DR6	Arginine-tRNA ligase <i>Methanosarcina acetivorans</i>

Predicción de estructura con HHPRED

Para robustecer los resultados a nivel secuencia se siguió el protocolo de HHPRED. Los resultados sugirieron distintas proteínas para distintas regiones de la estructura. Para el segmento correspondiente al Dominio 1 y 2 se encuentra la coincidencia de la enzima de adición del CCA de Clase I. En el dominio 4 se encuentran enzimas con un dominio HD, entre ellas la Fosfodiesterasa PgpH. El dominio 5 coincide con el dominio anticodón de la arginil tRNA sintetasa (argRS), la cual pertenece a la Clase I de aaRSs.

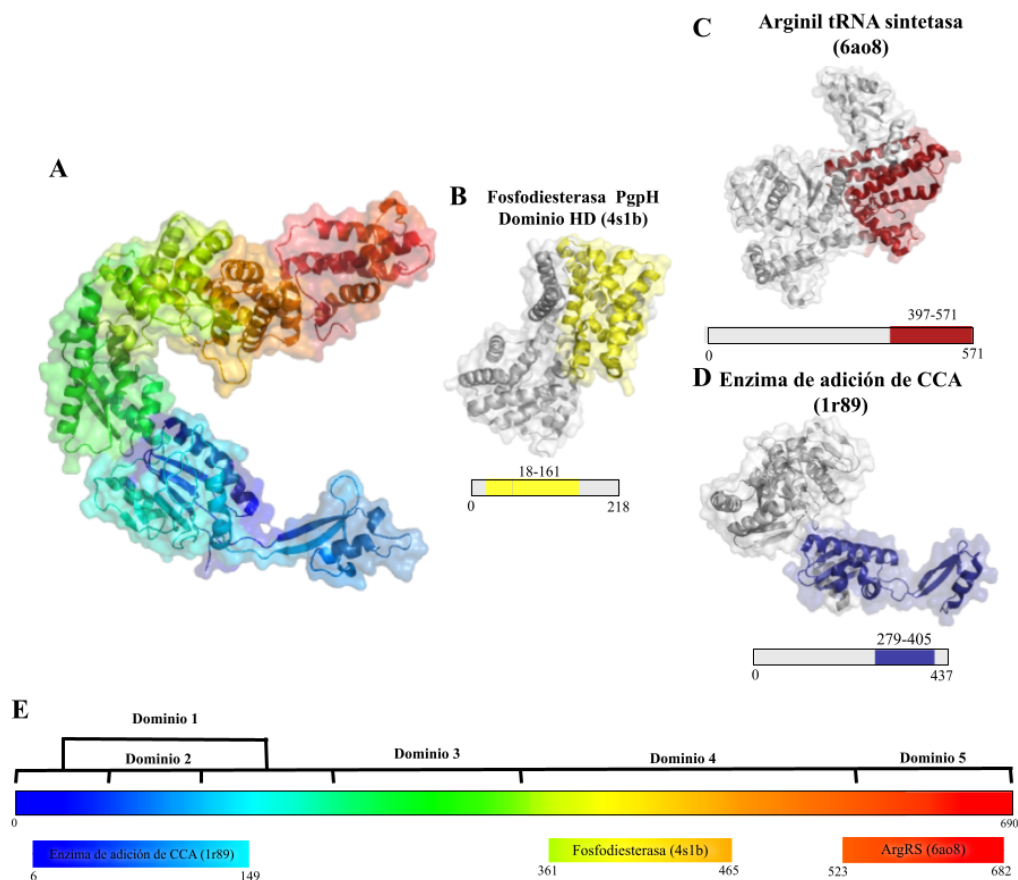


Figura 11. Resultados por búsqueda de secuencia con HHPRED. A. Modelo estructural de la β -bac-glyRS con los colores del arcoiris de N-terminal (en azul) a C-terminal (en rojo) B. Fosfodiesterasa PgpH (PDB: 4s1b). Segmento de la coincidencia en amarillo C. Dominio anticodón de la argRS (en rojo, PDB:6ao8) D. Enzima de adición del CCA. Segmento con similitud en azul (PDB: 1r89) E. Se representa la secuencia de la bac-glyRS. Se indican los dominios así como las coincidencias indicando con cuáles segmentos de la secuencia de la subunidad β corresponden.

Las enzimas de adición del CCA, como su nombre indica son añaden los nucleótidos citidina, citidina y adenosina al final de los tRNAs que no lo tienen codificado. Éste procesamiento es necesario para que posteriormente los tRNAs sean reconocidos por las aaRSs. El hecho que una enzima de adición de CCA y la argRS sean parte de los resultados de HHPRED indicaba que la subunidad β podría mostrar similitud con otras proteínas que unieran tRNAs.

En la **Figura 11** se muestran qué segmentos de las enzimas han sido sugeridos para su análisis. En el caso de la enzima de adición de CCA Clase I HHPRED propone usar del aminoácido 279 al 405 para modelar del 8 al 149 de la subunidad β de la bac-glyRS. El dominio HD de la Fosfodiesterasa PgpH va del aminoácido 18 al 161 como se muestra en la barra inferior a la estructura. Las probabilidades generadas con HHPRED se muestran en la **Tabla 6**.

Tabla 6. Tres mejores coincidencias de HHPRED por región.

Nombre	Código PDB	Probabilidad (%)	Puntuación
Enzima de adición del CCA	1r89	92.4	49.9
Fosfodiesterasa	4s1b	90.8	41.9
Arginil tRNA sintetasa	6ao8	90.6	51.3

Búsqueda por comparación de estructuras

La estructura está más conservada que la secuencia. Gracias a esto una superposición de estructuras podría soportar una hipótesis de homología lejana. Bajo esta premisa se procedió a hacer una búsqueda con DALI comparando la estructura de interés con las presentes en el Protein Data Bank. Los resultados mostraron varias polimerasas, enzimas de adición de CCA de Clase I y otras proteínas relacionadas con el procesamiento de tRNA.

Como se muestra en la **Tabla 7**, el RMSD y la puntuación Z se encuentran en áreas grises, donde no son lo suficientemente pequeños para declarar una homología a partir de estos aunque muestren parecido estructural. Sólo se muestra una selección pequeña de los resultados. Una versión más completa de los resultados se encuentra en la **Tabla Suplementaria 3**.

Tabla 7. Resultados seleccionados de DALI SEARCH.

Código PDB	Nombre	Puntuación Z	RMSD
6l42	RNA Polymerase	6.7	3.9
4x4r	CCA-Adding enzyme	6.6	3.2
5che-A	Glutamyl-tRNA reductase 1, Chloroplastic	4.9	3.4
5axk-B	tRNA-(HIS)-5'-Guanylyltransferase THG1	4.2	3.6
3wbz-E	Histidyl tRNA specific guanylyltransferase	4.1	3.6
3al0-B	Glutamyl-tRNA amidotransferase subunit A	4.0	3.9

Los alineamientos estructurales mostraron que la similitud estructural sólo se daba en regiones de entre 130-170 aminoácidos como se puede mostrar en la **Tabla Suplementaria 3**. Por esto se realizó una búsqueda de estructuras con cada dominio. **Tablas Suplementarias 4-8**. Esto permitió agrandar las coincidencias por sección. Las proteínas propuestas para el dominio 3 contenían una hélice con buena sobreposición con la presente en este dominio pero no tenían otras características en común.

Discusión

Ancestralidad y distribución de la glicil tRNA Sintetasa Bacteriana

Aunque en el árbol de Banfield y colaboradores [25] no se tenga un grupo externo para designar la raíz, al colocarla entre Arquea y Bacteria el árbol sugiere que el grupo Candida Phylum Radiation (CPR) estaría cercano al LUCA (Último Ancestro Común Universal por sus siglas en inglés). La presencia de la bac-glyRS no ha sido reportada en organismos CPR. Esto sugeriría la ausencia de la versión bacteriana en LUCA.

Sin embargo, la ramificación profunda del árbol que se usó de referencia no ha sido bien aceptada. Uno de los argumentos en contra es que el grupo CPR se compone mayoritariamente de simbioses y parásitos de genomas pequeños con altas tasas de transferencia horizontal de genes con sus organismos asociados. Banfield y colaboradores en un artículo posterior han sugerido que la raíz podría encontrarse entre linajes mayores como Proteobacteria, Actinobacteria, Firmicutes, o Cyanobacteria [7]. Por esto se prosiguió a buscar un estudio enfocado en enraizar el árbol y así definir cuál versión de glyRS tenían los organismos de los grupos cercanos.

Coleman y colaboradores [8] buscaron enraizar el árbol del Último Ancestro Común de Bacteria (LBCA por sus siglas en inglés). En este se usó una aproximación jerárquica filogenética, la cual les permitió colocar la raíz en el árbol sin la necesidad de un grupo externo. Con este método no se encontró soporte alguno para tener a los organismos de CPR cercanos a la raíz. En cambio en este árbol CPR forma parte del grupo Terrabacteria y la raíz del árbol entre Terrabacteria y Gracillicutes.

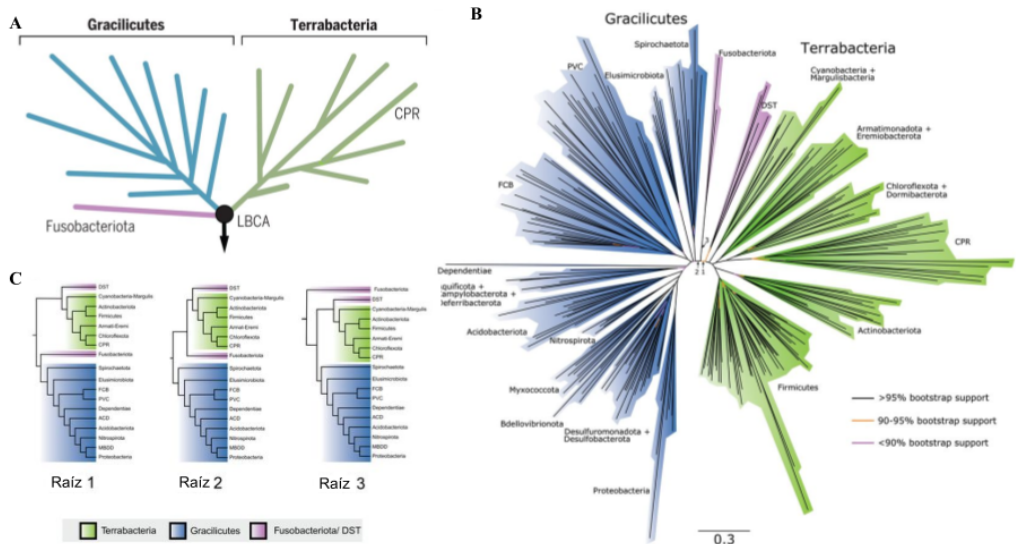


Figura 12. Último Ancestro Común de Bacteria. A. Esquema representando el árbol con raíz de Bacteria. B. Árbol construido sin grupo externo C. Se muestran las 3 raíces posibles del árbol. Recuperado y Modificado de Coleman et al (2021) [8].

En los datos reportados por Coleman y colaboradores [8] hay tres posibles raíces (**Figura 12C**) pero en todas ellas está involucrado el grupo Fusobacteria. Se procedió a verificar la presencia de la bac-glyRS en Fusobacteria, Spirochaeta y el grupo DST (Deinococcus-Thermus, Synergistetes, Thermotogae). En Fusobacteria todos los organismos cuentan con bac-glyRS, mientras que en los organismos Spirochaeta y DST se encuentran una u otra dependiendo de la especie. En la **Tabla 8** se puede ver un conteo de los organismos en nuestro muestreo.

Tabla 8. Presencia de bac-glyRS en grupos cercanos a la raíz del LBCA

DST	Deinococcus-Thermus	3/6	21/24
	Synergistetes	9/9	
	Thermotogae	9/9	
Spirochaeta		3/7	
Fusobacteria		8/8	

La presencia absoluta de la bac-glyRS en Fusobacteria es un indicador de la antigüedad de esta enzima. Es posible que comparar la distribución de la eu-glyRS pueda esclarecer cuál de estas dos versiones llevaba a cabo la esterificación de la glicina en el Último Ancestro Común Universal (LUCA por sus siglas en inglés).

Dos versiones para la misma función

Dentro de los alcances de los resultados se puede decir que se han dado menos reemplazos no homólogos bac-glyRS→eu-glyRS que en el sentido inverso. Esto en conjunto con la uniformidad en ciertos grupos indicaba a la existencia de un elemento de presión selectiva. Existe una mutación en el nucleótido 73, posición conocida como el discriminador, en el gly-tRNA. Se exploró la posibilidad de que existiera una relación entre el tipo de glyRS con la de esta mutación. En estudios preliminares no se pudo establecer dicha relación. Además que el modo de interacción del tRNA y el nucleótido 73 no se conoce actualmente [19].

Similitudes estructurales de la bac-glyRS con otras proteínas

Entre estos resultados de comparaciones estructurales cabe destacar los siguientes:

1. Enzima de adición de CCA

Como se muestra en la Figura 10, la superposición sólo se da la región de los dominios 1 y 2 de la subunidad β de la bac-glyRS (aminoácidos 1-170) y la región terminal de la enzima de adición de CCA (aminoácidos 270-410). Dicho segmento tiene un papel en el reconocimiento del brazo T Ψ C en el tRNA como se muestra en la **Figura 13** [47]. El alineamiento parece ser congruente conforme a lo predicho por HHPRED aunque por métodos de secuencia no se encontrara dicha coincidencia.

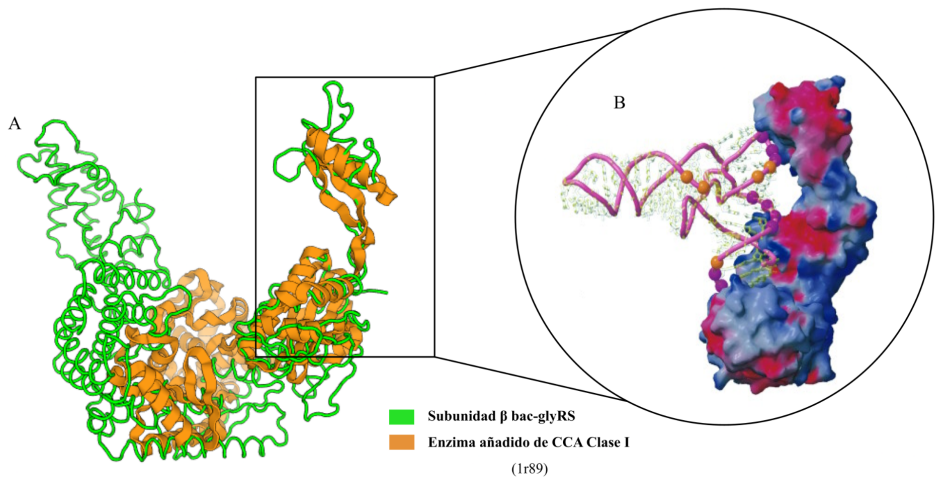


Figura 13. Relación Estructural con la enzima de Adición de CCA. A. Alineación estructural resultado de DALI SEARCH. Se da en la región inicial de la subunidad β de la bac-glyRS (en verde) y región final de la Enzima de adición de CCA Clase I (en naranja, PDB: 1r89). B. La región alineada tiene un papel en el reconocimiento del brazo TΨC en el tRNA. Modelo propuesto y recuperado de Xiong, Y (2003) [49].

En los tres dominios de la vida, los tRNAs tienen la secuencia CCA universalmente conservada en el extremo 3'. El CCA terminal es un prerrequisito indispensable para la aminoacilación llevada a cabo por las aaRSs. De acuerdo con bases de datos de tRNAs sólo entre el 60% y 70% de los tRNAs bacterianos cuentan con el CCA codificado en su secuencia [38]. En otros organismos, incluyendo eucariotas y ciertas especies de arqueas el CCA no está codificado por lo que debe ser añadido después de la transcripción.

La adición del CCA se lleva a cabo por enzimas con actividad ATP(CTP): tRNA nucleotidil transferasa, también conocida como enzima de adición de CCA. Estas enzimas son una RNA polimerasa especializada que reconoce todos los tRNAs disponibles. Son considerados miembros de la Superfamilia de las Polimerasas β . Comparaciones entre las enzimas de adición del CCAs llevaron a la clasificación de éstas en Clase I y II, las cuales están relacionadas sólo de manera lejana [38].

En ciertas bacterias la adición del CCA es realizada por dos enzimas distintas. Una enzima cataliza la adición de dos citocinas mientras que otra añade la adenosina. Éstas son conocidas como enzima de adición del CC y enzima de adición de A. Al igual que la enzima de adición del CCA de Eubacteria y Eucariontes, éstas pertenecen a la Clase II de la Superfamilia de polimerasas β . En las enzimas de adición del CC un evento de delección causó que se perdiera la capacidad de cambiar la especificidad del nucleótido, por ello sólo unen las citosinas al tRNA. Mientras que la enzima de adición de adenosina parece haber surgido de un ancestro que comenzó a ser selectivo por tRNAs que ya tuvieran las dos citocinas unidas [45].

Clase I

- Enzima de adición del CCA en arqueas
- Poly(A) polimerasas en Eucariontes
- DNA polimerasa β
- Desoxinucleotidil transferasa terminal (TdT)
- Nucleotidil Transferasa de kanamicina
- Adeniltransferasa de Estreptomycin

Clase II

- Enzima de adición del CCA en Eubacteria y Eucariontes
- Enzimas de adición de CC- y A- en Eubacteria
- poly(A) polimerasas en Eubacteria.

La similitud estructural de la bac-glyRS es sólo con las de Clase I. Aunque ambas Clases tienen parecido estructural éstas difieren en el dominio reconocimiento del brazo T Ψ C en el tRNA donde se da la sobreposición con los dominios 1 y 2 de la β -bac-glyRS. Como se muestra en la **Figura 14** ambas clases de Enzimas de Adición tienen un parecido estructural pero difieren en su zona N-terminal.

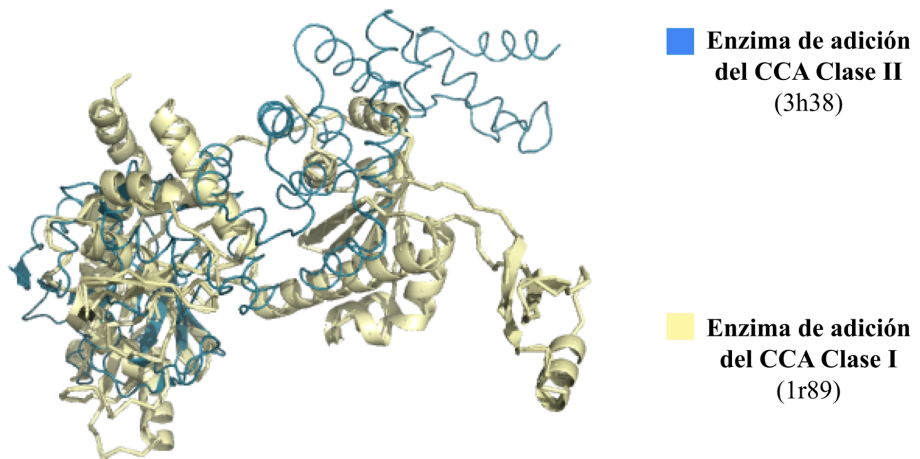


Figura 14. Las dos Clases de Enzimas de adición del CCA difieren en el dominio de reconocimiento del brazo TΨC en el tRNA. Sobreposición de un ejemplar de cada Clase de Enzima de adición de CCA. Clase I presente en arqueas (en amarillo, PDB: 1r89) y la Clase II presente en Eubacteria y Eucariontes (en azul, PDB: 3h38). En contraste éstas tienen un eje de rotación diferente entre dominio 1 y 2. Y difieren en el dominio de reconocimiento del brazo TΨC en el tRNA, el cuál para la Clase II se compone sólo de alfa hélices.

Esta misma región de las enzimas de adición Clase I había sido reportada similar a la proteína de procesamiento de rRNA Utp22 [31]. Utp 22 es un componente del pre-ribosoma 90S de la subunidad pequeña del procesosoma que realiza el procesamiento temprano del rRNA 18S. Éste a su vez es un componente de la subunidad pequeña (40S) del ribosoma eucarionte. La proteína Utp 22 no tiene actividad catalítica reportada.

2. Arginil tRNA sintetasa

El dominio 5 tiene similitud con la argRS en el dominio de unión al anticodón. Ésta es una aminoacil tRNA sintetasa de Clase I. Ya que este segmento interactúa con el tRNA es posible que en la bac-glyRs tuviese la misma función **Figura 15A**.

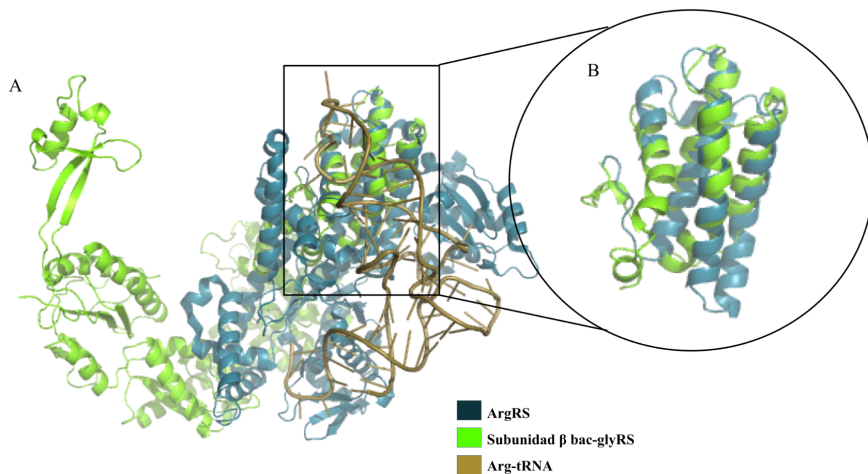


Figura 15. Alineación estructural ArgRS. A. Superposición de bac-glyRS (en verde) y ArgRS (en azul, PDB: 1f7u) unido al tRNA de arginina (en café). Como se puede observar la región de similitud es en la zona de unión al anticodón del tRNA. B. Superposición recortada alrededor de la zona de alineación entre ambas proteínas.

Otros resultados involucran el dominio 3 y 4. Estos muestran similitud estructural con enzimas fosfodiesterasas y miembros de la familia de las polimerasas.

Aunque se pudo establecer una relación estructural entre la glyRS heterotetramérica con otras proteínas, no es posible establecer una relación evolutiva entre ellas sin tener más datos sobre su parecido. Establecer un alineamiento a nivel de secuencia entre los fragmentos que mostraron una superposición sería importante para poder establecer si estas proteínas tuvieron ancestros en común.

La función de las proteínas encontradas parece recapitular las posibles funciones de la subunidad β . Aunque no se tiene la estructura del complejo bac-glyRS - gly-tRNA se espera que el extremo anticodón del tRNA repose en el dominio 5 de la β -bac-glyRS que tiene similitud con el dominio anticodón de la ArgRS. A su vez, el dominio 1 y 2 podrían interactuar con el brazo T Ψ C, el mismo papel que cumple dicho segmento en la enzima de adición del CCA. Determinar si estas regiones tienen similitud estructural debido a una convergencia evolutiva está más allá de las metodologías usadas en esta tesis.

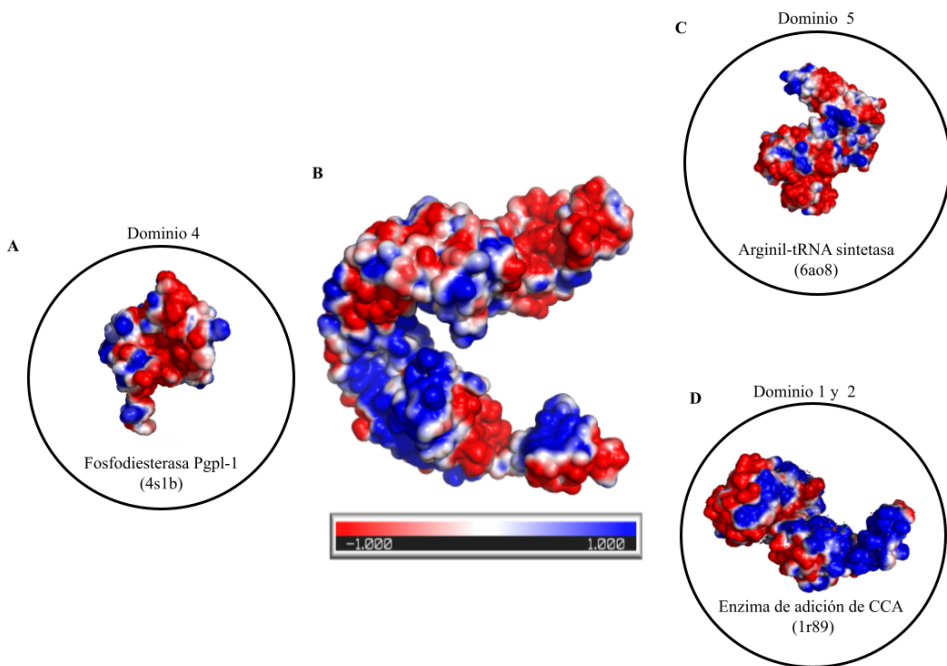


Figura 16. La superficie del potencial electrostático sugiere posicionamiento del gly-tRNA. Se muestran con un gradiente de colores de rojo (negativo, -1 kT/e) a azul (positivo, +1 kT/e) la superficie electrostática de las proteínas con similitud estructural a la β -bac-glyRS y la β -bac-glyRS (B, al centro). A. Para el dominio 4 se muestra la Fosfodiesterasa Pgp1-1 (PDB: 4s1b). C. Con similitud al dominio 5 la arginil-tRNA sintetasa (PDB:6ao8) y D. el Dominio 1 y 2 la Enzima de adición del CCA.

Barajeo de dominios y permutación de secuencias.

El estudio de la subunidad β pone sobre la mesa varias teorías de origen de proteínas. La teoría de barajeo de dominios propone que tras un proceso de recombinación los dominios de una proteína podrían reorganizarse y con esta nueva topología la proteína podría cambiar de función.

La mosaicidad en la bac-glyRS y sus candidatos homólogos parecen crear una narrativa de las funciones en esta proteína. Mientras que la subunidad α muestra similitud con la alaRS y realiza la catálisis, la subunidad β se alinea estructuralmente con proteínas que unen al tRNA en la región y orientación que se espera la interacción del gly-tRNA y la bac-glyRS. Como antes mencionado, las coincidencias podrían explicarse por convergencia evolutiva, al tener la misma función se podría haber llegado a la misma respuesta estructural pero

también podría explicarse mediante una relación evolutiva entre las proteínas de nuestros resultados y la bac-glyRS.

Es posible que algoritmos como HMMER o psiBLAST no encuentren a las enzimas de adición del CCA debido a la longitud de los segmentos y diferencias de topología. Específicamente el segmento que alinea entre las enzimas de adición de CCA y la bac-glyRS están en los extremos contrarios de estas proteínas. Este tipo de fenómeno ha sido observado en proteínas que sufrieron permutación en su secuencia [32].

La mejor forma de establecer si eventos de barajeo de dominios o permutación han sucedido es mediante el análisis de otros ejemplos dentro de la misma familia de enzimas. Ésta es una limitante ya que la subunidad α no tiene homólogos reportados hasta ahora. Varias coincidencias encontradas con DALI SEARCH fueron enzimas pertenecientes a la superfamilia de polimerasas β . Si se llegara a establecer una relación a nivel de secuencia entre los dominios similares descritos anteriormente podríamos entender como eventos ancestrales dieron origen a varias proteínas con plegamientos similares.

Conclusiones

- Por la distribución de la bac-glyRS podemos saber que es una proteína antigua. Pero no podemos definir cuál de las dos versiones de la glyRS estaba presente en LUCA.
- El modelo de LGT fue útil para identificar que los eventos de pérdida de la bac-glyRS fueron más comunes que las ganancias de la bac-glyRS.
- Las comparaciones a nivel estructural proveen información valiosa sobre la posible función de dominios de la bac-glyRS, aunque por sí solas no son suficientes para establecer una hipótesis de homología.
- Una vez que se tenga más información sobre la interacción del tRNA con la bac-glyRS se podría confirmar si las similitudes estructurales también son de función.

Perspectivas

- Aplicar esta metodología en el árbol de LBCA sería útil para descartar que la presencia en *Fusobacteria* fuese por un evento de transferencia horizontal.
- Buscar una similitud a nivel de secuencia de las proteínas encontradas por estructura es necesario para soportar o descartar las hipótesis de homología.
- Una vez que se cuente con la estructura del complejo gly-tRNA - bac-glyRS se podrían comparar las formas de interacción de éste y los dominios encontrados con sus respectivos tRNAs.
- Recurrir a bases de datos con dominios estructurales para incluir en el estudio a proteínas que pertenecen a la misma familia que las que encontramos por comparación de estructura.
- Hacer un árbol filogenético incluyendo a los posibles homólogos encontrados por estructura.

Referencias

1. Abibi, A., Ferguson, A. D., Fleming, P. R., Gao, N., Hajec, L. I., Hu, J., Buurman, E. T. (2014). The role of a novel auxiliary pocket in bacterial phenylalanyl-tRNA synthetase druggability. *Journal of Biological Chemistry*, 289(31), 21651–21662. doi:10.1074/jbc.m114.574061
2. Alva V, Söding J, Lupas AN. A vocabulary of ancient peptides at the origin of folded proteins. *Elife*. 2015 Dec 14;4:e09410. doi: 10.7554/eLife.09410. PMID: 26653858; PMCID: PMC4739770.
3. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol*. 1990 Oct 5;215(3):403-10. doi: 10.1016/S0022-2836(05)80360-2. PMID: 2231712.
4. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research*. 1997 Sep 1;25(17):3389-402. doi: 10.1093/nar/25.17.3389. PMID: 9254694; PMCID: PMC146917.
5. Aravind, L., Koonin, E. V. (1998). The HD domain defines a new superfamily of metal-dependent phosphohydrolases. *Trends in Biochemical Sciences*, 23(12), 469–472. doi:10.1016/s0968-0004(98)01293-6
6. Betat H, Mede T, Tretbar S, Steiner L, Stadler PF, Mörl M, Prohaska SJ. The ancestor of modern Holozoa acquired the CCA-adding enzyme from Alphaproteobacteria by horizontal gene transfer. *Nucleic Acids Research*. 2015 Aug 18;43(14):6739-46. doi: 10.1093/nar/gkv631. Epub 2015 Jun 27. PMID: 26117543; PMCID: PMC4538823.
7. Castelle, Cindy J.; Banfield, Jillian F. (2018). Major new microbial groups expand diversity and alter our understanding of the Tree of Life. *Cell*, 172(6), 1181–1197. doi:10.1016/j.cell.2018.02.016
8. Coleman GA, Davin AA, Mahendrarajah TA, Szánthó LL, Spang A, Hugenholtz P, Szöllösi GJ, Williams TA. A rooted phylogeny resolves early bacterial evolution. *Science*. 2021 May 7;372(6542):eabe0511. doi: 10.1126/science.abe0511. PMID: 33958449.
9. Cusack, S., Berthet-Colominas, C., Härtlein, M., Nassar, N., Leberman, R. (1990). A second class of synthetase structure revealed by X-ray analysis of *Escherichia coli* seryl-tRNA synthetase at 2.5 Å. *Nature*, 347(6290), 249–255. <https://doi.org/10.1038/347249a0>
10. Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, 32(5), 1792–1797. doi:10.1093/nar/gkh340
11. Erber L, Betat H, Mörl M. CCA-addition gone wild: unusual occurrence and phylogeny of four different tRNA nucleotidyl transferases in *Acanthamoeba castellanii*. *Molecular Biology and Evolution*. 2021 Mar 9;38(3):1006-1017. doi: 10.1093/molbev/msaa270. PMID: 33095240; PMCID: PMC7947759.

12. Erber L, Franz P, Betat H, Prohaska S, Mörl M. Divergent evolution of eukaryotic CC- and A-Adding Enzymes. *Int J Mol Sci.* 2020 Jan 10;21(2):462. doi: 10.3390/ijms21020462. PMID: 31936900; PMCID: PMC7014341.
13. Fournier, G. P., Andam, C. P., and Gogarten, J. P. (2015) Ancient horizontal gene transfer and the last common ancestors. *BMC Evolutionary Biology.* 15, 70
14. Ernst, F. G., Rickert, C., Bluschke, A., Betat, H., Steinhoff, H. J., Mörl, M. (2015). Domain movements during CCA-addition: a new function for motif C in the catalytic core of the human tRNA nucleotidyltransferases. *RNA biology,* 12(4), 435–446. <https://doi.org/10.1080/15476286.2015.1018502>
15. Fournier GP, Alm EJ. Ancestral reconstruction of a pre-LUCA aminoacyl-tRNA synthetase ancestor supports the late addition of Trp to the genetic code. *Journal of Molecular Evolution.* 2015 Apr;80(3-4):171-85. doi: 10.1007/s00239-015-9672-1. Epub 2015 Mar 20. PMID: 25791872.
16. Frickey T., Lupas AN. (2004) CLANS: a Java application for visualizing protein families based on pairwise similarity. *Bioinformatics* 20(18):3702-3704. PMID: 15284097
17. Gabler F, Nam SZ, Till S, Mirdita M, Steinegger M, Söding J, Lupas AN, Alva V. (2020) Protein sequence analysis using the MPI bioinformatics toolkit. *Current Protocols of Bioinformatics.* 2020 Dec;72(1):e108. doi: 10.1002/cpbi.108.
18. Gomez, M.A.R. and Ibba, M. (2020) Aminoacyl-tRNA synthetases. *RNA.* 26, 910-936. <https://doi.org/10.1261/rna.071720.119>
19. Giegé R., Eriani G. (2014) Transfer RNA recognition and aminoacylation by synthetases. In: eLS. John Wiley & Sons, Ltd: Chichester. 10.1002/9780470015902.a0000531.pub3
20. Guindon S., Dufayard J.F., Lefort V., Anisimova M., Hordijk W., Gascuel O. (2010) New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Systematic Biology,* 59(3):307-21.
21. Hildebrand, A., Remmert, M., Biegert, A., Söding, J. (2009). Fast and accurate automatic structure prediction with HHpred. *Proteins: Structure, Function, and Bioinformatics,* 77(S9), 128–132. doi:10.1002/prot.22499
22. Holm, L., Kääriäinen, S., Wilton, C., Plewczynski, D. (2006). Using Dali for structural comparison of proteins. *Current Protocols in Bioinformatics.* doi:10.1002/0471250953.bi0505s14
23. Holm, L. (2019). DALI and the persistence of protein shape. *Protein Science.* doi:10.1002/pro.3749
24. Hug, L.A., Castelle, C.J., Wrighton, K.C. et al. Community genomic analyses constrain the distribution of metabolic traits across the *Chloroflexi* phylum and indicate roles in sediment carbon cycling. *Microbiome* 1, 22 (2013). <https://doi.org/10.1186/2049-2618-1-22>
25. Hug, L. A., Baker, B. J., Anantharaman, K., Brown, C. T., Probst, A. J., Castelle, C. J., Banfield, J. F. (2016). A new view of the tree of life. *Nature Microbiology,* 1(5). doi:10.1038/nmicrobiol.2016.48
26. Ibba M., Söll D. (2000) Aminoacyl-tRNA synthesis. *Annual Reviews of Biochemistry.* 69:617-50

27. Khafif, M., Cottret, L., Balagué, C. et al. Identification and phylogenetic analyses of VAS_t, an uncharacterized protein domain associated with lipid-binding domains in Eukaryotes. *BMC Bioinformatics* 15, 222 (2014). <https://doi.org/10.1186/1471-2105-15-222>
28. Koonin, E. V., Novozhilov, A. S. (2009). Origin and evolution of the genetic code: the universal enigma. *IUBMB life*, 61(2), 99–111. <https://doi.org/10.1002/iub.146>
29. Letunic, I., Bork, P. (2011). Interactive Tree Of Life v2: online annotation and display of phylogenetic trees made easy. *Nucleic Acids Research*, 39(suppl), W475–W478. doi:10.1093/nar/gkr201
30. Letunic, I., Bork, P. (2016). Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Research*, 44(W1), W242–W245. <https://doi.org/10.1093/nar/gkw290>
31. Lin, J., Lu, J., Feng, Y., Sun, M., Ye, K. (2013). An RNA-binding complex involved in ribosome biogenesis contains a protein with homology to tRNA CCA-adding enzyme. *PLoS Biology*, 11(10), e1001669. doi:10.1371/journal.pbio.10016
32. Lupas A N, Ponting C P, Russell R B (2001). On the evolution of protein folds: are similar motifs in different protein folds the result of convergence, insertion, or relics of an ancient peptide world? . *J Mol Biol*, 314(2-3), 0–203. doi:10.1006/jsbi.2001.4393
33. Neuenfeldt A, Just A, Betat H, Mörl M. Evolution of tRNA nucleotidyltransferases: a small deletion generated CC-adding enzymes. *Proc Natl Acad Sci U S A*. 2008 Jun 10;105(23):7953-8. doi: 10.1073/pnas.0801971105. Epub 2008 Jun 3. PMID: 18523015; PMCID: PMC2430343.
34. Pak, D., Kim, Y., Burton, Z. F. (2018). Aminoacyl-tRNA synthetase evolution and sectoring of the genetic code. *Transcription*, 1–15. doi:10.1080/21541264.2018.1467
35. Ribas de Pouplana, L., Schimmel, P. (2001). Two classes of tRNA synthetases suggested by sterically compatible dockings on tRNA acceptor stem. *Cell*, 104(2), 191–193. doi:10.1016/s0092-8674(01)00204-5
36. Ribas de Pouplana, L. (2020). The evolution of aminoacyl-tRNA synthetases: From dawn to LUCA. *The Enzymes*. doi:10.1016/bs.enz.2020.08.001
37. S R Eddy, Profile hidden markov models., *Bioinformatics*, Volume 14, Issue 9, 1998, Pages 755–763, <https://doi.org/10.1093/bioinformatics/14.9.755>
38. Sprinzl M, Cramer F. The C-C-A end of tRNA and its role in protein biosynthesis. *Progress in Nucleic Acid Research and Molecular Biology*. 1979;22:1-69. doi: 10.1016/s0079-6603(08)60798-9. PMID: 392600.
39. Steinegger, M., Meier, M., Mirdita, M. et al. HH-suite3 for fast remote homology detection and deep protein annotation. *BMC Bioinformatics* 20, 473 (2019). <https://doi.org/10.1186/s12859-019-3019-7>
40. Stern, A., Mayrose, I., Penn, O., Shaul, S., Gophna, U., Pupko, T. (2010). An evolutionary analysis of lateral gene transfer in thymidylate synthase enzymes. *Systematic biology*, 59(2), 212–225. <https://doi.org/10.1093/sysbio/syp104>
41. Terada, T., Nureki, O., Ishitani, R. et al. Functional convergence of two lysyl-tRNA synthetases with unrelated topologies. *Nature Structural Molecular Biology* 9, 257–262 (2002). <https://doi.org/10.1038/nsb777>
42. Tiessen, A., Pérez-Rodríguez, P. Delaye-Arredondo, L.J. Mathematical modeling and comparison of protein size distribution in different plant, animal, fungal and microbial

- species reveals a negative correlation between protein size and protein number, thus providing insight into the evolution of proteomes. *BMC Res Notes* 5, 85 (2012). <https://doi.org/10.1186/1756-0500-5-85>
43. Okabe, M., Tomita, K., Ishitani, R., Ishii, R., Takeuchi, N., Arisaka, F., Nureki, O., Yokoyama, S. (2003). Divergent evolutions of trinucleotide polymerization revealed by an archaeal CCA-adding enzyme structure. *The EMBO journal*, 22(21), 5918–5927. <https://doi.org/10.1093/emboj/cdg563>
 44. Pereira J, Lupas A N. The ancestral KH peptide at the root of a domain family with three different folds, *Bioinformatics*, Volume 34, Issue 23, 01 December 2018, Pages 3961–3965, <https://doi.org/10.1093/bioinformatics/bty480>
 45. Smith, T. F., Hartman, H. (2015). The evolution of Class II Aminoacyl-tRNA synthetases and the first code. *FEBS Letters*, 589(23), 3499–3507. doi:10.1016/j.febslet.2015.10.006
 46. Valencia-Sánchez, M. I., Rodríguez-Hernández, A., Ferreira, R., Santamaría-Suárez, H. A., Arciniega, M., Dock-Bregeon, A.-C., Torres-Larios, A. (2016). Structural insights into the polyphyletic origins of glycyl tRNA synthetases. *Journal of Biological Chemistry*, 291(28), 14430–14446. doi:10.1074/jbc.m116.730382
 47. Xiong, Y., Li, F., Wang, J., Weiner, A. M., & Steitz, T. A. (2003). Crystal structures of an archaeal Class I CCA-Adding Enzyme and its nucleotide complexes. *Molecular Cell*, 12(5), 1165–1172. doi:10.1016/s1097-2765(03)00440-4
 48. Wang S-W, Bitbol A-F, Wingreen NS (2019) Revealing evolutionary constraints on proteins through sequence analysis. *PLoS Comput Biol* 15(4): e1007010. <https://doi.org/10.1371/journal.pcbi.1007010>
 49. Zimmermann L, Stephens A, Nam SZ, Rau D, Kübler J, Lozajic M, Gabler F, Söding J, Lupas AN, Alva V. J (2018) A completely reimplemented MPI bioinformatics toolkit with a new HHpred server at its core. *Mol Biol.* 2018 Jul 20. S0022-2836(17)30587-9.

Información Suplementaria

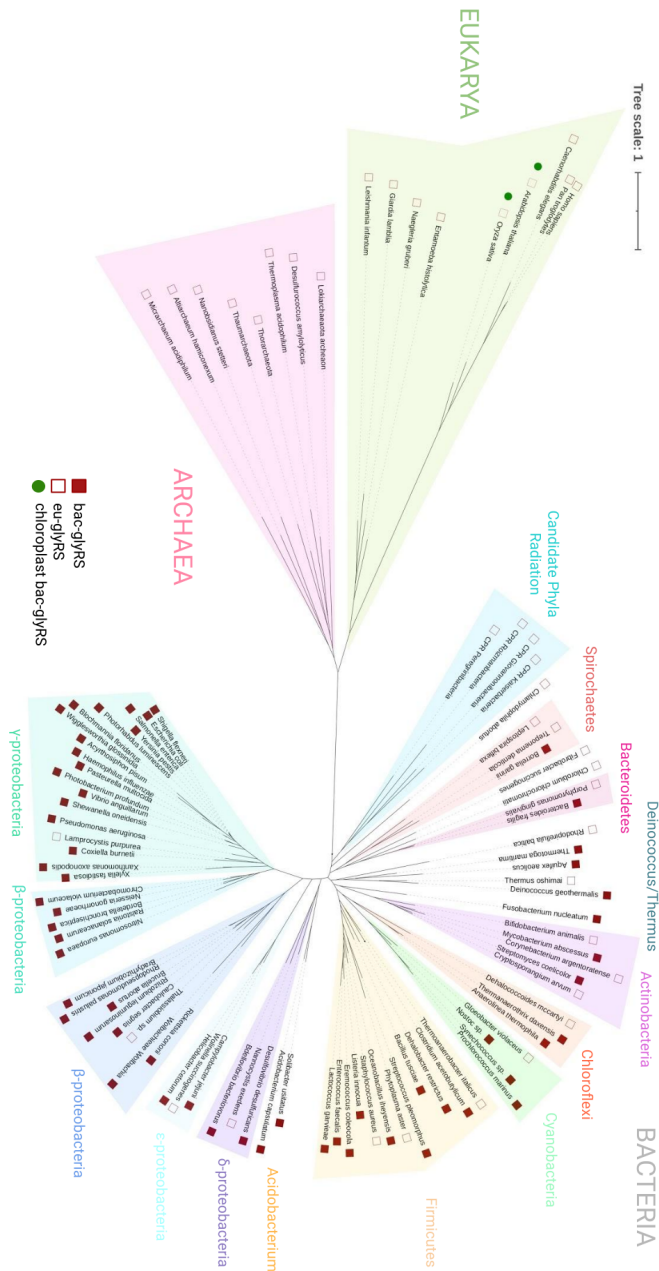


Figura Suplementaria 1. Presencia de las glicil-tRNA sintetasas en el Árbol de la Vida. 1872 nodos.

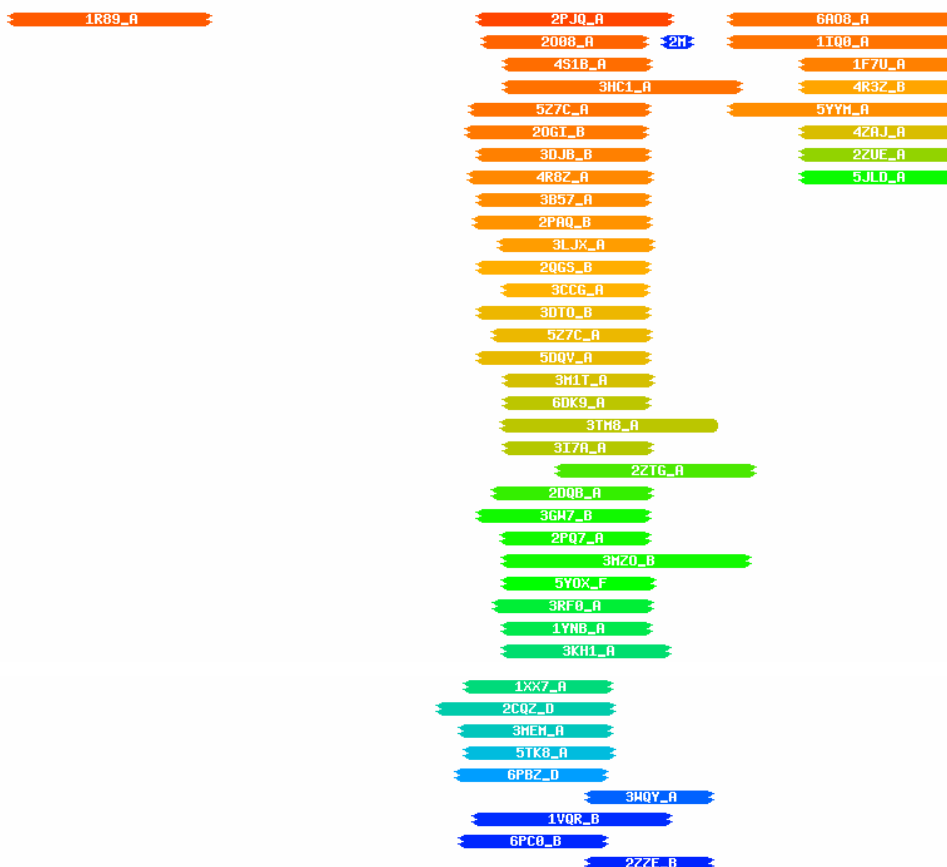


Figura Suplementaria 2. Resultados de HHPRED. Diagrama de coincidencias encontradas con HHPRED. Se muestran los códigos PDB y el indicador de cadena. La barra está colocada dependiendo de la región que encuentra en la subunidad β de la glicil-tRNA sintetasa de N-terminal a C-terminal.

Tabla Suplementaria 1. Análisis de Transferencia Horizontal de Genes. Eventos de LGT se ilustran con la probabilidad de ganancia de una versión de la enzima y pérdida de la otra versión. Para el reemplazo de eu-glyRS→bac-glyRS sólo se registró el caso de el nodo interno N6 del árbol. Para los casos significativos de bac-glyRS→eu-glyR se consideraron 19 casos de pérdida de bac-glyRS, para 11 de estos la ganancia de eu-glyRS también fue mayor a 0.5.

	Ganancia bac-glyRS	Pérdida eu-glyRS
N6	0.944	0.946
	Ganancia eu-glyRS	Pérdida bac-glyRS

Thermoanaerobacter italicus	0.761	0.993
Phytoplasma aster		0.743
Staphylococcus aureus	0.856	1
Dehalococcoides mccartyi	0.866	0.994
Gloeobacter violaceus	0.805	0.992
N33	0.805	0.519
Mycobacterium abscessus		0.893
Corynebacterium argenteratense		0.59
Thermus oshimai	0.671	0.966
Leptospira biflexa	0.511	0.755
Treponema denticola	0.524	0.748
Borrelia garinii		0.544
Porphyromonas gingivalis		0.739
Bacteroides fragilis		0.656
Chlorobium chlorochromatii		0.634
Fibrobacter succinogenes		0.721
Rhodopirellula baltica	0.672	0.919
Wolinella succinogenes	0.603	0.985
Thalassiosira sp	0.859	0.999

Tabla Suplementaria 2. Resultados HHPRED. En la tabla se muestran los primeros 30 resultados de HHPRED en respuesta a la secuencia de la subunidad β .

Código PDB y Cadena	Descripción	Prob	Valor E	Valor P	Score	SS	Cols	Query HMM	Template HMM
2PJQ_A	Uncharacterized protein	94.2	0.69	1.3E-05	46.9	12.2	134	342-480	12-155 (231)
1R89_A	tRNA nucleotidyltransferase	92.4	1.6	2.9E-05	49.9	12.6	127	6-149	279-405 (437)
2O08_A	BH1327 protein; PUTATIV	91.6	2	3.7E-05	41.7	10.6	112	346-462	4-131 (188)
4S1B_A	Lmo1466 protein; c-di-A	90.8	2.9	5.4E-05	41.9	11.3	105	361-465	18-161 (222)

6A08_A	Arginine--tRNA ligase	90.6	1.6	3E-05	51.3	10.5	158	523-682	397-571 (580)
3HC1_A	uncharacterized HDOD do	90.5	8.1	0.00015	40.8	15.0	159	361-530	119-303 (305)
5Z7C_A	3'3'-cGAMP-specific pho	90.4	2.8	5.3E-05	47.6	12.0	127	337-464	17-208 (463)
1IQ0_A	ARGINYL-TRNA SYNTHETASE	90.3	2	3.7E-05	50.6	10.9	157	523-682	405-583 (592)
2OGI_B	Hypothetical protein SA	89.9	5.5	0.0001	38.9	12.1	120	334-462	4-139 (196)
1F7U_A	ARGINYL-TRNA SYNTHETASE	89.2	1.6	3E-05	51.7	9.1	109	574-682	482-598 (607)
3DJB_B	Hydrolase, HD family; a	89.2	5.6	0.00011	39.7	11.9	117	342-464	7-127 (223)
4R8Z_A	Cyclic di-GMP phosphodi	88.5	4.8	9E-05	39.2	10.7	130	336-466	3-167 (218)
3B57_A	Lin1889 protein; Lin188	88.2	8.9	0.00017	37.7	12.4	118	342-464	7-128 (209)
5YYM_A	Arginine--tRNA ligase (88.1	4	7.5E-05	48.0	11.4	158	523-682	399-577 (586)
2PAQ_B	5'-deoxynucleotidase Yf	87.6	8.8	0.00017	37.8	12.0	119	339-465	2-145 (201)
3LJX_A	MmoQ Response regulator	86.8	5.7	0.00011	41.5	10.6	110	357-467	105-235 (288)
4R3Z_B	Aminoacyl tRNA synthase	86.1	2.1	4E-05	51.6	7.8	109	574-682	542-666 (675)
2QGS_B	Protein SE1688; alpha-h	85.3	12	0.00022	37.6	11.7	117	342-464	7-128 (225)
3CCG_A	HD superfamily hydrolas	85.0	11	0.00021	35.9	10.9	99	360-463	19-134 (190)
3DTO_B	BH2835 protein; all alp	84.0	17	0.00031	36.4	12.0	117	342-464	7-127 (223)
5Z7C_A	3'3'-cGAMP-specific pho	83.9	9.3	0.00017	43.4	11.2	113	353-465	286-417 (463)
5DQV_A	Uncharacterized protein	83.7	19	0.00035	35.3	12.1	118	342-464	13-135 (211)

4ZAJ_A	Arginine--tRNA ligase,	82.8	3.5	6.5E-05	48.7	7.5	109	574-682	455-579 (601)
3M1T_A	Putative phosphohydrola	82.5	16	0.0003	37.5	11.5	101	361-467	106-230 (275)
6DK9_A	DNA damage-inducible pr	81.1	12	0.00023	38.0	9.9	98	361-464	60-170 (234)
3TM8_A	Uncharacterized protein	81.0	37	0.0007	36.2	14.0	143	359-514	165-328 (328)
317A_A	Putative metal-dependen	80.6	21	0.0004	36.8	11.7	100	361-466	118-240 (281)
2ZUE_A	Arginyl-tRNA synthetase	78.6	9.7	0.00018	45.2	9.3	108	574-682	500-620 (629)
2ZTG_A	Alanyl-tRNA synthetase	74.4	32	0.00061	42.2	12.4	128	399-540	285-425 (739)
2DQB_A	Deoxyguanosinetrip hosph	73.1	45	0.00084	37.0	12.2	109	353-466	64-203 (376)

Tabla Suplementaria 3. Primeras 50 proteínas con similitud estructural reportadas por DALI para la subunidad β de la bac-glyRS.

Código PDB	Descripción	Puntuación Z	RMSD	# aa
4s1b-D	LMO1466 PROTEIN;	11.5	2.9	143
2pq7-A	PREDICTED HD SUPERFAMILY HYDROLASE;	10.7	3.1	132
3mem-A	PUTATIVE SIGNAL TRANSDUCTION PROTEIN;	10.4	7.1	172
3hc1-A	UNCHARACTERIZED HDOD DOMAIN PROTEIN;	10.2	3.1	149
4mcw-B	METAL DEPENDENT PHOSPHOHYDROLASE;	9.8	20.7	147
3nqw-A	CG11900;	9.7	3.3	141
3tmb-B	UNCHARACTERIZED PROTEIN BD1817;	9.6	13.0	143
1iq0-A	ARGINYL-TRNA SYNTHETASE;	9.3	6.2	154
3wfr-H	RNA (74-MER);	9.3	20.7	167

3mzo-B	LIN2634 PROTEIN;	9.2	3.6	146
5tk6-A	OXSA PROTEIN;	9.1	3.6	140
3i7a-A	PUTATIVE METAL-DEPENDENT PHOSPHOHYDROLASE;	9.0	3.7	153
6pc0-B	GUANOSINE PENTAPHOSPHATE PHOSPHOHYDROLASE;	8.9	17.3	157
1u6z-B	EXOPOLYPHOSPHATASE;	8.9	11.3	156
2gz4-A	HYPOTHETICAL PROTEIN ATU1052;	8.8	3.6	130
5z7c-A	3'3'-CGAMP-SPECIFIC PHOSPHODIESTERASE 3;	8.8	3.2	135
5yij-A	SDEA;	8.6	6.6	181
3hi0-A	PUTATIVE EXOPOLYPHOSPHATASE;	8.5	18.1	160
4mlm-A	PREDICTED HD PHOSPHOHYDROLASE PHNZ;	8.4	3.3	116
6dk9-F	DNA DAMAGE-INDUCIBLE PROTEIN;	8.3	3.3	128
3p3q-B	MMOQ;	8.3	3.9	133
7ppo-A	UBIQUITINATING/DEUBIQUITINATING ENZYME SDEA;	8.0	9.4	170
2ogi-B	HYPOTHETICAL PROTEIN SAG1661;	7.9	3.7	123
1ynb-A	HYPOTHETICAL PROTEIN AF1432;	7.9	3.8	132
6oiw-A	DEOXYGUANOSINETRIPHOSPHATE TRIPHOSPHOHYDROLASE;	7.9	4.3	136
3fnr-A	ARGINYL-TRNA SYNTHETASE;	7.9	7.3	141
3gw7-A	UNCHARACTERIZED PROTEIN YEDJ;	7.8	3.8	128
4l1j-B	HD DOMAIN-CONTAINING PROTEIN 2;	7.8	4.0	134
3kh1-A	PREDICTED METAL-DEPENDENT PHOSPHOHYDROLASE;	7.7	6.2	140
2o6i-B	HD DOMAIN PROTEIN;	7.6	9.4	149
2dqb-C	DEOXYGUANOSINETRIPHOSPHATE TRIPHOSPHOHYDROLASE, P	7.5	4.8	140
1vqr-A	HYPOTHETICAL PROTEIN CJ0248;	7.4	3.2	145
4wzi-B	CAMP-SPECIFIC 3',5'-CYCLIC PHOSPHODIESTERASE 4B;	7.3	10.2	173

6pwy-B	ZK177.8;	7.2	7.7	138
4yfl-A	LMO0812 PROTEIN;	7.1	3.6	128
6brh-B	DEOXYNUCLEOSIDE TRIPHOSPHATE TRIPHOSPHOHYDROLASE	6.9	4.3	135
6cpu-A	PHOSPHODIESTERASE;	6.9	11.2	170
1gax-A	TRNA(VAL);	6.9	6.6	169
5bq9-B	UNCHARACTERIZED PROTEIN;	6.8	14.1	141
5yox-A	HD DOMAIN-CONTAINING PROTEIN YGL101W;	6.8	3.4	126
2bnl-C	MODULATOR PROTEIN RSBP;	6.6	2.9	103
4x4r-A	CCA-ADDING ENZYME;	6.5	4.7	142
3wqy-A	ALANINE--TRNA LIGASE;	6.5	17.4	180
2huo-A	INOSITOL OXYGENASE;	6.3	3.0	116
6yxa-A	GTP PYROPHOSPHOKINASE;	6.2	5.0	162
6c66-G	CRISPR-ASSOCIATED HELICASE, CAS3 FAMILY;	6.1	5.7	137
1rqg-A	METHIONYL-TRNA SYNTHETASE;	6.0	8.5	170
6ifr-A	TYPE III-A CRISPR-ASSOCIATED PROTEIN CSM1;	6.0	17.1	130
2a6o-A	5'-D(*CP*CP*CP*CP*TP*AP*GP*CP*TP*T P*TP*AP*GP*CP*T	5.8	3.4	90
4q2d-A	ULE: CRISPR-ASSOCIATED HELICASE CAS3;	5.8	5.5	165

Tabla Suplementaria 4. Primeras 20 proteínas con similitud estructural reportadas por DALI para el Dominio 1 de la subunidad β de la bac-glyRS

Código PDB	Descripción	Puntuación Z	RMSD	# aa
3ovs-B	MOLECULE: CCA-ADDING ENZYME;	7.3	3.3	97
3ouy-B	MOLECULE: CCA-ADDING ENZYME;	6.8	3.4	97
6zyd-B	MOLECULE: LOW CONDUCTANCE MECHANOSENSITIVE CHANNEL YNAI,LOW	6.7	4.2	84
6zyd-F	MOLECULE: LOW CONDUCTANCE MECHANOSENSITIVE CHANNEL	6.7	4.2	84

	YNAI,LOW			
6zyd-D	MOLECULE: LOW CONDUCTANCE MECHANOSENSITIVE CHANNEL YNAI,LOW	6.7	4.1	84
2f4f-B	MOLECULE: TRANSPOSASE, PUTATIVE;	6.7	3.0	86
6vxn-A	MOLECULE: MECHANOSENSITIVE ION CHANNEL PROTEIN 1, MITOCHOND	6.6	2.8	82
6vxn-A	MOLECULE: MECHANOSENSITIVE ION CHANNEL PROTEIN 1, MITOCHOND	6.6	2.9	82
1r8b-A	MOLECULE: TRNA NUCLEOTIDYLTRANSFERASE;	6.6	3.3	97
6vxn-C	MOLECULE: MECHANOSENSITIVE ION CHANNEL PROTEIN 1, MITOCHOND	6.6	2.8	82
6vxn-D	MOLECULE: MECHANOSENSITIVE ION CHANNEL PROTEIN 1, MITOCHOND	6.6	2.8	82
6vxn-C	MOLECULE: MECHANOSENSITIVE ION CHANNEL PROTEIN 1, MITOCHOND	6.6	2.8	82
6vxp-F	MOLECULE: MECHANOSENSITIVE ION CHANNEL PROTEIN 1, MITOCHOND	6.6	2.8	82
6vxn-G	MOLECULE: MECHANOSENSITIVE ION CHANNEL PROTEIN 1, MITOCHOND	6.6	2.8	82
6vxn-E	MOLECULE: MECHANOSENSITIVE ION CHANNEL PROTEIN 1, MITOCHOND	6.6	2.8	82
1tfw-B	MOLECULE: 5'-R(*GP*CP*GP*GP*AP*UP*CP*CP*GP* CP*AP*C)-3';	6.6	3.3	97
6vxn-F	MOLECULE: MECHANOSENSITIVE ION CHANNEL PROTEIN 1, MITOCHOND	6.6	2.8	82
6vxp-B	MOLECULE: MECHANOSENSITIVE ION CHANNEL PROTEIN 1, MITOCHOND	6.6	2.8	82
6vxn-F	MOLECULE: MECHANOSENSITIVE ION CHANNEL PROTEIN 1, MITOCHOND	6.6	2.8	82
3ovs-B	MOLECULE: CCA-ADDING ENZYME;	7.3	3.3	97
3ouy-B	MOLECULE: CCA-ADDING ENZYME;	6.8	3.4	97
6zyd-B	MOLECULE: LOW CONDUCTANCE MECHANOSENSITIVE CHANNEL YNAI,LOW	6.7	4.2	84

6zyd-F	MOLECULE: LOW CONDUCTANCE MECHANOSENSITIVE CHANNEL YNAI,LOW	6.7	4.2	84
--------	---	-----	-----	----

Tabla Suplementaria 5. Primeras 20 proteínas con similitud estructural reportadas por DALI para el Dominio 2 de la subunidad β de la bac-glyRS

Código PDB	Descripción	Puntuación Z	RMSD	# aa
6qup-A	LYSM TYPE RECEPTOR KINASE;	4.3	4.1	54
4m5d-A	U3 SMALL NUCLEOLAR RNA-ASSOCIATED PROTEIN 22;	4.2	3.7	53
1p4d-B	TRAI PROTEIN;	3.1	2.9	57
2mz0-A	DEFENSIN-LIKE PROTEIN 32;	3.0	3.4	44
5mgz-B	8-DEMETHYLNVOBIOCIC ACID C(8)-METHYLTRANSFERASE;	3.0	3.4	49
5n8o-A	DNA HELICASE I;	2.9	3.2	58
6i69-A	30S RIBOSOMAL PROTEIN S6;	2.9	3.4	50
6i6o-C	30S RIBOSOMAL PROTEIN S6,30S RIBOSOMAL PROTEIN S6	2.9	3.0	50
6i6w-A	30S RIBOSOMAL PROTEIN S6,30S RIBOSOMAL PROTEIN S6	2.9	3.1	50
2bxj-A	30S RIBOSOMAL PROTEIN S6;	2.9	3.1	50
1cqj-A	RIBOSOMAL PROTEIN S6;	2.9	3.1	50
6i6e-A	30S RIBOSOMAL PROTEIN S6;	2.9	3.0	49
2n69-A	DEFENSIN-LIKE PROTEIN;	2.8	2.5	46
6i6o-B	30S RIBOSOMAL PROTEIN S6,30S RIBOSOMAL PROTEIN S6	2.8	3.0	50
1qjh-A	RIBOSOMAL PROTEIN S6;	2.8	3.3	50
6a5k-A	HISTONE-LYSINE N-METHYLTRANSFERASE, H3 LYSINE-9 S	2.8	3.9	48
1ris-A	RIBOSOMAL PROTEIN S6;	2.8	3.1	50
1zhq-F	MANNAN-BINDING LECTIN;	2.8	3.8	45
6i6u-A	30S RIBOSOMAL PROTEIN S6,30S RIBOSOMAL PROTEIN S6	2.8	3.5	52

1cqn-B	RIBOSOMAL PROTEIN S6;	2.8	3.1	50
2bxj-B	30S RIBOSOMAL PROTEIN S6;	2.8	3.1	50

Tabla Suplementaria 6. Primeras 20 proteínas con similitud estructural reportadas por DALI para el Dominio 3 de la subunidad β de la bac-glyRS

Código PDB	Descripción	Puntuación Z	RMSD	# aa
4fxd-B	DNA POLYMERASE ALPHA CATALYTIC SUBUNIT A;	6.3	3.0	66
6uen-A	RNA-DIRECTED RNA POLYMERASE L;	6.2	2.9	71
3k5o-A	DNA POLYMERASE II;	6.0	2.9	72
6pzk-A	RNA-DIRECTED RNA POLYMERASE L;	5.9	2.8	73
1q8i-A	DNA POLYMERASE II;	5.9	2.9	72
3ueb-E	PUTATIVE UNCHARACTERIZED PROTEIN;	5.8	3.2	77
6zyd-G	LOW CONDUCTANCE MECHANOSENSITIVE CHANNEL YNAI,LOW	5.8	4.8	76
4qw2-B	FRAGILE X MENTAL RETARDATION PROTEIN 1;	5.8	3.5	74
6fh2-B	PROTEIN-ARGININE KINASE;	5.8	6.4	83
6lyp-D	MECHANOSENSITIVE ION CHANNEL PROTEIN 1, MITOCHOND	5.7	4.6	74
6lyp-A	MECHANOSENSITIVE ION CHANNEL PROTEIN 1, MITOCHOND	5.7	2.9	70
6lyp-C	MECHANOSENSITIVE ION CHANNEL PROTEIN 1, MITOCHOND	5.7	4.7	74
6lyp-E	MECHANOSENSITIVE ION CHANNEL PROTEIN 1, MITOCHOND	5.7	4.6	74
6zyd-B	LOW CONDUCTANCE MECHANOSENSITIVE CHANNEL YNAI,LOW	5.7	4.8	76
6zyd-A	LOW CONDUCTANCE MECHANOSENSITIVE CHANNEL YNAI,LOW	5.7	4.8	76
6zyd-F	LOW CONDUCTANCE MECHANOSENSITIVE CHANNEL	5.7	4.8	76

	YNAI,LOW			
6zyd-E	LOW CONDUCTANCE MECHANOSENSITIVE CHANNEL YNAI,LOW	5.7	4.8	76
6zyd-D	LOW CONDUCTANCE MECHANOSENSITIVE CHANNEL YNAI,LOW	5.7	4.8	76
6zyd-C	LOW CONDUCTANCE MECHANOSENSITIVE CHANNEL YNAI,LOW	5.7	4.8	76
4zm5-B	CHAIN LENGTH DETERMINANT PROTEIN;	5.7	5.4	82
5u8e-A	ARGININE KINASE;	5.7	3.3	72
5zhq-B	ARGININE KINASE;	5.7	3.3	80
6lyp-B	MECHANOSENSITIVE ION CHANNEL PROTEIN 1, MITOCHOND	5.6	4.6	74

Tabla Suplementaria 7. Primeras 19 proteínas con similitud estructural reportadas por DALI para el Dominio 4 de la subunidad β de la bac-glyRS

Código PDB	Descripción	Puntuación Z	RMSD	# aa
3mem-A	PUTATIVE SIGNAL TRANSDUCTION PROTEIN;	9.6	2.9	133
2pq7-A	PREDICTED HD SUPERFAMILY HYDROLASE;	9.3	2.9	122
3tm8-B	UNCHARACTERIZED PROTEIN;	9.0	3.3	130
3tmc-A	UNCHARACTERIZED PROTEIN;	9.0	3.2	129
3tm8-A	UNCHARACTERIZED PROTEIN;	9.0	3.2	129
3tmc-B	UNCHARACTERIZED PROTEIN;	9.0	3.2	129
3tmb-B	UNCHARACTERIZED PROTEIN BD1817;	8.8	3.2	128
6yvc-B	BIFUNCTIONAL (P)PPGPP SYNTHETASE/GUANOSINE-3',5'-	8.6	3.2	134
6yvc-A	BIFUNCTIONAL (P)PPGPP SYNTHETASE/GUANOSINE-3',5'-	8.6	3.2	134
4r8z-B	CYCLIC DI-GMP PHOSPHODIESTERASE;	8.6	3.4	133
3hc1-A	UNCHARACTERIZED HDOD DOMAIN	8.4	3.1	134

	PROTEIN;			
4mcw-B	METAL DEPENDENT PHOSPHOHYDROLASE;	8.4	3.2	133
6pbz-A	GUANOSINE-5'-TRIPHOSPHATE,3'-DIPHOSPHATE PYROPHOS	8.3	4.1	126
4mdz-B	METAL DEPENDENT PHOSPHOHYDROLASE;	8.3	3.2	133
4me4-B	METAL DEPENDENT PHOSPHOHYDROLASE;	8.2	3.3	136
4r8z-A	CYCLIC DI-GMP PHOSPHODIESTERASE;	8.2	3.3	133
4mdz-A	METAL DEPENDENT PHOSPHOHYDROLASE;	8.2	3.2	133
6pbz-C	GUANOSINE-5'-TRIPHOSPHATE,3'-DIPHOSPHATE PYROPHOS	8.2	4.1	126
4mcw-A	METAL DEPENDENT PHOSPHOHYDROLASE;	8.2	3.3	136

Tabla Suplementaria 8. Primeras 22 proteínas con similitud estructural reportadas por DALI para el Dominio 5 de la subunidad β de la bac-glyRS

Código PDB	Descripción	Puntuación Z	RMSD	# aa
1bs2-A	PROTEIN (ARGINYL-TRNA SYNTHETASE);	10.0	3.1	103
4oby-A	ARGININE--TRNA LIGASE;	9.7	2.9	96
5yyn-C	ARGININE--TRNA LIGASE;	9.6	3.4	102
5b63-C	ARGININE--TRNA LIGASE;	9.6	3.4	102
1iq0-A	ARGINYL-TRNA SYNTHETASE;	9.4	2.9	99
6ao8-A	ARGININE--TRNA LIGASE;	9.1	3.2	98
1f7u-A	TRNA(ARG);	9.1	2.9	100
5yyn-A	ARGININE--TRNA LIGASE;	8.8	3.1	96
5b63-A	ARGININE--TRNA LIGASE;	8.8	3.1	96
5yym-A	ARGININE--TRNA LIGASE;	8.8	3.3	99
1f7v-A	TRNA(ARG);	8.7	3.0	100
5yym-B	ARGININE--TRNA LIGASE;	8.5	3.1	96
4zt7-B	METHIONYL-TRNA SYNTHETASE;	8.4	3.9	105

2zue-A	ARGINYL-TRNA SYNTHETASE;	8.3	2.9	96
4q2t-A	ARGININE--TRNA LIGASE, CYTOPLASMIC;	8.3	2.8	92
3fnr-A	ARGINYL-TRNA SYNTHETASE;	8.1	3.0	97
4q2x-B	ARGININE--TRNA LIGASE, CYTOPLASMIC;	8.1	2.9	94
5jld-A	ARGINYL-TRNA SYNTHETASE, PUTATIVE;	8.1	3.3	93
4q2y-B	ARGININE--TRNA LIGASE, CYTOPLASMIC;	8.1	2.9	91
4q2y-A	ARGININE--TRNA LIGASE, CYTOPLASMIC;	8.0	3.0	95
4zt3-A	METHIONYL-TRNA SYNTHETASE;	7.7	4.0	106
2byt-D	LEUCYL-TRNA SYNTHETASE;	7.7	3.5	97
2zuf-A	ARGINYL-TRNA SYNTHETASE;	7.7	3.1	97