



**UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO**

**Maestría y Doctorado en Ciencias Bioquímicas**

**ESTUDIOS ESTRUCTURALES SOBRE LA GLICIL-tRNA SINTETASA**

**TESIS**

QUE PARA OPTAR POR EL GRADO DE:

Doctor en Ciencias

PRESENTA:

MARCO IGOR VALENCIA SÁNCHEZ

TUTOR PRINCIPAL:

DR. ALFREDO TORRES LARIOS

*Instituto de Fisiología Celular*

MIEMBROS DEL COMITÉ TUTOR :  
DRA. ADELA RODRÍGUEZ ROMERO

*Instituto de Química*

DR. ENRIQUE MERINO PÉREZ

*Instituto de Biotecnología*

Ciudad de México. Septiembre, 2016



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.

## RECONOCIMIENTOS:

Esta tesis de doctorado se realizó bajo la dirección del Dr. Alfredo Torres Larios en el laboratorio 205 Sur del Departamento de Bioquímica y Biología Estructural del Instituto de Fisiología Celular de la Universidad Nacional Autónoma de México.

El Comité Tutorial que asesoró el desarrollo de esta tesis estuvo formado por:

Dr. Alfredo Torres Larios	Instituto de Fisiología Celular, UNAM
Dra. Adela Rodríguez Romero	Instituto de Química, UNAM
Dr. Enrique Merino Pérez	Instituto de Biotecnología, UNAM

Se reconoce la colaboración de:

Dra. Annia Rodríguez Hernández por su esencial colaboración en el proyecto en la realización de las reacciones enzimáticas y la producción de los tRNAs. Dr. Dmitry Svergun y Dr. Haydn Mertens por su asesoría en el trabajo con SAXS. Dr. Morten Grötlí por la síntesis del GSA<sub>d</sub>, la cual resultó vital en el proyecto. Dr. Marcelino Arciniega por su contribución en el Análisis de Componentes Principales.

Se reconoce el apoyo técnico en diversos aspectos del proyecto de Biol. Elizabeth Nallely Cabrera González, Dr. Hugo Aníbal Santamaría Suárez, y Dra. Martha Contreras.

Este trabajo fue posible gracias a los apoyos del Consejo Nacional de Ciencia y Tecnología (CONACYT), apoyos CB2015-253323 y PDCPN2014-47543, así como al Programa de Apoyo a Proyectos de Investigación e Innovación Tecnológica (PAPIIT-DGAPA) de la UNAM, apoyo IN202416. Durante los estudios de doctorado el alumno también obtuvo el sustento de una beca otorgada por CONACYT con No. de CVU 348053 (No. de Beca 354994). Se contó con el apoyo por parte del “Programa de Apoyo a los Estudios del Posgrado” (PAEP) de la UNAM para realizar la estancia en el EMBL estación de Hamburgo, Alemania, para la cual también se contó con un respaldo complementario de Beca Mixta otorgado por CONACyT (convocatoria No. 290936).

**El Jurado de Examen de Grado estuvo constituido por:**

<b>PRESIDENTE</b>	<b>Dr. Enrique Rudiño Piñera</b>
<b>VOCAL</b>	<b>Dra. Alejandra Hernández Santoyo</b>
<b>VOCAL</b>	<b>Dr. Alejandro Sosa Peinado</b>
<b>VOCAL</b>	<b>Dr. Antonio Eusebio Lazcano Araujo Reyes</b>
<b>SECRETARIO</b>	<b>Dr. Diego González Halphen</b>

## **AGRADECIMIENTOS:**

A María Luisa Valencia (mi madre) por su apoyo siempre.

Al pueblo de México por su lucha por educación de la que somos producto, desde la primaria hasta el posgrado y a la UNAM por la gran oportunidad de estudiar libre.

Al Dr. Alfredo Torres Larios por todo el apoyo para avanzar a la obtención del grado, el apoyo con asuntos no estipulados en la supervisión académica y el complicado aprendizaje de vida que ambos compartimos en todo el periodo del posgrado.

Al Dr. Enrique Merino Pérez por su apoyo moral y económico cuando otros pensaron que no debía continuar con el doctorado. A la Dra. Adela Rodríguez Romero por su apoyo moral en el mismo caso. Al Dr. Rogelio Rodríguez Sotres por su apoyo moral para permanecer en el primer semestre de la maestría.

A Gissa Gis por todo lo compartido, por volar juntos y el apoyo mutuo en los tiempos difíciles.

A los compañeros del IFC, del laboratorio 205 Sur: Alex Madrigal por las discusiones, el apoyo y todo lo compartido, Sergio por el mensaje en el techo, Hugo, Nancy, Daniel y la Dra. Annia por las conversaciones; de otros laboratorios: Memo, Eréndira, Moni, Martha Irais y Bety.

Grateful to Asia Pieprzyk for her friendship that has been invaluable for me this year.

Grateful to Dr. Dmitry Svergun for the opportunity of achieving my the stay in Hamburg, Dr. Haydn Mertens for his support with the SAXS work, gracias a Gustavo Vives por todo su apoyo durante la estancia en Alemania, Spyros for his support with some experimental test and to Annabel and Kate for his support in the wet-lab in DESY.

# ÍNDICE DE CONTENIDO

<b>Contenido</b>	<b>Página</b>
ÍNDICE DE CONTENIDO.....	5-7
ÍNDICE DE FIGURAS.....	8-9
ÍNDICE DE TABLAS.....	10
RESUMEN.....	11
ABSTRACT.....	12
ABREVIATURAS.....	13
1. INTRODUCCIÓN.....	14
1.1 Importancia y función de las aminoacil-tRNA sintetasas.....	14
1.2 Pasos de la reacción catalizada por las aaRSs.....	14
1.3 Clasificación realizada a partir de mecanismo, plegamiento y motivos conservados.....	. 15
1.4 El dominio catalítico central.....	18
1.5 Grandes diferencias entre los dos tipos de GlyRSs.....	18
1.6 Función de las subunidades y la activación del aminoácido.....	19
1.7 Polifilesis y el árbol filogenético universal.....	20
1.8 Clasificación estructural de las aaRSs.....	21
2. HIPÓTESIS.....	23
3. OBJETIVOS.....	24
4. MATERIALES Y MÉTODOS.....	25
4.1 Expresión y purificación.....	25
4.2 Síntesis del GSAd.....	26
4.3 SEC-MALS.....	26
4.4 Ensayos de actividad.....	27
4.5 Dispersión de rayos X de ángulo pequeño (SAXS) .....	28

4.6 Ensayo de termofluorimetría diferencial (termofluor, desplazamiento térmico).....	. 29
4.7 Cristalización.....	30
4.8 Colecta de datos.....	31
4.9 Reducción de datos.....	31
4.10 Resolución de la estructura de la $\alpha$ -AaGlyRS.....	32
4.11 Comparación del reconocimiento entre las GlyRSs.....	34
4.12 Alineamientos estructurales entre las aaRSs II.....	35
5. RESULTADOS.....	38
5.1 La subunidad $\alpha$ de la bacGlyRS es un dímero en solución.....	38
5.2 La subunidad $\alpha$ de la bacGlyRS por si sola presenta actividad de activación del aminoácido.....	. 40
5.3 Estructura tridimensional de la subunidad $\alpha$ de la bacGlyRS en complejo con un análogo del estado de transición.....	. 42
5.4 La unión del GSAd promueve cambios conformacionales en la subunidad $\alpha$ de la bacGlyRS.....	. 44
5.5 El reconocimiento de la glicina por la enzima bacteriana y la de origen arquea (eukGlyRS) ocurren de forma distinta.....	. 46
5.6 La $\alpha$ -bacGlyRS y la AlaRS comparten características que las diferencian de las otras aaRSs II.....	. 52
5.7 La bacGlyRS y la AlaRS constituyen una subclase de las aaRSs II.....	. 55
6. DISCUSIÓN.....	59
6.1 La subunidad $\alpha$ de la bacGlyRS es un dímero en solución.....	59
6.2 La subunidad $\alpha$ de la bacGlyRS puede llegar a catalizar la formación del glicil-adenilato por si sola.....	. 60
6.3 Efecto del GSAd sobre la subunidad $\alpha$ de la AaGlyRS.....	61
6.4 El reconocimiento de la glicina entre las GlyRS ocurre de forma distinta.....	. 62

6.5 Las moléculas de disolvente son parte de la red polar que reconoce la Gly en la bacGlyRS.....	64
6.6 La interacción Cation- $\pi$ del Trp podría ser esencial para reconocer la glicina.....	65
6.7 Mediante alineamiento estructural es posible inferir la relación estructural y filogenética de las GlyRSs.....	66
6.8 La bacGlyRS y la AlaRS se hallan estrechamente relacionadas.	68
6.9 Las aaRSs de clase II son agrupadas en cuatro subclases por alineamiento estructural.....	70
6.10 Las GlyRS tienen un origen polifilético de dos ancestros no contemporáneos.....	71
7. CONCLUSIONES.....	74
8. PERSPECTIVAS.....	76
9. ANEXO 1. RESULTADOS ADICIONALES.....	78
9.1 Avances en el estudio estructural del complejo bacGlyRS- tRNA.	78
9.2 Las subunidades $\alpha$ y $\beta$ forman forman la GlyRS completa al unirse con alta afinidad.....	78
9.3 La bacGlyRS tetramérica forma un complejo estable con el tRNA.....	79
9.4 En la estructura a baja resolución de la bacGlyRS de subunidades fusionadas puede observarse la disposición de las subunidades.....	80
9.5. Figuras de resultados adicionales.....	83
10. ANEXO 2. MÉTODOS Y RESULTADOS ADICIONALES .....	86
10.1 Portada del artículo resultado de este trabajo.....	100
11. BIBLIOGRAFÍA.....	101

## ÍNDICE DE FIGURAS.

<b>Figura</b>	<b>página</b>
Figura 1. La subunidad $\alpha$ de la AaGlyRS es un dímero en solución.....	39
Figura 2. La subunidad $\alpha$ de la AaGlyRS por si sola presenta actividad de activación del aminoácido.....	41
Figura 3. La unión del GSAd promueve cambios conformacionales en la subunidad $\alpha$ de la bacGlyRS.....	45
Figura 4. El reconocimiento de la glicina por la bacGlyRS y la eukGlyRS ocurren de forma distinta.....	49
Figura 5. La bacGlyRS presenta diferencias con la eukGlyRS en las interacciones con el aminoácido y semejanzas con la AlaRS. ....	51
Figura 6. La bacGlyRS y la AlaRS comparten características que las diferencian de las otras aaRSs II.....	54
Figura 7. La bacGlyRS y la AlaRS constituyen una subclase de las aaRSs II.	57-58
Figura 8. La interacción Cación- $\pi$ con el Trp podría ser esencial para reconocer la glicina.....	66
Figura 9. Las subunidades $\alpha$ y $\beta$ forman forman la GlyRS completa al unirse con alta afinidad.....	79
Figura 10. La AaGlyRS $\alpha\beta_2$ forma un complejo estable con el tRNA-Gly...	80
Figura 11. En la estructura a baja resolución de la bacGlyRS de subunidades fusionadas puede observarse la posible disposición de las subunidades....	82
Figura 12. Los cambios conformacionales promovidos por el GSAd corresponden a una estabilización térmica en la $\alpha$ -AaGlyRS.....	83
Figura 13. Dendrograma de agrupación por secuencia de regiones estructurales equivalentes de las aaRSs II.....	84
Figura 14. Secuencias del CCD de las aaRSs II alineadas con T-COFFEE.	84-85
Figura 15. Diferentes versiones de la bacGlyRS provenientes de distintos organismos fueron expresadas y purificadas para los estudios estructurales.	90
Figura 16. Estadísticas adicionales de la estructura de la $\alpha$ -AaGlyRS y la puntuación en el alineamiento estructural.....	93

Figura 17. Residuos conservados y estructuralmente equivalentes entre la bacGlyRS, la eukGlyRS y la AlaRS, y el reconocimiento del ATP.....	94-95
Figura 18. La bacGlyRS presenta semejanzas con la AlaRS en el reconocimiento del aminoácido y diferencias con la eukGlyRS.....	95-96
Figura 19. Estabilidad térmica de las mutantes de la AtGlyRS.....	99

## ÍNDICE DE TABLAS.

<b>Tabla</b>	<b>página</b>
Tabla 1. Características de las aaRS.....	15
Tabla 2. las aaRSs II son subclasificadas por algunas características comunes.....	17
Tabla 3. Estadísticas de colección de datos y de afinamiento.....	43
Tabla 4. El GSA <sub>d</sub> se encontró presente en las 5 moléculas de la unidad asimétrica .....	46
Tabla 5. Otras condiciones de cristalización de la AaGlyRS.....	91
Tabla 6. Distintas características de las mutantes del conector peptídico en la GlyRS de <i>Anaerolinea thermophila</i> .....	98

## RESUMEN

Las aminoacil-tRNA sintetetasas (aaRSs) se encuentran entre las proteínas más antiguas y diversas. Junto con el ribosoma, las aaRSs son responsables de mantener la fidelidad de la traducción del código genético mediante la verificación y adición de cada aminoácido a su tRNA correspondiente. Existe una sintetasa específica para cada aminoácido y cada una de ellas proviene de un solo ancestro común con la misma especificidad en los tres dominios de la vida. Sin embargo, existen dos tipos de glicil-tRNA sintetasa (GlyRS) que se habían propuesto como polifiléticos por su limitada identidad de secuencia entre sí (12 %), pero existía una controversia debido a que los dominios del núcleo catalítico (CCD) de las dos GlyRSs comparten el mismo plegamiento. Una GlyRS (GlyRS bacteriana) es un heterotetrámero  $\alpha_2\beta_2$  que se encuentra en la mayoría de las bacterias y cloroplastos; la otra GlyRS (GlyRS eucariótica) es un homodímero  $\alpha_2$  que se encuentra presente en las arqueas, citoplasma de eucariotas y algunas bacterias. En este trabajo determinamos la estructura del complejo de la subunidad  $\alpha$  de la GlyRS de la bacteria *Aquifex aeolicus* con un análogo del glicil-adenilato a una resolución de 2.81 Å. Demostramos que los residuos implicados en el mecanismo de reconocimiento de la glicina son completamente diferentes entre la GlyRS bacteriana y la GlyRS eucariótica. Debido a que las aaRSs presentan una identidad de secuencia menor al 25 %, mediante distintos alineamientos estructurales entre los CCDs de todas las aaRSs de clase II, el análisis de la variación estructural en sus componentes principales (PCA) y alineamientos de secuencia de regiones estructurales equivalentes, se definieron las relaciones entre las dos GlyRSs respecto a las otras aaRSs de clase II. Encontramos que las dos GlyRSs comparten más características con otras aaRSs que entre ellas, presentando entonces un caso de evolución convergente, ya que no comparten un ancestro común posterior al ancestro primigenio de todas las aaRSs de su clase. Proponemos que las dos GlyRSs son parálogos lejanos que convergieron en enzimas isofuncionales. También proponemos que la alanil-tRNA sintetasa y la subunidad  $\alpha$  de la GlyRS bacteriana divergieron de un antepasado más reciente, por lo que conformarían una nueva subclase de las aaRSs de clase II.

## ABSTRACT

Aminoacyl-tRNA synthetases (aaRSs) are among the most ancient and diverse proteins. Along with the ribosome, aaRSs maintain the fidelity of the genetic code by checking and linking their corresponding amino acid to their cognate tRNA.

There is one specific aaRS for each amino acid. Most aaRSs descend from a single common ancestor with the same specificity in the three kingdoms of life. However, there are two types of glycyl-tRNA synthetases (GlyRSs). Some propose that these enzymes have a polyphyletic origin because they have a very low sequence identity between them (12%). However, they share the same fold of their central core domain (CCD).

One type of GlyRS (bacterial GlyRS) is an  $\alpha_2\beta_2$  heterotetramer and it is present in most bacteria and chloroplasts; the other type of GlyRS (eukaryotic GlyRS) is an  $\alpha_2$  homodimer, which is present in archaea, eukaryotic cytoplasm and a few bacteria.

In this work, we solved the 2.81 Å crystal structure of the GlyRS  $\alpha$ -subunit from the bacteria *Aquifex aeolicus* in complex with an analog of glycyl-adenylate. We showed that the residues involved in the recognition mechanism of glycine in the bacterial and eukaryotic GlyRS are completely different.

To overcome the fact that the sequence identity between aaRSs is below 25%, we performed several structural alignments between the CCDs of class II aaRSs, analysis of the structural deviation of their principal components (PCA) and sequence alignments of equivalent structural regions. We thus established the relationship of the two GlyRSs with reference to the other class II aaRSs. We found that the two GlyRSs share more features with other aaRSs than between them, illustrating a puzzling case of convergent evolution, because they do not share a common ancestor besides of the very early, common class II aaRSs ancestor. We propose that the two GlyRSs are extremely distant paralogs that converged in isofunctional enzymes. We also propose that alanyl-tRNA synthetase and the  $\alpha$ -subunit of the bacterial GlyRS diverged from a most recent ancestor, and constitute a new subclass within class II aaRSs.

## ABREVIATURAS

**aa-AMP:** Aminoacil-AMP

**aa-tRNA<sup>aa</sup>:** Isoceptor del tRNA cargado con el respectivo aminoácido

**AaGlyRS:** Glicil-tRNA sintetasa de *Aquifex aeolicus*

**AtGlyRS:** Glicil-tRNA sintetasa de *Anaerolinea thermophila*

**aaRS:** Aminoacil-tRNA sintetasa

**ABD:** Dominio de Unión al Anticodón; “Anticodon Binding Domain”

**AlaRS:** Alanil-tRNA sintetasa

**$\alpha$ -AaGlyRS:** Subunidad  $\alpha$  de la glicil-tRNA sintetasa de *A. aeolicus*

**AMP:** Adenosín monofosfato; monofosfato de adenosina

**ATP:** Adenosín trifosfato; trifosfato de adenosina

**bacGlyRS:** GlyRS de origen bacteriano

**CCD:** Núcleo catalítico central; “Central Catalytic Core”

**eukGlyRS:** GlyRS presente en citoplasma de eucariotas (origen arquea)

**GlyRS:** Glicil-tRNA sintetasa

**GSAd:** glicil sulfamoil adenilato; 5'-O-[N-(L-glicil)sulfamoil] adenosina

**LB:** Medio de cultivo de Luria-Bertani

**MALS:** Dispersión de luz multi ángulo; “Multi-Angle Light Scattering”.

**PAGE:** Electroforesis en gel de poliacrilamida.

**PCA:** Análisis de Componentes Principales; “Principal Component Analysis”

**PDB:** Banco de datos de proteínas; “Protein Data Bank”

**PCR:** Reacción en cadena de la polimerasa; “Polimerase Chain Reaction”.

**PPi:** Pirofosfato; difosfato

**RMSD:** Raíz de la desviación de la media cuadrática; Desviación estándar; “Root-Mean-Square Deviation”.

**SAXS:** Dispersión de rayos X de ángulo pequeño; “Small Angle X-ray Scattering”

**SEC:** Cromatografía de exclusión de tamaño; “Size exclusion chromatography”

**tRNA:** Ácido ribonucleico de transferencia

**tRNA<sup>Gly</sup>:** Isoceptor del tRNA de glicina

# 1. INTRODUCCIÓN

## 1.1 Importancia y función de las aminoacil-tRNA sintetasas

La complejidad del proceso traduccional incluye la necesidad de unir correctamente los 20 residuos de aminoácidos exactamente en el orden especificado por un mRNA particular, para lograr la alta fidelidad del código genético.

La fidelidad de la traducción se verifica en dos etapas de reconocimiento independientes: a) La discriminación del aminoácido correcto y la unión covalente a su tRNA blanco. b) La selección en el ribosoma del correcto tRNA ya cargado especificado por cada codón en el mRNA y la entrega del aminoácido a la cadena naciente de proteína (Cavarelli y Moras, 1993).

Las aminoacil-tRNA sintetasas (aaRSs) son las responsables de la elección de los aminoácidos para ser unidos a sus tRNAs correspondientes y de catalizar la formación del enlace covalente (carga del tRNA). Las aaRSs se consideran dentro de las proteínas más antiguas, pues su origen precede a la raíz del árbol filogenético universal. Se piensa fueron unas de las primeras proteínas en sustituir las funciones de las ribozimas en el paso del mundo basado en RNA, y que su evolución está ligada a la evolución del código genético (O'Donoghue y Luthey-Schulten, 2003)

## 1.2 Pasos de la reacción catalizada por las aaRSs

Cada aaRS une el aminoácido (aa) a sus correspondientes isoaceptores de tRNA para formar un aminoacil-tRNA (tRNA<sup>aa</sup>). Este proceso requiere hidrólisis de ATP y se lleva a cabo en una reacción de 2 pasos secuenciales:

1. **Activación del aminoácido con ATP:** para formar el aminoacil-adenilato que permanece unido a la enzima. La reacción libera pirofosfato (PPi). Este paso es reversible.



2. **Aminoacilación del tRNA:** transferencia del aminoácido a su tRNA específico para la formación del aminoacil-tRNA+AMP.



### 1.3 Clasificación realizada a partir del mecanismo de acción, plegamiento y motivos conservados

Hay al menos una aaRS específica para cada uno de los 22 aminoácidos. Las aaRSs se agrupan en dos clases: la clase I y la clase II (tabla 1). Estas dos clases no tienen ninguna relación entre sí. En cambio dentro de cada clase, los miembros comparten varias características: el mecanismo de aminoacilación, el plegamiento del dominio catalítico central y la presencia de firmas consenso (tabla 1).

<b>Tabla 1. Características de las aminoacil-tRNA sintetasas</b>		
	<b>Clase I</b>	<b>Clase II (tabla 2)</b>
Aminoácidos que son reconocidos por estas sintetasas:	Arg, Cys, Ile, Leu, Met, Val Gln, Glu, Trp, Tyr, Lys I	His, Pro, Ser, Thr, Asn, Asp, Lys II, Phe, Gly, Ala, Sep, Pyl
<i>Unión del aminoácido inicialmente al tRNA en el extremo:</i>	2'OH	3'OH
<i>Secuencias consenso</i>	KMSKS y HIGH	Las enzimas de clase II poseen 3 secuencias consenso 1, 2 y 3 (Motivos).
<i>Plegamiento del dominio de unión a ATP tipo</i>	Tipo Rossmann	Plegamiento específico del dominio de unión a ATP
<i>Plegamiento</i>	Alternando 6 hebras $\beta$ paralelas con 4 cadenas $\alpha$ .	Hoja de 6 hebras $\beta$ antiparalelas, flanqueada por 3 cadenas $\alpha$
<i>Acerca al tRNA por</i>	Surco menor	Surco mayor
<i>Estado oligomérico (de la mayoría de las aaRSs de su clase)</i>	Monómeros	Dímeros

#### 1.3.1 Las aaRSs de clase I

Todos los aminoácidos son unidos al extremo 3'OH del tallo aceptor del tRNA, pero las aaRSs de tipo I inicialmente unen el aminoácido al extremo 2'OH (Norris y Berg, 1964;

Latherbarrow *et al.*, 1985); tienen un plegamiento central tipo Rossmann de unión al nucleótido, y las secuencias consenso KMSKS y HIGH (Moras, 1992).

### **1.3.2 Las aaRSs de Clase II**

En las aaRSs de la clase II (aaRSs II), el aminoácido se transfiere directamente al extremo 3'OH del tRNA (Norris y Berg, 1964; Latherbarrow *et al.*, 1985); presentan un plegamiento único de unión al ATP que tiene una organización de una hoja  $\beta$  de 6 hebras antiparalelas flanqueada por 3 hélices  $\alpha$ . Este plegamiento sólo es encontrado entre las mismas aaRSs II y la holoenzima de la sintetasa de biotina (Moras, 1992; O'Donoghue y Luthey-Schulten, 2003). La mayoría de las enzimas de esta clase tienen una estructura cuaternaria constituida de homodímeros  $\alpha_2$ . Las enzimas de clase II poseen 3 secuencias consenso (motivos) implicados en la asociación de subunidades (motivo 1) y en la catálisis (motivos 2 y 3) (Eriani *et al.*, 1990; Moras, 1992; Mazauric *et al.*, 1998). En general, las enzimas de esta clase se aproximan al tallo aceptor del tRNA por el surco mayor (Cavalleri y Moras, 1993)

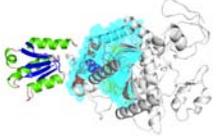
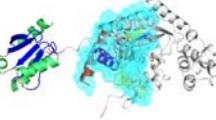
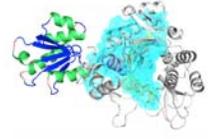
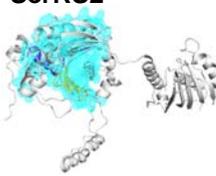
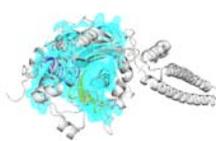
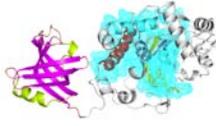
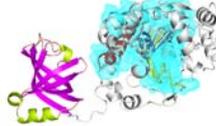
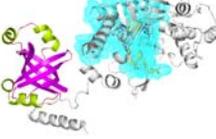
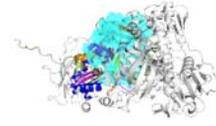
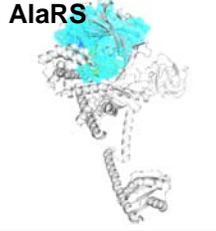
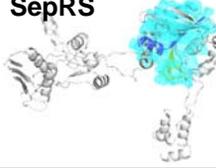
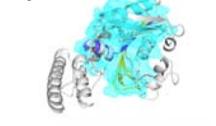
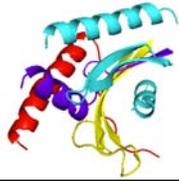
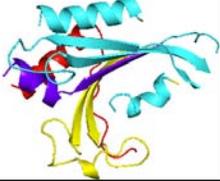
Las sintetetasas de clase II fueron subclasificadas por características comunes (tabla 2), en las subclases:

**Ila:** Tienen un dominio accesorio (idiosincrático) de unión al anticodón (ABD) común entre ellas en el extremo carboxilo terminal (tabla 2); esta subclase está formada por aaRSs que procesan aminoácidos pequeños y polares: Pro, Ser, Thr, His.

**Ilb:** Tienen en común un ABD amino terminal similar (tabla 2) y procesan aminoácidos cargados: Asn, Asp, Lys (Cusak, 1993).

**Ilc:** Anteriormente estaba mas bien definida por características distintas a las demás como la estructura cuaternaria: Phe, Gly, Ala, Sep y Pyl (Perona y Hadd, 2012) (para las excepciones y particularidades de estas sintetetasas ver apartado 1.7).

Tabla 2. Las aaRS II son subclasificadas por algunas características comunes

<b>Subclase IIA *</b>	<b>ThrRS</b> 	<b>ProRS</b> 	<b>HisRS</b> 	<b>eukGlyRS??</b> 
	<b>SerRS2</b> 	<b>SerRS1</b> 		
<b>Subclase IIB *</b>	<b>AspRS</b> 	<b>AsnRS</b> 	<b>LysIIRS</b> 	
<b>Subclase IIC *</b> ??	<b>PheRS</b> 	<b>AlaRS</b> 	<b>a-bacGlyRS??</b>	
	<b>SepRS</b> 	<b>PyIRS</b> 		
	<b>ThrRS (IIA)</b>	<b>AspRS (IIB)</b>	<b>PheRS</b>	<b>AlaRS</b>
<b>* CCD</b>				

\* El CCD se muestra rodeado por una malla color cian en las estructuras de todas las aaRSs II completas. En la parte inferior de la tabla se muestra el CCD, el motivo 1 se muestra en rojo, el motivo 2 en amarillo y el 3 en púrpura.

## 1.4 El dominio catalítico central

Todas las aaRSs son enzimas modulares. El núcleo catalítico o dominio catalítico central (CCD) contiene los determinantes para la activación del aminoácido y es la parte más conservada dentro de las aaRSs (tabla 2). También se piensa que es el más relacionado con una sintetasa primordial (Shiba *et al.*, 1994). Normalmente hay un segundo dominio conservado para aaRSs muy relacionadas, denominado idiosincrático o accesorio (tabla 2) (Shiba *et al.*, 1994; O'Donoghue y Luthey-Schulten, 2003).

## 1.5 Grandes diferencias entre los dos tipos de GlyRSs

La glicina es el aminoácido más pequeño que ocurre en la naturaleza, probablemente establecido de una forma muy temprana como componente de las proteínas en la evolución (Shiba *et al.*, 1994). La glicil-tRNA sintetasa (GlyRS) (EC 6.1.1.14) cataliza la unión de glicina a los isoaceptores de tRNA<sup>Gly</sup>.

Existen 2 distintas GlyRSs con una estructura cuaternaria diferente y una distribución distintiva en los dominios de la vida: la enzima homodimérica de subunidades  $\alpha_2$  se encuentra en el citoplasma de eucariotas (Shiba *et al.*, 1994), en arqueas y en algunas bacterias; la enzima heterotetramérica de subunidades  $\alpha_2\beta_2$  se encuentra en la mayoría de las bacterias y en cloroplastos (Mazauric *et al.*, 1996; Duchene *et al.*, 2000; Okamoto *et al.*, 2005).

Por lo anterior, en el texto se nombrará a la GlyRS  $\alpha_2\beta_2$  como de **origen bacteriano (bacGlyRS)**, y la  $\alpha_2$  presente en el citoplasma de eucariotas como de **origen arquea** (será nombrada **eukGlyRS** por corresponder con las publicaciones anteriores).

A pesar de que ambas GlyRSs tienen el plegamiento de unión a nucleótidos característico y al menos el motivo 3, que las identifican como aaRS de clase II, la secuencia de aminoácidos de la GlyRS tetramérica  $\alpha_2\beta_2$  (bacGlyRS), presenta una enorme diferencia comparada con la secuencia de la GlyRS dimérica  $\alpha_2$  (eukGlyRS) (Eriani *et al.*, 1990; Mazauric *et al.*, 1998), tanto que partir de sus secuencias, no se había podido inferir si descendían de genes homólogos (Shiba *et al.*, 1994).

Las diferencias entre las GlyRSs incluyen los motivos (secuencias consenso) 1 y 2,

implicados en la asociación de las subunidades y la catálisis (Eriani *et al.*, 1990; Mazauric *et al.*, 1998). Las razones de la evolución divergente de esta sintetasa, han permanecido desconocidas hasta este trabajo (Logan *et al.*, 1995; Mazauric *et al.*, 1996).

Una confusión al respecto de que el CCDs de las dos GlyRS podrían provenir uno del otro (ser ortólogos), fue extendida por una escueta interpretación de las estructuras cristalográficas existentes, que carecían de algún ligando en alguna conformación productiva o una clara densidad electrónica, o bien de una incompleta interpretación de los alineamientos estructurales (Tang y Huang, 2005., Tan *et al.*, 2012; Smith y Hartman, 2015).

Otra consecuencia importante de lo anterior es que se pensaba que la forma en que se reconocía la glicina por las dos GlyRSs era similar, pues a pesar de sus diferencias en la distribución entre los organismos, de estructura cuaternaria y de la baja identidad de sus secuencias, ello no implicaba que los residuos involucrados en el reconocimiento del aminoácido fueran distintos; los estudios estructurales y evolutivos anteriores asumían una misma forma de reconocimiento (Tang y Huang, 2005., Tan *et al.*, 2012; Smith y Hartman, 2015).

## **1.6 Función de las subunidades de la bacGlyRS y la activación del aminoácido**

La bacGlyRS tiene una estructura cuaternaria heterotetramérica  $\alpha_2\beta_2$  (Ostrem y Berg 1974). Se ha demostrado que la subunidad  $\alpha$  de 35 kDa contiene los elementos de unión de los sustratos ATP y glicina (Toth y Schimmel., 1990), y se confirmó que esta subunidad contenía el plegamiento del CCD (Tan *et al.*, 2012). También se había concluido que posiblemente las interacciones entre las subunidades  $\alpha$  son las responsables de mantener la estructura cuaternaria (McDonald *et al.*, 1980).

La subunidad  $\beta$  de la bacGlyRS de 77 kDa no comparte ninguna característica con los CCD de ninguna aaRSs II (Tang y Huang., 2005), pero esta subunidad presenta el mayor papel en la unión del tRNA, específicamente del anticodón del tRNA<sup>Gly</sup> (Nagel *et al.*, 1984). Por alineamientos de secuencia, la subunidad  $\beta$  se agrupó con la superfamilia DALR\_1 (Pfam, PF05746) , junto con el dominio de unión al anticodón (ABD) de la ArgRS (Tang y Huang., 2005), y se ha postulado que fue derivada de la superfamilia de las fosfohidrolasas HD con algunos de los CCDs de las aaRSs de clase I (Wolf *et al.*, 1999).

Sin embargo, fue aceptado durante mucho tiempo que la existencia de actividad catalítica requería estrictamente la participación de ambas subunidades (Ostrem y Berg 1974), posiblemente por estabilización conformacional al formar el tetrámero (McDonald *et al.*, 1980). En los trabajos de Toth y Schimmel *et al.*, (1990) se detectó cierta actividad del primer paso de la reacción en la subunidad  $\alpha$ , pero se pensó era un artificio de incorrecta separación en el análisis cromatográfico de las subunidades.

## 1.7 Polifilesis en las aaRSs y el árbol filogenético universal

El árbol filogenético universal o el patrón evolutivo canónico es resultado del flujo genético vertical y fue reconstruido con base a las relaciones filogenéticas de las moléculas de RNAr, por lo que es usado como marco de referencia para las relaciones evolutivas (O'Donoghue y Luthey-Schulten, 2003).

Existen tres reglas que cumplen la mayoría de las sintetasas: la regla del sitio activo, la regla de la clase y la regla de la monofilia. La regla del sitio activo se refiere a que todas las aaRSs de clase I aminoacilan el tRNA de forma inicial en el 2'OH y las de clase II directamente en el 3'OH. La regla de la clase consiste en que cada sintetasa pertenece a una clase pero no a ambas. Y la regla de la monofilia dice que todas las sintetasas para un mismo aminoácido provienen de un ancestro común en todos los dominios de la vida. Al analizar el dominio catalítico central, se concluye que la mayoría de las sintetasas claramente provienen de genes homólogos (O'Donoghue y Luthey-Schulten, 2003).

Un ejemplo es el caso de la Val, Ile y LeuRS; todas ellas son monofiléticas, por lo que posiblemente provienen de una duplicación de genes (Brown y Doolittle.,1995 ).

La PheRS es la única excepción a la regla del sitio activo ya que, siendo de clase II, lleva a cabo la aminoacilación en el 2'OH (Moras, 1992).

La Lys y la GlyRS violan la regla de la monofilia, siendo los dos ejemplos más claros de polifilia: sintetasas que tienen la misma especificidad, pero que no comparten directamente un ancestro común de la misma especificidad. La LysRS viola la regla de la monofilia y la de clase, pues existe una LysRS que pertenece a las aaRSs tipo I y otra de tipo II. A pesar de ello, tiene una distribución canónica respecto al árbol filogenético universal (O'Donoghue y Luthey-Schulten, 2003).

La GlyRS viola la regla de la monofilia. Por las características de cada una de las GlyRSs (apartado 1.5), la eukGlyRS ha sido agrupada en la subclase IIa (pues incluso posee el dominio idiosincrático característico de esta subclase) y la bacGlyRS ha sido agrupada en la clase IIc (tabla 2) (Friest *et al.*, 1996; O'Donoghue y Luthey-Schulten, 2003; Perona y Hadd, 2012; Smith y Hartman, 2015).

La SerRS II, la SepRS y la atípica PylRS presentan también distribuciones no canónicas. La SepRS y la PylRS habían sido colocadas en la misma subclase IIc que la PheRS en cuanto se determinaron sus estructuras, y debido a su aparente similitud en su oligomerización (Hauenstein *et al.*, 2008).

Para subrayar la teoría del endosimbionte como surgimiento del dominio eucariota, las plantas y algas eucariotas presentan en el citoplasma la eukGlyRS (de origen arquea) y los cloroplastos tienen la bacGlyRS (de origen bacteriano) (Duchene *et al.*, 2001; Huan y Tang 2005)

Todo lo anterior es intrigante, y nos conduce a tratar de conocer la historia compleja del desplazamiento enzimático y la transferencia de genes que caracteriza las vías evolutivas de las aaRSs polifiléticas, en particular de las GlyRSs.

## **1.8 Clasificación estructural de las aaRSs**

La comparación por secuencia de aaRSs específicas para distintos aminoácidos arroja resultados que caen por debajo del 25% de la identidad de secuencia, en la denominada zona de penumbra (O'Donoghue y Luthey-Schulten, 2003; Cusak *et al.*, 1993), por lo que la comparación por secuencia puede ser poco fidedigna y no existe una verdadera diferencia cuando se comparan secuencias de homólogos correctamente alineados y secuencias no relacionadas o aleatorias (Keiger *et al.*, 2000; O'Donoghue y Luthey-Schulten, 2003).

Chotia *et al.*, 1986, demostraron que, entre las proteínas, la estructura tridimensional se conserva más que la secuencia. Por eso y por las características de baja identidad entre las sintetetasas, el realizar alineamientos por estructura puede permitir construir una filogenia más exacta y completa para desentrañar el devenir evolutivo completo de las aaRSs (O'Donoghue y Luthey-Schulten, 2003).

Algunos de los primeros alineamientos de secuencias de las aaRSs que se realizaron,

como en el caso de Nagel y Doolittle 1991, tuvieron problemas con el umbral de la identidad. Lo valioso es que pudieron observar que el alineamiento de secuencia de ciertas regiones agrupaba a las aaRSs por especificidad de aminoácidos relacionados.

El más visionario análisis de las sintetasas a un nivel de especificidad lo realizó Woese *et al.* 2000. Una filogenia elaborada con algunos datos estructurales fue construida por Ribas de Pouplana *et al.*, 2001, en la que se realizaron alineamientos por secuencias de regiones estructurales equivalentes.

O'Donoghue y Luthey-Schulten, 2003 y Smith y Hartman, 2015 realizaron los intentos más cercanos para tratar de agrupar las sintetasas en subclases por alineamiento estructural, empleando las estructuras que habían sido publicadas hasta ese entonces. Sin embargo, la mencionada confusión respecto a la relación entre los CCDs de las GlyRS y el hecho de que no todas las estructuras de las aaRSs fueran conocidas o no incluirlas todas, les impidió construir una filogenia completa. O'Donoghue y Luthey-Schulten, 2003 se explayaron en indicar los beneficios de un alineamiento estructural y la importancia de considerar los huecos (gaps) en el alineamiento.

Dadas las enormes diferencias que llevaron a postular un distinto origen evolutivo de ambos tipos de GlyRS, se requería aportar evidencias para eliminar la suposición de que los CCD de las dos GlyRS eran ortólogos y de que existía identidad en el reconocimiento del aminoácido entre ellas, para así esclarecer la naturaleza de la polifilesis y elaborar una filogenia estructural más completa de las aaRSs.

## **2. HIPÓTESIS**

Si la glicil-tRNA sintetasa de origen bacteriano y la glicil-tRNA sintetasa de origen arquea tienen un ancestro común con la misma especificidad, compartirán el mismo reconocimiento de la glicina.

### **3. OBJETIVOS**

Determinar la estructura tridimensional de la subunidad  $\alpha$  de la GlyRS bacteriana en complejo con un análogo del glicil-adenilato.

Analizar los dos tipos de GlyRSs a nivel estructural para comparar la forma en que reconocen la glicina.

## 4. MATERIALES Y MÉTODOS.

### 4.1 Expresión y purificación

**4.1.1 Diseño y síntesis del gen de la subunidad  $\alpha$  de la GlyRS:** El gen de la subunidad  $\alpha$  de la GlyRS de *Aquifex aeolicus* ( $\alpha$ -AaGlyRS), optimizado para su expresión con la secuencia de los codones encontrados con alta frecuencia en *Escherichia coli*, fue mandado a sintetizar a la compañía Genscript y subclonado al vector de expresión pET23a. La secuencia codificaba además para 6xHis en el extremo amino terminal y el sitio de corte para la proteasa TEV.

**4.1.2 Expresión:** La construcción de la  $\alpha$ -AaGlyRS en pET23a se transformó en *E. coli* BL21 DE3 y se realizó el cultivo en 4L de medio LB-Vegitone (Sigma) con 100  $\mu$ g/mL de ampicilina con agitación a 250 rpm a 37°C, que al llegar a una OD<sub>600 nm</sub> de 1.0, fue inducido con IPTG (isopropil- $\beta$ -D-1-tiogalactopiranosido) 1 mM durante 10 horas. Se centrifugó el caldo de cultivo a 6,359 x g (rotor JA-10) y el botón de células se congeló a -20°C.

**4.1.3 Lisis:** El botón de células se descongeló en 30 mL de amortiguador A de cromatografía de afinidad a níquel (amortiguador de fosfato de potasio 100 mM pH 8.3 y KCl 250 mM) y se realizó la lisis por sonicación al 30% de amplitud durante 20 minutos con pulsos de 50 segundos de sonicación por 60 segundos de reposo. Se centrifugó el lisado celular a 14,000 x g y se recuperó el sobrenadante conteniendo la fracción soluble.

**4.1.4 Cromatografía de afinidad a níquel:** La fracción soluble se inyectó a una columna HisTrap FF Crude (GE) de 5 mL equilibrada con amortiguador A a 5 mL/min, la columna se lavó extensivamente con amortiguador A más 20 mM de imidazol, y la elución se realizó con un gradiente de imidazol de 60 mL de longitud de 20 a 300 mM de imidazol.

**4.1.5 Cromatografía de afinidad a níquel con alta fuerza iónica y en condiciones parcialmente astringentes:** Para obtener una purificación más astringente de la proteína, se realizó la purificación con resina de afinidad a níquel con ligeras modificaciones. El botón de células bacterianas se resuspendió en amortiguador A6.3 (amortiguador de fosfatos de sodio 50 mM pH 6.3, 100 mM KCl y 1 mM MgCl<sub>2</sub>) y se obtuvo la fracción soluble como se ha descrito anteriormente. Dicha fracción soluble se llevó a las concentraciones del amortiguador A50 (descrito a continuación) y se inyectó a un flujo de 5 mL/min a la columna

HisTrap FF Crude (GE) de 5 mL equilibrada con el amortiguador A50 (amortiguador de fosfato de sodio 100 mM pH 6.3, KCl 1 M, 0.5 M Urea, 50 mM de imidazol y 10% de glicerol), se lavó extensivamente con este amortiguador y se eluyó con un gradiente de imidazol de 50 a 500 mM. La proteína purificada en estas condiciones solo fue empleada para las reacciones enzimáticas.

**4.1.6 Tratamiento térmico:** La proteína purificada por afinidad a níquel, fue sometida a tratamiento térmico a 68°C por 20 minutos, y tras 5 minutos en hielo, se recuperó la fracción termoestable (soluble) mediante centrifugación a 13000 x g. En el caso del tratamiento en condiciones astringentes, el tratamiento térmico se realizó de forma similar pero a 75°C.

**4.1.7 Análisis por SDS-PAGE:** Para analizar la expresión de la  $\alpha$ -AaGlyRS una alícuota de proteína fue hervida durante 5 minutos en amortiguador de carga de Laemmli para proteínas, lo que fue esencial para disociar los dímeros de la subunidad; 16  $\mu$ g de proteína fueron cargados en el gel para electroforesis en poliacrilamida con sodio dodecilo sulfato (SDS-PAGE) al 10%, que fue teñido con azul de Coomassie .

**4.1.8 Concentración y desalado:** La subunidad  $\alpha$  de la AaGlyRS purificada, en los casos de los estudios estructurales, fue desalada con una columna de HiTrap Desalting (GE) de 5mL a un flujo de 5 mL/min en el amortiguador conteniendo Hepes 50 mM, NaCl 50 mM y DTT 1 mM y se concentró por ultrafiltración con un centricon de 30 kDa.

## 4.2 Síntesis del GSAd

El glicil sulfamoil adenilato, 5'-O-[N-(L-glicil)sulfamoil] adenosina, GSAd o G5A, fue sintetizado y purificado por un colaborador de nuestro grupo de investigación (Grötli *et al.*, según Redwan *et al.*, 2012). Se realizó una dilución del polvo en agua destilada a 100mM.

## 4.3 SEC-MALS

La proteína empleada para las reacciones enzimáticas fue purificada en condiciones

astringentes, y posteriormente analizada por cromatografía de exclusión de tamaño (SEC) acoplada a dispersión de luz de ángulo múltiple (MALS), con el equipo DAWN-Heleos II (Wyatt) y con el detector de índice de refracción Optilab T-rEX (Wyatt). En la columna Superdex S-200 10/300 (24 mL, GE) equilibrada con el amortiguador conteniendo Hepes 50 mM, NaCl 50 mM y DTT 1mM a un flujo de 0.5 mL/min, se inyectaron 100  $\mu$ L de la subunidad  $\alpha$  de la AaGlyRS a una concentración de 13 mg/mL o 6 mg/mL y se recuperó el eluido para emplearse en los ensayos de actividad. Los datos fueron procesados con el software del equipo para determinar el PM y el estado oligomérico en solución.

#### **4.4 Ensayos de actividad.**

**4.4.1 Reacción de activación del aminoácido:** Se cuantificó directamente la síntesis del glicil-adenilato (Gruic-Sovulj *et al.*, 2005), como una forma alterna al método anteriormente empleado para medirlo por la reacción inversa (por incorporación de [ $^{32}$ P]PPi).

La mezcla de reacción se realizó en el amortiguador de reacción (50 mM de Tris pH 8.0, 50 mM KCl) adicionando 10 mM MgCl<sub>2</sub>, 0.5 mM ATP frío, 0.25  $\mu$ Ci of [ $\alpha$ - $^{32}$ P]ATP y distintas concentraciones de glicina. Se pre-incubó la mezcla 5 min a 45°C y se comenzó la reacción por la adición de 40  $\mu$ M de  $\alpha$ -AaGlyRS. Se tomaron alícuotas de 1  $\mu$ L cada 10 minutos, deteniendo la reacción con 4  $\mu$ L de amortiguador de paro (SDS 0.1% y 400mM de acetato de sodio). De la reacción detenida se tomaron 2  $\mu$ L que fueron analizados por cromatografía de capa fina (TLC) en placas pre-lavadas de PEI- celulosa. Las machas fueron desarrolladas en acetato de amonio 100 mM / ácido acético al 5%.

Las placas de TLC fueron secadas y expuestas a una pantalla con capa de fósforo (general propose, Amersham Biosciences) toda la noche para monitorear la exigencia de Gly-[ $^{32}$ P]AMP, [ $^{32}$ P]AMP y [ $\alpha$ - $^{32}$ P]ATP. La imágenes fueron obtenidas sobre un dispositivo de captura de imágenes Amersham Storm 820 Imager, y cuantificadas con el programa ImageQuant T1 v2005. La fracción total de [ $^{32}$ P]AMP fue graficada contra el tiempo para obtener las velocidades iniciales. Las pendientes de los datos lineales fueron graficadas contra la concentración de glicina y ajustadas a la ecuación hiperbólica de Michaelis-Menten usando el programa Prism6.

La reacción se realizó tanto con la  $\alpha$ -AaGlyRS como con la AaGlyRS completa

reconstituida de ambas subunidades ( $\alpha_2\beta_2$ ) (anexo 1 figura 9, anexo 2 reconstitución de la AaGlyRS y figura 15).

**4.4.2 Reacción de aminoacilación del tRNA:** Para determinar la reacción de aminoacilación se utilizó el tRNA<sup>Gly</sup> de *A. aeolicus*. El tRNA<sup>Gly</sup> transcrito in vitro fue marcado radiactivamente en el extremo 3' intercambiando el último nucleótido por [ $\alpha$ -<sup>32</sup>P]ATP mediante la nucleotidil-tRNA transferasa de *Escherichia coli* (Ledoux y Uhlenbeck, 2008). El tRNA marcado fue purificado usando el kit de remoción de nucleótidos QIAquick (Quiagen), seguido de intercambio de amortiguador por ultrafiltración con un Amicon Ultra (Millipore) de 3 kDa.

La reacción fue ensayada en condiciones similares a las de activación de glicina pero empleando 5 mM de ATP frío en lugar de ATP marcado. Todas las reacciones fueron iniciadas por adición de la enzima. A los 5, 10 y 15 minutos, se tomaron alícuotas de 1  $\mu$ L en las que la reacción fue detenida con 10  $\mu$ L de amortiguador de paro (0.1% de SDS, 400 mM de acetato de sodio pH 5.2) con o sin la adición de 0.1 mM de la nucleasa P1 (Sigma) (Rodríguez-Hernández *et al.*, 2010). De la reacción detenida se tomaron 0.5  $\mu$ L que fueron analizadas por TLC de la misma forma como la reacción de activación de la glicina.

El tRNA aminoacilado con glicina al ser digerido con la nucleasa P1 produce Gly-[ $\alpha$ -<sup>32</sup>P]AMP (Rodríguez-Hernández, A. *et al.*, 2010), que pudo así ser comparado con el mismo producto resultado sólo del primer paso de la reacción, la activación de la glicina.

La reacción se realizó con la GlyRS de *A. aeolicus* reconstituida de sus subunidades ( $\alpha_2\beta_2$ ) y con una GlyRS bacteriana de subunidades codificadas en una sola cadena polipeptídica ( $(\alpha-\beta)_2$ ) de *Anaerolinea thermophila* (detalles de esta GlyRS en el cuerpo del anexo 2 y tabla 6 anexo 2).

## 4.5 Dispersión de rayos X de ángulo pequeño (SAXS)

La forma oligomérica de la subunidad  $\alpha$  de la GlyRS fue analizada por dispersión de rayos X de ángulo pequeño (SAXS). Se realizaron diluciones de la  $\alpha$ -AaGlyRS desde 10 mg/mL hasta 1.25 mg/ml, en el amortiguador E (Hepes 50 mM, NaCl 50 mM y DTT 1 mM) y se

colectó el patrón de dispersión en la línea P12 del anillo de almacenamiento Petra III del sincrotrón alemán (DESY), Hamburgo, Alemania. La intensidad de luz dispersada fue colectada como el vector de dispersión (transferencia de momento)  $s = 2 \sin \theta / \lambda$  a una longitud de onda de 0.124 nm a 23°C, con tiempos de exposición de 20x0.050 segundos.

El promedio de los datos fue normalizado y sustraído de la dispersión atribuida al disolvente. Los datos fueron procesados usando PRIMUS, de donde se calcularon los valores de la intensidad del haz directo  $I_0$  y de radio de giro  $R_g$  a partir de los patrones experimentales de SAXS usando la aproximación de Guinier; estos parámetros fueron también computados de la curva de dispersión completa usando la ley de Porod por el cálculo de la función de la distribución de distancias  $P(r)$  usando el programa GNOM (Svergun., 1992). Se usaron tres diferentes aproximaciones para determinar el peso molecular: el valor de  $I_0$ , el volumen de Porod (de la función  $P(r)$ ), y el volumen excluido (de los cálculos de DAMMIF) .

La estructura a baja resolución fue reconstruida *ab initio* por el programa DAMMIF (modelado por átomos marioneta con minimización por recocido simulado “versión rápida”) (Franke y Svergun, 2009) Los patrones teóricos de SAXS del dímero y el monómero fueron predichos con Crysol (Svergun *et al.*, 1995) a partir de la estructura cristalográfica y comparados gráficamente con los datos experimentales usando Sasplot (Konarev *et al.*, 2003). Los residuos flexibles faltantes de la etiqueta de 6XHis/TEV fueron modelados con Coral (Petoukhov *et al.*, 2012). La estructura cristalográfica con los residuos flexibles fue sobrepuesta con la envoltura *ab initio* usando Supcomb (Kozin y Svergun, 2001).

#### **4.6 Ensayo de termofluorometría diferencial (termofluor, desplazamiento térmico)**

Se realizaron ensayos de termofluorometría diferencial (de desplazamiento térmico) (Niesen *et al.*, 2007) para determinar la  $T_m$  y evaluar el efecto de la adición de los distintos ligados sobre la estabilidad de la proteína. A 1 mg de la  $\alpha$ -AaGlyRS, se agregaron distintas concentraciones de los ligandos GSAAd (0, 0.65, 1.25, 3.75, 7.5, 15, 30 mM) , ATP (0-10 mM) y Gly (0, 10, 50, 200 mM), sin Mg o con Mg 4 mM; en un volumen final de 10  $\mu$ L, y una concentración de 50X del fluoróforo hidrofóbico SYPRO orange (Sigma-Aldrich), en placas de 96 pozos y fueron leídas de 25 a 99.9°C con incrementos del 1% en el modo rápido de 40

minutos (equivalente a 1.8°C/min aproximadamente) en el equipo de PCR-tiempo real StepOne (Applied Biosystems).

#### 4.7 Cristalización

Se realizaron distintos escrutinios de cristalización empleando concentraciones de proteína de 15, 17, 21 y 25 mg/mL, a 18 y 28°C, de la proteína sola y en presencia de los ligandos ATP-Mg, ATP-Mg-Gly, GSAd solo o GSAd-Mg, con los reactivos del Crystal screen, Index screen y PEG-ion screen (Hampton Research). La presencia de cristales de proteína fue inspeccionada y corroborada bajo luz UV (295 nm, Crystal Light 100- Molecular Dimensions). La concentración de los ligandos fue de ATP 10 mM, de Mg 5mM en el caso de adicionarlo con ATP o 4 mM en el caso de adicionarlo con GSAd, GSAd 10 mM y glicina 10 mM.

Se obtuvieron cristales en algunas condiciones descritas en la tabla 5 (anexo 2) pero los que fueron empleados para la resolución de la estructura, se obtuvieron por la técnica de difusión de vapor en gota sentada con 1 µL la proteína a 15 mg/mL con ATP 10 mM/MgCl<sub>2</sub> 5 mM/Gly 10 mM y 1 µL de la solución de reservorio consistiendo de 30% de polietilenglicol monometil eter 2,000 y 100 mM de tiocianato de potasio, incubando las gotas durante 6 meses a 28°C. Transcurrido este periodo, las gotas fueron incubadas a 18°C y se obtuvieron los cristales en entre 1 a 2 semanas.

Los cristales que mejor difractaron, fueron sumergidos durante dos días en presencia de GSAd a 28 mM, mezclando 1.3 µL de la solución de la gota conteniendo los cristales, 2.5 µL de la solución del reservorio y 1.5 µL de GSAd 100 mM. Estos cristales fueron crioprotectidos preparando una solución idéntica a la del reservorio pero incluyendo 15% de glicerol.

Otras condiciones de cristalización menos relevantes y sus soluciones crioprotectoras se mencionan en el anexo 2 (tabla 5).

Para la crioprotección de los cristales, 7 µL de la solución madre se agregaron al pozo con la gota conteniendo los cristales, y con ello fueron transferidos a un portaobjetos silanizado (Hampton Research), donde se agregaron 7 µL de la solución de crioprotección. El montaje de los cristales se realizó pescándolos rápidamente con un asa (loop) de poliimida

LithoLoops Mesh (asas desde 0.4 mM hasta 40  $\mu$ M, con malla de 0.1 a 20  $\mu$ M, Molecular Dimensions) en el soporte del criovial magnético, bajo el microscopio estéreo (SMZ-U, Nikon) y el cristal montado fue sumergido en el nitrógeno líquido, tapado dentro del criovial y transferida al tanque de nitrógeno. Los cristales fueron transferidos en nitrógeno líquido hacia el sincrotrón en el dispositivo de transporte seco (DryShipper).

#### 4.8 Colecta de datos para cristalografía

La colecta de datos se realizó a una longitud de onda de 0.9785 Å a 100 K, en la línea 21-ID-G LS-CAT (Life Sciences Collaborative Access Team) del sincrotrón de la APS (Advanced Photon Source), del Argonne National Laboratory en Argonne, Illinois, USA.

Los crioviales con los cristales fueron transferidos al robot cargador automático de la estación, manteniendo siempre los crioviales sumergidos en nitrógeno líquido durante la transferencia. Todas las demás manipulaciones se realizaron a distancia desde la estación. Cada cristal fue centrado y se realizaron 2 exposiciones ("snap shots") para evaluar las características de difracción del cristal.

Para los datos que presentaron una difracción adecuada, se verificó la estrategia de colecta con el programa iMOSFLM, y se colectó el grupo de datos completo a una distancia del detector de 350 mm desde 253° a 1° con oscilaciones de 0.2°.

#### 4.9 Reducción de datos

El indizado e integración de los datos se realizó con iMOSFLM (Battye *et al.*, 2011) versión 7.1.3 en la suite de CCP4 o XDS versión Oct 15, 2015, siendo el más útil para la resolución de la estructura depositada iMOSFLM.

El grupo de datos empleado fue indizado y escalado en el grupo espacial ortorómbico P222 fijando la mosaicidad a 0.8°, y posteriormente reindizado con ReIndex al grupo espacial P2<sub>1</sub> 2<sub>1</sub> (Número 18), mismo que fue verificado con Pointless.

Debido a que de forma canónica se ha acordado reportar en cierto orden los grupos espaciales en los cristales, se verificó la posibilidad de trabajar en el grupo espacial P2<sub>1</sub> 2<sub>1</sub> 2 rotando los datos con ReIndex.

El escalamiento se realizó tanto con Scala (Evans, 2011) versión 3.3.22 como con Aimless versión 0.5.7 ambos de CCP4, eligiéndose los datos tratados con Aimless, realizando el corte del escalamiento a 2.81 Å (en la capa de más alta resolución 2.96-2.81) al juzgar los datos a la resolución en que el valor del  $I/\sigma$  fue mayor que 2 (2.4) en la capa de más alta resolución y el valor del  $R_{\text{merge}}$  global menor del 10% (global 8.4%).

#### 4.10 Resolución de la estructura de la $\alpha$ -AaGlyRS

Los programas empleados para trabajar con los datos pertenecieron a las suites de CCP4 versión 6.5.007 (Potterton *et al.*, 2003) y Phenix versión 1.10-2155-000 (Adams *et al.*, 2010).

**4.10.1 Determinación de las fases:** Para la resolución del problema de la fase se empleó Phaser en la suite de CCP4 (Phaser 2.5.7) o en Phenix (Phaser versión 2.6.0), así como Balbes versión 1.0.0.Nov\_16\_2011 (Long *et al.*, 2008) en CCP4, siendo seleccionados los resultados obtenidos por Balbes.

Para configurar la búsqueda de moléculas en la unidad asimétrica en Phaser, se determinó el índice de Matthews, el cual mostró la mayor probabilidad de tener 6 moléculas en la unidad asimétrica, considerando un porcentaje de disolvente de 49.34%. Sin embargo, en el reemplazo molecular sólo se encontraron 5 moléculas en la unidad asimétrica, con un porcentaje de disolvente de 57.71%.

Balbes encontró varias soluciones, eligiendo como modelo la estructura de la subunidad  $\alpha$  de la GlyRS de *Thermotoga marítima* en conformación apo con el código PDB 1j5w que presentó 64% de identidad con la  $\alpha$ -AaGlyRS. Phenix.phaser brindó estadísticas de TFZ=79.9 y LLGgain de =14878.239 para las 5 moléculas encontradas. Posterior al reemplazo molecular Balbes automáticamente realizó un afinamiento de cuerpo rígido con Refmac (R inicial/final 0.5290/0.4570,  $R_{\text{free}}$  inicial/final 0.5200/0.5020).

De forma inesperada una de las posibles soluciones encontrada por Balbes para el reemplazo molecular la obtuvo utilizando como molde la estructura de la alanil-tRNA sintetasa (2zzg), lo que nos llevó posteriormente a evaluar las similitudes locales de residuos y estructurales generales de estas dos sintetetasas.

Inmediatamente después del reemplazo molecular, se inspeccionó el modelo y los mapas de diferencias y de dobles diferencias de la solución encontrada, y pudo observarse en el mapa de diferencias contorneado a  $3\sigma$  la presencia de una densidad positiva en el sitio de unión del ATP en todas las moléculas de la  $\alpha$ -AaGlyRS en la unidad asimétrica.

**4.10.2 Construcción del modelo:** Mediante manipulaciones con Coot versión 0.8.1 (Emsley y Cowtan, 2004) se modificó el modelo arrojado por el reemplazo molecular para corresponder a la densidad electrónica, adicionando los residuos faltantes (61-65), ajustando a la densidad electrónica la totalidad de las cadenas laterales y regularizando la geometría de los residuos, hasta donde fue posible, centrándose en la cadena B que fue en la que la densidad electrónica era más clara. Posteriormente esta cadena fue sobrepuesta y reemplazadas las coordenadas de las otras cadenas por las de la cadena B construida. Se juzgó como adecuado el avance en la construcción del modelo cuando todos los residuos se encontraron dentro de las regiones favorecidas o permitidas en la gráfica de Ramachandran.

**4.10.3 Afinamiento:** El afinamiento inicial fue realizado con Phenix.refine empleando NCS y posteriormente con servidor PDB\_Redo (Joosten *et al.*, 2014), alternando con la construcción del modelo como se ha descrito.

**4.10.4 Localización del ligando:** 5 moléculas de ligando, una en cada molécula de la unidad asimétrica, fueron localizadas empleando inicialmente las constricciones encontradas por elBOW en Phenix para el G5A (GSAd), empleando el programa LigandFit en la suite de Phenix. Sin embargo, al realizar la validación, la geometría más adecuada para el ligando fue obtenida por las coordenadas/constricciones del GSAd presente en la estructura del fragmento catalítico de la alanil-tRNA sintetasa (código PDB 3hyo).

**4.10.5 Adición de moléculas de disolvente:** En el mapa de diferencias pudo verse claramente la presencia de dos moléculas de disolvente cercanas a la estructura correspondiente a la glicina en el GSAd. Las moléculas de disolvente fueron detectadas con Coot en el mapa de diferencias (DELWT PHDELWT) a 3.5 RMSD. Sin embargo, debido a la resolución de ésta estructura (2.81 Å) relativamente baja, se recomienda la adición de muy pocas moléculas de disolvente, y para demostrar la presencia de dichas moléculas de disolvente en todas las moléculas en la unidad asimétrica (excepto en la cadena D y E en que sólo una u otra de las moléculas de disolvente pudo ser vista por la densidad electrónica

menos clara presente para estas cadenas), sólo se asignaron dos moléculas de disolvente en la cadena B, donde la densidad electrónica estaba mejor definida.

**4.10.6 Mapa de omisión:** Se generó un mapa de omisión compuesto 2Fo-Fc de recocido simulado con Phenix (del 5% de los datos). En el contorno a  $1\sigma$  de este mapa, se observó sin lugar a dudas la presencia del ligando GSAd.

**4.10.7 Validación de la estructura:** La validación de la estructura se llevó a cabo inicialmente con Phenix (Xtriangle, polygon), pero fue corroborada directamente con las estadísticas arrojadas por el servidor de validación del wwPDB, previo a depositar la estructura. La estructura fue depositada con el código 5f5w en el PDB.

#### **4.11 Comparación del reconocimiento entre las GlyRSs.**

Para analizar los cambios conformacionales en la subunidad  $\alpha$  de la GlyRS bacteriana con la unión del GSAd, se realizó la superposición de la estructura en conformación apo (tanto con el PDB 1j5w de *Thermotoga maritima* como con el 3rgl de *Campylobacter jejuni*) con la estructura en presencia del ligando GSAd reportada en este trabajo (5f5w) con el programa Pymol versión 1.7.4.5. Para determinar de una forma cuantitativa el desplazamiento de cada uno de los residuos con la unión del GSAd en la  $\alpha$ -bacGlyRS se empleó el programa 3dss ("superposición estructural tridimensional").

Algunos de los alineamientos de coordenadas emplean el alineamiento de la secuencia como punto de referencia, como es el caso del programa de visualización Pymol. La comparación entre las dos distintas GlyRSs difícilmente puede ser realizada por alineamiento de secuencia solamente, por lo que se realizó la superposición de las coordenadas de las estructuras de la GlyRS bacteriana ( $\alpha$ -AaGlyRS, determinada en este trabajo) con la GlyRS eucariótica también en presencia del GSAd (*Homo sapiens*, código PDB 2zt8), usando la función que emplea el algoritmo SSM ("coincidencia por estructura secundaria") del programa Coot. También se realizó la superposición de las GlyRSs con la AlaRS de arquea en presencia del GSAd (2zzg) o en ausencia (2zzf PDB), con lo que fue posible comparar estructuralmente las posiciones equivalentes entre todas estas aaRS empleando Pymol.

## 4.12 Alineamientos estructurales entre las aaRSs II

Buscando establecer las relaciones estructurales entre las GlyRSs y de éstas con la AlaRS, más allá de sólo las relaciones de residuos del sitio activo, se compararon por alineamientos estructurales los CCDs de todas las aaRSs de clase II por distintos programas.

El CCD es la única porción conservada estructuralmente entre las aaRSs II, pues contiene el plegamiento característico y el sitio de unión al ATP y al aminoácido, por lo que la comparación entre las sintetasas de los distintos aminoácidos sólo es posible en la región del CCD. Una de las consideraciones más importantes para analizar las sintetasas, es tomar como homólogo el CCD entre todas las aaRSs II, es decir que existió un ancestro lejano común para todas ellas, debido a la conservación de su plegamiento. Los programas que realizan los alineamientos únicamente por secuencia son incapaces de detectar relaciones tan sencillas como agrupar las sintetasas de un mismo aminoácido provenientes distintos organismos, al compararlas con las sintetasas específicas para los otros aminoácidos.

La subunidad  $\alpha$  de la bacGlyRS contiene el CCD, por lo que representa a la bacGlyRS completa para las comparaciones con los CCDs de todas las aaRSs II. Para este trabajo se obtuvo el CCD de cada una de las aaRSs II localizando las regiones de su estructura correspondientes a los tres motivos característicos de las aaRSs II (motivos 1, 2 y 3) incluyendo las regiones intermedias y se retiraron todas las grandes inserciones (figura). Retirar las grandes inserciones fue importante para no sesgar el alineamiento estructural por la presencia de los dominios idiosincrásicos comunes entre las sintetasas de las distintas subclases. En el anexo 1 (figura 14) se muestra la secuencia completa de la región del CCD de una muestra representativa de las sintetasas.

Para poder definir la dirección de la evolución se requiere un árbol filogenético con raíz. Si un árbol filogenético no está enraizado se asume que no hay conocimiento de un ancestro común y a pesar de que muestra las posiciones relativas, no puede verse la dirección del camino evolutivo (Pevsner et al., 2015). Posterior a la publicación del artículo resultado de este trabajo (Valencia-Sánchez et al., 2016), algunos de los cladogramas fueron enraizados adicionando al alineamiento el equivalente del CCD de la holoenzima de la biotina sintetasa de *E. coli* (PDB 1bib) con resultados muy similares (figuras 7A, 7D y 13 anexo 1).

**4.12.1 Alineamientos estructurales con STAMP:** El programa Multiseq (“Alineamiento Múltiples de Secuencia”) (Russell y Barton, 1992) está incorporado como un programa conectado (“plugin”) en VMD versión 1.9.0, permite unificar el análisis por secuencia y estructura con un enfoque evolutivo. Multiseq emplea el programa STAMP para alineamiento estructural, STAMP (“alineamiento estructural múltiple de proteínas”) por medio de programación dinámica minimiza las distancias entre los C $\alpha$  (el RMSD) entre residuos alineados empleando rotaciones y translaciones de cuerpo rígido y mínimos cuadrados, agrupa las estructuras por pares y de cada nodo hace una salida del alineamiento estructural que reporta a la rama, hasta llegar a la raíz del árbol, construyendo así el alineamiento múltiple. Se realizó el alineamiento estructural con STAMP del CCD de 111 de las distintas aaRSs de clase II, y se construyeron dendrogramas en el mismo programa agrupando a las proteínas con base a RMSD, el valor de similitud estructural Qh y % de identidad de las regiones estructurales equivalentes (Russell y Barton, 1992).

**4.12.2 Matriz de puntuación de alineamientos estructurales con DALI:** Ya que DaliLite (alineamientos estructurales por parejas empleando matrices de distancias entre subestructuras y elementos de estructura secundaria) (Holm y Sander, 1995) realiza alineamientos estructurales considerados puramente estructurales e independientes de la secuencia, se realizaron en el servidor en línea alineamientos 1:1 entre las estructuras del CCD cada una de las distintas aaRSs II. Al comparar el valor de la puntuación arrojado por DALI, se observó que las sintetasas pertenecientes a cierta subclase conocida estaban sobre un valor de 18.5, por lo que se construyó una matriz agrupando las sintetasas que presentaban una puntuación Z por encima de este valor. También a partir de las sobreposiciones realizadas con DaliLite, fue posible establecer los residuos conservados en posiciones realmente equivalentes entre las dos GlyRSs y la AlaRS.

**4.12.3 Análisis de componentes principales (PCA) con Bio3d:** Se realizaron alineamientos estructurales y su interpretación por PCA con el programa Bio3D versión 2.2-0 (Grant *et al.*, 2006). Se emplearon los C $\alpha$  de los CCD de 42 estructuras incluyendo todas las aaRSs II conocidas con ligados y sin ellos. Con el programa Bio3D que corre sobre el lenguaje R, se alinearon las estructuras y se encontró el núcleo común invariante. Con base

al núcleo común se determinó el RMSD de desviación de cada una de las estructuras hacia este núcleo, se agruparon las estructuras por cercanía (RMSD) y se realizó un dendrograma con base en el agrupamiento. Basado en el alineamiento estructural de este núcleo invariante se realizó su análisis en los principales componentes (PCA) de las diferencias entre las estructuras de las aaRSs II, en el cual, la variación total de las estructuras respecto al núcleo común se descompone en fracciones de co-variación en todas las proteínas y se asigna a cada elemento de variación una jerarquía por orden decreciente según su efecto sobre la variación total (como componente 1, 2, 3, etc.). Posteriormente se agruparon y se obtuvo la gráfica de córfmeros de las estructuras en el subespacio bidimensional entre los dos principales componentes, y también se realizó un dendrograma basado en el agrupamiento por distancia según el componente principal PC1.

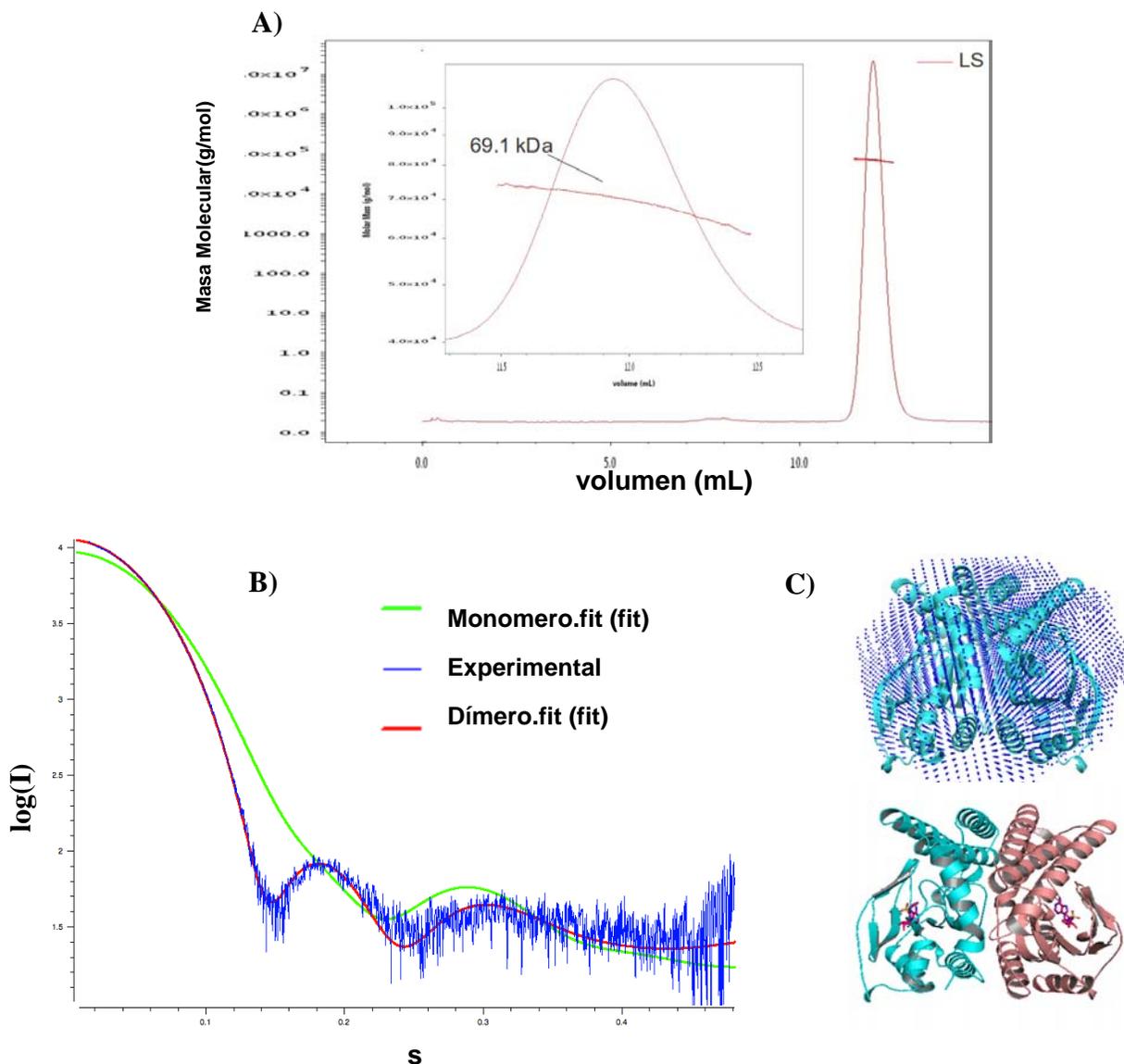
**4.12.4 Alineamientos de secuencia entre elementos estructurales equivalentes con T-COFFEE Expresso:** Las regiones equivalentes del núcleo invariable del CCD de las aaRSs II fueron evaluadas y recortadas manualmente en Pymol. Se obtuvieron las secuencias de dichas regiones en formato fasta y fueron alineadas con el servidor T-COFFEE, en su modalidad Expresso (Armougom *et al.*, 2006) para alineamiento de secuencias empleando información estructural. El alineamiento múltiple fue visualizado con Jalview versión 1.0 y se generaron los dendrogramas por distancia media (similitud medida con las matrices PAM250, BLOSUM6) o por vecinos (PAM250). Este procedimiento brindó un alineamiento por secuencia de elementos estructurales equivalentes. De forma paralela se construyó un dendrograma con PhyML (Han y Zmasek, 2009), empleando la salida de T-COFFEE con o sin la ponderación de las posiciones en el alineamiento por TCS (puntuación de consistencia transitoria) para ser estructuralmente correctas, y fue visualizado con Archaeopteryx incluido en el paquete (Zmasek y Eddy, 2001).

## 5. RESULTADOS

### 5.1 La subunidad $\alpha$ de la bacGlyRS es un dímero en solución

Para esclarecer la controversia existente del estado oligomérico en solución de la subunidad  $\alpha$  de la bacGlyRS realizamos su análisis por SEC-MALS (“cromatografía de exclusión molecular acoplada a dispersión de luz de ángulo múltiple”), así como por SAXS (“dispersión de rayos X de ángulo pequeño”).

Previo a los análisis de estado oligomérico, la pureza de la subunidad  $\alpha$  de la AaGlyRS fue verificada mediante SDS-PAGE (figura 15, anexo 2). El análisis por SEC-MALS de la  $\alpha$ -AaGlyRS mostró la presencia de una única especie de un peso molecular de 69.1 kDa correspondiente al dímero (figura 1A). También se llevó a cabo un estudio por SAXS, en donde se realizaron diluciones desde 10 mg/mL hasta 1.25 mg/mL, obteniéndose un peso molecular promedio de 68.8 kDa, y se corroboró la estructura dimérica en solución de esta subunidad al comparar la curva teórica de dispersión que produciría un monómero y un dímero con la curva del patrón de dispersión experimental (figura 1B).

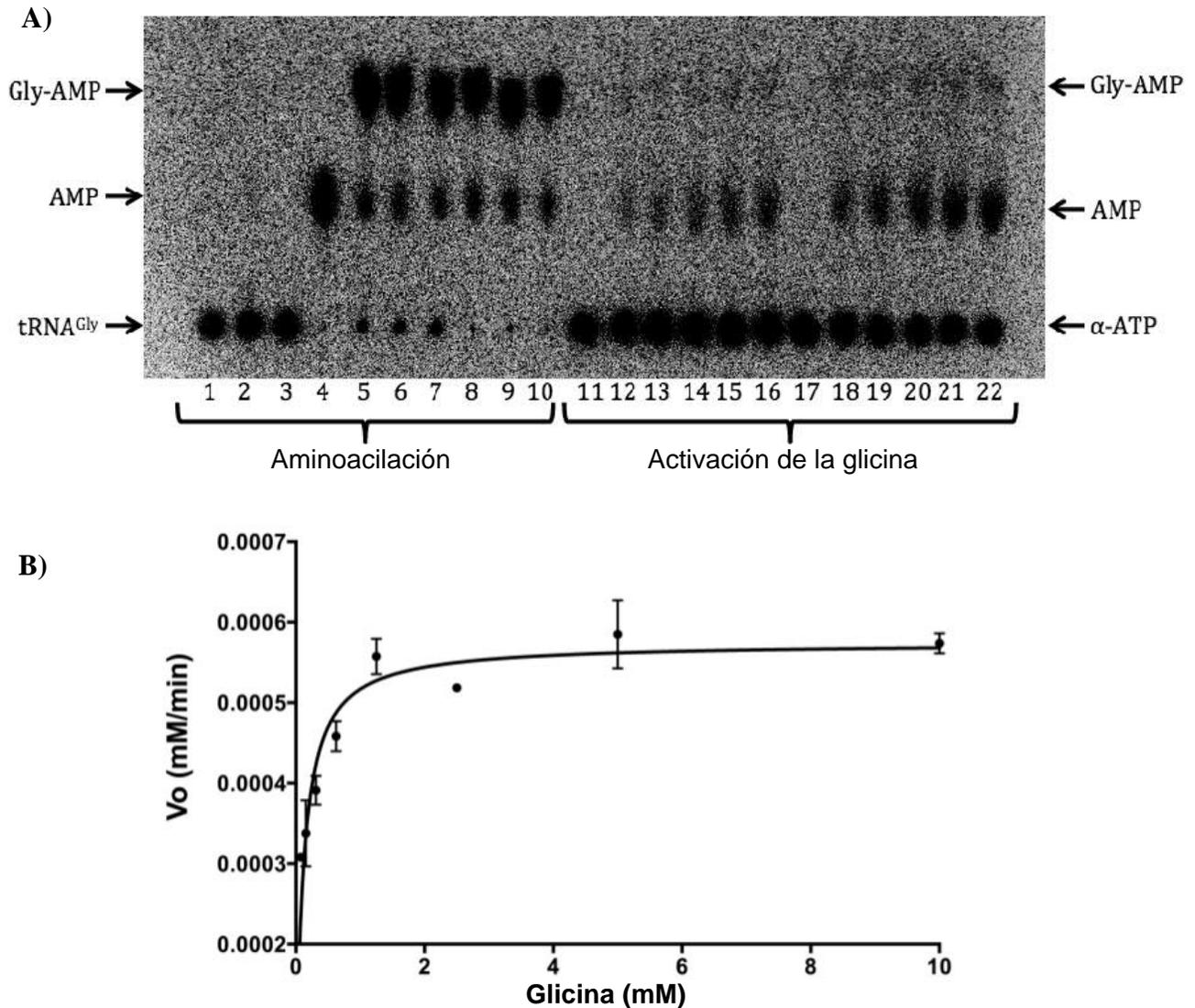


**Figura 1. La subunidad  $\alpha$  de la AaGlyRS es un d mero en soluci n.** **A)** An lisis por SEC-MALS que indica que la subunidad  $\alpha$  AaGlyRS presenta en soluci n una masa molecular de 69.1 kDa, correspondiente a un d mero. Se muestra la se al de dispersi n de luz graficando el peso molecular contra el volumen de eluci n de la columna Superdex S-75 10/300 con un flujo de 0.5 mL/min. Se inyectaron 100  $\mu$ L de muestra a una concentraci n de 6 mg/mL, en amortiguador HEPES 20 mM pH 7.0/NaCl 25 mM/ DTT 1 mM. **B)** An lisis por SAXS donde se muestra el ajuste de la curva te rica de dispersi n que producir a un d mero (rojo) y un mon mero (verde) a los datos experimentales (azul). La masa molecular determinada por distintos m todos proveniente del an lisis de los datos de SAXS fue un promedio de 68.8 kDa. Las caracter sticas de la muestra fueron similares a las del inciso A. Las curvas de dispersi n te ricas calculadas a partir de la estructura del d mero y el mon mero fueron determinadas con el programa CRYSOLOG (Svergun *et al.*, 1995) y el ajuste a los datos experimentales observado con el programa Sasplot. **C)** En la parte superior se muestra la envoltura determinada por modelado *ab initio* con el programa DAMMIF (Franke y Svergun, 2009) superpuesta con la estructura cristalogr fica del d mero de la subunidad  $\alpha$ -AaGlyRS (mostrada en la parte inferior).

## 5.2 La subunidad $\alpha$ de la Aa-GlyRS por sí sola presenta actividad de activación del aminoácido

La mayoría de los reportes de actividad de activación de la glicina por la bacGlyRS habían indicado que era necesaria la presencia de ambas subunidades de esta enzima para llevar a cabo el primer paso de la reacción, la activación del aminoácido con el ATP, y también el segundo paso, la aminoacilación del tRNA. Sin embargo, un miembro de nuestro grupo de investigación (Dra. Annia Rodríguez Hernández) encontró que la subunidad  $\alpha$  de la bacGlyRS presenta actividad en el primer paso de la reacción, presentando una  $K_m$  de 110  $\mu M$ , pero una  $K_{cat}$  de sólo  $2.3 \times 10^{-4} \text{ s}^{-1}$  (figura 2B).

La determinación de la activación del aminoácido por la subunidad  $\alpha$  de la AaGlyRS se llevó a cabo mediante una metodología que permite visualizar directamente la producción del aminoacil-adenilato (glicil-adenilato) (figura 2A), lo que constituyó una diferencia respecto a los ensayos reportados con anterioridad donde se determinaba la actividad del intercambio de ATP-PPi. Aprovechando que la reacción es reversible, se determina la incorporación de  $[^{32}P]$ -PPi en presencia de glicina en la reacción inversa. En el presente trabajo se determinaron las curvas de actividad enzimática y las constantes de actividad visualizando directamente la síntesis del glicil- $[^{32}P]$ AMP de la  $\alpha$ -AaGlyRS, pero cuantificando su hidrólisis espontánea en glicina y  $[^{32}P]$ AMP (figura 2A). Los controles del producto de la reacción (presuntamente el glicil- $[^{32}P]$ AMP en todos los casos) migraron todos de forma idéntica en la cromatografía de capa fina PEI-celulosa; el producto del 1er paso de la reacción con sólo la subunidad  $\alpha$ , el producto del mismo 1er paso de la reacción con la enzima completa reconstituida y el 2o paso de la reacción llevado a cabo con la enzima completa empleando el tRNA marcado digerido con la nucleasa P1 (métodos punto 4.4) (figura 2A).



**Figura 2. La subunidad  $\alpha$  de la AaGlyRS por sí sola presenta actividad de activación del aminoácido.** **A)** Análisis por cromatografía de capa fina en donde se muestran las bandas correspondientes a la migración de los productos de la reacción, carriles 1-10 de aminoacilación, carriles 11-22 de activación de la glicina. En el caso de la reacción de activación del aminoácido (11-22) se determinó la producción de glicil- $^{32}\text{P}$ -adenilato a partir de  $[\alpha\text{-}^{32}\text{P}]\text{ATP}$  (Gruic-Sovulj *et al.*, 2005). El glicil- $^{32}\text{P}$ -adenilato se hidroliza rápidamente en glicina y  $^{32}\text{P}$ -AMP. Como control positivo se empleó la reacción de aminoacilación el tRNA-CC- $[\alpha\text{-}^{32}\text{P}]\text{ATP}$  con glicina, que posteriormente es degradado con la RNasa P1, lo que produce glicil- $^{32}\text{P}$ -adenilato (Rodríguez-Hernandez, 2010), el cual también se hidroliza. Como los productos de las reacciones son idénticos pueden ser comparados directamente, demostrando que la subunidad  $\alpha$  por sí sola presenta activación de la glicina. **Reacciones de aminoacilación del tRNA.** Carriles 1-3 Controles respectivos (sin GlyRS, sin nucleasa y tiempo cero). Carril 4 reacción del tRNA en ausencia de glicina tratado con la nucleasa. Carriles 5-7 reacciones del tRNA con la AaGlyRS  $\alpha_2\beta_2$  reconstituida, puntos tomados cada 5 minutos. Carriles 8-10 reacciones del tRNA con la GlyRS completa de subunidades fusionadas de *Anaerolinea thermophila* ( $\alpha\text{-}\beta$ )<sub>2</sub> (detalles anexo 1 figura 11, anexo 2 tabla 6) puntos tomados cada 5 minutos. **Reacciones de activación de la glicina.** Carriles 11 y 17 controles respectivos (sin glicina), Carriles 12-16 sólo con la subunidad  $\alpha$ -AaGlyRS tomadas cada 10 minutos, carriles 18-22 con la AaGlyRS

$\alpha_2\beta_2$  completa tomadas a cada 10 minutos. **B)** Gráfica de Michaelis-Menten de las velocidades iniciales de las reacciones activación de la glicina con la subunidad  $\alpha$  de la AaGlyRS empleando [ $\alpha$ - $^{32}\text{P}$ ]ATP (continúa de la página anterior).

### **5.3 Estructura tridimensional de la subunidad $\alpha$ de la Aa-GlyRS en complejo con un análogo del estado de transición**

Se obtuvieron monocristales de la subunidad  $\alpha$  de la glicil-tRNA sintetasa de la bacteria *Aquifex aeolicus* ( $\alpha$ -AaGlyRS) que fueron sumergidos en presencia de un análogo del glicil-adenilato, el glicil-sulfamoil adenilato. Se realizó la colecta de datos de difracción de los cristales en la línea del sincrotrón de la Advanced Photon Source, USA (métodos punto 4.8) y se determinó la estructura a 2.81 Å de resolución de la  $\alpha$ -AaGlyRS en complejo con el GSAd por reemplazo molecular (las estadísticas se presentan en la tabla 3 y la figura 16 del anexo 2), empleando la estructura de la subunidad  $\alpha$  de la GlyRS en conformación apo de *Thermotoga maritima* (código PDB 1j5w) como modelo de búsqueda inicial. El cristal, perteneciente al grupo espacial P22<sub>1</sub>2<sub>1</sub> (por convención P2<sub>1</sub>2<sub>1</sub>2), con una celda unitaria de 101.8, 130.0, 145.5 Å y 90.0, 90.0, 90.0° (tabla 3), presentó 5 moléculas en la unidad asimétrica. En el mapa de diferencias (figura 3A) y posteriormente en el mapa de omisión (figura 3B) se pudo ver claramente la densidad electrónica positiva correspondiente al GSAd en todas las moléculas de la unidad asimétrica (tabla 4, figuras 3A y 3B).

<b>Tabla 3.</b> Estadísticas colección de datos y de afinamiento	
	$\alpha$ -AaGlyRS-GSAd
<b>Colecta de datos</b>	
Grupo espacial	P22 <sub>1</sub> 2 <sub>1</sub>
Dimensiones de la celda	
a, b, c (Å)	101.8, 130.0, 145.5
$\alpha, \beta, \gamma$ (°)	90.0, 90.0, 90.0
Resolución (Å)	83.43-2.81 (2.91-2.81) *
$R_{sym}$ o $R_{merge}$ **	0.084 (0.584)
$I/\sigma I$	9.7 (2.2)
Integridad (%)	100 (100)
Redundancia	4.4 (4.5)
CC <sub>(1/2)</sub> Coeficiente de correlación de medio-grupo de datos de (I) la media	0.990 (0.770)
Factor B de Wilson (Å <sup>2</sup> )	58.4
<b>Afinamiento</b>	
Resolución (Å)	80.19-2.81 (2.86-2.81)
No. de reflexiones	47741(2607)
$R_{work}/R_{free}$	0.239/0.252 (0.362/0.363)
No. de átomos	
Proteína	11740
Ligando/ion	135
Moléculas de disolvente	2
Factores B	
Proteína (por cadena)	45.1, 53.6, 56.9, 69.6, 73.1
Ligando/ion	40.8, 47.5, 52.6, 68.0, 74.6
Moléculas de disolvente	45.8
R.m.s desviaciones	
Longitudes de enlace (Å)	0.011
Ángulos de enlace (°)	1.462
<i>Estadísticas adicionales se muestran en la figura 16 del anexo 2</i>	

Se usó un solo cristal.

\*Los datos para la capa de más alta resolución se muestran en paréntesis

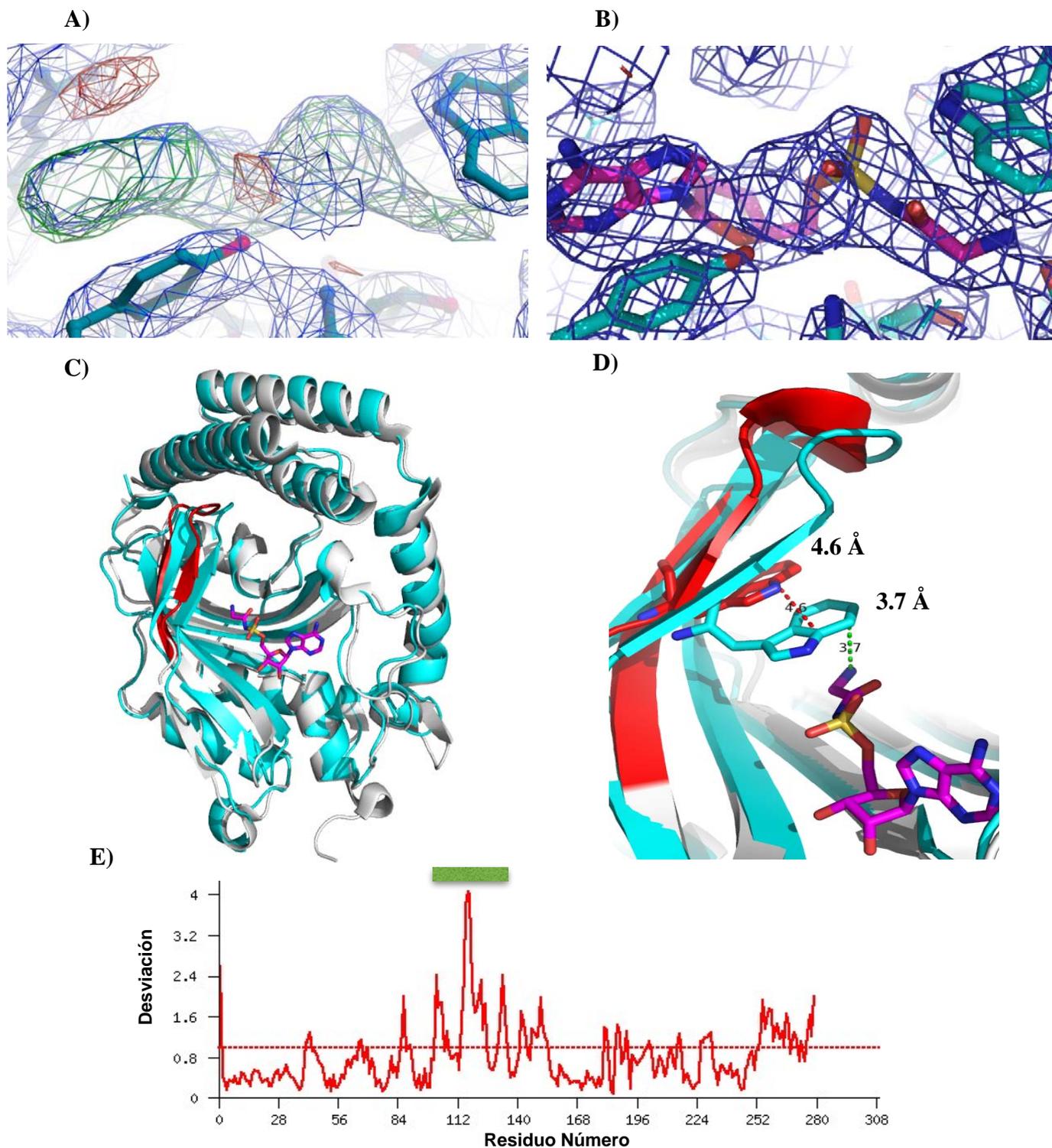
\*\* Las fórmulas se indican en la figura 16 del anexo 2

#### 5.4 La unión del GSAd promueve cambios conformacionales en la subunidad $\alpha$ de la bacGlyRS.

Al comparar la estructura de la  $\alpha$ -AaGlyRS en complejo con el GSAd resultado de este trabajo (que fue depositada con el código PDB 5f5w) con las estructuras de la subunidad que están en conformación apo (de *T. marítima* con código PDB 1j5w y de *Campylobacter jejuni* con código PDB 3rf1) (figura 3C), pudo observarse el desplazamiento de la región de los residuos 112-123, siendo el movimiento más significativo el del Trp 115, de aproximadamente 4.6 Å (figuras 3C a 3E). Gracias a este movimiento, el Trp 115 interacciona con la porción amino del grupo glicilo en el GSAd (figura 3D). El Trp 115 también forma un puente de hidrógeno con el equivalente del fosfato del AMP (O1S en el GSAd) (figura 3D).

Los cambios estructurales en la subunidad  $\alpha$  de la AaGlyRS al unirse el GSAd correspondieron a una estabilización térmica de la proteína que fue determinada mediante un análisis de desplazamiento térmico por fluorometría diferencial de barrido (anexo 1, figura 12). Esta estabilización fue de aproximadamente 12 °C en presencia del GSAd respecto a la proteína sin ligandos (anexo 1, figura 12).

Definir el reconocimiento de la glicina por la bacGlyRS no había sido posible debido a que en las estructuras en presencia de otros ligandos no se había producido ningún cambio conformacional. Gracias a la claridad del mapa de densidad electrónica en la región del ligando y a los cambios conformacionales producidos con la unión del análogo del aminoacil-adenilato (GSAd) (figura 3A, 3B y 3D, tabla 4), nos fue posible establecer cómo ocurre el reconocimiento de la glicina en la  $\alpha$ -AaGlyRS.



**Figura 3. La unión del GSAd promueve cambios conformacionales en la subunidad  $\alpha$  de la *bacGlyRS*.** A) Mapas de densidad electrónica Fo-Fc contorneado a  $3\sigma$  (verde y rojo) y 2Fo-Fc contorneado a  $1\sigma$  (azul), donde se observa la presencia de una densidad positiva (verde) en el sitio de unión del aminoacil-adenilato. B) Mapa de omisión compuesto 2Fo-Fc de recocido simulado de la densidad electrónica de la estructura de la subunidad  $\alpha$ -AaGlyRS en complejo con el GSAd

(magenta) contorneado a  $1\sigma$ , donde se muestra claramente la presencia del ligando. C) Superposición en listones de la subunidad  $\alpha$  de la bacGlyRS en conformación apo (en blanco) (PDB 1j5w) y en complejo con el GSAd (en cian) (PDB 5f5w determinada en este trabajo), donde se muestra la región del cambio conformacional más evidente (en rojo) provocado por la unión al GSAd (GSAd en magenta). D) Acercamiento a la región del mayor cambio conformacional de inciso C, donde se muestra el Trp115 que se desplaza alrededor de 4.6 Å para interactuar con el GSAd. E) Gráfica del RMSD por residuo entre las estructuras del inciso C, construida usando el servidor 3dss (continúa de la página anterior).

<b>G5A localizado por Phenix</b>	<b>Puntuación de ajuste</b>	<b>CC global</b>	<b>Átomos colocados</b>	<b>Factores B G5A</b>	<b>Factores B proteína</b>
Ligand_fit_1_2.pdb	87.8	0.88	27	40	45.1
Ligand_fit_1_1.pdb	89.7	0.89	27	47	53.6
Ligand_fit_1_3.pdb	87.0	0.88	27	52	56.9
Ligand_fit_1_4.pdb	77.2	0.83	27	68	69.6
Ligand_fit_1_5.pdb	74.6	0.83	27	74	73.1

## **5.5 El reconocimiento de la glicina por la GlyRS bacteriana y la GlyRS presente en eucariotas ocurren de forma distinta.**

Debido a que todos los elementos para el reconocimiento del aminoácido y el ATP se encuentran en la subunidad  $\alpha$  de la bacGlyRS, fue suficiente el análisis de esta subunidad en complejo con el análogo del glicil-adenilato para definir dicho reconocimiento. Las conclusiones pueden ser extensivas a la mayoría de los organismos que la presentan, por el alto grado de conservación de esta proteína, presentando una identidad mayor a 50% entre los phyla más distantes, pero de 100% en los residuos importantes.

Se compararon los residuos encargados del reconocimiento del aminoácido y el ATP entre las dos GlyRSs, que solo pudieron ser comparadas mediante alineamiento estructural de la estructura determinada en este trabajo y la eukGlyRS (GlyRS de *Homo sapiens*, código PDB 2ZT8) (figuras 4C y 4D), ya que entre ellas la identidad de secuencia es muy baja (12.5 %), que sería lo reportado en un alineamiento entre dos secuencias no relacionadas (Kleiger *et al.*, 2000). Las dos estructuras comparadas se encuentran en complejo con el GSAd (figuras 4A y 4B).

El reconocimiento del ATP por ambas GlyRSs es muy semejante a la de la mayoría de las aaRSs de clase II, y ambas GlyRSs comparten el plegamiento característico de dicha clase (figuras 4A y 4B), además de 7 residuos, de los cuales 5 llevan a cabo el reconocimiento del ATP y son altamente conservados en todas las aaRSs II (figuras 6B y 6D). Los residuos compartidos y además en posiciones estructuralmente equivalentes fueron evaluados a partir del alineamiento estructural entre las dos tipos de GlyRS con el programa DALI.

### **5.5.1 Reconocimiento de la glicina por las dos GlyRSs**

El reconocimiento de la glicina se lleva a cabo por la carga negativa de una red de glutámicos en la eukGlyRS, mientras en la bacGlyRS se lleva a cabo por una red polar formada por varios residuos distintos.

#### **5.5.1.1 Reconocimiento de la glicina por la GlyRS presente en eucariotas (eukGlyRS) (de origen arquea)**

Como fue definido por Arnez *et al.*, 1999, el reconocimiento de la glicina se lleva a cabo en el extremo amino por 3 residuos idénticos que son los Glu 522, 296 y 245, cuya carga es equilibrada por la Arg 410, sin equivalente en la bacGlyRS. Todos estos residuos son absolutamente conservados (figura 4D). El reconocimiento del extremo amino se basa en una cavidad con carga negativa formada sólo por residuos de glutámico (figura 4D).

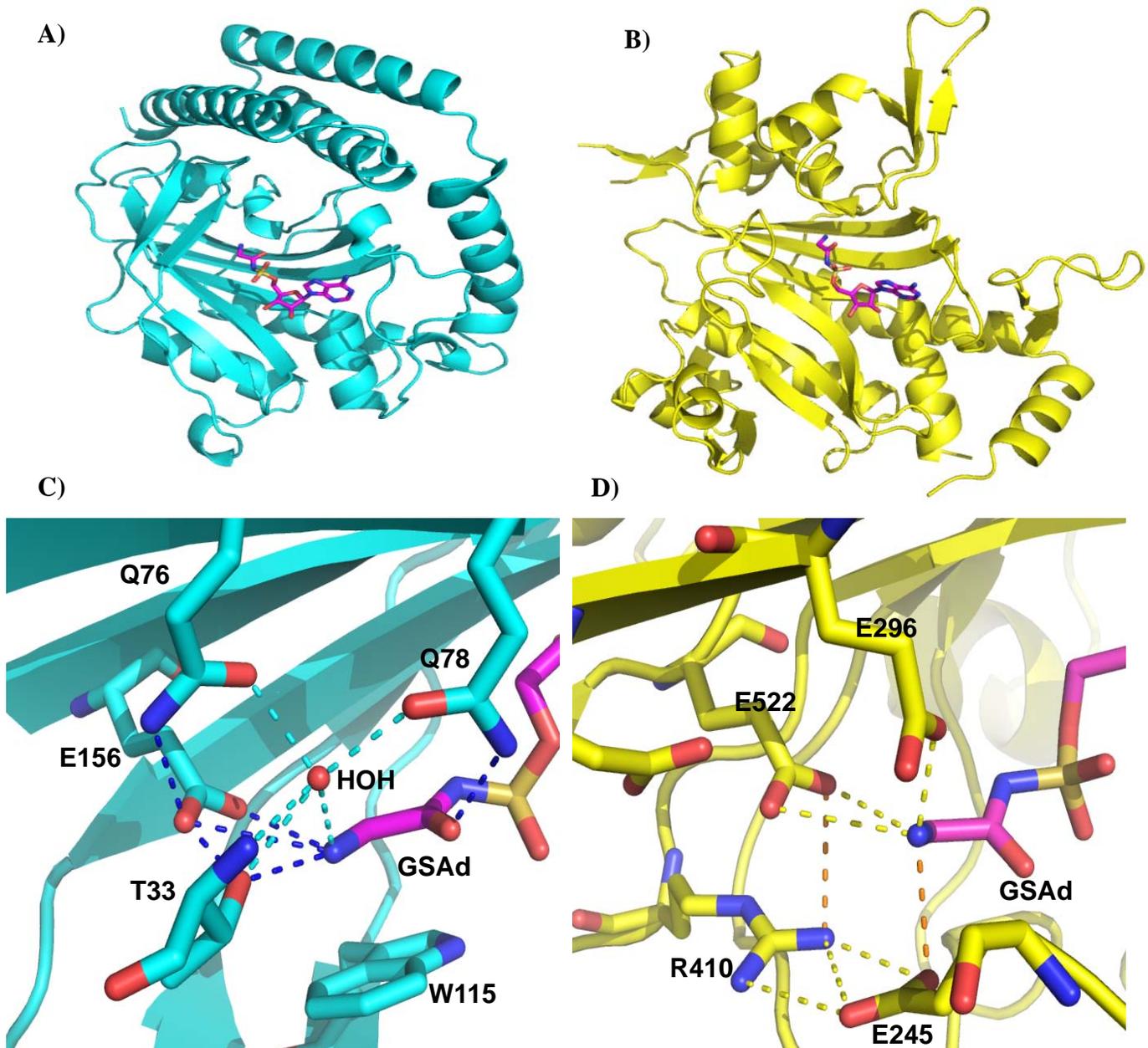
#### **5.5.1.2 Reconocimiento de la glicina por la GlyRS de origen bacteriano (bacGlyRS)**

El reconocimiento de la glicina se lleva a cabo por 5 residuos distintos muy conservados. En el extremo amino, la glicina es reconocida por el Trp 115 por una interacción catión- $\pi$ , por la Thr 33, el Glu 156, y por medio de dos moléculas de disolvente estabilizadas por puentes de hidrógeno con la Gln 76 y Gln 78 (figura 4C). La Gln 76 reconoce también a la glicina por el extremo carboxilo. El reconocimiento se basa en una red de interacciones polares (electrostáticas) diversas, destacando una interacción catión- $\pi$  y una red de puentes

de hidrógeno formada por residuos distintos (figura 4C).

Para el Glu 156, existe un equivalente tanto en la eukGlyRS (Glu 522) como en la AlaRS (Asp235) (figuras 4C, 4D, 5C y 5D).

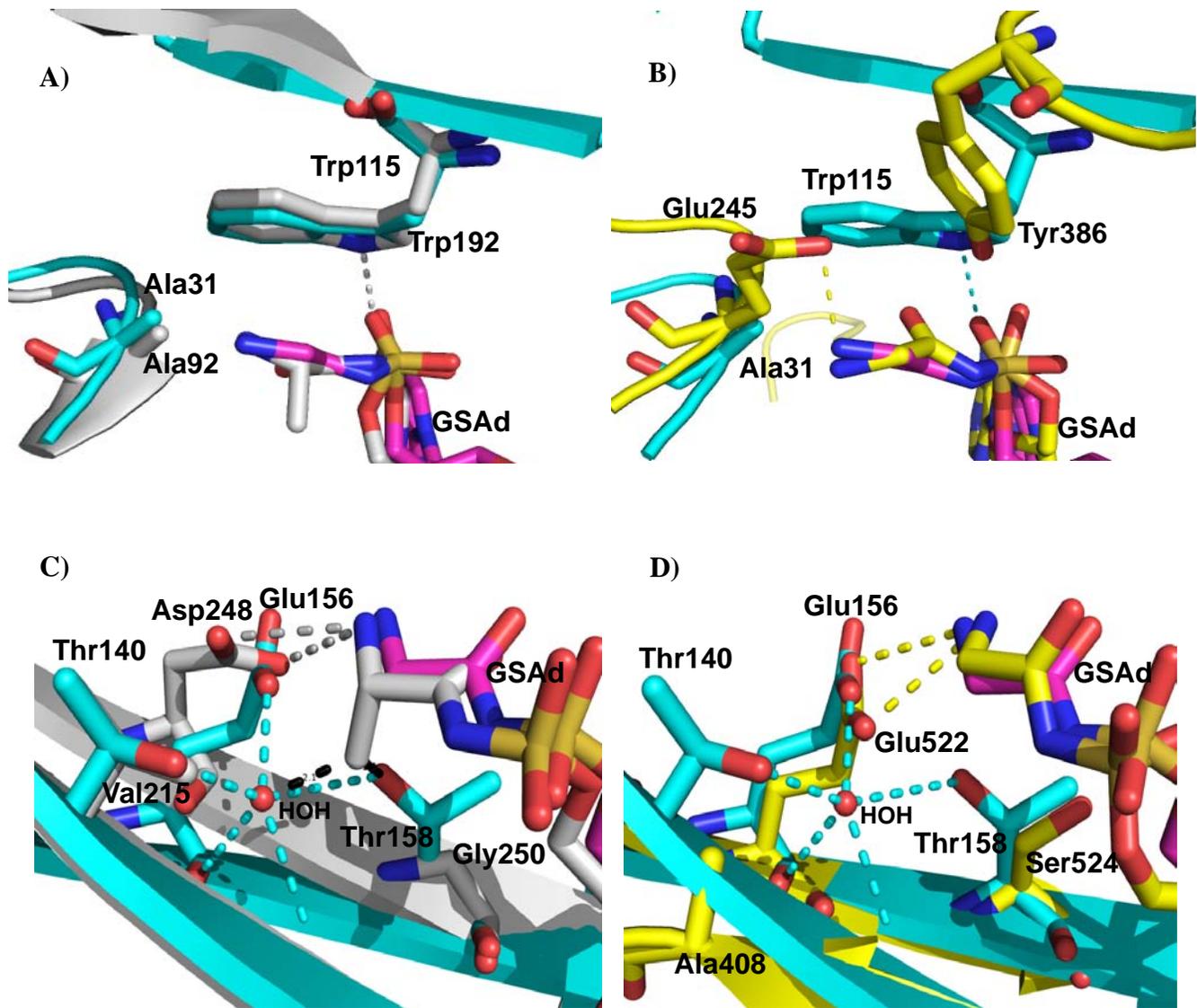
La interacción del triptofano 115 con la glicina, de tipo catión- $\pi$ , posiblemente incrementa la presencia del isómero protonado de la glicina (figura 4C, 5A, 8A) (Gallivan y Dougherty, 2000). Los cálculos teóricos de la energía de este tipo de interacción son de cerca de -5.5 kcal/mol en medio acuoso (figura 8B), mayor que la de un puente salino (-2.2 kcal/mol) (Gallivan y Dougherty, 2000).



**Figura 4. El reconocimiento de la glicina por la bacGlyRS y la eukGlyRS ocurren de forma distinta.** A) Estructura global de la subunidad  $\alpha$  de la bacGlyRS. B) Estructura global del dominio catalítico de la eukGlyRS C) Reconocimiento de la glicina en la bacGlyRS por medio de una interacción catión- $\pi$  con el Trp 115 y una red polar con los residuos Thr 33, Glu 156, Gln 78 y puentes de hidrógeno de una molécula de disolvente con la Gln 76. La Gln 78 también reconoce el extremo carboxilo de la glicina. D) Reconocimiento de la glicina en la eukGlyRS por medio de un pozo cargado negativo, mediante los residuos Glu 296, Glu 522 y Glu 245, que interactúan con la Arg 410. Se muestra la estructura en listones y en líneas gruesas en cian de la subunidad  $\alpha$  de la bacGlyRS determinada en este trabajo (incisos A y C). y en amarillo de la eukGlyRS de *Homo sapiens* (B incisos B y D) (código PDB 2zt8) en complejo con el análogo del glicil-adenilato (magenta). Todos los residuos de reconocimiento se hallan totalmente conservados.

### 5.5.2 Efecto adicional por impedimento estérico

Una forma adicional y esencial de discriminación del aminoácido se da mediante impedimento estérico en las GlyRSs, reduciendo el espacio en el fondo de la cavidad del sitio de unión del aminoácido (figura 5B y 5D). De esta forma se evita la unión de otros aminoácidos mayores que la glicina, como la alanina o la serina. Este impedimento estérico en el caso de la bacGlyRS se produce por dos treoninas (Thr 140 y Thr 158, figuras 5A y 5C), por lo que además de reducir el tamaño de la cavidad impiden la unión de la cadena lateral hidrofóbica de la alanina (figuras 5A y 5C). Para la eukGlyRS, el impedimento estérico puede producirse por un sólo residuo pequeño no conservado en la posición equivalente de la Thr 158 (figura 5D). En algunos organismos, la eukGlyRS llega a presentar una Ser (524) o Ala en esta posición (figura 5D).



**Figura 5. La bacGlyRS presenta diferencias con la eukGlyRS en las interacciones con la glicina y semejanzas con la AlaRS.** A) Superposición de la bacGlyRS con la AlaRS, donde se observa la interacción catión- $\pi$  del Trp 115 bacGlyRS (Trp 192 AlaRS) con el aminoácido y la Ala 31 bacGlyRS (Ala 99 en la AlaRS) cerrando el sitio. B) Contraste con la superposición de la bacGlyRS con la eukGlyRS. Se observa que posiciones similares son ocupadas por la Tyr 386 eukGlyRS en algunas de las aaRSs II y el Glu 254, con una interacción completamente distinta. C) y D) El Glu156 de la bacGlyRS corresponde con el Asp 248 en la AlaRS y con el Glu 522 en la eukGlyRS. Todos los residuos anteriores se hallan conservados. **Impedimento estérico.** C) La Thr 158 y Thr 140 presentan impedimento estérico y polar en la bacGlyRS para prevenir la unión de la alanina junto con una molécula de disolvente que se estabiliza por puentes de hidrógeno con los residuos circundantes (Thr 140, Glu 156, Thr 33). En la región equivalente de la AlaRS la Val 215 y la Gly 250 permiten la entrada del grupo metilo de la cadena lateral de la alanina. D) Los residuos equivalentes en la eukGlyRS, la Ala 408 (equivalente a la Thr 140 en la bacGlyRS) se halla totalmente conservada; mientras la Ser 524 (equivalente a la Thr 158 en la bacGlyRS) no es conservada y no se encuentra a la distancia para producir el mismo efecto, sin embargo presenta algún impedimento estérico también en este sitio para impedir la entrada de cadenas laterales de aminoácidos mayores. Se muestra la

estructura en listones la superposición de la subunidad  $\alpha$  de la bacGlyRS en cian, tanto con la AlaRS en blanco (A y C), como con la eukGlyRS en amarillo (B y D), ambas en complejo con un análogo de aminoacil-adenilato. Se muestra la cadena lateral del alanil-sulfamoil-adenilato (C) y el GSAAd (D) en líneas gruesas magenta y en negro las distancias entre los residuos que producen impedimento estérico (continúa de la página anterior).

## 5.6 La $\alpha$ -bacGlyRS y la AlaRS comparten características que las diferencian de las otras aaRSs II

Los elementos de reconocimiento del aminoácido y de oligomerización se encuentran presentes principalmente en los 3 motivos característicos de las aaRSs de clase II (figura 6A). Dichos elementos se encuentran en la subunidad  $\alpha$  de la bacGlyRS, por lo que fue posible comparar las diferencias con las demás aaRSs II.

Encontramos ciertas características que comparten la subunidad  $\alpha$  de la bacGlyRS y la AlaRS (código PDB zzg) (figura 6A) y que las diferencian respecto a las otras aaRS de clase II:

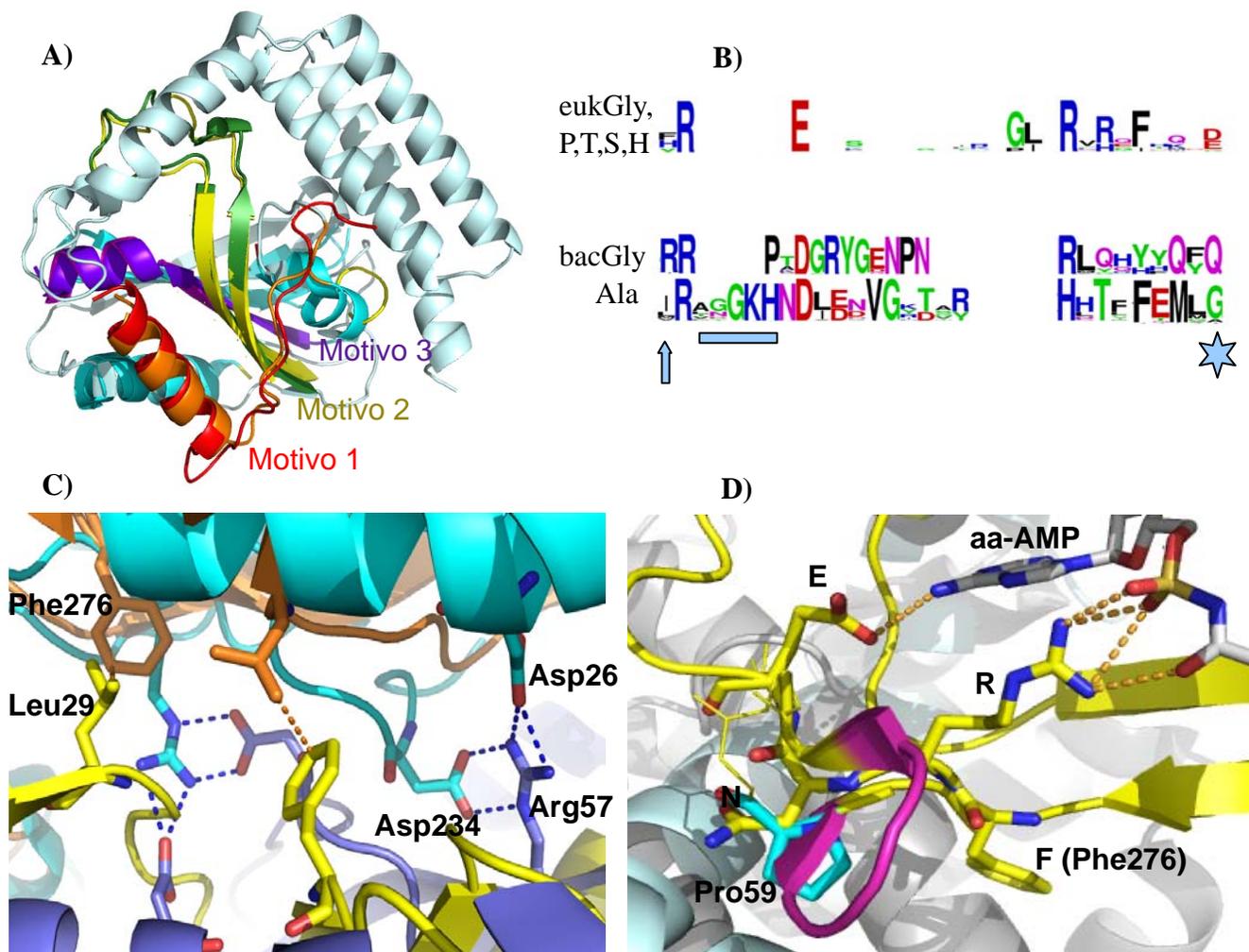
- Reconocimiento del aminoácido. El Trp 115 es equivalente al Trp 192 en la AlaRS (figura 5A), que a pesar de no ser parte de los tres motivos, proviene del mismo elemento estructural en estas dos enzimas. El elemento estructural del que se extiende el Trp en las otras aaRSs II es una asa denominada asa del aminoácido, pero en la AlaRS y bacGlyRS con la unión del ligando la región se estructura formando parte de una hebra  $\beta$  del plegamiento central. En muchas de las otras aaRSs II cercano a esta posición pero en otra disposición espacial con respecto al Trp 115, hay otro residuo aromático como Phe o Tyr moderadamente conservado, como la Tyr 386 totalmente conservada en la eukGlyRS (figura 5B).

- Reconocimiento del grupo amino. Existe una sustitución de un glutámico conservado que reconoce el grupo amino del aminoácido. Esta sustitución la comparten la bacGlyRS, AlaRS, HisRS, PylRS y SepRS. En el caso de la bacGlyRS, en esta posición, una Gln (78) interacciona con una molécula de disolvente (figura 6B). La eukGlyRS cuenta con un residuo Glu (296) canónico (figura 6B).

- Inserción de residuos dentro de la firma de las aaRSs II (FRxE). Existe una inserción dentro de la firma de residuos altamente conservados. En el caso de la bacGlyRS, presenta

la inserción de la Pro 59 (figura 6B y 6D) y en la AlaRS de 2 a 5 residuos (figura 6B y 6D), siendo las únicas aaRS con dicha inserción. Previamente se habían realizado estudios de mutación de la Pro 59->Leu, resultando que esta mutación puntual elimina la actividad de la  $\alpha$ -bacGlyRS (Toth y Schimmel, 1990, Val 265).

- Estrategias de dimerización: El residuo Phe al inicio de la firma característica de las aaRSs II (FRxE) en la mayoría de las sintetasas está implicado en la dimerización (figura 6B, 6C y 6D). La bacGlyRS dimeriza por medio de puentes salinos (el residuo Arg57 forma puentes salinos con los Asp26 y Asp234, mientras el Glu28 con la Lys238) (figura 6C). Las demás aaRSs II dimerizan por interacciones de residuos hidrofóbicos (Leu129-Phe276, en la eukGlyRS) (figura 6C). La AlaRS dimeriza de forma distinta de las otras aaRSs II (Swarijo *et al.*, 2004), mediante el dominio carboxilo terminal (Naganuma *et al.*, 2014).



## 5.7 La bacGlyRS y la AlaRS constituyen una subclase de las aaRSs II.

Para investigar si independientemente de las similitud en los residuos de reconocimiento del aminoácido había una mayor relación entre la AlaRS y la bacGlyRS, y que dicha relación podría agruparlas por separado de la eukGlyRS y de las otras sintetasas de clase II, se realizaron alineamientos estructurales de una muestra representativa de los CCDs las aaRSs II.

El dominio central (núcleo) catalítico (CCD) (figura 6A) fue definido en el punto 4.12. Los alineamientos estructurales entre el CCD de las aaRSs II se realizaron con el algoritmo STAMP en el programa Multiseq implementado como programa conectado en VMD, también realizando una matriz de puntuación resultado del alineamiento con el servidor DaliLite, así como de secuencia de regiones estructurales equivalentes con el programa 3D T-COFFEE expreso con la inferencia realizada en PhyML. Para llevar a cabo dicho análisis se partió de las estructuras publicadas para todas las aaRSs II, incluyendo la del presente trabajo.

Los cladogramas resultado de los alineamientos con STAMP (figura 7A) y de regiones estructurales equivalentes por T-COFFEE Expreso (figura 7D, anexo 1 figuras 13 y 14) mostraron el agrupamiento de los CCDs de las sintetasas de clase II en 4 clados (figuras 7A y 7D). Las sintetasas para un mismo aminoácido provenientes de distintos organismos y en presencia de distintos ligandos quedaron agrupadas en un solo nodo. Uno de los 4 grupos mencionados se encontró que estaba constituido sólo por la bacGlyRS y la AlaRS, mientras los otros tres grupos correspondieron exactamente con el agrupamiento en las tres subclases previamente definidas de las sintetasas de clase II. Mediante STAMP y T-COFFEE se realizaron alineamientos con la porción mínima del CCD (muy similar al núcleo invariante entre las sintetasas) recortada a mano, los cuales brindaron un resultado similar e incluso más claro de agrupamiento en 4 grupos, eliminando la posibilidad de un sesgo por motivos específicos de cada subclase de las sintetasas.

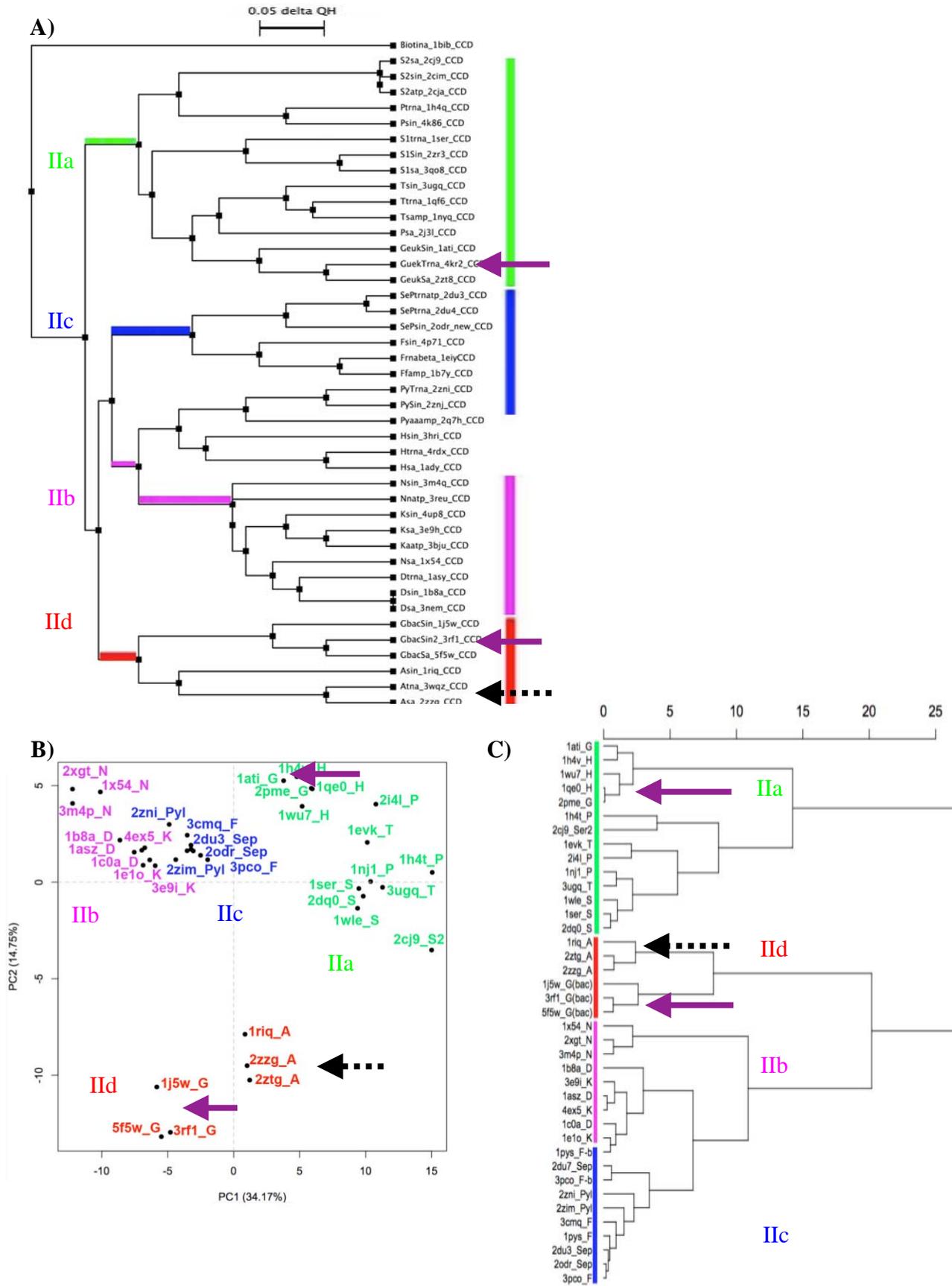
Los grupos encontrados por alineamiento estructural (figuras 7A y 7D) fueron:

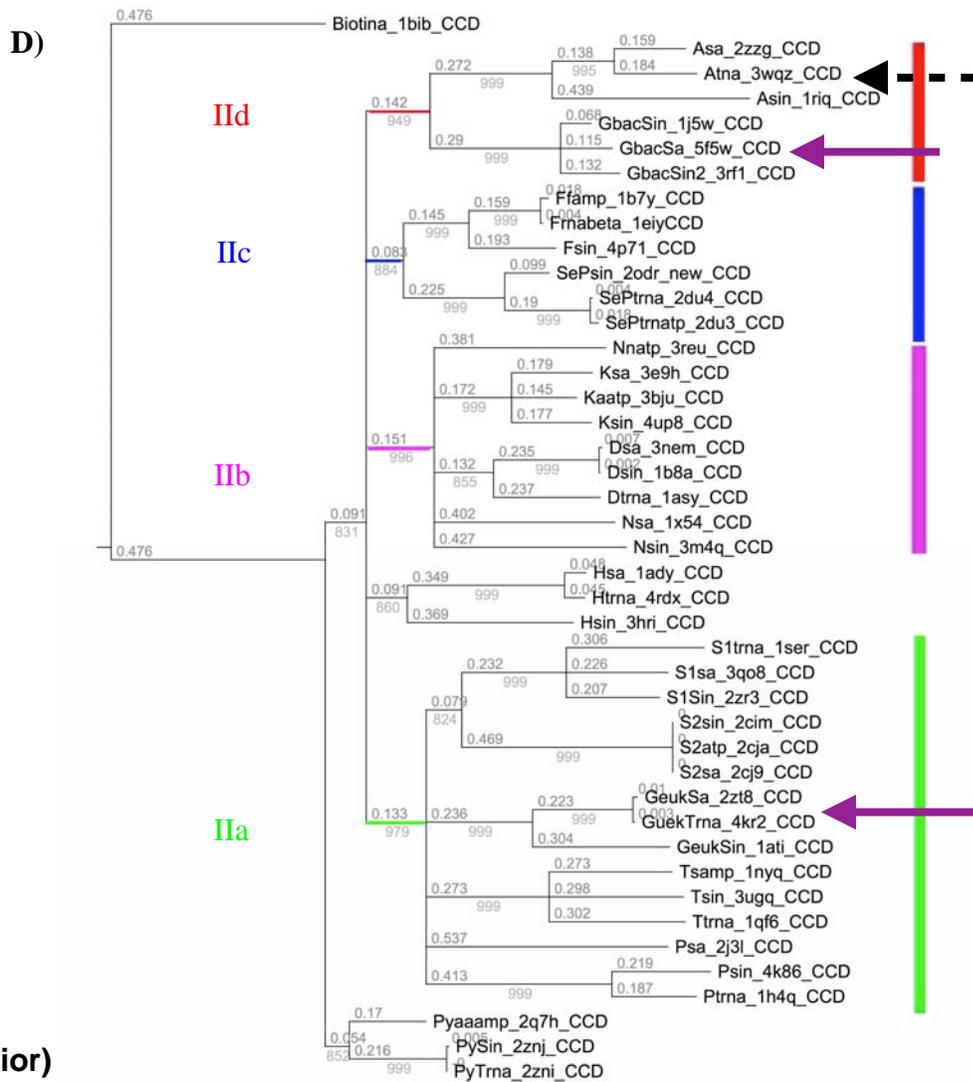
- El grupo nombrado IId: constituido por la AlaRS y la bacGlyRS.
- El grupo IIa: correspondiente a la subclase IIa, constituido por la ProRS, SerRS 1, SerRS 2, ThrRS, HisRS y muy importante de subrayar la eukGlyRS.

- El grupo IIb: correspondiente a la subclase IIb, constituido por la AspRS, LysRS II y la AsnRS.
- El grupo IIc: correspondiente a la subclase IIc, constituido por la PheRS, SepRS y dependiendo del tipo de alineamiento, la PylRS.

Los mismos 4 grupos fueron encontrados en la matriz construida con la puntuación arrojada por DaliLite con el valor de corte de 18.5.

Para evaluar el efecto de las distintas porciones de la estructura sobre el agrupamiento encontrado por alineamiento estructural, se descompuso la variación entre las estructuras en sus componentes, mediante un análisis de componentes principales (PCA) automatizado con el programa Bio3D (figuras 7B y 7C). A partir del CCD de la muestra representativa de todas las sintetasas de clase II el programa realizó un alineamiento estructural y calculó un núcleo invariante de carbonos  $\alpha$  de 80 residuos. La variación de las sintetasas respecto al núcleo invariante dentro del alineamiento estructural fue analizada en sus componentes. Los principales componentes 1 y 2 (PC1 y PC2) comprendieron el 34% y 14% de la variación estructural, y estos dos planos principales al ser proyectados en la gráfica de confórmeros permitió observar un agrupamiento en los 4 grupos (figuras 7B y 7C) muy similares a los encontrados con STAMP y T-COFFEE para todo el CCD. Los 4 grupos fueron también evidenciados mediante un cladograma con base a la distancia en el subespacio de los dos componentes principales (figuras 7B y 7C). Los resultados del PCA indican que el componente principal en la variación estructural entre las distintas sintetasas es el mayor responsable del agrupamiento de las sintetasas en las cuatro subclases que aquí presentamos.





(página anterior)

**Figura 7. La bacGlyRS y la AlaRS constituyen una subclase de las aaRSs II.** Se realizaron distintos alineamientos estructurales, PCA y alineamientos de secuencia entre todas las aaRSs de clase II. A raíz de las semejanzas encontradas entre la subunidad  $\alpha$ -bacGlyRS y la AlaRS, pudo observarse la formación de cuatro grupos de aaRSs II. Proponemos que la AlaRS y la subunidad  $\alpha$ -bacGlyRS forman una subclase separada (IIId) de la subclase IIc (Phe, Pyl, Sep), mientras la eukGlyRS se encuentra en la subclase IIa (eukGly, Pro, Ser 1 y 2, Thr y posiblemente la His). La subclase IIb se encuentra formada por Lys, Glu, Asp. Se muestran los dendrogramas resultado del alineamiento de los CCD de las aaRSs de clase II: A) estructural con el algoritmo STAMP implementado en el programa Multiseq de VMD con base al valor de Qh, que es una medida de la similitud estructural con una relación inversa a el RMSD (mide las distancias del C- $\alpha$  de cada residuo específico a los C- $\alpha$ s de todos demás residuos de la proteína asignando una penalidad a los intervalos) (fórmula en la figura 16, anexo 2), B) con base a PCA realizado con sólo los C- $\alpha$  del CCD en Bio3D presentado como gráfica de confórmeros, dados como una proyección de todas las estructuras en los planos principales según los mayores PC, y C) distancias de B en el subespacio de los PC mostradas como dendrograma. D) de secuencia realizado con 3D T-COFFEE y construido con PhyML con un bootstrap de 1000 repeticiones (secuencias del alineamiento completo en figura 14 anexo 1). En todos los casos las flechas moradas indican la posición en los alineamientos de cada una de las GlyRSs y las flechas punteadas de la AlaRS.

## 6. DISCUSIÓN

### 6.1 La subunidad $\alpha$ de la bacGlyRS es un dímero en solución

En los estudios realizados en este trabajo tanto por SEC-MALS como por SAXS se observó la presencia de un dímero como la especie oligomérica en solución de la subunidad  $\alpha$  de la Aa-GlyRS (figura 1). En los primeros reportes de velocidad de sedimentación de la subunidad  $\alpha$  de la bacGlyRS se había determinado también la presencia de un dímero en solución (Ostrem y Berg, 1974), o bien de un equilibrio monómero-dímero (McDonald *et al.*, 1980), pero recientemente se postuló que la especie en solución era un monómero para *C. jejuni* en un análisis por SEC (cromatografía de exclusión de tamaño) (Tan *et al.*, 2012).

La bacteria *Aquifex aeolicus* vive alrededor de los 95°C, por lo que existiría la posibilidad de que a temperatura ambiente fuera baja temperatura para esta proteína, y se produjera una interacción como en el caso de proteínas mesofílicas en las que es posible incrementar la presencia de un dímero a baja temperatura.

En la determinación del SAXS, a una concentración alrededor de 1.25 mg/mL, la curva teórica del dímero aparentemente dejó de ajustar exactamente a la experimental, y el  $R_g$  comenzó a variar. Sin embargo, a estas bajas concentraciones, la medición es demasiado ruidosa. Estos estudios indicaron la presencia de un dímero en solución (figura 1A y 1B), pero existe la posibilidad de un equilibrio dímero-monómero a bajas concentraciones. En todas las mediciones realizadas en solución se detectó un dímero y en ninguna se detectó un monómero. Cabe mencionar que la pureza de la subunidad  $\alpha$  de la bacGlyRS fue analizada por SDS-PAGE (anexo 2, figura 15).

El puente salino establecido entre residuos conservados observado en la estructura cristalográfica (5f5w) de la  $\alpha$ -AaGlyRS, aunada a los resultados de los estudios en solución, parecen indicar que la interacción que estabiliza la especie dimérica es significativa (resultados punto 5.6, figura 6C).

## 6.2 La subunidad $\alpha$ de la bacGlyRS puede llegar a catalizar la formación del glicil-adenilato por sí sola

Anteriormente no se había detectado actividad para ninguno de los dos pasos de la reacción con una sola de las subunidades (Ostrem y Berg, 1974; Toth y Schimmel, 1986; 1990). La medición del primer paso en la subunidad  $\alpha$ -bacGlyRS se había determinado en dichos trabajos por incorporación de  $^{32}\text{P}\text{P}\text{i}$  ( $^{32}\text{P}$ ]-PPi) en el ATP dependiente de la concentración de glicina.

La Dra. Annia Rodríguez de nuestro grupo de investigación realizó la prueba de detección del glicil-adenilato (figura 2A y 2B), tanto con sólo la subunidad  $\alpha$ -AaGlyRS, como con la presencia de la subunidad  $\beta$ , así como con una enzima de subunidades  $\alpha$ - $\beta$  naturalmente fusionadas (figura 2A, la descripción de esta GlyRS se encuentra en el anexo 1 figura 11, anexo 2 figura 15 y tabla 6), y aunque el producto de la reacción se incrementó enormemente con la presencia de ambas subunidades y el tRNA (posiblemente por inclinar la reacción reversible a los productos) (figura 2A), la migración de los productos del primer paso de la reacción sólo con la subunidad  $\alpha$  y con la Aa-GlyRS reconstituida completa fue similar en la cromatografía de capa fina (figura 2A).

Cabe mencionar que la medición del producto de la reacción, así como la determinación de los parámetros cinéticos de la  $\alpha$ -AaGlyRS se realizaron con la medición de la banda correspondiente al  $^{32}\text{P}$ -AMP (figura 2A y 2B), resultado de la hidrólisis del glicil-adenilato, ya que su hidrólisis en Gly y AMP es muy rápida y no podía medirse adecuadamente la banda de glicil-adenilato.

La  $K_{\text{cat}}$  para la activación del aminoácido de la  $\alpha$ -AaGlyRS fue muy baja ( $2.3 \times 10^{-4} \text{ s}^{-1}$ ) comparada con otras sintetetasas, incluyendo la de eukGlyRS por la glicina de *Thermus thermophilus* de  $5 \text{ s}^{-1}$  a  $37^\circ\text{C}$  ( $50 \text{ s}^{-1}$  a  $70^\circ\text{C}$ ) (Mazauric et al., 1996) y la AlaRS por alanina de  $4.5 \text{ s}^{-1}$  (que incluso presenta una  $K_{\text{cat}}$  de  $3.7 \text{ s}^{-1}$  por la glicina) (Tsui y Fersht, 1981). Se había reportado una  $K_{\text{cat}}$  para intercambio de PPi (también de activación del aminoácido) para la bacGlyRS con ambas subunidades de *E. coli* de  $11 \text{ s}^{-1}$  en ausencia del tRNA y hasta de  $27.5 \text{ s}^{-1}$  en presencia de tRNA (Ostrem y Berg, 1974). Sin embargo, el valor reportado para *E. coli* sería uno de los números de recambio más altos para las sintetetasas, por lo que un estudio exhaustivo debe ser realizado, ya que los números de los autores para la  $K_{\text{cat}}$  del segundo

paso varían en dos órdenes de magnitud entre  $36 \text{ s}^{-1}$  (Ostrem y Berg, 1974) y  $0.36 \text{ s}^{-1}$  (Mazauric et al., 1998). Una posible explicación para el bajo valor que encontramos se hallaría en la presencia de colas de histidinas en ambas subunidades estudiadas en este trabajo, considerando que podrían afectar la entrada del sustrato, la interacción entre las subunidades o la interacción de estas con el tRNA, pues aparentemente todas estas interacciones tienen efecto en la activación del aminoácido en la bacGlyRS. Los estudios posteriores deberán de explicar qué cambios ocurren sobre la catálisis con la unión de la subunidad  $\beta$  y con el tRNA.

La  $K_m$  para la glicina de la  $\alpha$ -AaGlyRS ( $110 \mu\text{M}$ ) resultó similar a la encontrada para la enzima completa de *E. coli* de entre  $30 \mu\text{M}$  sin tRNA y de  $140 \mu\text{M}$  en presencia de tRNA, (Ostrem y Berg, 1974). El significado de este valor para la subunidad  $\alpha$  sería que la afinidad de unión de los sustratos es similar a la de la enzima completa, pero posiblemente el cambio conformacional de la subunidad  $\alpha$  necesario para catalizar eficientemente la producción del glicil-adenilato (cambio que es el observado en nuestra estructura) se facilita en presencia de la subunidad  $\beta$ ; el efecto del cambio conformacional sería el incremento de la eficiencia de recambio de la enzima. La  $K_m$  de la subunidad  $\alpha$ -AaGlyRS por la glicina se halla en el rango de todas las sintetasas.

Es de importancia que la  $K_{cat}$  para el segundo paso (aminoacilación, aproximadamente  $0.4 \text{ s}^{-1}$ ) y la  $K_m$  por el tRNA (aproximadamente  $0.2 \mu\text{M}$ ) para ambos tipos de GlyRS es casi idéntica (Mazauric et al., 1998), lo que indica que aparentemente ambas enzimas son igualmente eficientes en la catálisis cuando tienen todos los componentes de la reacción.

### **6.3 Efecto del GSAd sobre la subunidad $\alpha$ de la AaGlyRS**

El co-cristal de la subunidad  $\alpha$ -AaGlyRS conteniendo ATP  $10 \text{ mM}$ -glicina  $5 \text{ mM}$ , fue sumergido por 2 días en presencia de  $28 \text{ mM}$  de GSAd. Como dato adicional, todos los co-cristales que sólo contenían ATP-Mg difractaron a una resolución menor a  $4 \text{ \AA}$  (tabla 5, anexo 2), y únicamente los co-cristales que tuvieron presencia del GSAd presentaron un cambio en el hábito cristalino e incrementaron su resolución. Sin embargo, el mejor resultado para obtener datos colectables fue sumergir los cristales con el GSAd. La unión del GSAd a la  $\alpha$ -

AaGlyRS correspondió claramente a una estabilización térmica (anexo1, figura 12) que no se presentó con los ligandos ATP, Mg o Gly, como fue determinado mediante el ensayo de desplazamiento térmico por fluorometría de barrido diferencial (anexo1, figura 12).

Una densidad electrónica positiva fue observada claramente en el mapa de diferencia en el sitio de unión del ligando y confirmada en el mapa de omisión (figura 3A). Este resultado permitió asignar claramente la presencia del GSAd (figura 3A). Las estructuras de la subunidad  $\alpha$ -bacGlyRS reportadas con anterioridad, en complejo con ATP y/o glicina presentaron algunas deficiencias: el ATP en una conformación inadecuada, las coordenadas de la glicina asignada en sólo un monómero de forma dudosa dentro de la densidad electrónica, y ningún cambio conformacional (Tan *et al.*, 2012). Debido a que el GSAd es un análogo del glicil-adenilato que no es hidrolizable por la acción de las aaRS, pero cuyo reconocimiento por la enzima es prácticamente idéntico al sustrato natural, presentó la mayor ventaja para obtener esta estructura. El cambio conformacional que ocurrió en la subunidad  $\alpha$  de la AaGlyRS con la unión del GSAd (figuras 3C-3E) permitió definir el reconocimiento de la glicina en la bacGlyRS.

#### **6.4 El reconocimiento de la glicina entre las GlyRS ocurre de forma distinta**

Nuestra afirmación de que las dos GlyRS reconocen a la glicina de manera distinta parte de que el reconocimiento en la eukGlyRS se lleva a cabo por residuos idénticos cargados negativamente (tres glutámicos) (figura 4D) y en la bacGlyRS se da por medio de residuos distintos que forman una red de interacciones polares (electrostáticas) diversas (por una red de puentes de hidrógeno con residuos polares sin carga y dos moléculas de solvente, un residuo cargado, y una interacción catión- $\pi$ ) (figura 4C).

Comparación de algunos residuos cuya equivalencia entre las dos GlyRSs podría ser controvertida:

- Glu 156 de la bacGlyRS por Glu 522 de la eukGlyRS (figuras 5C y 5D). Este residuo es el único que presentaría una carga negativa en el caso de la bacGlyRS equivalente a la que presentan tres glutámicos en la eukGlyRS (figura 4C). El Glu 156 corresponde con Asp 248 en la AlaRS (figura 5C), en la cual cumple la misma función. Precisamente en las estructuras

existentes de la AlaRS puede observarse que este residuo permite la unión de los aminoácidos Ala, Gly y Ser (Guo *et al.*, 2009). Este residuo es compartido entre las dos GlyRS y es conservado en las tres aaRS (las dos GlyRSs y la AlaRS) (figuras 5C y 5D).

- Glu 296 en la eukGlyRS por Gln 78 en la bacGlyRS (figuras 4C y 4D). En el caso de estos residuos, el reconocimiento químico que efectúan sería distinto, aún si una GlyRS derivara de la otra, pues una de ellas habría evolucionado para reconocer a la glicina sólo por carga y la otra por una red polar. Además no se trata de la posición exacta, pues el lugar del Glu 296 se halla ocupado por la molécula de disolvente que une la Gln 78.

- Thr 158, Thr 140 y Thr 33 en la bacGlyRS, por Ser 524 y Ala 408, Ala 247 en la eukGlyRS (figura 5D). En el caso de la eukGlyRS, la Ala 247 y Ala 408 están totalmente conservadas, mientras la Ser 524 (figura 5D) se halla aproximadamente un 80% conservada, por lo que es probable que se trate de una evolución convergente para llevar a cabo el impedimento estérico en aquellos organismos que presenten serina. Además de que la cadena principal de estos residuos se halla alejada de la de las posiciones equivalentes en la otra enzima.

- Por otra parte, existen residuos en cada una de las GlyRS sin equivalente entre ellos. Este es el caso de Glu 245 en la eukGlyRS (figura 5B), que interacciona con la glicina en una región relativamente cercana al Trp 115 en la bacGlyRS (figura 5B) pero con una interacción completamente distinta; en esta posición, la bacGlyRS presenta una Ala conservada (figura 5B); en la eukGlyRS, la Arg 410 (figura 4C y 4D) sin ningún residuo equivalente y Glu 298 (figura 4D), cuya única función es reducir el tamaño del pozo negativo. En el caso de la bacGlyRS, como se ha comentado, un elemento importante del reconocimiento es el Trp 115 (figura 5B) con su fuerte interacción catión- $\pi$  que no presenta ningún equivalente en la eukGlyRS.

Una vez resuelto que el reconocimiento del aminoácido ocurre de forma distinta, se procedió a evaluar si la nueva estructura permitía establecer alguna relación filogenética entre las dos GlyRSs.

## 6.5 Las moléculas de disolvente son parte de la red polar que reconoce la glicina en la bacGlyRS

En el mapa de densidad electrónica de diferencias claramente se observó la presencia de dos moléculas de disolvente como picos positivos de densidad electrónica. En la cadena B, que es la cadena en que el modelo presenta un mejor ajuste a la densidad electrónica, se asignaron las dos moléculas de disolvente. Los dos picos positivos de densidad electrónica se observaron claramente en todos los monómeros de la unidad asimétrica, excepto en las cadenas D y E en que la menor densidad electrónico solo permitió observar alguna de las dos moléculas de disolvente.

Estas moléculas de disolvente se hallan a una distancia para formar puentes de hidrógeno con la glicina, con la Thr 140, Glu 156, Gln 138 y Thr158 una de ellas (figura 5B), y con el glicina (GSAd), Gln 76, Gln 78, Thr 33 la otra (figura 4C). Por las distancias presentaron la posibilidad de formar puentes de hidrógeno bifurcados, los cuales en realidad se dice que existen como el camino de baja energía para el cambio entre un puente de hidrógeno con una molécula al formarlo con otra (Lagee y Haynes, 2006).

Las moléculas de disolvente terminarían por demostrar la naturaleza tan distinta del reconocimiento de ambas GlyRSs (punto 5.5.1), y subrayar la importante contribución de la red de puentes de hidrógeno a las distintas interacciones polares que proveen el mecanismo de reconocimiento de la bacGlyRS.

Estas moléculas de disolvente también pueden ser observadas en la misma posición en las estructuras de otras aaRSs II, pero no ha sido descrita su importancia (AlaRS, SerRS).

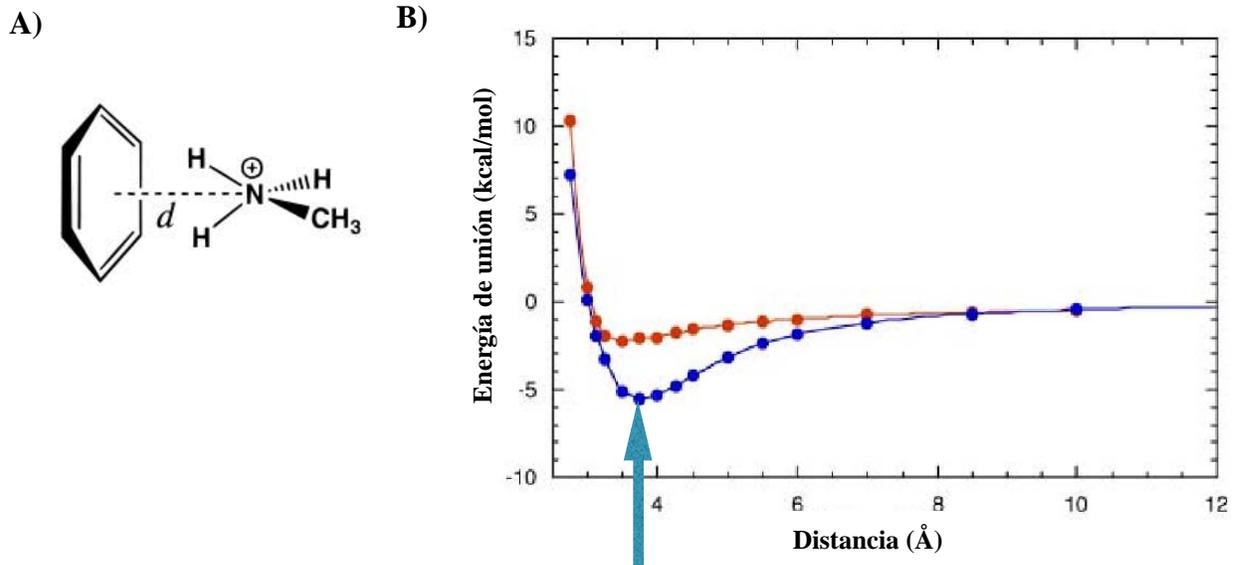
Si las moléculas de disolvente descritas no participaran en el reconocimiento de la glicina, sería suficiente con los residuos descritos que interaccionan directamente (Trp 115, Thr 33, Glu 156 y Gln 76) (figura 4C) y los que presentan el impedimento estérico (Thr 158, Thr 140) (figura 5B), para explicar el reconocimiento de la Gly. Sin embargo, debido a que dichos residuos forman una red polar, es altamente probable la presencia de las moléculas de disolvente. El demostrar su función sería posible al retirar alguno de los residuos clave en su estabilización, pero que no intervienen en otra forma de reconocimiento de la glicina, por ejemplo la Gln 78 (figura 4C) o Thr 140 (figura 5B).

## 6.6 La interacción Cation- $\pi$ del Trp podría ser esencial para reconocer la glicina

La interacción Cation- $\pi$  (también llamada amino-aromática) se forma por el potencial electrostático debido a los electrones deslocalizados sobre un anillo aromático y un grupo catiónico (cadena lateral con nitrógeno amonio) que se hallan a una distancia óptima de 3.8 Å (entre 3.4 y 4 Å es de -5 kcal/mol cuando se encuentra expuesto al disolvente) (figuras 3D, 4C y 8B). Los casos más estudiados en proteínas son los formados entre Lys o Arg actuando como cationes e interaccionando con los anillos del Trp, con Tyr o Phe, que tienen una elevada frecuencia en las proteínas. La interacción con el Trp es la que ocurre preferencialmente, 26% de todos los Trp's se encuentran formando este tipo de interacciones (Gallivan y Dougherty, 1999).

Los estudios de simulación química (Gallivan y Dougherty, 2000), que miden la energía de la interacción se han realizado con metil-amonio ( $\text{CH}_3\text{-NH}_4^+$ ) (figura 8A), en donde se han calculado varias de sus propiedades. Por lo que a pesar de no haber muchos reportes de la Gly o la Ala, es bastante probable por la disposición y las distancias que sea el tipo de interacción que se forma. La interacción entre los aminoácidos y el anillo no es una unión de apilamiento clásica por la naturaleza del grupo amino.

Esta interacción Cation- $\pi$  en medio acuoso es más fuerte que una interacción de puente salino (figura 8B). Además al ocurrir la interacción se incrementa la presencia del isómero protonado del grupo metil-amonio, por lo que sería primordial en la interacción de la Gly en la bacGlyRS y Ala en la AlaRS (figuras 3D, 4C y 5A).



**Figura 8. La interacción Cation-  $\pi$  con el Trp podría ser esencial para reconocer la glicina.** A) Algunos cálculos computacionales (estudios de simulación química) se han realizado para evaluar la interacción del metil-amonio con benceno. B) La energía de unión en medio acuoso según la distancia en el inciso A (señalada como d) de una interacción cation- $\pi$  (azul) es mayor que la de un puente salino (rojo) (tomado de Gallivan y Dougherty, 2000), la flecha indica la distancia entre el extremo amino de la glicina y el anillo aromático del Trp 115 en la bacGlyRS (ver figuras 3A y 5A).

### 6.7 Mediante alineamiento estructural fue posible inferir la relación estructural y filogenética de las GlyRSs

Todas las aaRSs II comparten el plegamiento del CCD (tablas 1 y 2, figura 6A), a pesar de que como en el caso de las dos GlyRS, su secuencia es tan distinta como dos proteínas no relacionadas, es aceptado que todas las aaRSs II son antiguos parálogos que provienen de algún ancestro primigenio (Gabaldón y Kooning, 2013; O'Donoghue y Luthey-Schulten, 2003). A nivel de reconocimiento, las aaRSs II comparten los residuos que reconocen el ATP (5 aminoácidos en el caso de las GlyRSs, figura 17 anexo 2), pero las dos GlyRSs no comparten los residuos que reconocen el aminoácido, excepto por un Glu (Glu 156 de la bacGlyRS) (figuras 4C y 5D, figura 18 anexo 2).

El alineamiento estructural con STAMP y el análisis PCA con Bio3D de una muestra representativa de los CCDs de todas las aaRSs II, fue de utilidad para evidenciar que cada una de las GlyRSs se agrupan dentro de dos conjuntos distintos, dos distintas subclases

(figuras 7A, 7B y 7C). Ambos programas emplean sólo los C $\alpha$  para evaluar la distancia entre las estructuras, por lo que realizaron el agrupamiento de las sintetasas sin importar la conservación de los residuos en la secuencia (figuras 7A a 7C). Además, el PCA permitió descomponer el efecto de las distintas porciones de la estructura sobre el agrupamiento, donde ambas GlyRSs fueron agrupadas en conjuntos diferentes con base a los dos componentes de mayor variación entre las estructuras. Pero la mayor importancia se encuentra en la selección correcta de las regiones estructurales equivalentes para ser alineadas, en este caso el CCD.

Debido a que comparten el plegamiento central y catalizan la misma reacción, algunos autores pensaban que la bacGlyRS y la eukGlyRS eran un par de genes que divergieron entre ellos. Sin embargo los alineamientos estructurales y de secuencia de regiones estructurales equivalentes, agruparon todas las aaRSs II en cuatro subclases distintas (figuras 7A, 7B, 7C y 7D), separando las dos GlyRSs en dos subclases diferentes (la bacGlyRS en la subclase IId y la eukGlyRS en la subclase IIa) y agrupando la bacGlyRS con la AlaRS en la misma subclase. Esto corrobora que los dos tipos de GlyRS no derivan uno del otro y que posiblemente la AlaRS y la bacGlyRS provienen de un ancestro común más reciente (figuras 7A, 7B, 7C y 7D). Adicionalmente la eukGlyRS posee el dominio idiosincrático para reconocer el anticodón característico de las aaRS de subclase IIa (tabla 2), mientras la bacGlyRS emplea la subunidad  $\beta$  para esta función, la cual no tiene ninguna relación con las sintetasas de clase II (Wolf *et al.*, 1999).

Distintos autores reportaron que los alineamientos estructurales podrían ser esenciales para lograr inferencias en proteínas tan divergentes como las aaRSs (Chotia *et al.*, 1986, Nagel y Doolittle 1991; Russell y Barton, 1992; Ribas de Pouplana *et al.*, 2001; O'Donoghue y Luthey-Schulten 2003; Smith y Hartman, 2015); sin embargo, no contaban con toda la información estructural que hoy existe. Smith y Hartman, 2015, realizaron un estudio bioinformático empleando el CCD de las aaRSs, e identificaron algún agrupamiento estructural entre la bacGlyRS y AlaRS, pero el trabajo que realizaron se limitó a pocas aaRSs específicas para algunos aminoácidos. Aunque la carencia de información respecto a las relaciones de las GlyRSs limitó la interpretación de sus resultados, Smith y Hartman, 2015, lograron un vislumbre de las relaciones filogenéticas de las aaRSs y sus posibles ancestros relacionados con el surgimiento del código genético. La rigurosa filogenia construida por

O'Donoghue y Luthey-Schulten, 2003 con base a variación estructural agrupó de forma excelsa los CCDs de las aaRSs de acuerdo a las subclases definidas hasta ese momento, pero no contaba con todas las estructuras de las sintetasas que hoy existen. Cabe mencionar que el grupo de Luthey-Schulten realizó la implementación del algoritmo STAMP en el programa Multiseq empleado en este trabajo para los alineamientos estructurales.

En todos los casos para la información derivada de los alineamientos estructurales de las sintetasas (incluyendo el presente trabajo), el agrupamiento en las subclases empleando la estructura del CCD ha sido corroborado por la concordancia con las subclases previamente definidas (Cusak, 1993; Perona y Hadd, 2012) por similitud en los dominios idiosincráticos y mecanismos bioquímicos.

Al comparar los dendrogramas de proteínas tan divergentes como las aaRSs II, entre los alineamientos estructurales y los alineamientos de secuencia de regiones estructurales equivalentes (figuras 7A y 7D), pudo observarse que los alineamientos de estructura y secuencia pueden llegar a brindar resultados similares si se eligen las regiones equivalentes correctas. Lo anterior permite tener una alternativa para realizar un análisis de conservación de secuencia en familias de proteínas con identidad abajo de la zona de penumbra, seleccionando sólo las secuencias de regiones estructurales comunes.

## **6.8 La bacGlyRS y la AlaRS se hallan estrechamente relacionadas.**

La similitud más evidente entre la subunidad  $\alpha$  de la bacGlyRS y la AlaRS es la presencia del Trp exactamente en posiciones equivalentes (Trp 115 en la bacGlyRS y Trp 192 en la AlaRS) (figura 5A). En este trabajo se observó que la interacción catión- $\pi$  del Trp y el cambio conformacional al unirse al estado de transición, también ocurrían en la AlaRS (figura 3D y 5A). Este triptofano en ambas enzimas cambia de posición, posiblemente para dar lugar a la formación del respectivo aminoacil-adenilato, presentando una interacción semejante con el extremo amino del aminoácido (figura 5A). También la bacGlyRS y la AlaRS comparten la Ala 31 (Ala 31 en la bacGlyRS y Ala 99 en la AlaRS) totalmente conservada que se haya a 4Å de este Trp y que ayuda a cerrar la cavidad (figura 5A); el equivalente de esta Ala en el caso de la eukGlyRS es el Glu 245 (figura 5B).

Además del ya mencionado equivalente del Glu (Glu 156 en la bacGlyRS, Asp 248 en

la AlaRS) (figuras 5C y 5D), y de los 5 aminoácidos compartidos para el reconocimiento del ATP conservados en la mayoría de las aaRSs II (figura 6B y 6D), también se hallan conservados al menos otros 12 residuos entre la AlaRS y la bacGlyRS (anexo 2, figura 18).

Pero, además de ello, también se encuentra el hecho de que la AlaRS aminoacila por error el tRNA<sup>Ala</sup> con glicina con una alta frecuencia de 1:170, incluso mayor que con serina (1:300) (Tsui y Fersht, 1981) y que para corregirlo presenta una edición de los tRNA erróneamente cargados con glicina también con una alta tasa (Tsui y Fersht, 1981).

Este trabajo aporta evidencias a lo que había sido sólo predicho por algunos autores (Swarijo *et al.*, 2004, Smith y Hartman, 2015), para entender la estrecha relación que existe entre los CCD de estas dos aaRSs (AlaRS y bacGlyRS). La paradoja de los errores de aminoacilación de la AlaRS, queda esclarecida al cotejar que los residuos que reconocen el aminoácido en la AlaRS se encuentran presentes en la bacGlyRS, siendo esta última la que, mediante impedimento estérico, limita la entrada de la alanina (figura 5A y 5C).

Como se demostró por alineamiento estructural con STAMP-Multiseq y PCA con simplemente la cadena principal (los C $\alpha$ ) de los CCDs, fue posible encontrar un agrupamiento de la bacGlyRS y la AlaRS (figuras 7A, 7B y 7C). También este agrupamiento se encontró por alineamiento estructural y alineamiento de secuencia de regiones estructurales equivalentes (7D).

La subunidad  $\alpha$ -bacGlyRS pareciera contener una versión mínima de los elementos necesarios en el CCD al compararla con la AlaRS, por lo que es posible visualizar el mecanismo de discernimiento entre los aminoácidos glicina y la alanina al comparar el sitio de unión del aminoácido de ambas aaRS. Esto nos lleva a proponer cambios en la AlaRS que permitirían interconvertir la especificidad entre ambas enzimas con una capacidad catalítica significativa (figuras 5A, 5C y 6A).

Los principales cambios que permitirían cambiar la especificidad en la bacGlyRS serían tres: Thr 33, Thr 140 y Thr 158, por los residuos correspondientes en la AlaRS (Ile101, Val 215, Gly250) (figura 5C y 4C); los residuos de Thr 140 y 158 son esenciales por impedir la entrada de cualquier aminoácido (sustrato) con cadenas laterales más grandes y en particular con grupos hidrofóbicos como la alanina. El cambio que quizá podría hacer la mayor diferencia es la Thr 140 --> Val (figura 5C) por ser la región donde la cadena lateral de la Ala se ubica, y es este quizá el principal residuo que estabiliza la molécula de disolvente

en esta parte de la molécula (figura 5C). La Thr 158 --> Gly, también sería de importancia por presentar el impedimento estérico directamente (figura 5C); a este respecto, Guo *et al.*, 2009 realizaron una mutación reemplazando la Gly 250 por Ala en la AlaRS para tratar de disminuir la afinidad por la Gly y la Ser, pero en el presente este trabajo mostramos que lo ideal sería el reemplazo por Thr. En el caso de la Thr 33 (figura 4C), la enzima de bacterias tiene una Met que no impide la unión de la respectiva molécula de disolvente en la AlaRS, pero dicho cambio quizá sería crítico en el caso de la AlaRS de arqueas que presenta una Ile. Un cambio adicional que podría evaluarse es el reemplazo en la bacGlyRS de la Gln 78—>Ala 149, ya que la principal función de la Gln 78 es la estabilización de la molécula de disolvente (figura 4C).

## **6.9 Las aaRSs de clase II son agrupadas en cuatro subclases por alineamiento estructural**

En este trabajo propusimos una subclase adicional de las aaRSs II, pues las aaRSs II originalmente se encontraban agrupadas en 3 familias: IIa, IIb y IIc, compartiendo características comunes. Sin embargo la familia IIc contenía las aaRSs más diversas: la GlyRS, la AlaRS y la PheRS.

Ya que el agrupamiento aquí presentado de las aaRSs II fue consistente en los alineamientos estructurales por STAMP-Multiseq-VMD, el análisis del alineamiento estructural por PCA, y de secuencia de regiones estructurales equivalentes en los CCD, propusimos una clasificación en 4 subclases de aaRSs II (figuras 7A, 7B, 7C y 7D). Este agrupamiento ya había sido obtenido por otros autores, sin embargo por la carencia de datos de las estructuras de la AlaRS, de la bacGlyRS y la evaluación de la diferencia en el reconocimiento con la eukGlyRS (figuras 4C y 4D), no se habían enunciado de una forma directa (O'Donoghue *et al.*, 2003; Smith y Hartman, 2015). Los grupos son las sintetasas de las subclases IIa (eukGly, His, Thr, Pro, Ser1 y Ser2), IIb (Asp, Asn, LysII), IIc (Phe, Sep) y IId (Ala, bacGly) (figuras 7A, 7B, 7C y 7D).

En el agrupamiento anterior se creía que había alguna relación entre la PheRS y la bacGlyRS, ya que se asumía que la estructura oligomérica  $\alpha_2\beta_2$  de ambas era muy similar. Sin embargo hay extensas diferencias de la PheRS con la bacGlyRS; la primera es que la

subunidad  $\beta$  de la PheRS (Perona y Hadd, 2012) es una duplicación de su subunidad  $\alpha$ , mientras en el caso de la bacGlyRS sus dos subunidades son completamente distintas en secuencia y estructura y tienen distintas funciones; además, la PheRS aminoacila el tRNA en el 2'OH de forma similar a como lo hace la SepRS (que han sido agrupadas juntas con nuestro análisis), mientras la bacGlyRS lo hace en el 3'OH canónico para las aaRSs II, razón que subraya el agrupamiento en nuestro trabajo de la PheRS y la bacGlyRS en una distinta subclase.

Uno de estos autores incluso postula el surgimiento de ciertas sintetetasas primordiales, en este caso 4 aaRSs II, de las que derivaron las otras aaRSs II relacionadas (Smith y Hartman, 2015). En los estudios anteriores solamente se habían realizado alineamientos estructurales, por lo que el empleo de la metodología de PCA aportó una herramienta importante para visualizar el posible desplazamiento estructural que seguirían para originarse cada una de las sintetetasas a partir de un CCD promedio invariante hacia los cuatro grupos (figura 7B) y cómo derivaron a sus respectivas aaRSs II dentro de cada grupo.

Un dato adicional es que en SCOP, PFAM y otras bases de datos basadas en estructura, las aaRSs II ya se habían clasificado en 6 familias estructurales (tRNA\_synt\_2, 2b, c, d, e, f), los cuales corresponden a los grupos que aquí definimos. En las familias estructurales reportadas en las bases de datos, la AlaRS, la subunidad  $\alpha$  de la bacGlyRS y la subunidad  $\beta$  de la bacGlyRS se encuentran reportados como grupos estructurales separados. Lo aquí presentado permite entender las características que colocan a la subunidad  $\alpha$ -bacGlyRS en un grupo separado junto con la AlaRS (figura 7), un posible camino evolutivo que las relaciona, presentando a la bacGlyRS como quizá una versión divergente simplificada de la AlaRS. Los datos presentados también permiten entender el por qué la eukGlyRS se agrupa en una subclase distinta a la bacGlyRS, presentando a las GlyRSs como un par de sintetetasas que a pesar de seguir un camino evolutivo distinto, convergieron para reconocer y catalizar la reacción con el mismo aminoácido.

## **6.10 Escenarios del posible origen y sucesión entre las dos GlyRSs polifiléticas.**

Con base en los hallazgos aquí reportados, proponemos algunos posibles escenarios para la evolución de los CCDs de las dos GlyRSs:

- Que un tipo de GlyRS provenga de la otra GlyRS o de un ancestro común con la misma especificidad. La evidencia que presentamos de diferencias en los elementos de reconocimiento del aminoácido y de pertenencia a subclases distintas por agrupamiento estructural parece eliminar esta posibilidad.
- Que ambas GlyRS surgieron independientemente desde el origen de las aaRSs primigenias. Implicaría que los organismos se separaron en los dominios arquea y bacteria antes (o de forma paralela) del surgimiento de las dos versiones de la GlyRS, y por tanto en el caso de las bacterias una pre-AlaRS o la AlaRS realizaba la adición de alanina y la glicina en bacterias, y posteriormente por duplicación de genes se especializó en AlaRS y bacGlyRS. También implica un evento posterior de transferencia horizontal de información hacia algunas bacterias de la eukGlyRS.
- Que una de las GlyRS (la eukGlyRS) surgió primero en todos los dominios de la vida. Posteriormente surgió la otra versión  $\alpha_2\beta_2$  (la bacGlyRS) a partir de la AlaRS y reemplazó a la primera sólo en las bacterias. En este escenario, el surgimiento de la bacGlyRS sería paralelo (o muy cercano) a la división en los dos dominios de la vida arquea y bacteria.
- El escenario opuesto del anterior: Que la bacGlyRS y la AlaRS fueron de las sintetasas primordiales, derivadas de un gen común. Posteriormente surgió una nueva versión de la GlyRS (la eukGlyRS) que sustituyó a la primera versión (la bacGlyRS) en las arqueas y por tanto en el citoplasma de eucariotas posteriormente. Este mecanismo explicaría la persistencia de la paradoja de reconocimiento de la AlaRS y el que la versión más nueva de la GlyRS (la eukGlyRS) sea más específica para la glicina sin el reconocimiento cruzado presentado por la AlaRS.

Todos los escenarios cumplirían con la idea de que la glicina fue uno de los primeros aminoácidos en aparecer y en ser empleado en las proteínas, pero los últimos tres escenarios explicarían las semejanzas entre la bacGlyRS y la AlaRS, y las diferencias con la eukGlyRS, presentando así una posible historia evolutiva.

Por varios años se había indicado que las dos GlyRSs eran proteínas homólogas, pero pensando en que provenían de un ancestro común con la misma especificidad, es decir que sus CCD eran ortólogos. En realidad, las GlyRSs sí son proteínas homólogas porque los CCDs de todas las aaRSs II son homólogos que comparten un ancestro primigenio, pero las GlyRSs son parálogos, pues aunque convergieron para ser enzimas isofuncionales

proviene de distintos caminos evolutivos.

De comprobarse con la determinación de su estructura que la subunidad  $\beta$  de la bacGlyRS fue adquirida como una versión del dominio de unión al anticodón de la ArgRS (aaRS clase I), derivadas ambas de la superfamilia de fosfohidrolasas HD (con la firma DALR), aunado con la evidencia que presentamos aquí de que la subunidad  $\alpha$  de la bacGlyRS proviene de una divergencia a partir de la AlaRS, sería uno de los casos más interesantes del origen de una aaRS dentro de estas antiguas proteínas.

## **7. CONCLUSIONES**

### **7.1 Las glicil-tRNA sintetasa bacteriana y la glicil-tRNA sintetasa presente en eucariotas difieren en el mecanismo de reconocimiento de la glicina.**

El reconocimiento de la glicina en la glicil-tRNA sintetasa presente en eucariotas se lleva a cabo por residuos idénticos con carga negativa (tres glutámicos) y en la glicil-tRNA sintetasa bacteriana se da por medio de distintos residuos que forman una red de interacciones polares diversas (por una red de puentes de hidrógeno con residuos polares sin carga y moléculas de solvente, un residuo cargado y una interacción catión- $\pi$ ).

### **7.2 Las dos glicil-tRNA sintetasa son parálogos lejanos que convergieron en enzimas isofuncionales.**

Cada una de las glicil-tRNA sintetasa comparte mas en el mecanismo de reconocimiento del aminoácido con otras sintetasa que entre ellas, por lo que no provienen de un ancestro común con la misma especificidad, pues aunque comparten el plegamiento del dominio catalítico central, también tienen una secuencia distinta y distinta distribución en los dominios de la vida.

Las dos glicil-tRNA sintetasa no comparten un ancestro común posterior al ancestro primigenio de todas las aminoacil-tRNA sintetasa de clase II.

### **7.3 La alanil-tRNA sintetasa y la glicil-tRNA sintetasa bacteriana se hallan estrechamente relacionadas y comparten características que las distinguen de las otras aaRS de clase II. Proponemos que estas dos sintetasa conformarían una nueva subclase dentro de las aminoacil-tRNA sintetasa de clase II.**

La alanil-tRNA sintetasa y la glicil-tRNA sintetasa bacteriana se distinguen por compartir el papel de un residuo de triptofano en el reconocimiento de su aminoácido respectivo por medio de una interacción catión- $\pi$ , la inserción de residuos dentro de la firma altamente conservada de las aaRSs II, una estrategia de dimerización distinta a las otras aaRSs II,

además de que la AlaRS y la bacGlyRS son capaces de reconocer la glicina por un mecanismo similar y aminoacilar con ella sus distintos tRNAs blanco. El alineamiento estructural de los CCDs de las aaRSs II las agrupa en cuatro clados, 3 corresponden a las subclases previamente definidas y la AlaRS y la bacGlyRS forman un clado por si solas.

## 8. PERSPECTIVAS

La principal perspectiva es obtener la estructura de la bacGlyRS completa en complejo con el tRNA y su análisis estructural, evolutivo y bioquímico, así como sus posibles aplicaciones para el diseño de antibióticos.

Con la estructura de la bacGlyRS completa podrán esclarecerse preguntas como las relaciones de la subunidad  $\beta$  con otras proteínas incluyendo las sintetasas de clase I; con la determinación de la estructura del complejo se esclarecerá la forma de reconocimiento del tRNA por esta sintetasa, identificando los elementos discriminatorios. El análisis de las relaciones estructurales de la enzima completa permitirá contribuir a terminar de armar el devenir evolutivo de esta sintetasa y de las aaRSs II en general, camino en el cual este trabajo realizó una aportación. Otra pregunta por contestar serán los cambios conformacionales que sufren ambas subunidades para formar la enzima completa y/o para unir el tRNA.

También quedan por esclarecerse experimentalmente algunas incógnitas respecto a la actividad enzimática de la bacGlyRS, como verificar el número de recambio en la activación del aminoácido para la enzima completa  $\alpha 2\beta 2$  (con una diferencia de cerca de dos órdenes de magnitud en los reportes anteriores), así como corroborar este valor para la enzima en complejo con el tRNA reportado para incrementarse 4 veces. De importancia también será cotejar los resultados de la actividad a los resultados de los cambios conformacionales en la estructura. Una vez confirmadas las constantes cinéticas de la bacGlyRS completa se podrán comparar nuevamente con las de la eukGlyRS y discutir si aportan datos sobre sus implicaciones en la convergencia de dos GlyRS en distintos dominios de la vida, en particular de un posible sustitución de alguno de los genes por el otro en uno de los dominios, la persistencia de la otra en el otro dominio, y la subfuncionalización de una AlaRS o pre-AlaRS para dar origen a la bacGlyRS.

Comprobar el papel que desempeñan el Trp 115 de la bacGlyRS y las moléculas de solvente en la unión y estabilización de la glicina y en la formación del glicil-adenilato, mediante estudios de mutagénesis y el análisis de la mutantes mediante ensayos de actividad y estructurales. Lo anterior también podrá ser evaluado mediante cálculos de dinámica molecular y simulaciones, para proponer un mecanismo para la catálisis con la

intervención de estos elementos.

Todos los estudios anteriores podrán ser extrapolados o incluir a la bacGlyRS de subunidades fusionadas presente en el cloroplasto y sus implicaciones sobre el metabolismo de algas y plantas. Esta es una enzima que es esencial hasta la germinación de las semillas en plantas. Los resultados de su estudio podrían ser aplicados incluso en el desarrollo de herbicidas o antibióticos contra infecciones con algas del género *Prototheca* por ejemplo. Los estudios de esta versión fusionada de la bacGlyRS contribuirán en la ingeniería de proteínas al desarrollo de proteínas de fusión por conectores peptídicos, como en el caso del conector que empleamos en este trabajo para mejorar las características de cristalización de la bacGlyRS derivado del fago RB69 (anexo 2, tabla 6).

Con el incremento en la resistencia a los antibióticos, lo descrito en las diferencias entre la bacGlyRS y la GlyRS presente en eucariotas (eukGlyRS), puede emplearse en el desarrollo de antibióticos contra bacterias de los géneros *Vibrio* (cólera), *Clostridium* (antrax, tétanos, gangrena, sepsis), *Yersinia* (peste), *Pseudomonas* (sepsis y neumonía. *Pseudomonas* es una bacteria especialmente resistente a varios antibióticos), *Haemophilus* (gripe de origen bacteriano en niños, meningitis), *Neisseria* (meningitis, gonorrea), *Helicobacter*, *Chlamydia* (clamydiosis), *Escherichia* (infecciones urinarias y existe una variedad llamada enterohemorrágica) y *Legionella* (enfermedad del legionario). Los fármacos dirigidos contra la bacGlyRS buscarían evitar la activación del aminoácido al unirse a la región de reconocimiento, evitar la aminoacilación, o la dimerización de la subunidad  $\alpha$  (la dimerización es distinta a la de las otras sintetasas) o su interacción con la subunidad  $\beta$ .

## **9. ANEXO 1.**

### **RESULTADOS ADICIONALES**

#### **9.1 Avances en el estudio estructural del complejo bacGlyRS– tRNA.**

La mayor perspectiva es la determinación de la estructura de la bacGlyRS completa con el tRNA, pues sólo en la estructura completa podrán esclarecerse preguntas como el origen de la subunidad  $\beta$  y la forma de reconocimiento del tRNA, y la relación que existe entre estos elementos con las otras sintetasas. En este apartado se muestran los avances para la obtención de dichas estructuras.

#### **9.2 Las subunidades $\alpha$ y $\beta$ forman forman la GlyRS completa al unirse con alta afinidad.**

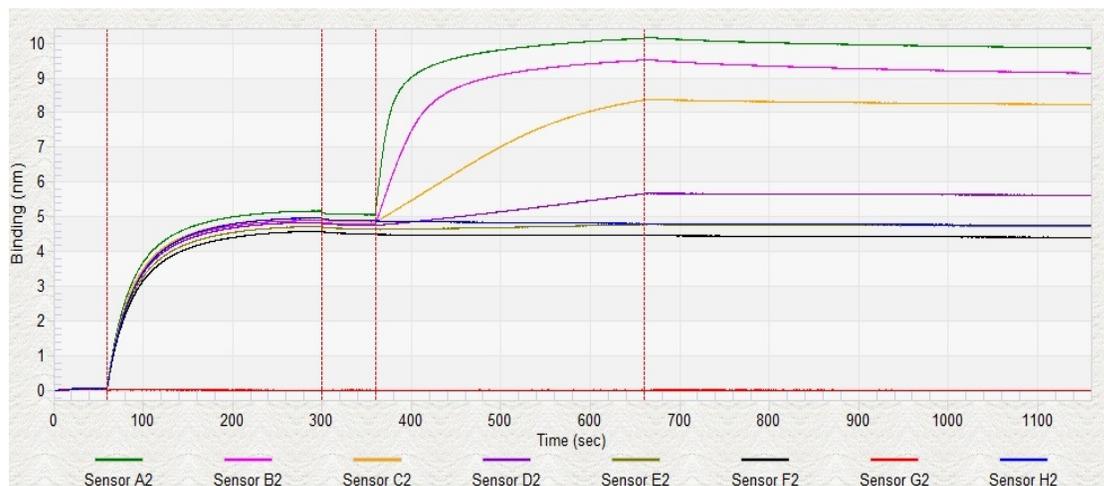
Se demostró que las subunidades  $\alpha$  y  $\beta$  de la bacGlyRS forman un complejo con una Kd de alrededor de 700 nM mediante interferometría de biocapa (figura 9). La subunidad  $\alpha$  fue biotinilada y unida al biosensor de streptavidina, y se determinó la tasa de asociación-disociación de la subunidad  $\beta$  a varias concentraciones, para obtener la constante de disociación Kd, como medida de la afinidad (figura 9).

La existencia y tamaño del complejo  $\alpha$ - $\beta$  fue comprobado también por SDS-PAGE desnaturizante y nativa, del complejo reconstituido y purificado por SEC (anexo 2, figura 15).

Asociación  
subunidad  $\alpha$   
al biosensor

Asociación  
subunidad  $\beta$

Disociación  
subunidad  $\beta$



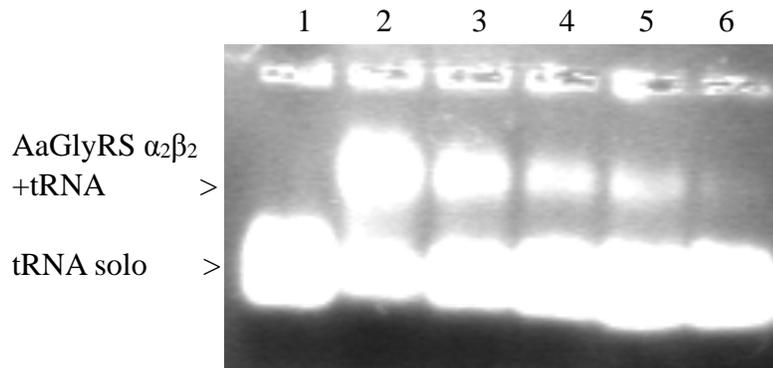
**Figura 9. Las subunidades  $\alpha$  y  $\beta$  forman la GlyRS completa al unirse con alta afinidad.** Se muestran los trazos en la gráfica de la interferometría de biocapa (realizada en el equipo Octet-red, ForteBio) de la asociación de la subunidad  $\alpha$  biotinilada al biosensor de estreptavidina, y un segundo paso de asociación de distintas concentraciones de la subunidad  $\beta$  a la subunidad  $\alpha$ , y la disociación de la subunidad  $\beta$  incluyendo los controles respectivos. El programa del equipo procesó los datos de la fracción unida y las concentraciones para determinar la constante de disociación  $K_d$  de 700 nM entre las subunidades.

### 9.3 La bacGlyRS tetramérica forma un complejo estable con el tRNA.

Se demostró la formación de un complejo estable de la AaGlyRS tetramérica (reconstituida de subunidades  $\alpha$  y  $\beta$ ) con el tRNA-Gly de *A. aeolicus* (figura 10). El tRNA de *Aquifex aeolicus* fue sintetizado a partir de un templado (armado de 4 oligonucleótidos) de DNA que lo codificaba y su transcripción *in vitro* por un integrante de nuestro grupo de investigación (Dra. Annia Rodríguez Hernández). Se encontraron las condiciones de corrida en gel de agarosa (en condiciones nativas) para observar claramente el retardamiento o co-migración del tRNA con enzima tetramérica.

Se realizó una curva de concentraciones de proteína manteniendo constante la concentración de tRNA. El tRNA fue retardado, co-migrando al formar un complejo en cantidades proporcionales a la cantidad de proteína adicionada (figura 10). La estabilidad del complejo pudo verse reflejada ya que el retardamiento observado fue de una banda de tRNA que co-migra completamente con la proteína y no un barrido del mismo (figura 10).

Resultados preliminares indican que la afinidad de la interacción se halla en el orden de una Kd de 100  $\mu$ M.



**Figura 10. La AaGlyRS  $\alpha_2\beta_2$  forma un complejo estable con el tRNA.** Se evaluó la estabilidad de la interacción del tRNA<sup>Gly</sup> con la bacGlyRS  $\alpha_2\beta_2$  mediante retardamiento en gel de agarosa, observándose su co-migración y el retardo proporcional a la concentración. Se muestra el gel de agarosa teñido con bromuro de etidio, Carril 1: tRNA de *A. thermophila* sin proteína. Carriles 2 al 6: Cantidades descendentes de proteína de la AaGlyRS  $\alpha_2\beta_2$  manteniendo constante el monto de tRNA<sup>Gly</sup>.

#### 9.4 En la estructura a baja resolución de la bacGlyRS de subunidades fusionadas puede observarse la disposición de las subunidades

En algunas pruebas de cristalización previos a este trabajo (Torres-Larios *et al.*, 2002) se había observado la disociación de las subunidades  $\alpha$ - $\beta$  de la GlyRS de *E. coli* en algunas condiciones de cristalización.

Por este antecedente, uno de los caminos que seguimos para tratar de obtener la estructura de la bacGlyRS completa fue emplear una enzima en que las subunidades  $\alpha$  y  $\beta$  son codificadas como un solo polipéptido. Las subunidades se encuentran fusionadas mediante un conector peptídico de una longitud de unos 20 aminoácidos. Esta enzima de subunidades fusionadas está distribuida en algunos organismos que van desde los relacionados con el grupo de las Chlamidias hasta todos los cloroplastos de plantas y algas. Encontramos un organismo termofílico que presentaba este tipo de enzima de subunidades fusionadas, la GlyRS de *Anaerolinea thermophila* (AtGlyRS).

La proteína AtGlyRS fue expresada en *E. coli* y purificada. Sin embargo, se observó que la proteína sufría un corte proteolítico que la separaba en las subunidades. Este corte lo estudiamos más a fondo mediante secuenciación de Edman de la porción amino terminal de

los fragmentos derivados del corte. Realizamos también distintas mutantes con distintos conectores peptídicos reemplazando el conector entre las subunidades de la AtGlyRS. Uno de los conectores peptídicos fue derivado de la secuencia de un conector peptídico de la polimerasa de un fago. Sin embargo, en todos los casos la degradación proteolítica ocurrió en la región del conector peptídico sin importar su secuencia.

Un miembro de nuestro grupo de investigación encontró que la mutante AtGlyRS con el conector peptídico del fago en complejo con el tRNA-Gly de *A. aeolicus* formaba cristales, pero en la difracción fueron altamente anisotrópicos. Sin embargo, el resultado es que este conector peptídico del fago facilitó su cristalización.

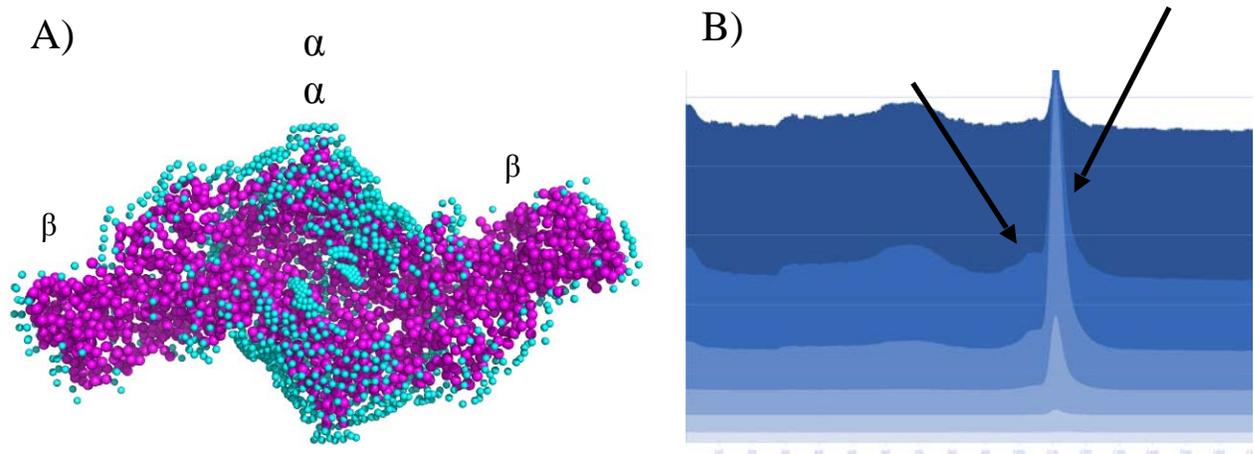
#### **9.4.1 Estructura a baja resolución de la AtGlyRS.**

Se realizaron distintos experimentos de dispersión de rayos X de ángulo pequeño (SAXS) para determinar la estructura a baja resolución de la GlyRS de subunidades fusionadas de *Anaerolinea thermophila* (figura 11A) (detalles en el cuerpo del anexo 2, la figura 15 y tabla 6).

Debido a que siempre estuvo presente un pequeño monto de agregado de peso molecular cercano al doble de la enzima completa, mediante cromatografía de exclusión de tamaño (SEC) acoplada a SAXS (SEC-SAXS) con adquisición continua de difracción se adquirieron exposiciones sin la presencia del agregado (figura 11B).

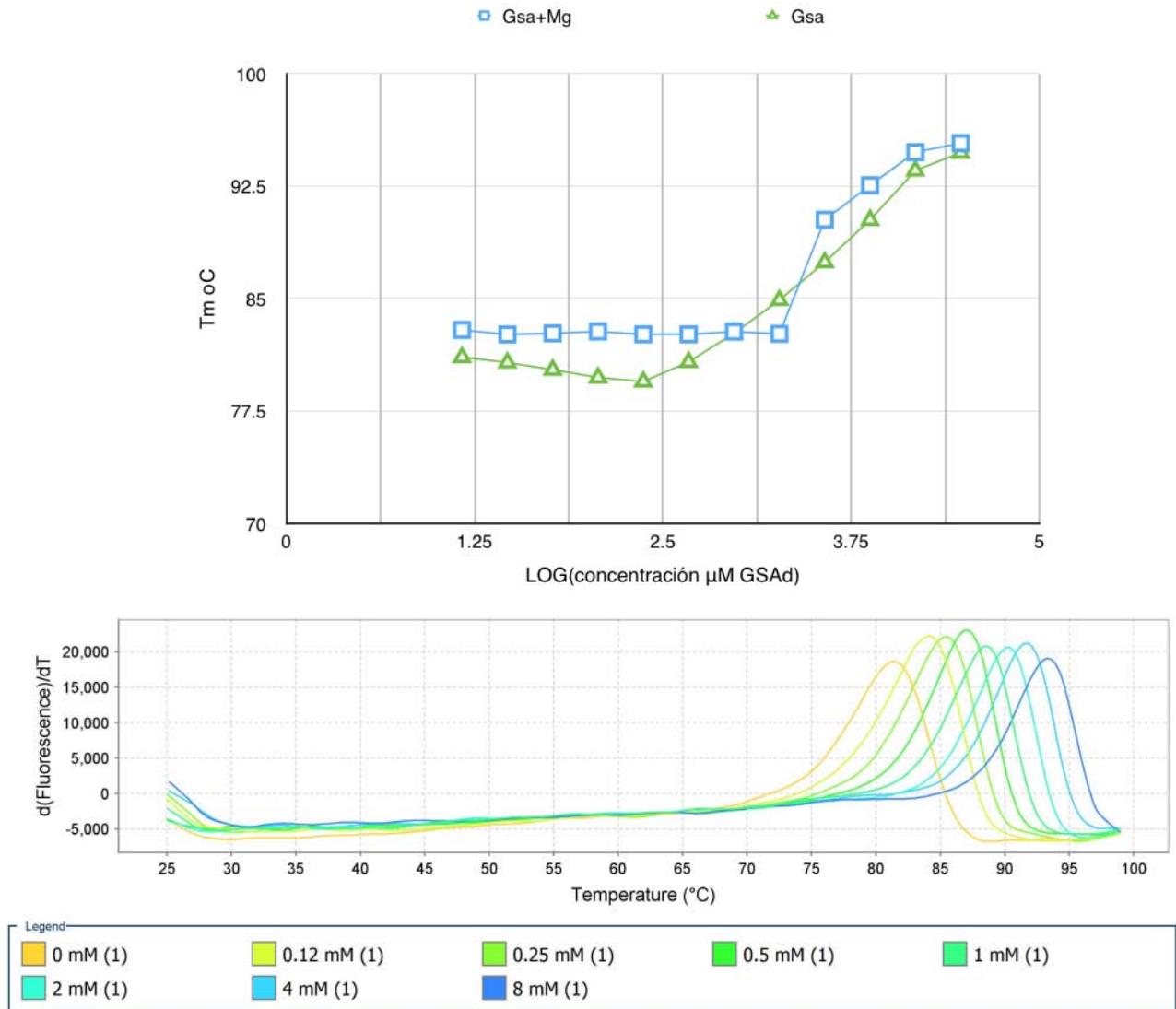
La opción alterna fue realizar la adquisición de datos por SAXS inmediatamente después del SEC, sin estar acoplado pero de forma inmediata. De esta forma se determinó la envoltura de la AtGlyRS silvestre mediante SAXS y se determinó *ab initio* su estructura a baja resolución, mediante el tratamiento de datos con el programa GASBOR (figura 11A) (Petukov *et al.*, 2012).

En la estructura a baja resolución *ab initio* de la AtGlyRS se puede observar la posible disposición de las subunidades  $\alpha$  y  $\beta$  en la bacGlyRS; las subunidades  $\alpha$  al centro de la enzima responsables de la oligomerización y la subunidades  $\beta$  proyectándose hacia el exterior (figura 11A).

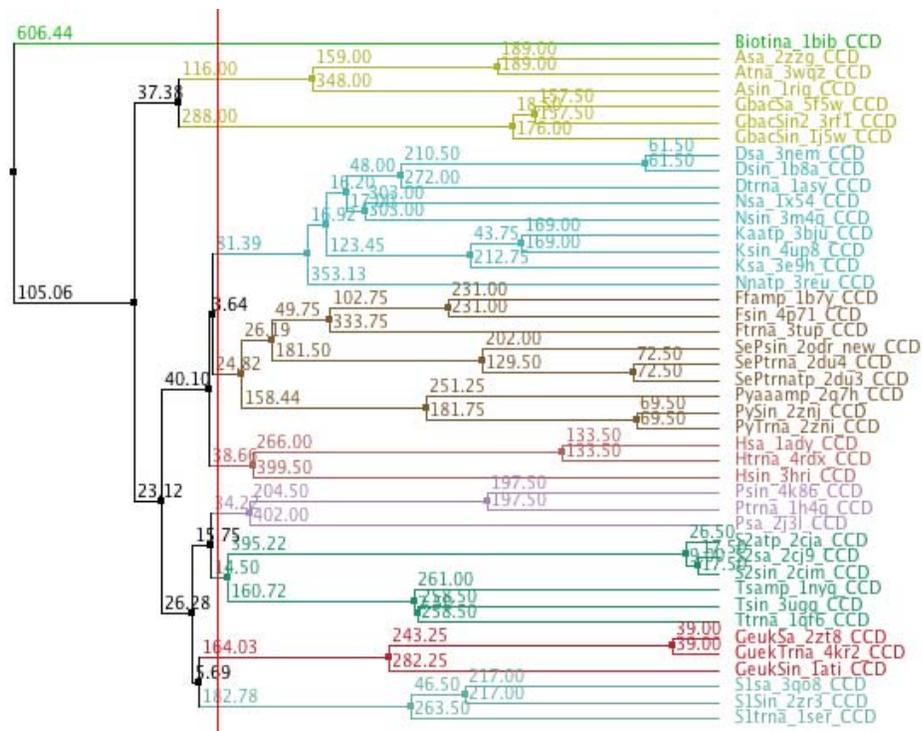


**Figura 11. En la estructura a baja resolución de la bacGlyRS de subunidades fusionadas puede observarse una posible disposición de las subunidades.** A) Se obtuvo la estructura a baja resolución mediante modelado *ab initio* de la AtGlyRS silvestre ( $\alpha$ - $\beta$ )<sub>2</sub> con el programa GASBOR, que fue el que ajustó mejor a los datos experimentales de SAXS. La posible disposición de las subunidades en la AtGlyRS donde la subunidad  $\alpha$  se halla al centro aportando la interacción del dímero, y las subunidades  $\beta$  se proyectan hacia afuera. B) El problema de una mínima polidispersidad presente en todas las preparaciones de la AtGlyRS producía una alta señal de dispersión, que fue eliminada mediante SEC-SAXS; se muestra la gráfica de la colecta con una medición continua de las exposiciones a la salida de la columna.

## 9.5. Figuras de resultados adicionales



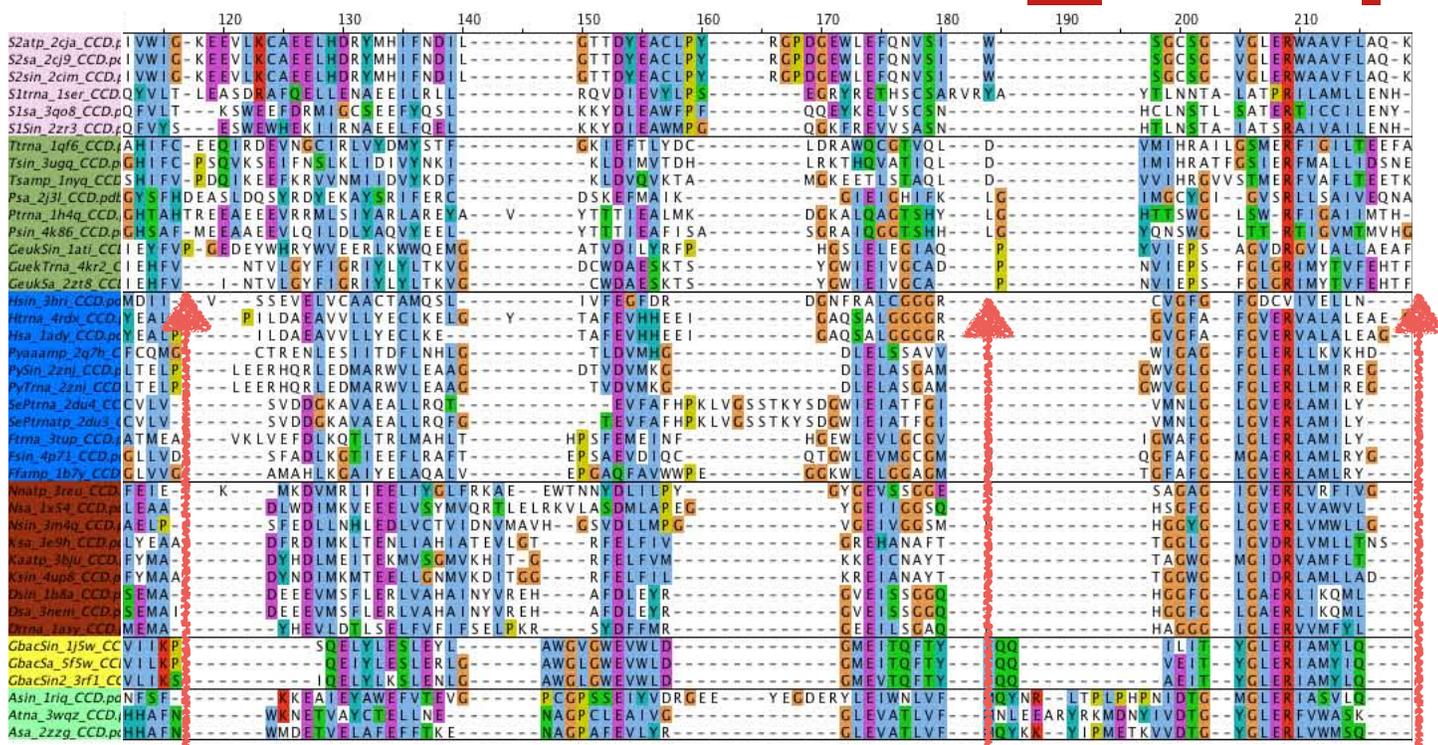
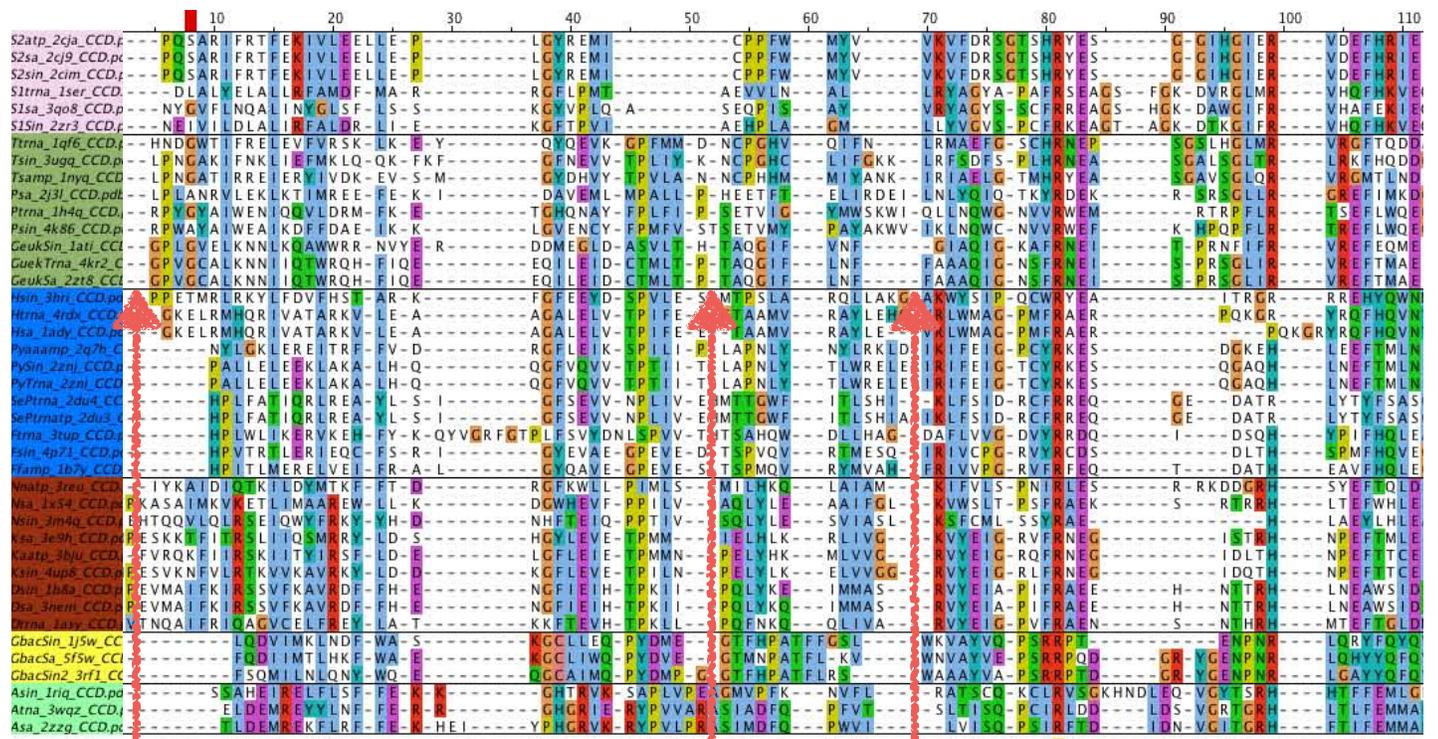
**Figura 12. Los cambios conformacionales promovidos por el GSAd corresponden a una estabilización térmica en la  $\alpha$ -AaGlyRS.** A) Desplazamiento térmico de la Tm de la  $\alpha$ -bacGlyRS según la concentración presente de GSAd solo o en presencia de MgCl<sub>2</sub>. B) Ensayo de desplazamiento térmico en presencia del GSAd y MgCl<sub>2</sub>.



**Figura 13. Dendrograma de agrupación por secuencia de regiones estructurales equivalentes de las aaRSs II.** Las regiones del núcleo invariante del CCD de las aaRSs fueron alineadas por secuencia tomando en cuenta la información estructural mediante el servidor T-COFFEE y el dendrograma fue construido con Jalview por distancia media con la matriz BLOSUM62. El dendrograma fue enraizado con la holoenzima de la sintetasa de biotina con la que las aaRSs II comparten el plegamiento del CCD.

**(Siguiete página)**

**Figura 14. Secuencias del CCD de las aaRSs II alineadas con T-COFFEE.** Secuencias del alineamiento reportado en la figura 13. Las flechas indican la posición de los motivos 1, 2 y 3, y las barras rojas secuencias conservadas incluyendo la firma en las aaRSs II **(Siguiete página)**



## 10. ANEXO 2. MÉTODOS Y RESULTADOS ADICIONALES

### Diseño del gen de la GlyRS de *A. aeolicus*

**Origen del gen:** *Aquifex aeolicus* es una bacteria Gram negativa quimiolitótrofa obligada, microaerófila, que es una de las dos bacterias más hipertermófilas conocidas, cuya temperatura de crecimiento se halla en 85°C, aislada en el parque de Yellowstone, USA (Deckert *et al.*, 1998). Se considera como una de las bacterias que más temprano divergió en la evolución y que ocupa una de las ramas más cortas y profundas en el árbol de la vida en el dominio de las bacterias (Deckert *et al.*, 1998).

En su metabolismo de la traducción, y en particular de las aaRS, presenta características muy importantes como que carece de GlnRS y AsnRS, empleando para estas funciones una transamidasa (Deckert *et al.*, 1998). Mas aún, su versión de la MetRS y LeuRS están codificadas por dos cadenas polipeptídicas muy separadas en el cromosoma, lo que resulta importante pues las formas canónicas de estas sintetetas sólo se conocen codificadas en una sola cadena polipeptídica en otros organismos, por lo que la organización de las subunidades de la MetRS y LeuRS de *A. aeolicus* posiblemente refleja la posterior organización en dominios de las proteínas homólogas de los otros organismos (Deckert *et al.*, 1998).

**Optimización de la secuencia:** El gen de las subunidades  $\alpha$  y  $\beta$  de la GlyRS de *A. aeolicus* fue optimizado con los codones encontrados con más alta frecuencia de uso en *E. coli* y sintetizado por la compañía Genscript entre las secuencias de restricción, y con la secuencia que codifica para una etiqueta de 6His separada por la secuencia de corte de la proteasa del virus del jaspeado del tabaco (TEV). El diseño original que fue sintetizado contenía la secuencia codificante para las subunidades  $\alpha$  y  $\beta$  fusionadas por un conector peptídico en el vector pUC57.

El gen de la GlyRS de *A. aeolicus* fue subclonado en el vector pET23a (3592 pb, Novagen/Millipore) empleando enzimas de restricción.

### **Mutación por cambio rápido (quick-change) para separar las subunidades $\alpha$ y $\beta$ .**

Las subunidades fueron separadas mediante cambio rápido por PCR (reacción en cadena de la polimerasa) en la que todo el vector es amplificado; ya que los oligonucleótidos contienen una versión mutante de la región a la que se aparean, los amplicones son versiones

conteniendo la mutación de la construcción amplificada.

**Oligonucleótidos:** Los oligonucleótidos para cambio rápido por PCR se diseñaron para alinearse a una temperatura arriba de 72°C, con una longitud alrededor de 40 nucleótidos, conteniendo en la región central la secuencia mutante y buscando que no se aparearan entre el oligonucleótido directo y reverso o consigo mismos.

**Subunidad  $\alpha$ :** La subunidad  $\alpha$  fue separada mediante la introducción de un codón de paro.

Oligonucleótido directo  $\alpha$ :

ggtgtttgaaaacgtggcgcgaccttaactgaaagaaggcgcggtg

Oligonucleótido reverso  $\alpha$ :

caccgcgcttctttcagttaggctcgcgcccacgtttcaaacacc

**Subunidad  $\beta$ :** Se diseñaron los oligonucleótidos para producir una horquilla en el vector al amplificar y poder separar la subunidad  $\beta$ , para poder así conservar las 6His pero deletar la subunidad  $\alpha$ .

Histidinas en el extremo amino

Oligonucleótido directo  $\beta$  NH:

CatcatgaaaacctgtatcttcagagcAtgaccaacgaactgctgattGaaattggcaccgaagaactgccg

Oligonucleótido reverso  $\beta$  NH:

CggcagttcttcggtgccaatttcAatcagcagttcgttggtcatGctctgaaaatacaggtttcatgatg

Histidinas en el extremo carboxilo

Oligonucleótido directo  $\beta$  COOH:

GttaaacttaagaaggagatatacatAtgaccaacgaactgctgattGaaattggcaccgaagaactgccgg

Oligonucleótido reverso  $\beta$  COOH:

CggcagttcttcggtgccaatttcAatcagcagttcgttggtcatAtgtatatctccttcttaaagtaaac

**PCR para cambio rápido:** La mezcla de reacción contenía: 100 ng vector, mezcla de dNTP 's 200  $\mu$ M, Oligonucleótidos 200 nM y 2 unidades de Vent DNA polimerasa (New England Biolabs) por cada 50 $\mu$ L de reacción, en el amortiguador de la polimerasa que ya contiene

magnesio.

El programa de extensión empleado fue:

-T desnaturalización inicial 94°C X 20 segundos

14 ciclos de

-T desnaturalización 94°C X 30 segundos

-T alineamiento 55°C X 1 minuto

-T extensión 72°C X 7 minutos

y T extensión final 72°C X 8 minutos.

Para la mutagénesis por cambio rápido:

-Se realizó la amplificación de todo el vector con el oligonucleótido directo por 5 ciclos

-Se agregó el oligonucleótido reverso y se realizó la amplificación de todo el vector por 14 ciclos, de forma que los amplicones resultado contendrían la mutación.

-Se digirió el vector parental (sin la mutación) con 20 unidades de nucleasa *Dpn I* por al menos 2 horas, esta enzima corta sólo en las secuencias con nucleótidos metilados.

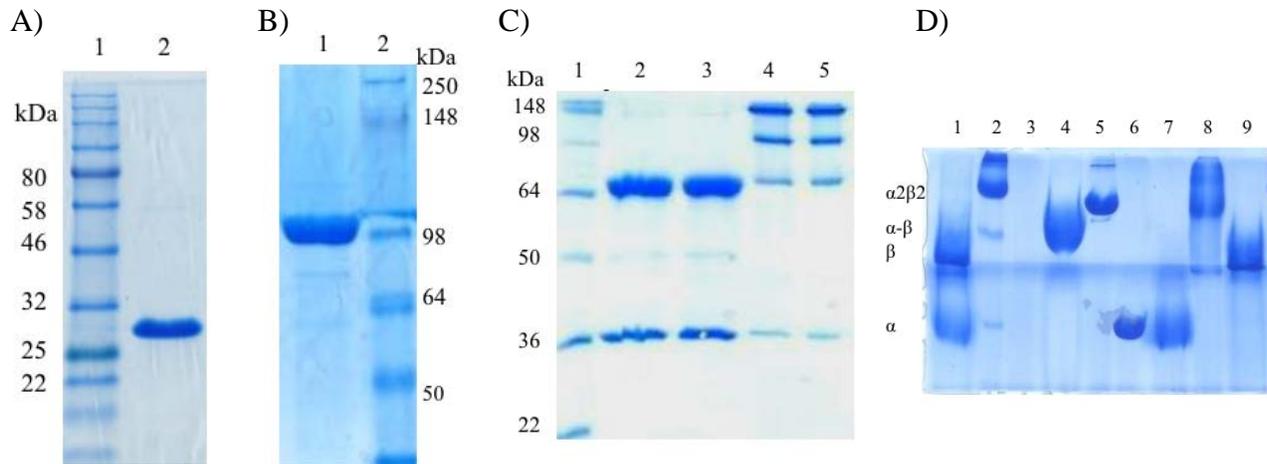
**Transformación en cepas de expresión:** En un tubo de microcentrífuga, 50 µL de células competentes fueron incubadas con 10 µL de la reacción de PCR de cambio rápido por 30 minutos en hielo, se les aplicó el choque térmico por 45 segundos a 42°C, y se incubaron nuevamente en hielo por 2 minutos. Para la recuperación de las bacterias transformadas se agregaron 500 µL de medio LB templado y se incubó el tubo de microcentrífuga en agitación a 37°C por 1 hora. Tras la incubación se plaquearon las bacterias transformadas en placas con agar-LB con ampicilina a 100µg/mL. Es importante mencionar que el DNA fue transformado directamente en cepas de expresión de *E. coli* BL21(DE3) (también fueron ensayadas las cepas BL21CodonPlus(DE3)-RIL (Agilent Technologies) y Rosetta(DE3) (Novagen).

**Pruebas de expresión:** Se realizó la prueba de expresión de 5 colonias y se extrajo el DNA. Para la expresión 20mL de cultivo con LB-Vegitone (Sigma) conteniendo 100µg/mL de ampicilina de la clona se indujeron a una OD600 de 0.6 con IPTG 1mM durante 3 horas, y el botón celular se congeló. Las células fueron lisiadas por sonicación en el amortiguador A de níquel y se realizó una purificación con resina de afinidad a níquel en lote con 30 µL de perlas de agarosa níquel-NTA (Quiagen). Las proteínas adheridas durante 10 minutos a la resina fueron lavadas y eluidas con 30µL de amortiguador A con 500 mM de imidazol. Las

perlas fueron centrifugadas y el eluido se analizó por SDS-PAGE teñido con azul de Coomassie (figura 15, anexo 2). El DNA de las clonas positivas para la expresión se extrajo de 10 mL de cultivo durante toda la noche (O.N.) de las mismas colonias empleadas para expresión, empleando el kit de extracción de DNA genómico de Qiagen.

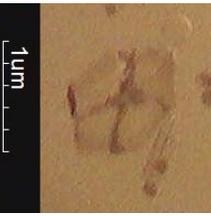
**Secuenciación:** 1 µg de DNA eluido en agua inyectable de las clonas que mejor expresaron cada una de las proteínas se mandó a secuenciar a la compañía HTSeq con los oligonucleótidos para el promotor y terminador de T7, para verificar que las subunidades estuvieran en el marco de lectura correcto y no contuvieran mutaciones no deseadas.

**Optimización de la expresión:** La expresión de la subunidad  $\alpha$  y  $\beta$  fue optimizada determinando el efecto de distintas concentraciones de IPTG (250 µM-2mM), distintos tiempos de inducción (de 1 a 16 hrs) y O.D.<sub>600 nm</sub> de inducción (0.3 -1.2). Las diferencias en la expresión pudieron evidenciarse directamente cargando 15 µL de cultivo cargados con amortiguador de Laemmli de proteínas en SDS-PAGE (figura 15, anexo 2).



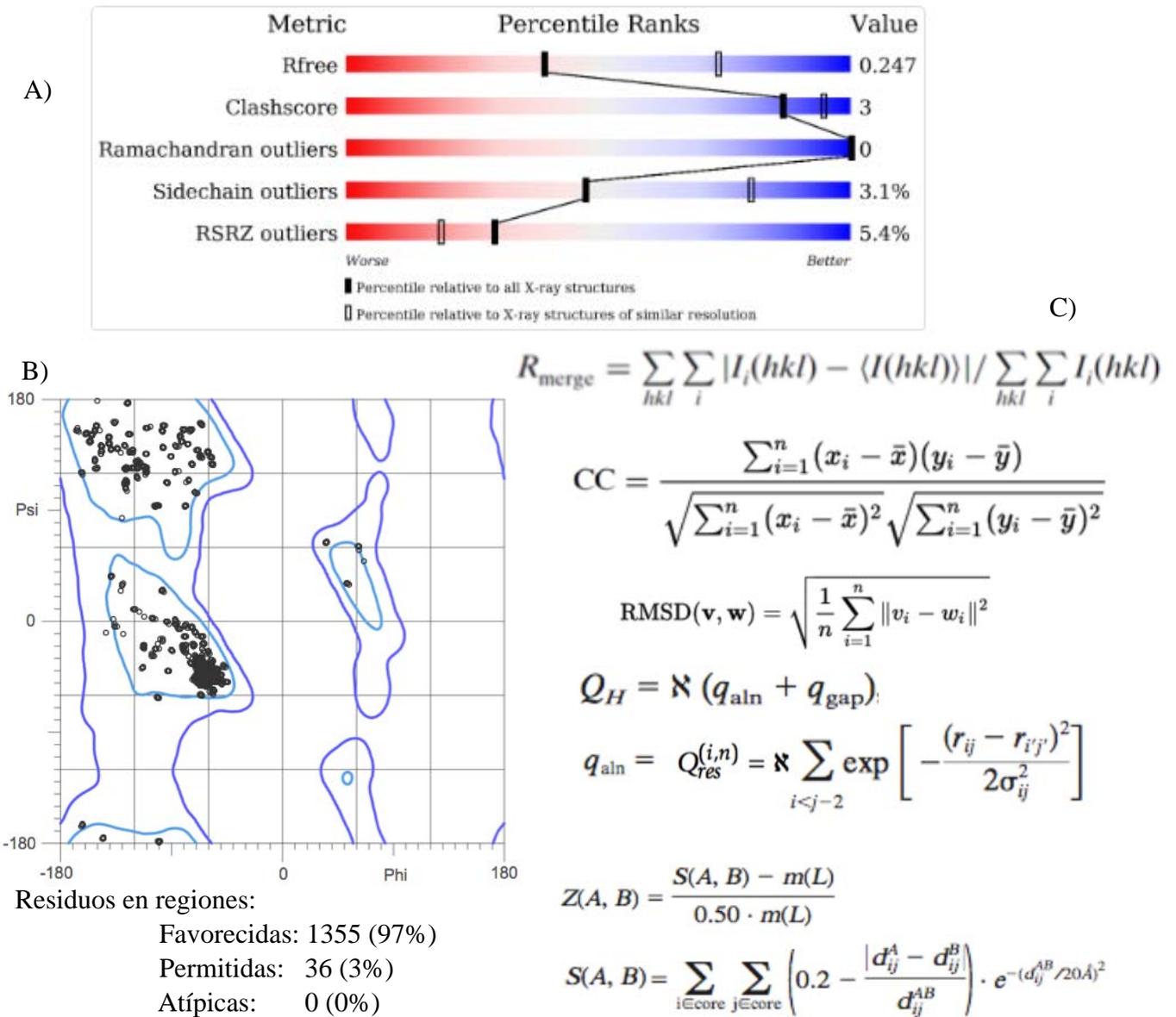
**Figura 15. Diferentes versiones de la bacGlyRS provenientes de distintos organismos fueron expresadas y purificadas para los estudios estructurales** A) Análisis por SDS-PAGE de la  $\alpha$ -AaGlyRS hervida para desnaturalizar el dímero. Carril 1: Marcador de peso molecular. Carril 2: Monómero de la  $\alpha$ -AaGlyRS que migra en aproximadamente 30 kDa. B) Análisis por SDS-PAGE de la GlyRS de *A. thermophila* (AtGlyRS) de subunidades fusionadas ( $\alpha$ - $\beta$ )<sub>2</sub>. Carril 1: monómero de la AtGlyRS con una longitud de 998 aminoácidos. Carril 2: Marcador de peso molecular. C) Análisis por SDS-PAGE de la reconstitución de la  $\alpha$ 2 $\beta$ 2 bacGlyRS. Carril 1: Marcador de peso molecular. Carriles 2 y 3: AaGlyRS  $\alpha$ 2 $\beta$ 2 hervida. Se puede observar la banda del monómero de la subunidad  $\alpha$  y la del monómero de la subunidad  $\beta$  (migra aproximadamente 75 kDa). Carriles 4 y 5 la AaGlyRS  $\alpha$ 2 $\beta$ 2 sin hervir, mostrando la sobrevivencia del dímero y el tetrámero. D) Análisis por PAGE-azul nativa (BN-PAGE) empleando azul de Coomassie para facilitar la migración de las proteínas en condiciones nativas de las diferentes GlyRS y sus subunidades. Carriles 1 y 2): AaGlyRS  $\alpha$ 2 $\beta$ 2 hervida en presencia de DTT para separar y desnaturalizar las subunidades (carril 1) o en condiciones nativas (carril 2). Carriles 3): Vacío. Carril 4 y 5: AtGlyRS de subunidades fusionadas ( $\alpha$ - $\beta$ )<sub>2</sub> hervida separar y desnaturalizar el dímero (carril 4) o en condiciones nativas (carril 5). Carriles 6 y 7): Subunidad  $\alpha$  de la AaGlyRS en condiciones nativas (carril 7) o hervida para separar y desnaturalizar las subunidades. Carriles 8 y 9) subunidad  $\beta$  de la AaGlyRS en condiciones nativas en ausencia de DTT (carril 8) o hervida en presencia de DTT (carril 9).

**Tabla 5. Otras condiciones de cristalización de la  $\alpha$ -AaGlyRS**

Ligandos y comentarios	Solución del reservorio	Temperaturas	Hábito cristalino	Difracción	Cristal
Co-cristales con ATP-Mg con o sin Gly	12% de polietilen glicol (PEG) 20,000, MES 100mM pH 6.5	28oC (cuando se usó 12% de PEG). 18oC (cuando se usó 16% de PEG).	Columnar. De erizos a prismas maclados dependiendo la concentración de proteína, de Mg, DTT y Gly	Abajo de 4 Å  crioprotección con 25% etilenglicol	
Co-cristales con GSAd	22% w/v ácido acrílico, 100mM Hepes pH 7.5, 20mM de MgCl	18°	Columnar. De erizos a prismas maclados.	No probado	
Sin ligando	Varias condiciones	28°	Columnar a prismas con los extremos maclados.	2.8 Å, no procesable	
Co-cristales con GSAd-Mg. Mismos resultados con microsembrado.	12% de PEG 20,000, MES 100mM desde pH 5.5 hasta 7.5	28°	De estrellado a placas. De erizos a placas macladas. Tamaño dependiente del pH.	2.2 Å, no indizable  crioprotección con 25% etilenglicol	
Co-cristales con ATP-Mg	10mM CTAB, 10mM MgCl 500mM NaCl. + varios aditivos	18°	Cristales únicos con forma rombohédrica	Abajo de 4 Å  crioprotección adicionando 25% PEG400	
Co-cristales con ATP-Mg con Gly. Sumergidos con GSAd.	Formato de Mg 100mM, 15% PEG 3350	28° trasladados a 18°	Cristales únicos piramidales	2.78 Å, P2  crioprotección llevando a 30% el PEG3350	
Co-cristales con ATP-Mg con Gly. Sumergidos con GSAd. Grupo de datos de la estructura.	100mM tiocianato de potasio, 30% de PEG mmE 2000	28° trasladados a 18°	Cristales únicos.	2.81 Å, P22121  crioprotección con 15% de glicerol	

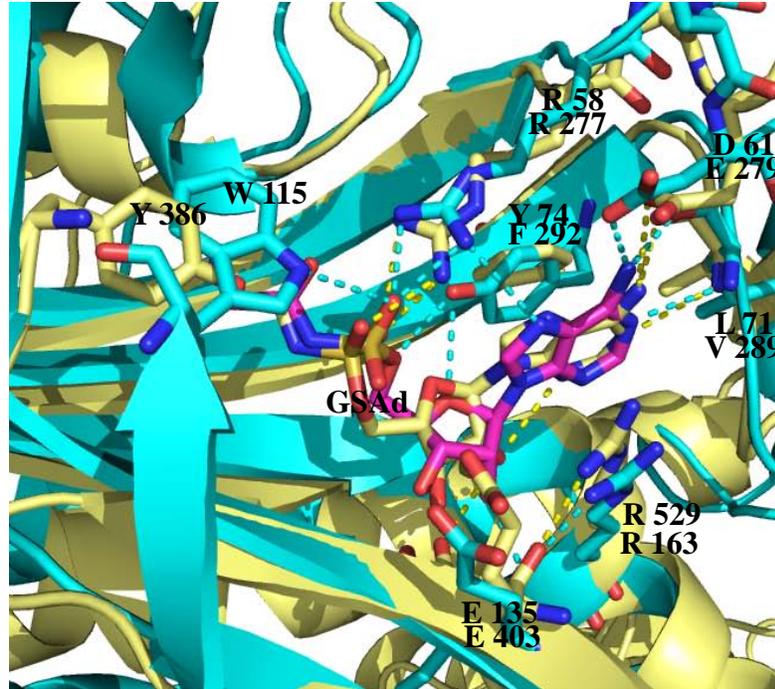
Crecimiento en arriba de 6 meses	<b>Subunidad <math>\beta</math></b> AaGlyRS 1.6M de citrato de sodio, pH 6.5	28°	De cuasicristales a cristales globulares	Abajo de 8 Å	
----------------------------------	--	-----	--	--------------	---

Algunas de las pruebas preliminares de difracción de los co-cristales que difractaron abajo de 4 Å obtenidos con 12% de polietilen glicol (PEG) 20,000, MES 100mM pH 6.5 en el reservorio (condición 1 de la tabla 5), se realizaron en el difractorómetro de rayos X Rigaku del Laboratorio Nacional de Estructura de Biomacromoléculas en el Instituto de Química.

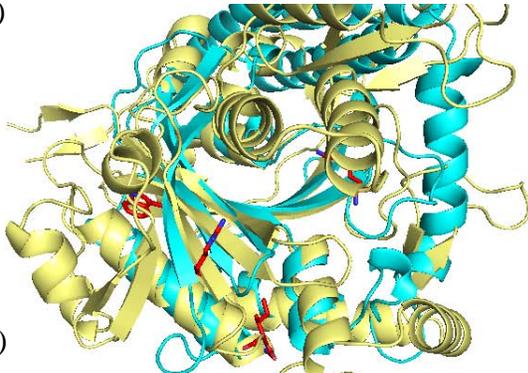


**Figura 16. Estadísticas adicionales de la estructura de la  $\alpha$ -AaGlyRS y la puntuación en el alineamiento estructural.** A) Estadísticas finales de la estructura de la  $\alpha$ -AaGlyRS depositada en el PDB con el código 5f5w. B) Gráfico de Ramachandran de A. C) Distintas fórmulas de algunas estadísticas de la reducción de datos (Rmerge y CC) y de la puntuación de los alineamientos estructurales reportados en este trabajo (RMSD, Qh y puntuación Z).

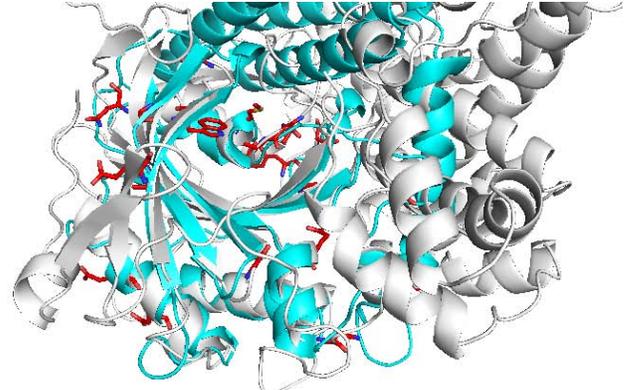
A)



B)



C)



D)

```

DSSP  LHHHLHHHL-----LILLLEEEEEEEELHhLHLLLL--LLEEEEEEEEEE
Query  GAGTMNpATFLK-----vlgKKPWNvAYVEPSRRPqdGRYGENP-NRLQHYQFQVILK  82
ident  |||
Sbjct  PETAQG-IFLNfkrlllefngqKLPFAAAQIGNSFRN--EISPRSGIIRVREFTMAEIEHF  237
DSSP  LLLHHH-HHLLHhhhhhhllllLEEEEEEEEBELL--LHLLHLLHLLLEEEEEEEEEE

DSSP  L-----LHLLHHHHHH
Query  P-----APRNPQEIYL  93
ident  |||
Sbjct  VdpsekdhpkfqnvdhlhlylsakaqvsqgsarkmrlgdaveqgvinntVLGYFIGRIY  297
DSSP  ElhhhhllllhhhhllllllleeehhhhhhllllleellhhhhhhllllllhHHHHHHHHHH

DSSP  HHHHLLHhLLEEEEE---EEEEEhhleEEEEEEEL---LEEEEEEEEEE-E
Query  ESLELRGINPleHDIRFVED---DWESptlgaWGLGWEVWLD---GMEITQFTYFQ-Q  144
ident  |||
Sbjct  LYLTkVGIsp--DKLRFQhmenemaHY---ACDCWAESKtsygnIEIVGCADRScY  350
DSSP  HHHHLLH--HHEEEELllllL---LLEEEEBEelleEEEEELHLLH

DSSP  ELLE-----eLLEEEEEEEHHHHHHLL-----Llll
Query  AGGL-----dIDEISVEITYGLERIAMYIQD-----kdev  174
ident  |||
Sbjct  DLSCharatkvpIvaekpyveevVPNVIEPSFGLGRIMYTVFEhtfhvregdeqrffsf  410
DSSP  HHHHHhhhhllllleeeellleeLLEEEELHHHHHHHHHHhhleeeellllllleeeell

```

E)

```

Query  IIMTLHKFWAEK---GCLIWQPYDV-----EVGAGTMNpatFLKVLG---kkpWNV  50
ident  |||
Sbjct  MREKFLRFPEKHeiyPHGRVKRYPVlprwrdvdlLVGASIMDF-qPWVISGeadppaNPL  119
DSSP  HHHHHHHHHHLLHLLLEELLLLLLllllllllllLHLLHHHHH--LHLLHLLHLLLE

DSSP  EEEEEELHHL-LHLLHLLLEEEEEEEEEE-----LHLL-LHHHHHHHHH-HLLL
Query  AYVEPSRRPQDG-RYGENPNRLQHYQFQVILK-----PAPR-NPQEIYLESLE-RLGI  101
ident  |||
Sbjct  VISQPSIRFtDIdnVGIAGRHTIFEMMAHAFNypgkplYWMdETVELAFEFFtKELKM  179
DSSP  EEEEEELLLLLLHLLHLLLEEEEEEEEEEe11111LHhHHHHHHHHHHHLLHLL

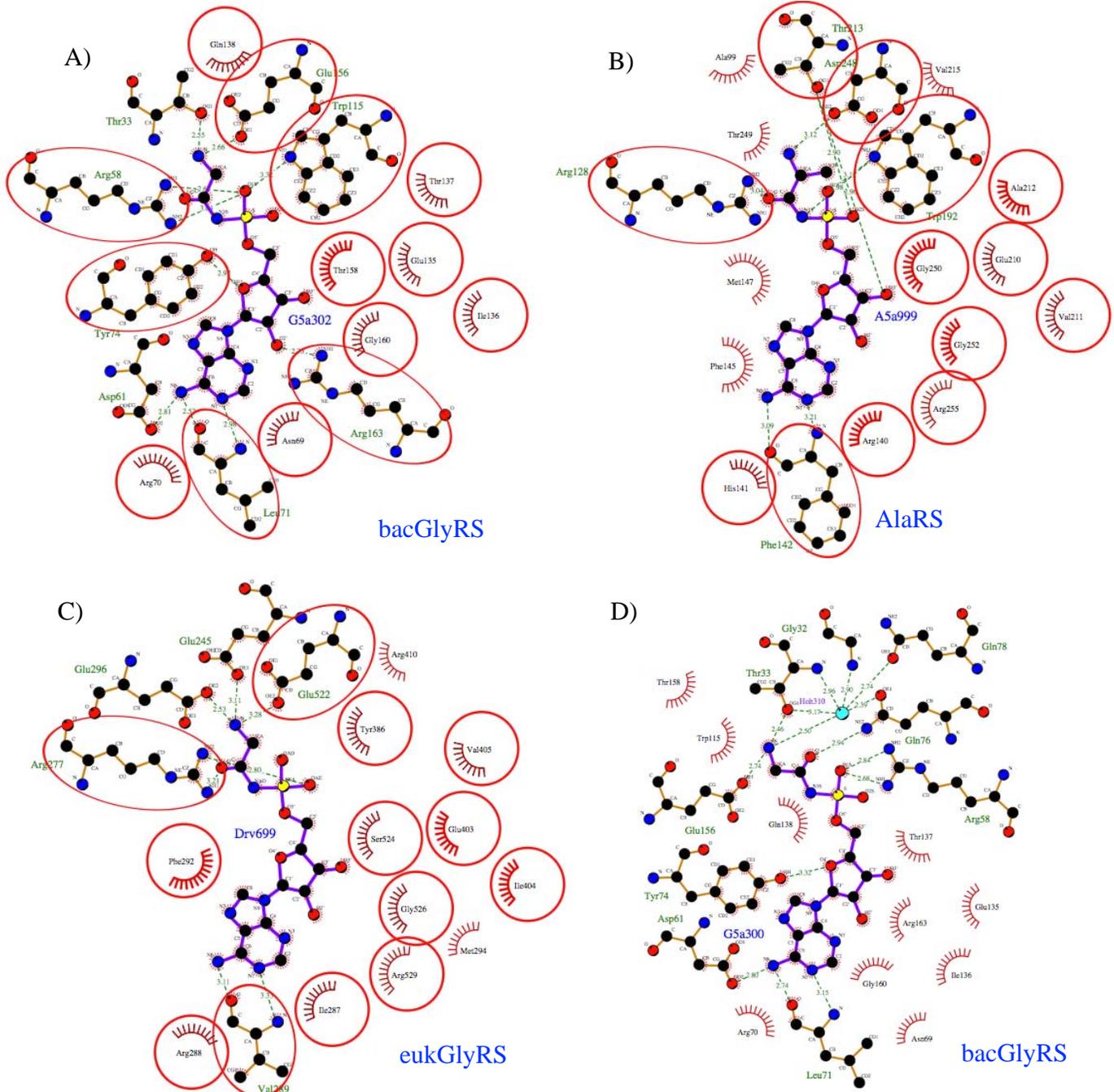
DSSP  LhhLLEEEEEEEehhHLEEEEEEEELLEEEEEEEEEEEL-----l
Query  NPleHDIRFVEDDWEspLGAWGLGWEVWLDGMEITQFTYFQAG-----g  147
ident  |||
Sbjct  KP--EDITFKENPWAG--GCNAGPAFVLYRGLVATLVFMQYRKapeqdvvvikgeky  235
DSSP  LL--LLEEEEEEEEEE---LLEEEEEEEELLEEEEEEEEEEEL11111leelllee

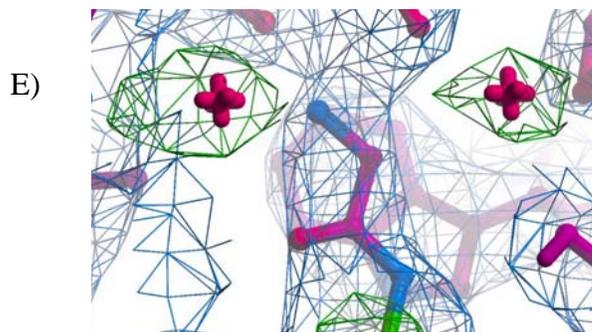
DSSP  EELLLLLLEEEEEEEHHHHHHHLLHLLHLLLEeelleeHHH-----
Query  LDLEISVEITYGLERIAMYIQDKDSVFDIewkegitYGEI-----  188
ident  |||
Sbjct  IPMETKVVVDtGYGLERLVWMSQGTPTAYDAV-----LGYVveplkkmagiekidekilm  289
DSSP  EEEEEEEEEEHHHHHHHLLHLLHLLH---LHHhhhhhhhhllllllhhhh

```

Figura 17. Residuos conservados y estructuralmente equivalentes entre la bacGlyRS, la eukGlyRS y la AlaRS, y el reconocimiento del ATP. A) El reconocimiento del ATP por la

bacGlyRS (cian) (código PDB 5f5w) y la eukGlyRS (amarillo) (código PDB 2zt8) es similar al de todas las aaRSs II. B) Residuos compartidos (líneas rojas) únicamente entre la bacGlyRS (listones en cían) y la eukGlyRS (listones en amarillo) . C) Residuos compartidos (líneas rojas) únicamente entre la bacGlyRS (listones en cían) y la AlaRS (listones en blanco). D) alineamiento de todos los residuos de B equivalentes en secuencia y estructura reportados por DaliLite. E) alineamiento de todos los residuos de C equivalentes en secuencia y estructura reportados por DaliLite.





(página anterior)

**Figura 18. La bacGlyRS presenta semejanzas con la AlaRS en las interacciones con el aminoácido y diferencias con la eukGlyRS.** Diagramas esquemáticos de las interacciones con el análogo del aminoacil-adenilato generados con el programa LigPlotPlus ver 1.4.5 de: A) La subunidad  $\alpha$ -bacGlyRS (código PDB 5f5w, cadena C). B) La AlaRS (código PDB 2zzg). C) La eukGlyRS (código PDB 2zt8). D) De la subunidad  $\alpha$ -bacGlyRS tomando en cuenta las interacciones con las moléculas de solvente (código PDB 5f5w, cadena B). E) Mapas de densidad electrónica 2Fo-Fc (azul) contorneado a  $1\sigma$  y Fo-Fc (verde y rojo) contorneado a  $3\sigma$  donde se observa la densidad electrónica positiva (verde) correspondiente a las dos moléculas de solvente asignadas.

### **Expresión, purificación y características de las GlyRS de subunidades fusionadas.**

Se trabajó con una versión de la bacGlyRS de un organismo que presenta de forma natural ambas subunidades codificadas en una sola cadena polipeptídica de *Anaerolinea thermophila* (AtGlyRS), de forma que es el equivalente a la fusión de las dos subunidades.

*A. thermophila* es una bacteria Gram negativa filamentosa, siendo un organismo anaerobio estricto, termófilo con un crecimiento óptimo a 50°C, aislado de los lodos de aguas residuales industriales en Japón (Yamada *et al.*, 2006).

El gen que codificaba para la GlyRS de *A. thermophila* (AtGlyRS) fue diseñado para contener la etiqueta de 6 histidinas, y sintetizada por la compañía DNA 2.0 con los codones optimizados para expresión en condiciones de inanición (déficit de nutrientes), clonada en el vector pJex04 bajo el control del promotor del fago T5 que es transcrito por la RNA polimerasa celular.

**Expresión y purificación de la AtGlyRS:** La construcción con el gen fue transformada en *E. coli* BL21(DE3) y se realizaron 4L de cultivo en LB-Vegitone con 50  $\mu\text{g}/\text{mL}$  de kanamicina, que fueron inducidos a una O.D.<sub>600 nm</sub> de 1 con IPTG 1mM durante 12 horas a 30°C. Se obtuvo el botón celular por centrifugación como se ha descrito y se congeló el botón celular.

Las células fueron lisadas por sonicación como se describió anteriormente en amortiguador A de cromatografía de afinidad a níquel adicionado con PMSF 1 mM y una pastilla de inhibidor de proteasas Complete EDTA free (Roche). Tras la lisis se obtuvo la fracción soluble por centrifugación y fue purificada por cromatografía de afinidad a níquel como se ha descrito para la subunidad  $\alpha$ , adicionando en todos los pasos inhibidores de proteasas.

La proteína purificada por afinidad a níquel fue precipitada agregando sulfato de amonio hasta una concentración de 1M, y se realizó una cromatografía de interacción hidrofóbica con una columna empacada en casa butil-Toyopearl (10X100mm), equilibrada con el amortiguador A de cromatografía hidrofóbica (Tris 50mM pH 8.8) y fue eluida en una gradiente de sulfato de amonio de 2 M a 0. En todos los pasos de la purificación se adicionó inhibidores de proteasas. La proteína purificada por cromatografía de interacción hidrofóbica se concentró por ultrafiltración con un centricon de 100 kDa a alrededor de 15 mg/mL y se le realizó una separación por cromatografía de exclusión de tamaño (SEC) con la columna Superdex S-200 10X300 (GE) equilibrada con el amortiguador Tris 20mM pH 7.5, NaCl 50 mM y PMSF 0.5 mM. Una fracción conteniendo un único pico se colectó y se concentró por ultrafiltración con un centricon de 100 kDa a alrededor de 15 mg/mL.

**Polidispersión y corte proteolítico en la AtGlyRS:** Posterior a la purificación, como criterio de pureza la proteína se analizó por SDS-PAGE (figura 15, anexo 2) y la fase reversa con que se purificó la proteína donde se observó un solo pico. Al analizar la AtGlyRS por SDS-PAGE se observó una digestión en dos fragmentos de tamaño similar a las subunidades de la bacGlyRS, a pesar de la adición de inhibidores de proteasas y exhaustiva limpieza de las columnas cromatográficas con 0.5 M de NaOH; por lo que se siguieron dos estrategias para tratar de resolver esta degradación, ya que se consideraba que la fusión de las subunidades facilitaría la obtención de la estructura de la bacGlyRS completa; una de las estrategias consistió en mutar el conector peptídico entre las subunidades por alguna secuencia peptídica que incrementara la estabilidad frente a las proteasas y la otra estrategia fue secuenciar por degradación de Edman el extremo amino terminal de los fragmentos resultado de la degradación realizado por la compañía Alphasize.

Por otra parte, otro problema encontrado fue la presencia de una especie oligomérica de más alto peso molecular, que producía por tanto polidispersión de la AtGlyRS, así como agregación en ausencia de DTT. El problema con la agregación podría ser trivial, pero ya que

los agentes reductores son antagonistas del inhibidor de proteasas PMSF, se convirtió en un problema importante a considerar. Esta especie de alto peso molecular podía ser separada por SEC, pero siempre se encontró presente en una muy pequeña proporción. Para resolver el problema de la especie oligomérica de alto peso molecular se mutaron todas las cisteínas de la AtGlyRS, lo cual resolvió el problema de la agregación permitiendo agregar los inhibidores de proteasas.

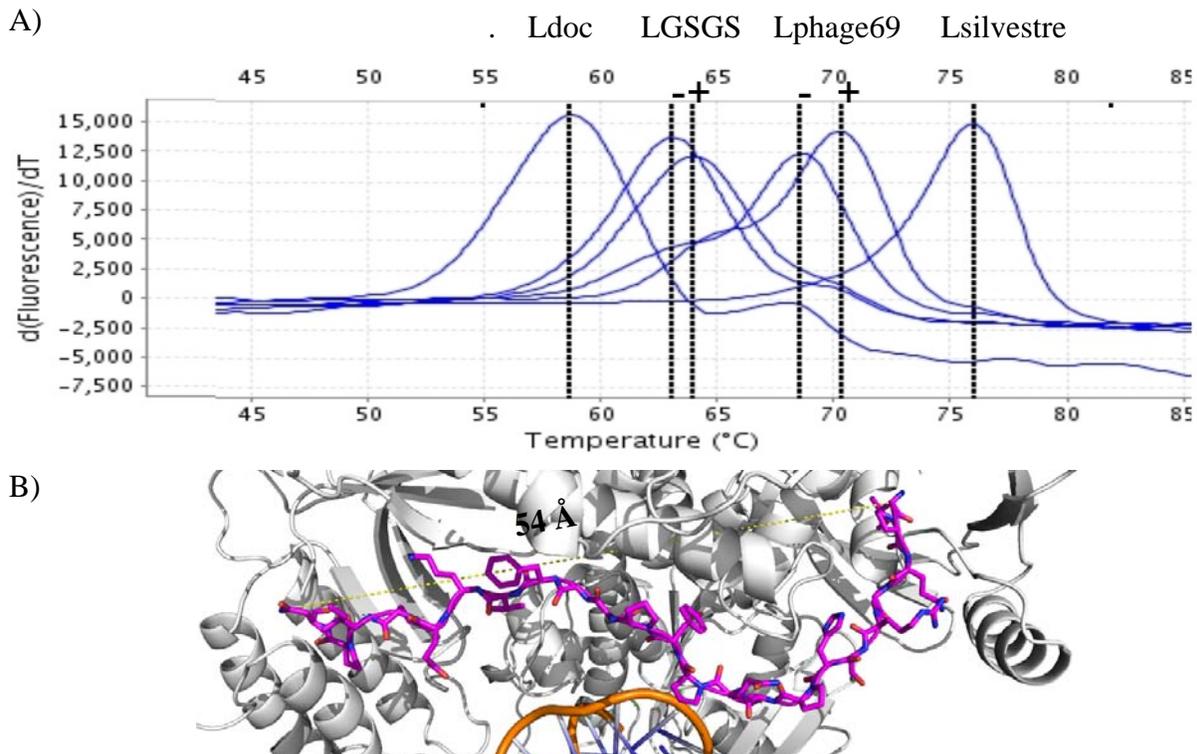
**Obtención de las mutantes de la AtGlyRS:** Para la mutación de los conectores peptídicos entre las subunidades, se diseñaron 5 versiones diferentes de conector peptídico. Uno de ellos de una longitud y estructura secundaria equivalente al conector peptídico silvestre de la AtGlyRS fue diseñado basado en un conector peptídico presente en múltiples cristales de la DNA polimerasa del colifago RB69 (Lphage69). Este conector peptídico resultó atractivo por haber sido previamente cristalizado y porque se asumió que la polimerasa del fago debería presentar resistencia a las proteasas de su hospedero *E. coli*. Las otras mutantes de la AtGlyRS con distintos conectores peptídicos fueron obtenidas por PCR-cambio rápido (quick-change) con la metodología que se ha descrito anteriormente.

<b>Tabla 6. Distintas características de las mutantes del conector peptídico en la GlyRS de <i>Anaerolinea thermophila</i></b>			
Conector peptídico	Secuencia de aminoácidos	Tm (oC)	
silvestre	WLKEGAVPQADTVSLAVSGTVSQPDL	76	
Lphage69	WLPQGRSHPVQPYPGAFVKEPIP	69	
Ldoc	WLPQGRS	59	
LGSGS	GGSGGS	63	
Lmin32	GGSGGS	nd	Delección 290-432
Lmin1	WLPQGRSHPVQGGSGGS	nd	Delección 301-432

La secuencia de las mutantes fue verificada por secuenciación como se ha descrito anteriormente, y fueron expresadas y purificadas de forma similar a como se describió para la versión silvestre. Al analizarse las proteínas purificadas por SDS-PAGE, de forma inesperada las mutantes Lphage69 y LGSGS presentaron la misma degradación proteolítica que la versión silvestre de la AtGlyRS.

La secuenciación del extremo amino terminal de los fragmentos de la degradación de la AtGlyRS silvestre y Lphage69 mostró que efectivamente el corte ocurría entre las regiones equivalentes de las subunidades de la bacGlyRS. El corte proteolítico ocurría en el conector peptídico sin importar su secuencia.

La estabilidad térmica de las mutantes de la AtGlyRS fue determinada mediante el ensayo de termofluorimetría diferencial que se describió anteriormente en presencia o en ausencia del GSAd, y pudo observarse la influencia del conector peptídico entre las subunidades en la estabilidad.



**Figura 19. Estabilidad térmica de las mutantes de la AtGlyRS.** A) Ensayo de desplazamiento térmico de 1  $\mu$ g de las mutantes de la AtGlyRS de la tabla 6. El signo + indica la presencia del GSAd y el signo - la ausencia. B) Conector peptídico de la DNA polimerasa del fago RB69.

**Cristalización de la mutante AtGlyRS-Lphage69 con el tRNA de *A. aeolicus*:** Un integrante de nuestro grupo de investigación (Dra. Annia Rodríguez) realizó pruebas de cristalización de las mutantes de la AtGlyRS en complejo con el tRNA de *A. aeolicus*, y observó que la mutante Lphage69 mejoró la tendencia a la cristalización de la AtGlyRS, sin

Structural Insights into the Polyphyletic Origins of Glycyl tRNA Synthetases

Marco Igor Valencia-Sánchez<sup>1</sup>, Annia Rodríguez-Hernández<sup>1,2</sup>, Ruben Ferreira<sup>3</sup>, Hugo Aníbal Santamaría-Suárez<sup>1</sup>, Marcelino Arciniega<sup>1</sup>, Anne-Catherine Dock-Bregeon<sup>4</sup>, Dino Moras<sup>5</sup>, Brice Beinstainer<sup>5</sup>, Haydyn Mertens<sup>6</sup>, Dmitri Svergun<sup>6</sup>, Luis Brieba de Castro<sup>2</sup>, Morten Grøtli<sup>3</sup>, and Alfredo Torres-Larios<sup>1\*</sup>

From <sup>1</sup>Departamento de Bioquímica y Biología Estructural, Instituto de Fisiología Celular, Universidad Nacional Autónoma de México, Circuito Exterior s/n, Ciudad Universitaria, Apartado postal 70-243, Mexico City, 04510, México. <sup>2</sup>Laboratorio Nacional de Genómica para la Biodiversidad, Centro de Investigación y Estudios Avanzados del Instituto Politécnico Nacional, Irapuato, Guanajuato, 04510, México. <sup>3</sup>Department of Chemistry and Molecular Biology, University of Gothenburg, SE-412 96 Gothenburg, Sweden. <sup>4</sup>Laboratoire de Biologie Intégrative des Milieux Marins, 29680 Roscoff, France CNRS UMR8227. <sup>5</sup>Centre for Integrative Biology, Department of Integrated Structural Biology, Institute of Genetics and of Molecular and Cellular Biology (IGBMC), 1 rue Laurent Fries, Illkirch, France CNRS UMR 7104. <sup>6</sup>European Molecular Biology Laboratory, Hamburg Outstation, c/o DESY, Notkestrasse 85, Hamburg, 22603, Germany.

Running title: *Insights into the Polyphyletic Origins of GlyRSs*

To whom correspondence should be addressed: Alfredo Torres-Larios, PhD; Circuito Exterior s/n, Ciudad Universitaria. Apartado postal 70-243, México D.F. 04510, México. Fax 011-52-55-5622-5630; E-mail: [torres@ifc.unam.mx](mailto:torres@ifc.unam.mx).

**Keywords:** aminoacyl tRNA synthetase, structure-function, crystal structure, substrate specificity, molecular evolution

**ABSTRACT**

Glycyl tRNA synthetase (GlyRS) provides a unique case among class II aminoacyl tRNA synthetases (aaRSs), with two clearly widespread types of enzymes: a dimeric ( $\alpha_2$ ) species present in some bacteria, archaea, and eukaryotes, and a heterotetrameric form ( $\alpha_2\beta_2$ ) present in most bacteria. Although the differences between both types of GlyRS at the anticodon binding domain level are evident, the extent and implications of the variations in the catalytic domain have not been described and it is unclear whether the mechanism of amino acid recognition is also dissimilar. Here we show that the  $\alpha$ -subunit of the  $\alpha_2\beta_2$  GlyRS from the bacterium *Aquifex aeolicus* is able to perform the first step of the aminoacylation reaction, which involves the activation of the amino

acid with ATP. The crystal structure of the  $\alpha$ -subunit in the complex with an analog of glycyl adenylate at 2.8 Å resolution presents a conformational arrangement that properly positions the cognate amino acid. This work shows that glycine is recognized by a subset of different residues in the two types of GlyRS. A structural and sequence analysis of class II catalytic domains shows that bacterial GlyRS is closely related to AlaRS, which led us to define a new subclassification of these ancient enzymes and to propose an evolutionary path of  $\alpha_2\beta_2$  GlyRS, convergent with  $\alpha_2$  GlyRS and divergent from AlaRS, thus providing a possible explanation for the puzzling existence of two proteins sharing the same fold and function, but not a common ancestor.

## 11. BIBLIOGRAFÍA

- Armougom, F., Moretti, S., Poirot, O., Audic, S., Dumas, P., Schaeli, B., Keduas, V., Notredame, C. (2006) Espresso: automatic incorporation of structural information in multiple sequence alignments using 3D-Coffee. *Nucleic Acids Res* 34, W604-608
- Arnez, J. G., Dock-Bregeon, A. C., Moras, D. (1999) Glycyl-tRNA synthetase uses a negatively charged pit for specific recognition and activation of glycine. *J Mol Biol* 286: 1449-1459
- Battye, T. G., Kontogiannis, L., Johnson, O., Powell, H. R., Leslie, A. G. (2011) iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM. *Acta Crystallogr D Biol Crystallogr* 67: 271-281.
- Berg, P., Bergmann, F. H., Ofengand, E. J., Dieckmann, M. (1961) Enzymic Synthesis of Amino Acyl Derivatives of Ribonucleic Acid. *J Biol Chem* 236: 343-356.
- Cavarelli J y Moras D. (1993) Recognition of tRNAs by aminoacyl-tRNA synthetases. *FASEB J*. 7(1):79-86.
- Chothia, C. y Lesk. A. M. (1986). The relation between the divergence of sequence and structure in proteins. *EMBO J*. 5:823–826.
- Cusack S, Leberman R, Hartlein M (1991). Sequence, structural and evolutionary relationships between class 2 aminoacyl-tRNA synthetases. *Nucleic Acids Res*. 19 (13): 3489–3498.
- Deckert, G., Warren, P. V., Gaasterland, T., Young, W. G., Lenox, A. L., Graham, D. E., Huber, R. (1998). The complete genome of the hyperthermophilic bacterium *Aquifex aeolicus*. *Nature*, 392(6674), 353-358.
- Duchêne, A. M., Peeters, N., Dietrich, A., Cosset, A., Small, I. D., y Wintz, H. (2001). Overlapping destinations for two dual targeted glycyl-tRNA synthetases in *Arabidopsis thaliana* and *Phaseolus vulgaris*. *J Biol Chem*. 276(18):15275-15283.
- Emsley, P., y Cowtan, K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* 60: 2126-2132.
- Eriani, G., Delarue, M., Poch, O., Gangloff, J., Moras, D. (1990) Partition of tRNA synthetases into two classes based on mutually exclusive sets of sequence motifs. *Nature* 347:

- Evans, P. (2011) An introduction to data reduction: space-group determination, scaling and intensity statistics. *Acta Crystallogr D Biol Crystallogr* 67, 282-292
- Folk, W. y Berg, P. (1970). Characterization of Altered Forms of Glycyl Transfer Ribonucleic Acid Synthetase and the Effects of Such Alterations on Aminoacyl Transfer Ribonucleic Acid Synthesis In Vivo. *J Bacteriol.* 102(1): 204–212.
- Franke, D., y Svergun, D. I. (2009) DAMMIF, a program for rapid ab-initio shape determination in small-angle scattering. *J Appl Crystallogr* 42: 342-346
- Freist, W., Logan, D. T., Gauss, D. H. (1996) Glycyl-tRNA synthetase. *Biol Chem H-S*, 377(6), 343-356.
- Gabaldon, T., y Koonin, E. V. (2013) Functional and evolutionary implications of gene orthology. *Nat Rev Genet*, 14(5): 360-366.
- Grant, B. J., Rodrigues, A. P. C., ElSawy, K. M., McCammon, J. A., Caves, L. S. D. (2006) Bio3d: an R package for the comparative analysis of protein structures. *Bioinformatics* 22: 2695-2696
- Gruic-Sovulj, I., Uter, N., Bullock, T., Perona, J. J. (2005) tRNA-dependent aminoacyl-adenylate hydrolysis by a nonediting class I aminoacyl-tRNA synthetase. *J Biol Chem* 280: 23978-23986.
- Guo, M., Chong, Y. E., Shapiro, R., Beebe, K., Yang, X. L., Schimmel, P. (2009) Paradox of mistranslation of serine for alanine caused by AlaRS recognition dilemma. *Nature* 462: 808-812
- Han, M. V., y Zmasek, C. M. (2009) phyloXML: XML for evolutionary biology and comparative genomics. *Bmc Bioinformatics* 10: 356.
- Hauenstein, S. I., Hou, Y. M., y Perona, J. J. (2008). The homotetrameric phosphoseryl-tRNA synthetase from *Methanosarcina mazei* exhibits half-of-the-sites activity. *J Biol Chem*, 283(32): 21997-22006.
- Holm, L., y Sander, C. (1995). Dali: a network tool for protein structure comparison. *Trends in Biochem Sci* 20(11): 478-480.
- Joosten, R. P., Long, F., Murshudov, G. N., Perrakis, A. (2014) The PDB\_REDOserver for macromolecular structure model optimization. *IUCrJ* 1: 2.
- Kleiger, G., Beamer, L. J., Grothe, R., Mallick, P., y Eisenberg, D. (2000). The 1.7 Å crystal

- structure of BPI: a study of how two dissimilar amino acid sequences can adopt the same fold. *J Mol Biol* 299(4): 1019-1034.
- Konarev, P. V., Volkov, V. V., Sokolova, A. V., Koch, M. H. J., Svergun, D. I. (2003) PRIMUS: a Windows PC-based system for small-angle scattering data analysis. *J Appl Crystallogr* 36(5): 1277-1282.
- Kozin, M. B., y Svergun, D. I. (2001) Automated matching of high- and low-resolution structural models. *J Appl Crystallogr* 34: 33-41.
- Leatherbarrow, *et al.* (1985). Transition-state stabilization in the mechanism of tyrosyl-tRNA synthetase revealed by protein engineering. *Proc Natl Acad Sci U S A.* 82(23): 7840–7844.
- Ledoux, S., y Uhlenbeck, O. C. (2008) [3'-32P]-labeling tRNA with nucleotidyltransferase for assaying aminoacylation and peptide bond formation. *Methods* 44: 74-80
- Logan, D. T., Mazauric, M. H., Kern, D., Moras, D. (1995) Crystal-Structure of Glycyl-Transfer-Rna Synthetase from *Thermus-Thermophilus*. *Embo Journal* 14: 4156- 4167
- Long, F., Vagin, A. A., Young, P., Murshudov, G. N. (2008) BALBES: a molecular-replacement pipeline. *Acta Crystallogr D Biol Crystallogr* 64: 125-132.
- Mazauric MH, *et al.* (1998). Glycyl-tRNA synthetase from *Thermus thermophilus*--wide structural divergence with other prokaryotic glycyl-tRNA synthetases and functional inter-relation with prokaryotic and eukaryotic glycylation systems. *Eur J Biochem.* 251(3):744-57.
- Mazauric, M. H., Reinbolt, J., Lorber, B., Ebel, C., Keith, G., Giege, R., Kern, D. (1996) An example of non-conservation of oligomeric structure in prokaryotic aminoacyl-tRNA synthetases. Biochemical and structural properties of glycyl-tRNA synthetase from *Thermus thermophilus*. *Eur J Biochem* 241: 814-826
- McDonald, T., Breite, L., Pangburn, K. L., Hom, S., Manser, J., Nagel, G. M. (1980). Overproduction, purification, and subunit structure of *Escherichia coli* glycyl transfer ribonucleic acid synthetase. *Biochemistry.* 19(7): 1402-9.
- Moras, D. (1992). Structural and functional relationships between aminoacyl-tRNA synthetases. *Trends in biochemical sciences,* 17(4): 159-164.
- Nagel, G. M., Cumberledge, S., Johnson, M. S., Petrella, E., Weber, B. H. (1984) The  $\beta$  subunit of *E. coli* glycyl-tRNA synthetase plays a major role in tRNA recognition. *Nucleic*

Acids Res 12, 4377-4384.

- Nagel, G. M., y Doolittle, R. F. (1991). Evolution and relatedness in two aminoacyl-tRNA synthetase families. *Proceedings of the National Academy of Sciences*, 88(18): 8121-8125.
- Nelson y Cox, (2005). *Principios de bioquímica Lehninger*. Edit. Omega 4ª ed., Barcelona. 1264 pp.
- Niesen, F. H., Berglund, H., Vedadi, M. (2007) The use of differential scanningfluorimetry to detect ligand interactions that promote protein stability. *Nat Protoc* 2: 2212-2221.
- Norris AT y Berg P. (1964) Mechanism of aminoacyl RNA synthesis: Studies with isolated aminoacyl adenylate complexes of isoleucyl RNA synthetase. *Proc Natl Acad Sci USA*. 52: 330-7.
- O'Donoghue, P., y Luthey-Schulten, Z. (2003) On the evolution of structure in aminoacyl-tRNA synthetases. *Microbiol Mol Biol Rev* 67: 550-573.
- Okamoto, K., Kuno, A., Hasegawa, T. (2005) Recognition sites of glycine tRNA for glycylyl-tRNA synthetase from hyperthermophilic archaeon, *Aeropyrum pernix* K1. *Nucl Acid S* 49: 299-300.
- Ostrem, D. L., y Berg, P. (1970) Glycyl-Transfer Rna Synthetase - an Oligomeric Protein Containing Dissimilar Subunits. *Proc Natl Acad Sci U S A*. 67(4):1967-74.
- Ostrem DL y Berg P. (1974) Glycyl transfer ribonucleic acid synthetase from *Escherichia coli*: purification, properties, and substrate binding. *Biochemistry* 13(7): 1338-1348.
- Pevsner, J. (2015). *Bioinformatics and functional genomics*. Edit. John Wiley & Sons. 3a ed. Hoboken, N.J. 753pp.
- Perona, J. J., y Hadd, A. (2012). Structural diversity and protein engineering of the aminoacyl-tRNA synthetases. *Biochemistry*, 51(44): 8705-8729.
- Petoukhov, M. V., Franke, D., Shkumatov, A. V., Tria, G., Kikhney, A. G., Gajda, M., Gorba, C., Mertens, H. D. T., Konarev, P. V., Svergun, D. I. (2012) New developments in the ATSAS program package for small-angle scattering data analysis. *J Appl Crystallogr* 45: 342-350.
- Potterton E., Briggs P., Turkenburg M., Dodson E. (2003). A graphical user interface to the CCP4 program suite *Acta Cryst*. D59: 1131-1137.
- Redwan, I. N., Ingemyr, H. J., Ljungdahl, T., Lawson, C. P., Grötli, M. (2012) Solid- Phase

- Synthesis of 5'-O-[N-(Acyl)sulfamoyl]adenosine Derivatives. *Eur J Org Chem*: 3665-3669.
- Rodriguez-Hernandez, A., Bhaskaran, H., Hadd, A., Perona, J. J. (2010) Synthesis of Glu-tRNA(Gln) by engineered and natural aminoacyl-tRNA synthetases. *Biochemistry* 49: 6727-6736.
- Russell, R. B. y Barton, G. J. (1992) Multiple protein sequence alignment from tertiary structure comparison: assignment of global and residue confidence levels. *Proteins* 14: 309-323
- Shiba, K., Schimmel, P., Motegi, H., Noda, T. (1994) Human glycyl-tRNA synthetase. Wide divergence of primary structure from bacterial counterpart and species-specific aminoacylation. *J Biol Chem* 269: 30049-30055
- Smith, T. F., y Hartman, H. (2015) The evolution of Class II Aminoacyl-tRNA synthetases and the first code. *FEBS Lett* 589: 3499-3507
- Svergun, D. I. (1992) Determination of the Regularization Parameter in Indirect-Transform Methods Using Perceptual Criteria. *J Appl Crystallogr* 25(4): 495-503.
- Svergun, D., Barberato, C., y Koch, M. H. J. (1995) CRY SOL - A program to evaluate x-ray solution scattering of biological macromolecules from atomic coordinates. *J Appl Crystallogr* 28: 768-773.
- Swairjo, M. A., Otero, F. J., Yang, X. L., Lovato, M. A., Skene, R. J., McRee, D. E., Ribas de Pouplana, L., Schimmel, P. (2004) Alanyl-tRNA synthetase crystal structure and design for acceptor-stem recognition. *Mol Cell* 13: 829-841.
- Tan, K., Zhou, M., Zhang, R., Anderson, W. F., Joachimiak, A. (2012) The crystal structures of the  $\alpha$ -subunit of the  $\alpha(2)\beta(2)$  tetrameric Glycyl-tRNA synthetase. *J Struct Funct Genomics* 13: 233-239.
- Tang y Huang. (2005) Evolution of different oligomeric glycyl-tRNA synthetases. *FEBS letters*, 579(6): 1441-144.
- Torres-larios *et al.*, (2002). Tesis doctoral. Universidad Louis Pasteur de Strasburgo.
- Toth MJ y Schimmel P. (1986) Internal structural features of E. coli glycyl-tRNA synthetase examined by subunit polypeptide chain fusions. *J Biol Chem*. 261(15):6643-6.
- Toth, M. J., y Schimmel, P. (1990) A Mutation in the Small ( $\alpha$ ) Subunit of Glycyl-Transfer Rna-Synthetase Affects Amino-Acid Activation and Subunit Association Parameters. *J Biol*

Chem 265: 1005-1009.

- Toth, M. J., y Schimmel, P. (1990) Deletions in the Large ( $\beta$ ) Subunit of a Hetero- Oligomeric Aminoacyl-Transfer Rna-Synthetase. *J Biol Chem* 265: 1000-1004.
- Tsui, W. C., y Fersht, A. R. (1981). Probing the principles of amino acid selection using the alanyl-tRNA synthetase from *Escherichia coli*. *Nuc Acids Res* 9(18): 4627-4637.
- Valencia-Sánchez, M. I., Rodríguez-Hernández, A., Ferreira, R., Santamaria-Suárez, H. A., Arciniega, M., Dock-Bregeon, A. C., Moras D., Beinsteiner B., Mertens H., Svergun D., Brieba L.G., Grøtli M., Torres-Larios A. (2016). Structural Insights into the Polyphyletic Origins of Glycyl tRNA Synthetases. *J Biol Chem* 291(28):14430-46.
- Woese, C. R., Olsen, G. J., Ibba, M., Soll, D. (2000) Aminoacyl-tRNA synthetases, the genetic code, and the evolutionary process. *Microbiol Mol Biol Rev* 64: 202-236.
- Zmasek, C. M., y Eddy, S. R. (2001). ATV: display and manipulation of annotated phylogenetic trees. *Bioinformatics* 17(4): 383-384.