



UNAM

*Licenciatura en Investigación
Biomédica Básica*

Instituto de Investigaciones Biomédicas.

Instituto de Fisiología Celular.

Facultad de Medicina.

*PROBABLE PROTEOMA ANCESTRAL DEL
MUNDO DE LAS RIBONUCLEOPROTEÍNAS*

T E S I S

que, para obtener el grado de

Licenciada en Investigación Biomédica Básica,

presenta:

Miryam Palacios Pérez

Director de Tesis

Dr. Marco Antonio José Valenzuela



México Distrito Federal.
Ciudad Universitaria.
20 de Noviembre del 2014.



Universidad Nacional
Autónoma de México



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

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

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



UNAM



Licenciatura en Investigación Biomédica Básica



Instituto de Investigaciones Biomédicas.



Instituto de
Fisiología
Celular

Facultad de Medicina



**PROBABLE PROTEOMA ANCESTRAL
DEL MUNDO DE LAS RIBONUCLEOPROTEÍNAS**

TESIS

AGRADECIMIENTOS ACADÉMICOS

- Dr. Marco Antonio José Valenzuela por la orientación recibida a lo largo de toda la elaboración de la presente tesis, en lo referente a la provisión y discusión de muchas de las referencias, aunque principalmente por la continua y aguda discusión acerca de los resultados obtenidos y en lo referente a las perspectivas del proyecto.
- Doctores: Arturo Becerra, Marco A. José, Carmen Gómez, Pedro Miramontes e Imelda López, cuyas agudas e importantes observaciones me permitieron mejorar la presentación de los resultados de esta tesis.
- L.I. Juan Román Bobadilla por la oportuna asistencia técnica y el apoyo en infraestructura computacional.
- Ing. Tzipe Govezensky, por ayudarme a profundizar en los antecedentes del presente trabajo y por su apoyo en la adecuada presentación de los resultados.
- A todos los integrantes del Grupo de Biología Teórica por las agudas observaciones y críticas en las diferentes fases del proyecto.

AGRADECIMIENTOS INSTITUCIONALES.

- Grupo de Biología Teórica en el Departamento Inmunología del Instituto de Investigaciones Biomédicas, Universidad Nacional Autónoma de México.



Porque cada uno los elementos antes citados me permitieron el satisfactorio desarrollo y conclusión de la presente tesis.

Gracias.

A mi mamá, quien desde muy temprana edad me dio a conocer que nada ocurre por accidente, nada en la naturaleza se desperdicia y siempre hay que rescatar las cosas buenas: “En la Vida hay que Ser Positiva”, me ha venido recordando desde que tengo memoria (y tal vez desde antes): LucyPZ.

A mi papá, quien estando a mi lado me ha ayudado en lo que ha podido, y quien me ha apoyado aún sin estar conmigo: IPC.

Al amor de mi Vida, quien me ha hecho ser mejor persona cada vez, él también es mi familia, mi mejor amigo, amable y amoroso: JMChS.

A toda mi familia (pequeña, mediana y grande) por su enormísimo apoyo y comprensión.

A mis dos abuelitas por la gran sabiduría que me han transmitido mediante sus consejos, siempre tan atinados y ciertos.

A mis primas, primos y sobrinos por el mutuo apoyo.

A mis tías y tíos por las críticas objetivas y justas peticiones.

A AyE, algo antagonista a mí, pero tal vez (de algún modo) complementaria.

A los mejores amigos que muchas personas necesitan: Cintia PG, Felipe GC, Elizabeth QV, Emma OT, Andrea AG.

A Marco, por ser una extraordinaria persona, un extraordinario jefe y un gran amigo, además de un paradigma excelso para mí.

A Tzipe, por ser amable y escuchar, por su templanza y corazón.

A Víctor, por ser mi cuate y dejarme ser su “cuate”.

A ‘Chús’, mi hermanito académico medio perdido por el momento.

A Luz, por ser un reflejo.

A todo el Grupo de Biología Teórica, por el apoyo y comprensión, por los consejos académicos y personales; por atender mis ideas y corregirme cuando erré; por transmitirme su ánimo, ganas y conocimiento (Nabor, Elisa, Homero, Emilio, Gabo, Alejandro y Noemí): “¡Qué bueno es sentirse en casa!” he expresado en varios momentos desde que llegué...

A Chente por ser tan generoso, apoyarme tanto, confiar siempre en mí y sobre todo brindándome seguridad en ciertos momentos. A los cometines (Pau, Chervis, Vale, Nahum, Pablo y Martincillo), a Emilio Rojas y a MariCarmen.

A Irmix (jefa muy exigente, quien me enseñó a ser metódica y transparente), a Kakker (quien se dio cuenta de mi gran curiosidad y con la paciencia suficiente para escuchar y explicarme), a Adrián † (¡el médico más biomédico, “eres” increíble!, tan único) y a Rodolfo (quien me hizo ver cómo es ser “gran-de-mente” y un biomédico genial).

A todos los profesores que he tenido en la carrera, especialmente GGO, AZD, CarGom, LRZ, ILó, LE y VFELMGSS, a mis compañeros de carrera como Danny, AuroCast y Pittol.

A la biblioteca del IIB; Mari, MCariño, Julieta y Lic.LucyBO.

A l@s integrantes de aquel grupo, ustedes saben quiénes son.

Al mejor guía político en el México actual, su legado y su muy amadísima familia.

A las personas que me hicieron madurar, o me ayudaron a hacerlo, en un sentido u otro.

A nuestra muy querida Universidad Nacional Autónoma de México, orgullo histórico de nuestro suelo mexicano y Patrimonio de la Humanidad.

A la Licenciatura en Investigación Biomédica Básica, donde encontré un lugar...

**A La Ciencia,
Al Amor.**

In nomini Amor et

In nomini Scientia ☺.

PROBABLE PROTEOMA ANCESTRAL DEL MUNDO DE LAS RIBONUCLEOPROTEÍNAS

ÍNDICE GENERAL

- I. Presentación. --- Pág. 9.
 - 1.1 Conceptos clave. --- Pág. 11.

- II. Marco teórico (introducción). --- Pág. 12.
 - 2.1 Mundos ancestrales. --- Pág. 13.
 - 2.2 LUCA. --- Pág. 17.
 - 2.3 Código genético y aminoácidos ancestrales. --- Pág. 19.
 - 2.3.1 Código RNY. --- Pág. 20.
 - 2.3.2 Hipótesis diferentes a la teoría RNY. --- Pág. 22.
 - 2.4 Organismo elegido. --- Pág. 23.

- III. Pregunta de investigación e Hipótesis. --- Pág. 24.

- IV. Objetivos. --- Pág. 25.
 - Objetivo general.
 - Objetivos particulares.

- V. Metodología. --- Pág. 26.
 - 5.1 Pasos previos. --- Pág. 26.
 - 5.2 Metodología propia. --- Pág. 27.

- VI. Resultados y Análisis. --- Pág. 28.
 - 6.1 Nivel de corte. --- Pág. 29.
 - 6.2 Clasificaciones. --- Pág. 31.

6.3. Análisis. --- Pág. 32.

6.3.1 Sentido biológico, prueba de hipótesis. --- Pág. 36.

6.3.1.1 RNA, procesamiento y modificaciones. --- Pág. 37.

6.3.1.2 DNA, procesamiento y modificaciones. --- Pág. 68.

6.3.1.3 Metabolismo. --- Pág. 82.

Energético. --- Pág. 85.

Carbohidratos. --- Pág. 103.

Aminoácidos. --- Pág. 118.

Nucleótidos. --- Pág. 132.

Lípidos. --- Pág. 141.

Coenzimas. --- Pág. 147.

Iones inorgánicos. --- Pág. 152.

No incluidas en COGs. --- Pág. 155.

6.3.1.4 Proteínas, procesamiento y modificaciones. ---
Pág. 166.

6.3.1.5 Estructuras membranales y extramembranales. --
- Pág. 190.

6.3.1.6 Transportadores y Canales. --- Pág. 224.

6.3.1.7 Transducción. --- Pág. 247.

6.3.1.8 Resistencia y supervivencia. --- Pág. 259.

6.3.1.9 Proteínas hipotéticas y/o conservadas. --- Pág.
269.

6.3.2 Genoma aleatorio, control. --- Pág. 277.

VII. Discusión. --- Pág. 278.

Contribución fundamental de la tesis. --- Pág. 289.

VIII. Conclusiones y Prospectivas. --- Págs. 294 y 295.

IX. Apéndice. --- Pág. 296.

X. Referencias bibliohemerográficas. --- Pág. 297.

ÍNDICE DE FIGURAS

- F1. Cronología consenso desde eventos prebióticos hasta LUCA. Pág. 9.
- F2. Una ribozima habría catalizado la síntesis de péptidos a partir de aa que existían en el medio y paulatinamente habría comenzado la codificación de esos primeros aa en el RNA. Pág. 15.
- F3. Escenario evolutivo propuesto por Forterre, en que se indican 3 edades con base en las moléculas bióticas fundamentales. Pág. 16.
- F4. Pasos detallados en el origen y evolución de la vida, desde la formación misma de la Tierra hasta el surgimiento de LUCA. Pág. 17.
- F5. Esquema que ilustra como el proceso ramificado de la evolución de las especies se ha visto intervenido constantemente por la transferencia horizontal de genes (HGT). Pág. 18.
- F6. Comparación de las frecuencias relativas, con respecto al valor-E (en Log_{10}), del genoma con sentido biológico y del aleatorio. Pág. 30.
- F7. Comparación de las frecuencias relativas, con respecto al **S**, del genoma con sentido biológico y del aleatorio. Pág. 30.
- F8. Reacciones químicas que muestran la síntesis de moléculas biológicas a partir de compuestos abióticos. Pág. 279.
- F9. Esquema que muestra la estructura general de un fosfolípido (A), la estructura de una bicapa lipídica de fosfolípidos (B) y la estructura de un liposoma con bicapa lipídica (C). Pág. 280.
- F10: Circularización de una bicapa lipídica, esquemas de la vista sagital del proceso. Pág. 281.
- F11. Primeros pasos de la evolución biológica. Pág. 282.
- F12. Esquema de la replicación del RNA. Pág. 283.
- F13. Representación de los distintos tipos de proteínas, donde el **'sitio de unión y estabilización del cofactor'** (CSBS) –el primordio, codificado por tripletes RNY–, se indica por una región más clara en cada figura, *i.e.* bindoma ancestral. Pág. 292.
- Apéndice. Programa que llama a BLAST. Pág. 295.

ÍNDICE DE TABLAS

T1. Proteínas de **SagA909**, con regiones codificadas por tripletes RNY y que se ajustan al nivel de corte establecido. **Págs.** 33, 34 y 35.

T2. Proteínas del genoma aleatorio de **SagA909**, con regiones codificadas por tripletes RNY y que se ajustan al nivel de corte establecido. **Pág.** 35.

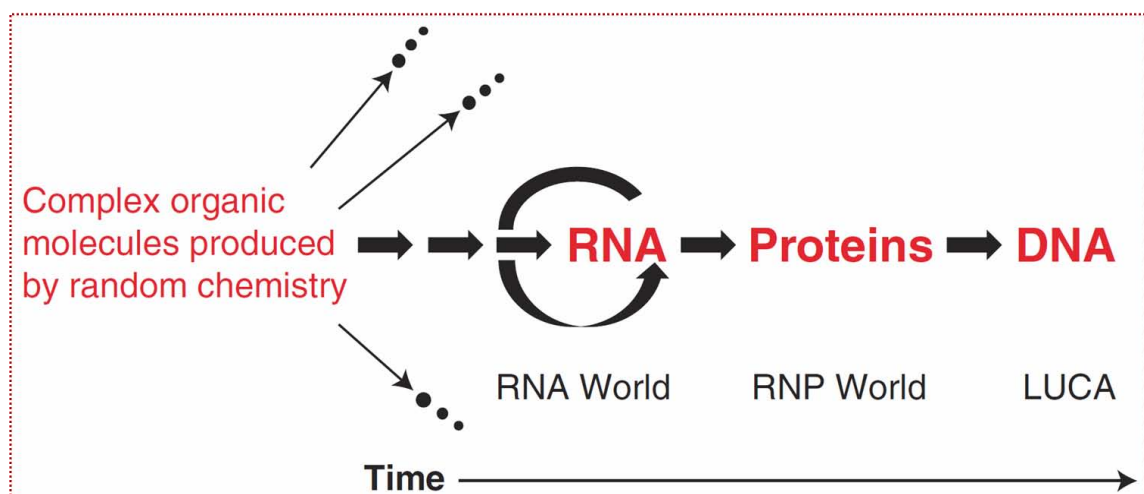
T3. Tabla que muestra los resultados para la prueba de hipótesis, se indica el tipo de proteínas diferentes con su código de colores correspondiente; así como, entre paréntesis, el número de resultados en cada conjunto. **Pág.** 36.

A continuación se enlistan las tablas de las proteínas por cada categoría **indicada en la sección '6.3 Análisis'**. **Se indica la categoría a la que pertenecen.** En primer lugar se menciona la tabla obtenida a partir de la general (tablas pares) y a continuación la tabla que resume los resultados encontrados (tablas impares).

- Proteínas del procesamiento de RNA. **T4. Pág.** 38. **T5. Pág.** 67.
- Proteínas del procesamiento de DNA. **T6. Pág.** 69. **T7. Pág.** 81.
- Proteínas del metabolismo. **T8. Pág.** 85. **T9. Pág.** 165.
- Proteínas que procesan o modifican proteínas **T10. Pág.** 166. **T11. Pág.** 189.
- Proteínas involucradas en la formación de estructuras membranales y extramembranales **T12. Pág.** 191. **T13. Pág.** 223.
- Transportadores y canales. **T14. Pág.** 227. **T15. Pág.** 246.
- Proteínas involucradas en transducción. **T16. Pág.** 248. **T17. Pág.** 258.
- Proteínas involucradas en resistencia o supervivencia de **SagA909**. **T18. Pág.** 259. **T19. Pág.** 268.
- Proteínas o dominios hipotéticos(as) y/o conservados(as). **T20. Pág.** 270. **T21. Pág.** 276.
- Resultados del genoma aleatorio. **T22. Pág.** 277. **T23. Pág.** 277.

I. PRESENTACIÓN

Varios eventos evolutivos ocurrieron antes que surgiese el Último Ancestro Universal Común (LUCA, por sus siglas en inglés). Una de las cronologías más aceptadas, con base en las moléculas hereditarias de cada etapa evolutivaⁱ, es la descrita por la figura (F1):



F1. Cronología consenso desde eventos prebióticos hasta LUCA. De CECH T (2011), The RNA Worlds in Context, Cold Spring Harb. Perspect. Biol. doi: 10.1101/cshperspect.a006742.

El presente trabajo es el primero hasta donde sabemos, en el que se muestra una aproximación formal hacia aquellas proteínas ancestrales que han preservado su codificación primordial hasta nuestros días, proteínas que existían en una población de organismos antes que LUCA siquiera emergiese.

La presente tesis está dividida de manera *semejante* a un artículo de investigación científica. Presento en primer lugar el marco teórico donde se indican los conceptos a los que aduce la presente tesis. En particular describo las propuestas que existen con relación al Mundo de RNA y al Mundo de las Ribonucleoproteínas, que son los periodos evolutivos en los cuales están

ⁱ DNA y ADN se utilizarán de indistintamente a lo largo del texto, lo mismo para RNA y ARN.

enmarcados los resultados de la presente investigación. Dado que es importante que los organismos ancestrales hayan tenido un sistema para la transmisión de caracteres hereditarios, es importante hablar del código genético primitivo, con referencia particular a los tripletes más ancestrales cuya estructura es probablemente del tipo RNY; se hace además una descripción del tipo de aminoácidos que son codificados por este tipo de tripletes. Posteriormente esbozo un panorama general de lo que se ha definido formalmente como LUCA. Finalmente despliego una sucinta explicación acerca de la biología del organismo *Streptococcus agalactiae* A909, que fue el procarionte elegido para el desarrollo del presente trabajo.

A continuación doy a conocer explícitamente los objetivos del presente trabajo como tercera sección y tomando en cuenta todo lo anterior, planteo la pregunta de investigación e hipótesis en la cuarta parte.

Como quinto título apunto la metodología enfocada al tratamiento del problema, describiendo *grosso modo* la secuencia de pasos que otros miembros del Grupo de Biología Teórica han seguido para extraer del genoma de un organismo únicamente los tripletes RNY y el genoma aleatorio correspondiente; describo la sistemática utilizada para obtener todas las proteínas codificadas por estos tripletes (RNY), así como las proteínas del control aleatorio.

Como parte del presente trabajo se hizo una comparación numérica de los parámetros obtenidos para las proteínas de la prueba de hipótesis y de su control con el fin de establecer un nivel de corte para los resultados; describo el procedimiento que he seguido para clasificar todas las proteínas; finalmente menciono sus características generales, así como las hipótesis que puedan existir con respecto a su origen y evolución.

El capítulo sexto de resultados y análisis se divide en tres partes. Primeramente se exponen los resultados del análisis numérico realizado para

establecer los valores de corte. En segundo lugar proporciono las listas completas de las proteínas que se obtuvieron tanto para la prueba de hipótesis como para el control, tomando en cuenta sólo aquellas que se adecúan al nivel de corte establecido. Como última sección de este capítulo proporciono la lista general de las proteínas encontradas; éstas las agrupo de acuerdo con la función celular que actualmente desempeñan y las describo *grosso modo*, añadiendo en cada caso las explicaciones más plausibles acerca de su existencia ancestral con base en la literatura existente.

En el capítulo siete de discusión aporto una visión global del proteoma analizado, que correspondería al de un organismo que probablemente existió antes del famoso LUCA.

Proporciono las conclusiones del trabajo como capítulo número ocho, así como las perspectivas del mismo con el fin de hacer una clara contribución a los modelos que existen sobre la evolución del código genético y su fenotipo.

El apéndice del apartado nueve corresponde a un programa utilizado para obtener los datos a partir de los cuales se desarrolló la presente tesis, así como un breve glosario de los términos más relevantes en la presente tesis.

Están citadas las referencias bibliohemerográficas, que revisé durante la elaboración de la presente TESIS, como décimo punto.

*** Cada uno de los apartados y subapartados pueden comprenderse independientemente del resto, empero es un trabajo unificado.**

1.1 CONCEPTOS CLAVE:

Mundo de las Ribonucleoproteínas, codones RNY, bioquímica y fisiología celular, origen y evolución de proteínas, organismos previos a LUCA.

II. MARCO TEÓRICO

El organismo denominado LUCA se componía de DNA y varias proteínas esenciales; previo a LUCA debió existir una población de organismos o progenotes de mediana complejidad cuyo código genético proviene del Mundo del RNA, a partir del cual evolucionó el código genético estándar.

El proteoma correspondiente al código genético ancestral muy probablemente está conservado; es posible descubrir entonces que las proteínas de aquel sistema forman parte vital de la compleja organización de las formas de vida actuales y por tanto no resulta sorprendente encontrar vestigios de su presencia en los genomas de organismos actuales.

Es generalmente útil comenzar cronológicamente desde lo más grande hacia lo más pequeño y, comenzaremos entonces desde el *Big Bang* para dar un marco histórico de referencia. El proyecto WMAP de la NASA ha estimado la edad del Universo en $(13,7 \pm 0.13) \times 10^9$ años (~13,700 millones de años); la edad del nuestro Sol alrededor de 5,000 millones de años y la edad de la Tierra en algunos millones menos (poco más de 4,500 millones de años)¹.

Con respecto al origen de la vida, las diferentes teorías que existen coinciden con la existencia de un caldo primigenio o al menos un medio acuoso con elementos prebióticos para la formación de algún tipo de proto-organismos (3.7 mil millones de años, también llamado el origen Hadeano). La primera propuesta de este tipo fue elaborada por Oparin-Haldane. Posteriormente Miller-Urey demostraron experimentalmente la formación de aminoácidos a partir de los elementos existentes en una atmósfera primitiva reductora, utilizando como fuente de energía descargas eléctricas. Recientemente Lazcano-Miller y colaboradores re-ensayaron dicha hipótesis modificando las condiciones del experimento². Es difícil precisar la

fecha de inicio de la vida en la Tierra puesto que existen controversias sobre la definición de vida ^(vid.3).

2.1 MUNDOS ANCESTRALES.

Existen varias propuestas que tratan de explicar el origen y evolución de cada rasgo biológico, que definiría a los primeros organismos, a partir de compuestos prebióticos y hasta el surgimiento de la vida tal como la concebimos^{4,5,6}.

* Mundo de los lípidos. Los lípidos de las membranas celulares pueden fácilmente sintetizarse a partir de compuestos prebióticos, esos lípidos pueden agruparse en micelas y formar liposomas; dentro de esas estructuras, según se ha mostrado experimentalmente, pueden formarse diversos ciclos de tipo hereditario o metabólico^{7,8}. Las membranas son además un componente importante para que los diferentes compuestos estén confinados con el fin de que haya mayor probabilidad que interactúen y al mismo tiempo permiten un pequeño intercambio con el medio que permitiría que los recursos no se agoten, asimismo los componentes estarían protegidos de agentes destructores. Con base en las características de las moléculas oleicas, Daniel Segré y colaboradores han acuñado el término Mundo de los Lípidos⁹. Eörs Szathmáry ha señalado que la membrana biológica y las redes metabólicas surgieron y evolucionaron de forma paralela¹⁰.

* Mundo del Hierro-Azufre. **Conocido también como "Metabolismo de Superficie", según esta propuesta una** forma primitiva de metabolismo basada en sulfuros de hierro y níquel precedió a la genética; de acuerdo con Wächterhäuser la membrana se generó paulatinamente, mientras algunas moléculas polianiónicas comenzaban a interactuar en ciclos metabólicos y posteriormente hereditarios¹¹.

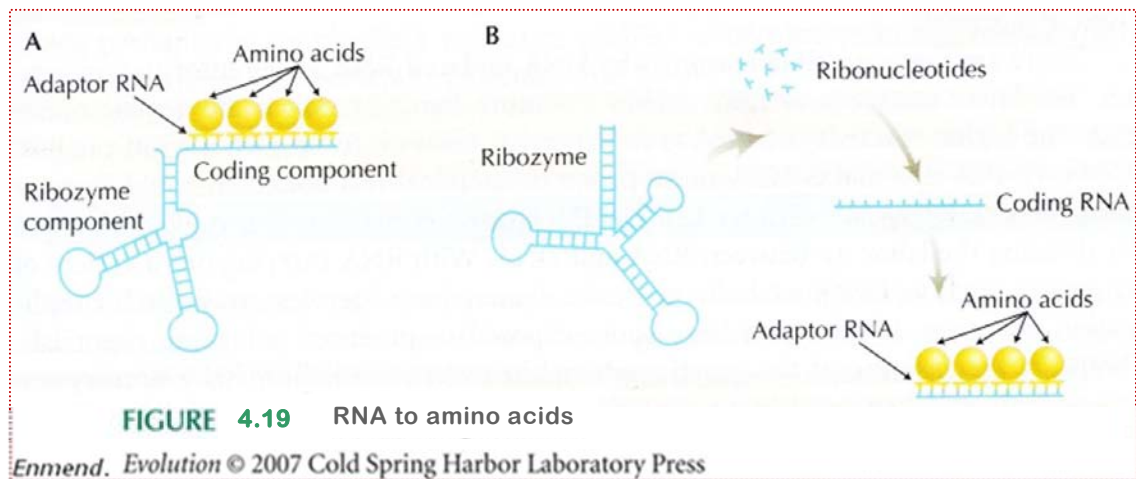
* Mundo de la Coenzima. Hipótesis que propone que las coenzimas fueron las primeras moléculas biológicas^{12,13} de tipo metabólico e incluso hereditario que dieran origen a los organismos¹⁴. El soporte experimental ha tenido como base los aptámeros¹⁵ de RNA-coenzimas^{16,17,18}. Con base en todo lo anteriormente expuesto Alexei Sharov acuñó el término Mundo de la Coenzima y ha refinado el modelo¹⁹.

* Mundo del RNA. La naturaleza multifuncional del RNA es la principal evidencia de que éste pudo haber desempeñado un papel angular en el pasado evolutivo^{20,ii}. La idea inicial se remite a Carl Woese, Francis Crick y Leslie Orgel quienes plantearon, hace alrededor de 50 años de manera independiente, que debió existir una molécula adaptadora que funcionara como genotipo y como fenotipo, más tarde se descubrió que la molécula propuesta debió ser el RNA de transferencia (tRNA)¹⁹; no obstante el planteamiento inicial fue ignorado hasta que Thomas Cech y Sydney Altman descubrieron en 1982 que un intrón del Grupo I de *Tetrahymena thermophila* es autocatalítico²¹ con lo cual se demostró que el RNA posee estas capacidades y se estableció entonces el vocablo *ribozima*; en 1986 Walter Gilbert acuñó el término *Mundo del RNA*²² y desde entonces varias teorías y experimentos se han desarrollado al respecto (*vid.* 23 y refs. citadas).

* Mundo de las Ribonucleoproteínas (RNPs). Intuitivamente es el periodo evolutivo que siguió al Mundo del RNA, cuando los componentes principales de péptidos y de RNA, *i.e.* aminoácidos (aa) y nucleótidos (nt) que habían surgido en el ambiente prebiótico, comenzaron a coevolucionar. Una de las propuestas es que las ribozimas catalizaban la polimerización de oligopéptidos y a su vez los aminoácidos (aa) conferían estabilidad a las estructuras de RNA de manera similar a como lo hacen en la actualidad, de hecho se postula que los aa pudieron haber fungido como

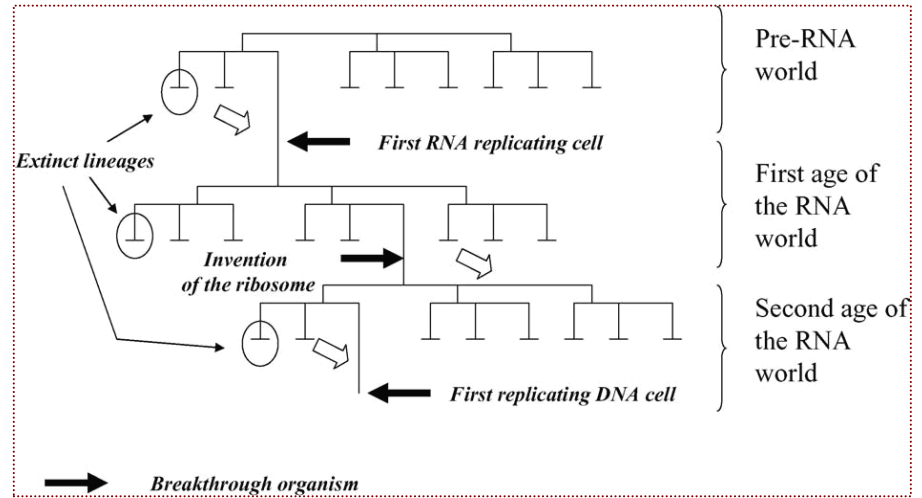
ⁱⁱ Aunque varios experimentos indican que el RNA puede polimerizarse en condiciones prebióticas, éstos habían fallado en mostrar la síntesis *de novo* de sus componentes, incluso varios autores lo señalan como "la pesadilla de la química prebiótica" (Cfr. CECH T (2011), *The RNA Worlds in Context*, Cold Spring Harb. Perspect. Biol. doi: 10.1101/cshperspect.a006742); no obstante, recientemente se ha logrado sintetizar pirimidinas *de novo* a partir de química prebiótica (POWNER MW *et al.* Nature 459, doi:10.1038/nature08013).

cofactores de los RNA, incrementando las actividades catalíticas de estos últimos²⁴; es posible que dicha interacción comenzara entonces con la formación de los peptidil-tRNA, ya que en los organismos actuales es el complejo molecular mediante el cual se lleva a cabo el cambio de lenguaje de nucleótidos a aminoácidos⁴. La interacción inicial mencionada se encuentra esquematizada en la figura (F2):



F2. Una ribozima habría catalizado la síntesis de péptidos a partir de aa que existían en el medio y paulatinamente habría comenzado la codificación de esos primeros aa en el RNA.

De acuerdo con Patrick Forterre la evolución ocurrió en varias etapas separadas por cuellos de botella, como se esquematiza en la figura (F3): la primera corresponde a un ambiente prebiótico (pre-RNA), seguida por el Mundo del RNA (primera edad del Mundo del RNA), que evoluciona hacia un Mundo de RNA+proteína (segunda edad del Mundo del RNA o Mundo de las RNPs), que culmina entonces en un mundo basado en DNA²⁵; Forterre e Eugene Koonin proponen que los virus fueron partículas parasitarias que aceleraron los procesos evolutivos; ellos sugieren que transferencias independientes, en cada edad, del material genético de los virus hacia las células, produjo dos tipos (procarionte y eucarionte), divididos en los tres dominios celulares conocidos actualmente (*Eubacteria*, *Archaea* y *Eucaria*^{24,26}).



Op. Cit. ²⁵

F3. Escenario evolutivo propuesto por Forterre, en que se indican 3 edades con base en las moléculas bióticas fundamentales.

La propuesta del Mundo de las RNPs se apoya experimentalmente en los diferentes ejemplos de complejos de RNA-proteínas que se encuentran en diversos organismos actuales; por ejemplo hay casos en los cuales pocos nucleótidos sin estructura o incluso un solo nucleótido pueden servir de unión a una proteína, por otro lado hay casos en que péptidos pequeños o incluso unos pocos aminoácidos interaccionan específicamente con RNA²⁷.

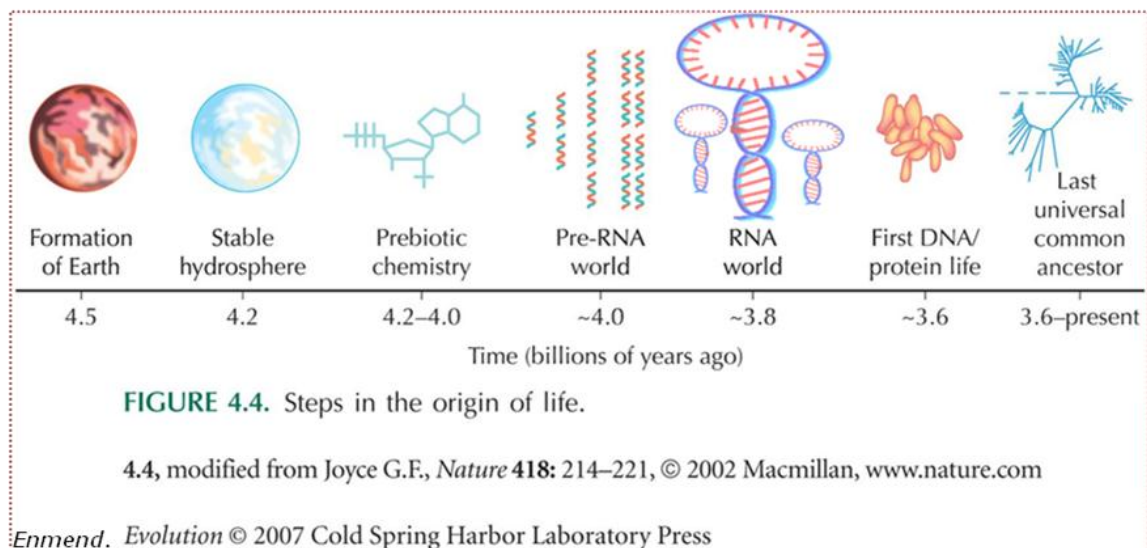
Hace relativamente poco tiempo se ha propuesto que existió un ambiente de RNA+proteínas cerca de los albores del origen de la vida, lo que es bastante probable ya que los aminoácidos son mucho más fáciles de sintetizar en un ambiente prebiótico que los nucleótidos²⁷. Eörs Szathmáry ha propuesto que en un inicio los aminoácidos fueron cofactores de las ribozimas²⁸. Massimo Di Giulio propone que nunca existió un periodo de RNA puro, sino que desde el principio hubo tRNAs unidos covalentemente a pequeños péptidos y que estos complejos podían catalizar diversas reacciones bioorgánicas con lo que inició el metabolismo²⁹, en su hipótesis Di Giulio señala que el código genético surgió como consecuencia de la interacción de los peptidil-tRNA entre sí en la última fase del Mundo de las RNPs²⁹.

2.2 LUCA.

El paso evolutivo posterior al Mundo del RNA se dio cuando el DNA, más estable termodinámicamente, surge como la molécula que preserva la información genética; las proteínas ahora fungen como las principales moléculas catalizadoras de las reacciones bioorgánicas, sin eliminar la catálisis del RNA.

El producto al que convergieron al menos algunos de los eventos evolutivos anteriormente descritos fue un organismo, o más probablemente una población de organismos ancestrales, cuya información hereditaria está cifrada en DNA que se traduce a proteínas con base en el código genético estándar (CGE) actual^{4,30}. Al dicho organismo virtual se le ha denominado Último Ancestro Común Universal (LUCA)³¹, el cual dio origen a los tres grandes reinos en que se clasifican los organismos celulares en la actualidad, de tal forma que *Eubacteria*, *Archaea* y *Eucaria* comparten el mismo código genético³².

Las grandes transiciones evolutivas, desde el ambiente prebiótico hasta el surgimiento de LUCA, se pueden esquematizar como lo muestra la figura (F4):



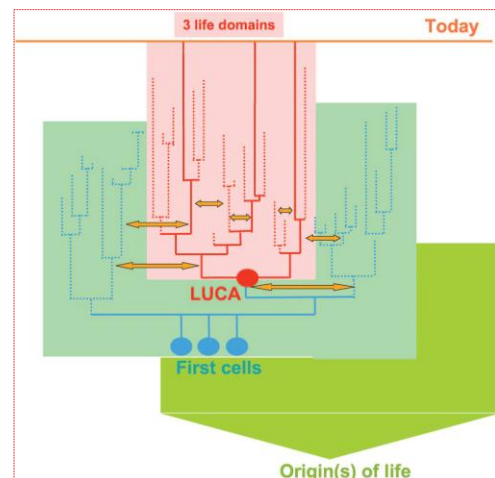
F4. Pasos detallados en el origen y evolución de la vida, desde la formación misma de la Tierra hasta el surgimiento de LUCA.

Hasta ahora hemos descrito cronológicamente como pudo haberse originado y desarrollado la vida; por otro lado, se han efectuado extrapolaciones hacia atrás en el tiempo para basar los estudios realizados acerca del LUCA en los organismos actuales⁴ y de ello resulta que la evolución de las especies puede representarse como un proceso ramificado³³; al comparar las secuencias las subunidades del RNA ribosomal (rRNA) 16S rRNA de procariontes y 18S rRNA de eucariontes³⁴ y además las proteínas que comparten dichos organismos^{35,36}, se ha modelado un esquema filogenético denominado "árbol de la vida", en el que LUCA se ubicaría cerca de la raíz, pues correspondería al ancestro de los tres grandes dominios celulares³⁷.

Hay varias teorías acerca de la naturaleza de LUCA^{38,39}, cuestionando si era un organismo simple o complejo⁴⁰ o cuál era su material hereditario (genotipo)⁴¹; en general las características celulares (fenotipo) que tenía^{42,43,44,45}; si se trataba de un organismo, o de un grupo homogéneo o de una comunidad más bien ecléctica, pero cuyo desarrollo semejaba una unidad biológica⁴⁶; también se ha tratado el tema del grado de transferencia horizontal de genes (HGT), lo que podría cambiar incluso el concepto de árbol de la vida, como queda ilustrado en la siguiente figura (F5)^{47,48}; incluso se ha llegado a tratar de plantear las características esenciales de la vida extraterrestre con base en el conjeturado contenido de LUCA⁴⁹. Se observa entonces que, no obstante las teorías acerca del LUCA, aún no existe un consenso acerca de su naturaleza ni aun de su existencia^{50,51}.

*Op. Cit.*³⁷

F5. Esquema que ilustra como el proceso ramificado de la evolución de las especies se ha visto intervenido constantemente por la transferencia horizontal de genes (HGT).



Se ha intentado dilucidar mediante diversos métodos cuáles son los elementos bióticos más antiguos; varios métodos se han seguido para tratar de esclarecer cómo fue el origen de la vida en la Tierra y la composición genética y/o proteica de los primeros organismos hasta el surgimiento de LUCA; sin embargo para cualquier análisis evolutivo, máxime desde hace tantos millones de años, es menester tomar en cuenta una gran cantidad de factores ^(cfr.52).

Como podemos apreciar, es posible que los diferentes mundos ancestrales que se plantean pudiesen haber coexistido de algún modo en los albores de las condiciones pre-celulares, sin embargo un modo de rastrear algunos de los componentes más antiguos es analizar las proteínas que eran codificadas por un código genético ancestral.

2.3 CÓDIGO GENÉTICO Y AMINOÁCIDOS ANCESTRALES.

Una de las principales transiciones evolutivas es el origen del código genético en el contexto del Mundo de las RNPs, es decir cuando nucleótidos y aminoácidos comenzaron a interactuar^{4,25,27,29}.

Francis Crick propuso que el código genético era resultado de un **"accidente congelado"**⁵³; se ha mostrado que pese a que la asignación de codones es redundante, ésta no es aleatoria⁴. Actualmente se considera que el código genético es estándar y no universal, debido a que en algunos organismos se han encontrado ligeras variaciones en la asignación de codones⁴; se ha propuesto que las más de 20

variantes conocidas hasta ahora⁵⁴ son adaptaciones a estilos de vida particulares debido a diferentes mecanismos de reasignación de codones^{55,56,57}.

Se ha propuesto que el CGE se originó en el Mundo del RNA. Diversas teorías sugieren que el código genético fue incorporando los codones, y sus respectivos aminoácidos, en distintos tiempos evolutivos.

2.3.1 CÓDIGO RNY.

En 1976 Francis Crick propuso que los tripletes más antiguos, de un código genético libre de comas, eran del tipo RRY (R: purinas, Y: pirimidinas)⁵⁸, sin embargo tal estructura sería termodinámicamente inestable debido al exceso de purinas⁵⁹. En 1978 Manfred Eigen y Peter Schuster publicaron la tercera parte de una serie de magnánimos artículos acerca de hiperciclos evolutivos que generarían secuencias con mayor contenido de información a partir de cuasi-especies de RNA; con base en **el modelo de hiperciclos evolutivos se propuso que las secuencias "semilla" debían cumplir con tres características: las estructuras secundarias de RNA que se generen pueden ser reconocidas más fácilmente por enzimas debido a que ambas hebras son equivalentes; la simetría de éstas permite además su eficiente replicación y "por último los hiperciclos pueden surgir con mayor facilidad de una cuasi-especie si ésta, debido a su simetría, ofrece dos funciones complementarias"**. El grupo de tripletes con patrón RNYⁱⁱⁱ contiene tanto un codón como su inverso complementario, cumpliendo con las características antes mencionadas, por lo tanto Eigen concluye que los tripletes de este tipo (RNY) son los más ancestrales⁶⁰. Jean Lehmann llegó recientemente a las mismas conclusiones con base en un modelo estereoquímico⁶¹.

ⁱⁱⁱ Donde **R** corresponde a las bases A y G (purinas), **N** simboliza cualquiera de las cuatro bases en el RNA (A, G, C, U) y la **Y** representa C y U (pirimidinas).

John Shepherd publicó dos estudios en los que analiza los patrones de correlación en secuencias de DNA que codifican proteínas, observando una periodicidad cada tres bases nucleotídicas; propuso que estas oscilaciones se generaron a partir de la evolución de cadenas ancestrales de tripletes tipo RNY, siendo dichos patrones vestigios del Mundo del RNA^{62,63}. Eigen publicó en 1985 un trabajo en el cual analizó 200 diferentes secuencias de 5SrRNA y encontró que la frecuencia de los tripletes decae en el orden RNY>RNR>YNY>YNR⁶⁴. En 1986 Tze-Fei Wong y Robert Cedergren rebatieron la posibilidad de la existencia de vestigios de tiempos tan remotos y mencionan que el sesgo hacia codones del tipo RNY se debe a la selección natural y para nada refleja una codificación ancestral⁶⁵. Thomas Jukes ha señalado que la prevalencia de codones del tipo RNY en los genomas puede deberse a la preferencia por los tRNA correspondientes⁶⁶.

Con el fin de esclarecer la controversia que existe en torno al origen de las repeticiones RNY y con respecto al origen mismo del código genético, José MV y colaboradores analizaron genomas completos de diversos procariontes y encontraron que varias propiedades estadísticas de las series de distancias desde genomas ancestrales (ver adelante), hasta genomas que siguen el CGE no han sido alteradas *i.e.* existe una invariancia de escala que se ha preservado por al menos 3 mil millones de años⁶⁷. El grupo propuso dos vías evolutivas a partir del código primitivo RNY, denominadas extendido I y extendido II (Ext.I y Ext.II), a partir de las cuales puede generarse el CGE; esas vías se obtienen mediante distintas operaciones algebraicas, que biológicamente corresponden o bien a los cambios de marco de lectura (Ext.I) o bien a transversiones en la primera o tercera base de cada codón (Ext.II)⁶⁷ (y refs. 68,69,70,71). De manera notable, José MV y su grupo mostraron que además dichos procesos evolutivos pueden representarse en tres dimensiones, representación llamada Hotel Genético del CGE, en la que es posible visualizar que el RNY forma un "departamento" dentro del "Hotel"^(68-71 y 72).

El código genético del Mundo del RNA codificaría sólo para ocho aminoácidos en el Mundo de las RNPs: GGY (Gly, **G**), GCY (Ala, **A**), GAY (Asp, **D**), GUY (Val, **V**), AGY (Ser, **S**), ACY (Thr, **T**), AAY (Asn, **N**) y AUY (Ile, **I**)^{iv(73)}.

La hipótesis de que esos aminoácidos fueron los primeros en ser codificados es respaldada por un análisis de Trifonov, con el que busca generar un consenso entre más de cien teorías acerca de la evolución del CGE; la cronología resultante cumple con las reglas fundamentales de termoestabilidad, complementariedad y procesividad^{74,75}.

Mientras que en lo general, Eigen y colaboradores han estimado que "el código genético no es más antiguo, pero casi tan antiguo como nuestro planeta"⁷⁶.

2.3.2 HIPÓTESIS DIFERENTES A LA TEORÍA RNY.

Además del código RNY, se han propuesto distintas hipótesis con respecto al origen y evolución del código genético:

* La recreación prebiótica de Miller⁷⁷ indica a **AGDV** como los aa primordiales.

* Otros trabajos señalan que el código genético coevolucionó con el metabolismo, por ejemplo Wong⁷⁸, Taylor y Coates⁷⁹, Davis⁸⁰ y DiGiulio⁸¹ señalan que la asignación de codones es reflejo de las rutas biosintéticas de los aa.

* Según Ikehara el origen de la vida comenzó con la pseudo-replicación de los aa **GA[D/E]V**, correspondientes a un código GNC⁸².

* Guimarãez indica que el código genético se originó por interacciones cíclicas de proto-tRNAs y aminoácidos; en ese modelo autorreferencial, los tRNAs formaban

^{iv} **G**: neutro, hidrófilo, alifático; **A**: neutro, hidrófobo, alifático; **D**: ácido o con carga negativa; **V**: neutro, hidrófilo, alifático; **S**: hidroxiaminoácido, neutro, hidrófilo; **T**: hidroxiaminoácido, neutro, hidrófilo; **I**: neutro, hidrófobo, alifático; **N**: neutro, hidrófilo, amida de aminoácido ácido⁴⁷. A lo largo de la presente tesis, se utilizará la abreviatura en tres letras y en una letra para todos los aminoácidos (aa); marcando con **negritas** los que sean codificados por tripletes RNY.

dímeros a través de sus anticodones y portaban aa en el otro extremo, el contacto cercano entre aa favorecía la formación de péptidos, que a su vez estabilizaban las estructuras diméricas de los tRNAs, formando una estructura ribonucleoproteica⁸³.

* Marc Delarue⁸⁴, Rodin y Rodin⁸⁵ plantean que la asignación de codones se dio por las características de las aminoacil-tRNA sintetasas.

* El código YRY(N)₆YRY es también mencionado como ancestral porque es un motivo muy recurrente en las secuencias de proteínas de diversos organismos⁸⁶.

2.4 ORGANISMO ELEGIDO.

Para este trabajo se han utilizado previamente programas computacionales *ad hoc* para obtener un genoma ancestral, en RNY, a partir del genoma del Firmicutes *Streptococcus agalactiae* (*Sag*) cepa *A909* (*SagA909*).

SagA909 es un anaerobio facultativo, mesófilo, Gram-positivo (G+)^v, con bajo contenido GC en su genoma, que pertenece al serotipo Ic del Grupo B de *Streptococcus* (GBS). La bacteria *Sag* es la causa principal de meningitis y sepsis en recién nacidos, es la causa de infecciones graves en adultos inmunocomprometidos y es capaz de colonizar asintóticamente la piel humana y las membranas mucosas; también está implicada en casos de mastitis en vacas⁸⁷.

En *Sag* iniciaron los trabajos de pan-genoma; un pan-genoma es la colección de genomas que caracterizan una especie y contiene ~80% del contenido de cualquiera de ellos, aun cuando nuevos genomas secuenciados arrojen genes únicos^{88,89,90}; con el objeto de buscar el pan-genoma ancestral, se eligió inicialmente *SagA909* y aunque mi objetivo es distinto, trabajé con esos datos (sig. apartados).

^v G+: organismos Gram-positivos, G-: organismos Gram-negativos.

III. PREGUNTA DE INVESTIGACIÓN e HIPÓTESIS

Se desconoce la ruta evolutiva que dio origen a los primeros organismos, sin embargo un modo de averiguar cuáles podrían ser elementos ancestrales es encontrar vestigios de su presencia en organismos actuales.

El trabajo presentado aquí se basa en la búsqueda de vestigios de proteínas ancestrales, surgidas en el Mundo de las Ribonucleoproteínas, en el proteoma actual de un organismo, a partir de su codificación en los ancestrales tripletes RNY del Mundo del RNA.

PREGUNTA DE INVESTIGACIÓN:

¿Es posible identificar vestigios de proteínas ancestrales en el proteoma actual de un organismo?

HIPÓTESIS:

Si el código RNY es ancestral y surge en el Mundo del RNA, entonces las proteínas que son codificadas por este tipo de tripletes habrían surgido al inicio del Mundo de las Ribonucleoproteínas y podemos encontrar sus vestigios en el proteoma de un organismo actual.

IV. OBJETIVOS

OBJETIVO GENERAL.

Mostrar que en un organismo contemporáneo (el 'organismo elegido' *SagA909*.) es posible encontrar vestigios de proteínas ancestrales, que provienen del Mundo de las Ribonucleoproteínas.

OBJETIVOS PARTICULARES.

- o Disponiendo de la lista completa de proteínas de *SagA909* que son codificadas al menos en parte por tripletes RNY, realizar comparaciones numéricas de sus parámetros estadísticos con proteínas derivadas de un control, con el fin de establecer valores de corte (ver 'metodología').
- o Clasificar las proteínas, que de acuerdo con sus parámetros estadísticos sean significativas, de manera comprensible de acuerdo con su función celular y *grosso modo* describirla.
- o Exponer los diversos argumentos que pudieran existir para suponer que efectivamente cada una de esas proteínas, o porciones de éstas, pueden ser ancestrales.
- o Realizar un análisis global del proteoma analizado, mismo que probablemente poseía al menos un riboproteo-organismo (o progenote) que existió antes de LUCA.
- o La investigación realizada nos permitirá modelar de mejor manera el proteoma ancestral mínimo.

V. METODOLOGÍA

El estudio que he llevado a cabo ha sido con base en trabajos previos realizados en el Grupo de Biología Teórica (al cual pertenezco y del cual es titular el Dr. Marco Antonio José Valenzuela).

5.1 PASOS PREVIOS.

Siguiendo la metodología realizada por José MV y colaboradores⁶⁷, se obtuvo mediante programas computacionales realizados *ad hoc*, una secuencia que contiene exclusivamente codones del tipo RNY a partir del proteoma (ORFeoma) original de *SagA909*; ya que el objetivo inicial era plantear el pan-genoma ancestral (ver 'organismo elegido') y, que aun cuando no es el objetivo de la presente tesis, son los datos con los que se ha trabajado con el fin de obtener un conocimiento sobre la probable naturaleza de los péptidos ancestrales.

Se obtuvieron las secuencias nucleotídicas de los ORFs de la bacteria *SagA909* mediante el archivo "ffn" del FTP de la página del NCBI, con la dirección:

ftp://ftp.ncbi.nih.gov/genbank/genomes/Bacteria/Streptococcus_agalactiae_A909_uid3_26/CP000114.ffn

Para que la estructura original de las cadenas se conservara, se ensamblaron todos los ORF reportados, uno tras otro tal cual el orden original de orientación en el (OOO), anotando el complementario de aquellos ORF que se encuentran en la hebra rezagada y dejando intactos los de la hebra líder; de esta manera queda una estructura como si todos los ORFs se leyeran en la hebra líder.

Como control negativo se generó una secuencia aleatoria a partir del genoma OOO, desordenando los nucleótidos, con el fin de eliminar el sentido biológico de la secuencia.

Para obtener una secuencia compuesta exclusivamente de RNY, ya sea con el sentido biológico o a partir de la cadena aleatoria, se descartaron del genoma OOO los tripletes que no correspondían.

Por ejemplo, si la secuencia OOO es: ATG ATA CTA ATG AAG TAT AGI GCI ATI TTA TTA
ATA TGT AGC GTI TTA AAI TTT ...

Entonces, el genoma RNY es: AGT GCT ATT AGC GTT AAT ...

Las secuencias resultantes RNY (con el sentido biológico y la del control negativo) se compararon por separado con la secuencia genómica original de *SagA909*, mediante el procedimiento BLAST en su versión instalada, para obtener únicamente proteínas codificadas por ese tipo de tripletes (el programa que 'llama' a BLAST se ofrece como apéndice en el apartado IX). Para el archivo de salida no se restringieron los valores de los parámetros (ni el valor-E, ni el "score", ni la longitud de la secuencia, ni el porcentaje de identidad; explicación en el sub-apartado 'valor de corte' en la sección 'Resultados y Análisis'). Las proteínas coincidentes se mostraron en un formato de hoja de cálculo, en este caso una lista de la secuencia con sentido biológico y otra lista de la secuencia aleatoria.

5.2 METODOLOGÍA PROPIA.

* Se realizó una comparación numérica de las proteínas que resultan del procedimiento BLAST, de la secuencia con sentido biológico que se utiliza ahora como prueba de hipótesis y del control, con el fin de establecer valores de corte.

* Clasifiqué todas las proteínas, identificadas por contener regiones codificadas por tripletes RNY, de manera semejante a la base de datos COG⁹¹ y utilicé la base de datos KEGG⁹² para los casos en que se requería una clasificación más fina y con el fin de comprender los procesos metabólicos de manera plausible.

* Investigué las funciones celulares de cada una de las proteínas con sentido biológico, para la prueba de hipótesis, en libros de texto clásicos^{93,94,95} y en las ligas ofrecidas por la base de datos Protein Cluster⁹⁶, tanto de literatura como de otros recursos; en algunos casos agrupé las proteínas bajo el criterio de similitud funcional. Todo ello con el propósito de ofrecer una explicación plausible al hecho de que fragmentos aún reconocibles de esas proteínas estén codificados por tripletes RNY y que por tanto, bajo nuestro supuesto, sean ancestrales; bien porque se hayan realizado trabajos al respecto o al inferirlo a partir de lo que se haya publicado. Los aminoácidos codificados por tripletes RNY, que aparezcan en los **motivos** ("motifs") de las proteínas analizadas, los resalto en letras **negritas**, e.g. **GxxTEDxGP**. Es necesario mencionar que modifiqué *post hoc* la categoría COG original de algunas proteínas, con base precisamente en su funcionamiento y para reflejar de manera más adecuada su relevancia evolutiva en el Mundo de las RNPs.

* He analizado de manera semejante los resultados del genoma aleatorio o control.

* Finalmente aportó una visión global del proteoma o sus primordios y sus características funcionales dentro de un riboproteo-organismo que habría existido antes que LUCA.

VI. RESULTADOS y ANÁLISIS

El trabajo presentado aquí se basa en la búsqueda de vestigios de proteínas ancestrales surgidas en el Mundo de las Ribonucleoproteínas, en el proteoma de un organismo contemporáneo (ver 'organismo elegido' y 'metodología'), a partir de su codificación en los ancestrales tripletes RNY del Mundo del RNA.

6.1 NIVEL DE CORTE.

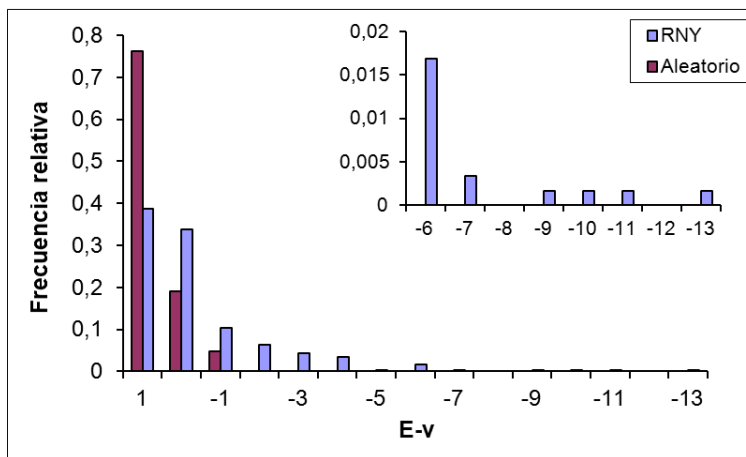
Con el fin establecer un nivel de corte para las proteínas obtenidas de ambos genomas en RNY, el que tiene sentido biológico que servirá para la prueba de hipótesis y el aleatorio que funge como control; se ha realizado una comparación numérica de las frecuencias relativas de las proteínas de ambas listas, para cada uno de los parámetros obtenidos del programa BLAST⁹⁷, reportando la frecuencia relativa de las proteínas con respecto al 'valor esperado' (valor-E)^{vi} y al 'Score' (**S**)^{vii}.

Para el valor-E se obtuvieron los negativos de sus logaritmos (Log_{10}) con el fin de que la comparación fuera fácilmente visualizada; podemos observar en la figura **F6** que la frecuencia relativa de proteínas en el genoma aleatorio se ubica en $E > 0.05$. El parámetro **S** también fue evaluado y como podemos observar en la figura **F7**, la frecuencia relativa de proteínas en el genoma aleatorio se ubica en $S \geq 24$.

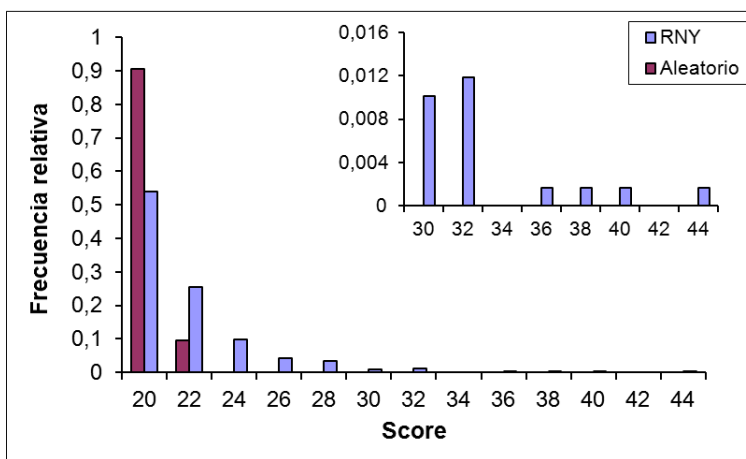
^{vi} El valor-E es la probabilidad de que la coincidencia encontrada entre dos secuencias sea resultado del azar y mientras menor sea el valor-E, se considera una mejor coincidencia; para valores de $E < 0.01$, el valor-E es equivalente al valor-P estadístico.

^{vii} El **S** es un valor numérico que describe la calidad de un alineamiento y mientras mayor sea el **S**, se considera que hay mayor similitud entre ambas secuencias.

F6. Comparación de las frecuencias relativas, con respecto al valor-E (en Log_{10}), del genoma con sentido biológico y del aleatorio.



F7. Comparación de las frecuencias relativas, con respecto al S , del genoma con sentido biológico y del aleatorio.



En ocasiones el parámetro S va a depender de si se toman en cuenta la suma de todos los fragmentos coincidentes entre ambas secuencias, es por ello que el parámetro valor-E es el que se utiliza como estándar para reportar el significado estadístico de un alineamiento; de tal manera que el umbral para el nivel de corte queda establecido en $E \leq 0.05$, aun cuando para la presente tesis ambos valores son válidos.

6.2 CLASIFICACIONES Y CATEGORÍAS.

Inicialmente clasifiqué todos los resultados bajo cuatro criterios generales: **proteínas individuales**, **proteínas que pertenecen a una familia**, **proteínas putativas** y **proteínas/dominios hipotéticas(os)**. Posteriormente descarté los resultados que rebasaron el nivel de corte establecido. Finalmente dividí las proteínas, que sí se ajustaban al nivel de corte, con base en su categoría y algunas por subcategoría funcional, *grosso modo* conforme a las bases de datos COG y KEGG. Cabe señalar que la nominación de todas las proteínas es la misma del archivo FTP de **SagA909**.

- o **RNA, estructura y modificaciones**. Categoría que agrupa las proteínas relacionadas con la traducción, estructura del ribosoma y modificación del RNA.
- o **DNA, estructura y modificaciones**. Categoría donde he incluido factores transcripcionales; así como proteínas relacionadas con la modificación estructural y procesamiento del DNA, tanto de la bacteria como de su bacteriófago asociado.
- o **Metabolismo**. Aquí agrupé todas las proteínas involucradas en alguna ruta metabólica (metabolismo energético, de carbohidratos, de nucleótidos, de lípidos, de aminoácidos, de coenzimas, o de iones inorgánicos, inclusive las no incluidas en algún COG pero que de acuerdo con su función o la base de datos KEGG sí forman parte de algún tipo de metabolismo específico).
- o **Proteínas, estructura y modificaciones**. Categoría donde he agrupado todas las proteínas relacionadas con la modificación y procesamiento de otras proteínas o sus grupos funcionales.
- o **Estructuras membranales y extramembranales**. En esta categoría están incluidas las proteínas involucradas en la síntesis de componentes de la pared y/o de la membrana celular; aquí también incluí **las "proteínas putativas de membrana"** pues, aunque pueden fungir como transportadores o componentes del sistema de transducción de señales, no se menciona su función específica.

- o **Transportadores y Canales**. Todas las proteínas de esta categoría están involucradas en el transporte de moléculas hacia dentro de la célula o fuera de ella.
- o **Transducción**. Las proteínas de esta categoría están involucradas en la transducción de señales.
- o **Resistencia y Supervivencia**. Agrupamiento de proteínas diversas que confieren resistencia ante agentes ambientales, proteínas de defensa celular y proteínas que proporcionan actualmente características patógenas al microorganismo.
- o **Proteínas/dominios hipotéticas(os) y/o conservadas(os)**. Es la categoría de proteínas o dominios que no están plenamente caracterizadas(os); ya sea porque su función se desconoce o porque no se tiene la certeza que sean proteínas o dominios, mientras que conservada(o) implica que se encuentra el mismo fragmento en al menos 3 organismos filogenéticamente distantes. Ambas clasificaciones sólo coinciden en esta categoría de 'proteínas/dominios hipotéticos'.

6.3 ANÁLISIS.

A continuación se presentan dos tablas, en (T1) se muestran las proteínas que permitirán la prueba de hipótesis y en (T2) los resultados del control negativo. Todas las tablas presentan los cuatro criterios generales, referidos anteriormente con su respectivo código de colores; así como la indicación de la categoría y subcategoría funcional de cada resultado, también con su distintivo código de colores; adicionalmente, se indica en las tablas la nomenclatura derivada de la 'metodología'.

T1. Proteínas de *SagA909*, con regiones codificadas por tripletes RNY y que se ajustan al nivel de corte establecido $E < 0.05$. Proteínas clasificadas bajo los cuatro criterios generales y posteriormente con base en su categoría y subcategoría conforme a las bases de datos COG y KEGG, según se describe en el apartado anterior 'Clasificaciones y categorías'. En gris se indican las menciones repetidas.

Cur.Or.	Sc.	E-v	ID	%	Category	Subcat(Metab)	match 1	match 2
1	43	2.00E-14	54/57	94%	Metab.	Ener.		
2	37	7.00E-11	40/41	97%	Prot.		ATP synthase F1, alpha subunit (67.77 %)	
3	32	7.00E-08	32/32	100%	DNA		preprotein translocase, SecY subunit, putative (31.41 %)	
4	31	3.00E-07	31/31	100%	RNA		ribosomal protein L6 (70.88 %)	
5	31	3.00E-07	42/45	93%	Cons./HYP.		conserved hypothetical protein (98.97 %)	
6	31	3.00E-07	37/39	94%	Rest./Striv.		transferrin, hexapeptide repeat family (59.40 %)	
7	31	3.00E-07	31/31	100%	Transp./Ch.		ABC transporter, substrate-binding protein, putative (32.57 %)	
8	30	1.00E-06	30/30	100%	Metab.	Carb.	D-mannanate oxidoreductase Uxub, putative (4.89 %)	
9	30	1.00E-06	30/30	100%	Metab.	Aa.	glyoxylase family protein (48.91 %)	
10	30	1.00E-06	30/30	100%	DNA		DNA topoisomerase I (90.14 %)	
11	30	1.00E-06	30/30	100%	Metab.	Ion	sulfatase family protein (91.93 %)	
12	30	1.00E-06	30/30	100%	Transp./Ch.		glycine betaine/proline ABC transporter, permease/substrate-binding protein (70.30 %)	
13	30	1.00E-06	29/29	100%	Metab./Wall		ATP-dependent Nucleoside Diphosphate box family (50.32 %)	
14	28	2.00E-05	28/28	100%	Metab./Wall		membrane protein, putative (22.01 %)	
15	28	2.00E-05	28/28	100%	Metab.	Nucl.	GMP synthase (95.39 %)	
16	28	2.00E-05	28/28	100%	RNA		NOL1/NOP2/sun family putative RNA methylase (89.01 %)	
17	28	2.00E-05	28/28	100%	Cons./HYP.		conserved hypothetical protein (77.22 %)	
18	28	2.00E-05	28/28	100%	DNA		DNA-binding response regulator (68.66 %)	
19	28	2.00E-05	28/28	100%	DNA	Carb.	phosphonmutase (93.64 %)	
20	28	2.00E-05	39/42	92%	Metab.	Aa.	amino transferase, classes I and II (50.21 %)	
21	27	6.00E-05	27/27	100%	Metab.	Nucl.	hypoxanthine phosphoribosyltransferase (29.70 %)	
22	27	6.00E-05	27/27	100%	Cons./HYP.		conserved hypothetical protein (82.76 %)	
23	27	6.00E-05	27/27	100%	Metab.	Lip.	enoyl-CoA hydratase/isomerase family protein (65.87 %)	
24	27	6.00E-05	44/50	88%	Transp./Ch.		membrane protein oxaA, putative (68.59 %)	
25	27	6.00E-05	27/27	100%	Memb./Wall		ABC transporter, ATP-binding protein CytA (5.06 %)	ABC transporter, permease protein CytB (1.03 %)
26	27	6.00E-05	27/27	100%	Memb./Wall		lipoprotein, putative (68.64 %)	
27	27	6.00E-05	27/27	100%	Metab.	Carb.	UDP-N-acetylglucosamine 1-carboxyvinyltransferase (43.45 %)	
28	27	6.00E-05	27/27	100%	Metab.	Carb.	glycogen synthase (14.62 %)	
29	27	6.00E-05	27/27	100%	Metab.	Ener.	NADH oxidase, water-forming (67.23 %)	
30	27	6.00E-05	27/27	100%	Memb./Wall		streptococcal histidine triad family protein (89.71 %)	
31	27	6.00E-05	27/27	100%	Transp./Ch.		polar amino acid uptake (PAAT) family ABC transporter, amino acid-binding protein (8.81 %)	
32	26	2.00E-04	26/26	100%	Memb./Wall		glycosyl transferase family protein (19.74 %)	
33	26	2.00E-04	26/26	100%	Transp./Ch.		major facilitator family transporter (26.42 %)	
34	26	2.00E-04	29/30	96%	RNA	Carb.	ATP-dependent RNA helicase, DEAD/DEAF box family (27.70 %)	
35	26	2.00E-04	37/40	92%	Metab.		HAD-superfamily hydrolase, subfamily T1A (95.71 %)	
36	2.00E-04	26/26	100%	Transp./Ch.		iron chelate uptake ABC transporter, permease protein (1.14 %)	iron chelate uptake ABC transporter, permease protein (1.75 %)	
37	2.00E-04	26/26	100%	Transp./Wall		iron chelate uptake ABC transporter, permease protein (1.14 %)		
38	2.00E-04	26/26	100%	RNA		iron chelate uptake ABC transporter, permease protein (1.14 %)		
39	25	0.001	25/25	100%	Metab./Ch.		trna-type domain protein, tG (69.72 %)	
40	25	0.001	25/25	100%	DNA		peptide/protein/ribosomal uptake (pup1) family ABC transporter, substrate-binding protein (30.65 %)	
41	25	0.001	34/37	91%	Metab.	Lip.	site-specific recombinase, phage integrase family (4.42 %)	
42	25	0.001	25/25	100%	Metab.		acyl-CoA carboxylase, biotin carboxylase (24.53 %)	
43	25	0.001	25/25	100%	Transp./Ch.		glycerol-3-phosphate dehydrogenase (NAD(P)+) (29.63 %)	
44	25	0.001	36/39	92%	RNA	Ener.	permease, putative (43.78 %)	
45	25	0.001	25/25	100%	DNA		threonyl-tRNA synthetase (1.18 %)	
46	25	0.001	25/25	100%	Prot.		cell wall surface anchor family protein (49.54 %)	
47	25	0.001	25/25	100%	DNA		DNA polymerase III, delta subunit (23.92 %)	
48	25	0.001	25/25	100%	Cons./HYP.		hypothetical protein (25.60 %)	
49	25	0.001	25/25	100%	Cons./HYP.		ABC transporter, permease protein, carbohydrate uptake 2 (CUT2) family, putative (98.64 %)	
50	25	0.001	25/25	100%	RNA		appr-1-p processing enzyme family domain protein (45.72 %)	
51	25	0.001	25/25	100%	Metab.	Other	4-oxalocrotonate tautomerase family enzyme (0.00 %)	
52	25	0.001	25/25	100%	Memb./Wall		membrane protein, putative (42.31 %)	
53	25	0.001	25/25	100%	RNA		polyA polymerase (51.89 %)	
54	25	0.001	34/37	91%	Transp./Ch.		ABC transporter, ATP-binding protein (39.92 %)	
55	25	0.001	25/25	100%	Memb./Wall		dHB protein (76.60 %)	
56	24	0.004	24/24	100%	DNA		plasmid recombination enzyme (35.43 %)	
57	24	0.004	24/24	100%	Transp./Ch.		polar amino acid uptake (PAAT) family ABC transporter, permease protein (63.89 %)	
58	24	0.004	24/24	100%	Transp./Ch.		sensor/histidine kinase (20.00 %)	
59	24	0.004	24/24	100%	Transp./Ch.		serine/threonine protein kinase Sdk1 (80.97 %)	
60	24	0.004	24/24	100%	Transp./Ch.		Biot family protein (33.06 %)	
61	24	0.004	24/24	100%	Transp./Wall		valyl-tRNA synthetase (14.05 %)	
62	24	0.004	24/24	100%	RNA		Ser/Thr protein phosphatase family protein (94.53 %)	
63	24	0.004	35/38	92%	Memb./Wall		membrane protein, putative (61.02 %)	
64	24	0.004	24/24	100%	Cons./HYP.		conserved domain protein (32.97 %)	

T1 (continuación).

129	22	0.06	25/26	96%	Metab.	Aa.			
130	22	0.06	22/22	100%	Cons./Hyp.				
131	22	0.06	25/26	96%	DNA				
132	22	0.06	25/26	96%	Metab.	Nucl.			
133	22	0.06	22/22	100%	Cons./Hyp.				
134	22	0.06	22/22	100%	DNA				
135	22	0.06	22/22	100%	DNA				
136	22	0.06	22/22	100%	DNA				
137	22	0.06	22/22	100%	RNA				
138	22	0.06	22/22	100%	Cons./Hyp.				
139	22	0.06	22/22	100%	Prote.				
140	22	0.06	22/22	100%	Metab.	Multi./Gral.			
141	22	0.06	22/22	100%	Transp./Ch.				
142	22	0.06	33/36	91%	DNA				
143	22	0.06	22/22	100%	DNA				

T2. Proteínas del genoma aleatorio de *SagA909*, con regiones codificadas por tripletes RNY y que se ajustan al nivel de corte establecido $E < 0.05$. Proteínas clasificadas bajo los cuatro criterios generales y posteriormente con base en su categoría y subcategoría conforme a las bases de datos COG y KEGG, según se describe en el apartado 'Clasificaciones y categorías'.

Curr.Or.	Sc.	E-v	ID	%	Category	match 1
1	22	0.044	22/22	100%	DNA	transcriptional regulator, putative (58.38 %)

Las tablas T1 y T2 siguen la nomenclatura:

Curr.Or.: Ordenamiento consecutivo de las proteínas.

Sc: Parámetro denominado 'Score'.

E-v: Parámetro denominado 'valor-E'.

ID: Número de aa idénticos dentro del fragmento codificado por tripletes RNY.

%: Porcentaje derivado del parámetro ID.

Category: Categoría de cada proteína.

Subcat.: Subcategorías de las proteínas del metabolismo.

Match 1 y **Match 2:** Proteínas que contienen regiones codificadas por tripletes RNY.

6.3.1 SENTIDO BIOLÓGICO, PRUEBA DE HIPÓTESIS.

La tabla (T3) indica el tipo de proteínas diferentes por categoría y subcategoría de la secuencia con sentido biológico; cada resultado distinto se cuenta una sola vez, pues el hecho de que algunas proteínas o dominios se mencionen repetidamente depende del número de copias que existan en el genoma original (*SagA909*). Como indiqué previamente (en 'metodología'), modifiqué la categoría COG de algunas enzimas con base en su funcionamiento y descarté las menciones repetidas.

Categorías	Subcategorías
RNA, estructura y modificaciones (15).	
DNA, estructura y modificaciones (18).	
Metabolismo (M.) (29).	M. energético (5). M. de carbohidratos (7). M. de nucleótidos (4). M. de lípidos (2). M. de aminoácidos (5). M. de coenzimas (2). M. de iones inorgánicos (1). No incluidas en COGs (3).
Proteínas, estructura y modificaciones (7).	
Membrana y estructuras extramembranales (15).	
Transportadores y Canales (21).	
Transducción (3).	
Resistencia y Supervivencia (3).	
Hipótesis y/o Conservadas (12).	

T3. Tabla que muestra los resultados para la prueba de hipótesis, se indica el tipo de proteínas diferentes con su código de colores correspondiente; así como, entre paréntesis, el número de resultados en cada conjunto.

A continuación muestro cada una de las proteínas, agrupadas con base en su categoría funcional y por proceso metabólico respectivo. Al inicio de cada categoría presento una tabla, derivada de la general (T1), de las proteínas que en ella encontraremos; posteriormente ofrezco las explicaciones más plausibles que puedan fundamentar la existencia de esas proteínas o dominios en el Mundo de las RNPs, ya

sea porque existen trabajos al respecto o por inferencia propia con base en la literatura; finalmente presento una tabla adicional, a manera de resumen, de los resultados encontrados en cada categoría. Todas las tablas tienen los mismos códigos de colores señalados en '**clasificaciones y categorías**'.

6.3.1.1 RNA, estructura y modificaciones.

Algunos investigadores han definido el metabolismo del RNA como el compendio de todos los procesos celulares que involucran esta molécula y se han realizado exhaustivas comparaciones genómicas con el fin de mostrar el origen y evolución de los diversos dominios asociados a ribonucleótidos; según el estudio los dominios de proteínas involucradas en funciones como la traducción, transcripción y algunas formas de modificación al RNA, se originaron incluso antes de la aparición del LUCA y fueron reclutados hacia otros procesos o bien se realizaron innovaciones a partir de estos en los diferentes linajes evolutivos⁹⁸. Se ha propuesto además que la gran mayoría de las proteínas que actualmente interaccionan con el RNA ya eran parte del repertorio celular funcional del LUCA y que, bajo la hipótesis del Mundo del RNA, fueron de las primeras proteínas en ser codificadas, en especial los componentes del proceso de traducción actual, o al menos sus predecesores^{99,100}.

De las proteínas del metabolismo del RNA que se propone son ancestrales, únicamente algunas contienen región es codificadas por tripletes RNY de lo cual puedo inferir que al menos esas regiones surgen ulteriormente al Mundo del RNA, pero en los albores del Mundo de las RNPs (en T4 y T5). Entre las proteínas cuya secuencia es codificada por tripletes RNY, podemos observar varias proteínas que participan actualmente en el proceso de madurez y ensamblaje del ribosoma como proteínas ribosomales, proteínas de modificación y helicasas; por otro lado

encontramos tanto aminoacil-tRNA sintetasas como factores de elongación y/o terminación de la traducción, todos ellos con participación directa en el proceso de decodificación de la información genética; también hay otras proteínas a las que se identifica dentro de esta categoría porque unen ribonucleótidos o ribonucleósidos; **en el caso particular de la enzima 'DNA primasa' cambié su categoría COG original,** que la relacionaba con DNA para relacionarla con RNA, pues al revisar la información al respecto pude percatarme que esta proteína es en realidad una RNA polimerasa. Desde el punto de vista evolutivo, es por demás interesante que varias de las proteínas que contienen regiones codificadas por tripletes RNY, tienen contactos con el rRNA, ya sea durante el ensamblaje y maduración del ribosoma o bien durante el proceso de traducción⁹⁹.

Order	Sc.	E-v	ID	%	PROTEIN / FAMILY / DOMAIN
1	31	3.00E-07	31/31	100%	ribosomal protein L6 (74.07 %)
2	29	4.00E-06	29/29	100%	ATP-dependent RNA helicase, DEAD/DEAH box family (50.32 %)
3	28	2.00E-05	28/28	100%	NOL1/NOP2/sun family putative RNA methylase (89.01 %)
4	26	2.00E-04	29/30	96%	ATP-dependent RNA helicase, DEAD/DEAH box family (27.70 %)
5	26	2.00E-04	26/26	100%	translation elongation factor G (69.73 %)
6	25	0.001	36/39	92%	threonyl-tRNA synthetase (1.18 %)
7	25	0.001	25/25	100%	appr-1-p processing enzyme family domain protein (45.72 %)
6	25	0.001	25/25	100%	polyA polymerase (51.89 %)
8	24	0.004	24/24	100%	valyl-tRNA synthetase (14.05 %)
9	24	0.004	24/24	100%	DNA primase (81.53 %)
10	24	0.004	24/24	100%	glucose-inhibited division protein B (56.66 %)
11	24	0.004	24/24	100%	ribosomal protein S18 (97.07 %)
12	23	0.015	23/23	100%	asparaginyl-tRNA synthetase (93.61 %)
13	23	0.015	23/23	100%	modification methylase, HemK family (26.51 %)
14	22	0.06	22/22	100%	ATP-dependent RNA helicase, DEAD/DEAH box family (22.38 %)
15	22	0.06	22/22	100%	methyltransferase domain protein (74.58 %)
16	22	0.06	22/22	100%	ribosome recycling factor (2.87 %)

T4. Proteínas, del procesamiento de RNA, que contienen regiones codificadas por tripletes RNY; nomenclatura y código de colores igual que T1 y T2.

o **PROTEÍNAS RIBOSOMALES, L6 y S18.**

o Las capacidades catalíticas del RNA, no así las informacionales, aumentaron cuando se sintetizaron los primeros péptidos, pues éstos le conferían mayor estabilidad, eficiencia catalítica y plasticidad funcional al primero; tanto el RNA como las proteínas evolucionaron cooperativamente hasta que finalmente surge el ribosoma moderno^{cfr.29,99}. Se sabe que los residuos universalmente conservados se

encuentran generalmente en áreas que son funcionalmente significativas, particularmente cerca del sitio de unión del mRNA en la subunidad pequeña y en el centro de transferencia del peptidil en la subunidad grande, incluso estas regiones altamente conservadas pueden definir un ribosoma mínimo¹⁰¹.

- o La mayoría de las proteínas ribosomales tienen algún grado de conservación en los tres dominios de la vida, lo que indica que estaban presentes en el LUCA^{100,cfr.102}; sin embargo dado que este tipo de proteínas participan activa y directamente en el proceso de formación de un ribosoma estable, es posible que su origen se remonte al periodo de transición del Mundo del RNA al Mundo de las RNPs^{cfr.103}.

- o Las primeras proteínas ribosomales, es decir con regiones codificadas por tripletes RNY, sirvieron de plataforma estructural y estabilizadora a las moléculas de rRNA, tRNA y mRNA para una correcta interacción. De todas las proteínas ribosomales que contiene **SagA909**, sólo se identificaron 2 que contienen tripletes del tipo RNY en su secuencia. La cuestión interesante es que las proteínas identificadas están en contacto con el dominio central de la subunidad 30S (S18¹⁰⁴) o forman parte de la región de decodificación del ribosoma (L6).

- **S18** forma parte de la subunidad 30S y es una proteína altamente básica, propiedad directamente relacionada con su capacidad para unir RNA; en particular se le ha implicado en la unión del aminoacil-tRNA porque está situada en el sitio de decodificación del mRNA¹⁰⁵. El dominio central del 16SrRNA, el dominio más conservado e importante para el proceso de traducción, contiene un sitio específico de unión para la proteína S18, pero ésta aparentemente no forma extensiones con la subunidad grande del ribosoma¹⁰⁶, lo que sugiere que únicamente confiere estabilidad al 16SrRNA; adicionalmente la S18 parece estar involucrada en los cambios conformacionales que ocurren durante la activación del ribosoma y durante

las interacciones importantes con el factor G de elongación de la traducción (ver más adelante)¹⁰⁷. En procariontes, el dominio N-terminal de la proteína S18 interacciona directamente con la hélice Shine-Dalgarno (SD) del 16SrRNA (la región SD sirve para posicionar el ribosoma en el mRNA a través de una interacción con la región anti-SD)¹⁰⁸. Es probable que esa región, de interacciones con el RNA, haya sido la que se identificó como ancestral en la proteína S18.

- **L6** está compuesta principalmente de aminoácidos extendidos y forma parte de la subunidad 50S¹⁰⁹, donde interactúa con la molécula 23SrRNA¹¹⁰. La proteína L6, que parece ser ubicua, contiene dos dominios con pliegues casi idénticos, implicando que surgieron a partir de un evento de duplicación de un gen ancestral. Los sitios de unión al RNA parecen estar localizados en el dominio C-terminal, mientras que el N-terminal contiene sitios potenciales para las interacciones proteína-proteína¹¹¹. L6 se une a un pequeño fragmento del 23SrRNA, mediante una depresión hidrófoba en el dominio N-terminal¹¹². La proteína L6 también interacciona con el factor G de elongación de la traducción (más adelante) y con el dominio sarcina/ricina del 23SrRNA¹¹³. L6 se encuentra localizada en un sitio **denominado "sitio de unión de factores" en la subunidad 50S**¹¹⁴. Es posible que el dominio C-terminal es el que se identificó como ancestral, debido a que es el que reconoce RNA, el dominio N-terminal es poco probable que sea ancestral debido a que no se encontraron vestigios de más proteínas ancestrales en la subunidad 50S y la proteína S18 (arriba) no interacciona con L6.

- o Se ha planteado que las primeras proteínas ribosomales que existieron no necesariamente se encontraban en el mismo lugar que se observa actualmente en un ribosoma moderno y que tales cambios pudieron corresponder a la adición

funcional de otras proteínas ribosomales a lo largo de la evolución¹⁰⁰. Actualmente tanto S18 como L6 no son de las primeras proteínas en unirse al RNA ribosomal durante el proceso de ensamblaje del ribosoma moderno¹⁰¹; sin embargo, dado que son las únicas proteínas que contienen regiones codificadas por tripletes RNY, es posible que la parte ancestral de cada una fungiese como el primer estabilizador de la estructura terciaria de las moléculas primitivas del 16SrRNA y del 23SrRNA, respectivamente.

- o **HELICASAS DE RNA.**

- o Las moléculas de RNA pueden formar estructuras secundarias y terciarias relativamente estables, la función original de las helicasas era justamente participar en el despliegue de esas estructuras, a fin de que los procesos catalíticos e informacionales propios del RNA se llevasen a cabo^{cfr.115}.

- o Las helicasas de RNA de la familia **DEAD/DEAH** son nucleotidiltransferasas, estructuralmente homólogas a las helicasas de DNA y están asociadas con todos los procesos que involucran moléculas de RNA. Contienen de 7 a 8 dominios altamente conservados que definen el núcleo de la proteína. Algunos de los miembros de la familia muestran promiscuidad en los sustratos y pueden funcionar con dúplex de RNA-RNA, RNA-DNA o incluso DNA-DNA, aunque algunas helicasas pueden mostrar especificidad por ciertas secuencias o sustratos. Ala y Gly son los dos típicos aa de las proteínas **DEAD** [mismos que son codificados por tripletes RNY y fueron de los primeros en sintetizarse prebióticamente]. Aparentemente la energía liberada por la hidrólisis de un NTP (como ATP) es indispensable para que la RNA helicasa lleve a cabo su función¹¹⁶.

- o Los dominios por los que está conformada una RNA helicasa son: I) uno los fosfatos de un NTP; Ia) uno sustratos a través de la estructura de azúcares-fosfatos [RNA en este caso]; Ib) uno sustratos, pero usualmente no está presente; II) Una

los fosfatos β y γ a través de Mg^{2+} y coordina la hidrólisis del NTP con una molécula de agua, este dominio toma la forma de **DEAD**, **DEAH** o **DExH** de donde se toma el nombre para cada subgrupo; III) une el fosfato γ y enlaza la hidrólisis del NTP con la actividad de helicasa; IV) Une sustratos [RNA], es semejante a un dominio de las DNA helicasas, V) une sustratos a través de la estructura de azúcares-fosfatos [del RNA] y puede interaccionar con el NTP; VI) une el fosfato γ y transforma la unión del NTP en hidrólisis, con el movimiento de sus dominios 1 y 2^{117,118}. La mayoría de las helicasas **DEAD** están involucradas en la biogénesis del ribosoma para evitar que las moléculas de RNA queden plegadas en trampas cinéticas conformacionales. La direccionalidad de estas proteínas puede ser **5'->3'** o **3'->5'**, pueden actuar de manera no-procesiva disociando menos de 10pb, o de modo procesivo actuando en largas moléculas de RNA de doble cadena. La disponibilidad del ATP y Mg^{2+} pueden regular la actividad general de la helicasa¹¹⁸.

- o Las proteínas **DExH/D** pueden ser meramente helicasas de RNA o bien reconocer directamente las interacciones en un complejo ribonucleoproteico y desestructurarlas (RNPa) ¹¹⁹.

- o Se ha descubierto que hay ausencia de estas RNA helicasas en organismos que viven a altas temperatura y que en general las RNA helicasas son dispensables en la gran mayoría de los procariontes, sobre todo si se trata de organismos parásitos o cultivados en condiciones de laboratorio¹¹⁸, posteriormente se descubrió que las RNA helicasas pueden requerirse para la adaptación celular a condiciones ambientales alteradas de temperatura, luz y osmolaridad, permitiendo la regulación de la expresión de ciertos mRNAs o la actividad de RNAs funcionales sólo bajo tales circunstancias¹²⁰. De hecho es posible que la primer función de las helicasas de la familia **DEAD/DEAH** estuviese relacionada precisamente con la adaptación a las condiciones de estrés ambiental que imperaban en la Tierra primitiva.

o Algunas RNA helicasas catalizan alineamiento de moléculas RNA¹²¹, lo que pudo haber sido muy útil si recordamos que las proteínas codificadas por tripletes RNY se sitúan en el marco de la transición entre el Mundo del RNA y el Mundo de las RNPs. Otras helicasas catalizan la relajación de estructuras secundarias o terciarias de ssRNA o dsRNA¹¹⁹, lo que pudo haber ayudado en su modulación y por tanto la regulación de los procesos que catalizaran tales moléculas; se han descrito algunas **proteínas DEAD que promueven el mecanismo denominado "splicing"** (empalme) de los intrones de grupo I y grupo II¹¹⁹, lo que nuevamente nos remite a un escenario evolutivamente antiguo. Algunas helicasas son en realidad RNPasas¹¹⁹, lo que directamente nos remonta a la transición de un Mundo de RNA puro a uno con coevolución de RNA y proteínas y por tanto puedo sugerir que ambas funciones, las haya llevado a cabo una proteína **DEAD/DEAH** ancestral.

o Se ha establecido que el ATP es hidrolizado por las RNA helicasas, sin embargo es evidente que ninguno de los dominios de una proteína **DEAD/DEAH** es específico para dicha molécula, lo que sugiere que en su inicio posiblemente utilizaran cualquier NTP disponible (ATP o GTP). De manera adicional, recientemente se ha descubierto que: 1) las proteínas **DEAD** pueden separar completamente un dúplex de RNA utilizando una sola molécula de ATP¹²²; 2) la hidrólisis del ATP es requerida para liberar la helicasa (reciclarla), pero no se necesita para el relajamiento del dúplex¹²³ y 3) adicionalmente a las actividades dependientes de ATP, como las de estimulación de alineación de cadenas, desplazamiento de cadenas y actividad como chaperonas de RNA, la familia **DEAD** lleva a cabo funciones independientes de ATP¹²⁴.

o Las RNA helicasas participan activamente en la remodelación de las interacciones RNA-RNA y los complejos RNP durante la biogénesis del ribosoma¹⁴⁶ y es ahí donde precisamente radicaría la explicación de su antigüedad y su conservación evolutiva. Por otro lado en un intento por descubrir el origen evolutivo

de las helicasas **DEAD**, se han alineado masivamente las secuencias genéticas que codifican para dichas proteínas en diversos organismos, como resultado se ha encontrado que el dominio DbpA-RBP, de unión a DNA y RNA, es el dominio ancestral a partir del cual han surgido las RNA helicasas mediante eventos de duplicación y expansión y reducción en diferentes linajes¹²⁵.

- o Como podemos observar los dominios conservados en las helicasas **DEAD** unen partes de moléculas nucleotídicas, ya sea como RNA, NTPs o complejos ribonucleoproteicos en diversas circunstancias, todas esas formas pudieron estar presentes en los albores del Mundo de las RNPs y la acción sobre cualquiera de ellas había conferido ventaja evolutiva al sistema; adicionalmente, el único motivo conservado del que se menciona su secuencia y que además contiene aa codificados por tripletes RNY es **DEAD/DEAH/DExH** así que pudo haber sido éste, o bien el dominio DbpA-RBP o algún otro no indicado pero que une ribonucleótidos, el identificado como ancestral por la metodología descrita en la presente tesis; por todo lo anterior, no es posible discernir cuál es la región codificada por tripletes RNY en dichas helicasas de RNA, aunque ese módulo conformaría el primordio de tales enzimas y probablemente era un sitio de unión a nucleótidos.

- o **PROTEÍNAS MODIFICADORAS DEL RNA.**

- o La adición de un grupo metilo a moléculas biológicamente activas causa cambios en las propiedades fisicoquímicas de éstas al introducirles cargas positivas o bien les induce cambios estructurales. La metilación es catalizada por metiltransferasas (MT) que principalmente unen S-adenosilmetionina (S-AdoMet) y transfieren el grupo metilo a una amplia variedad de sustratos, como nucleósidos de rRNA durante la biogénesis del ribosoma, o bien factores que intervienen en el proceso de traducción, todo lo cual sugiere que posiblemente esas S-AdoMet MT son ancestrales.

o Muchas de las S-AdoMet MT tienen un dominio catalítico muy conservado que probablemente evolucionó a partir de un ancestro común o bien es resultado de una notable convergencia evolutiva. Las fusiones genéticas han añadido dominios de reconocimiento que dieron origen a la variedad de MT¹²⁶. La conservación del sitio catalítico no parece verse reflejado a nivel de secuencia, excepto por dos regiones críticas que no toleran inserciones: una de esas regiones conservadas está localizada en la zona N-terminal de la MT, es rica en Gly y se presume que es aquí donde se une la S-AdoMet¹²⁷, recordemos que la Gly es uno de los aminoácidos prebióticos y está codificado por tripletes RNY; la otra región conservada es un bucle ácido que interacciona con los hidroxilos ribosa^{ld.127}, como reminiscencia del Mundo del RNA.

o Las S-AdoMet MT que se identificaron como ancestrales, pertenecen a la clase I, la que plausiblemente se deriva de las proteínas con plegamiento Rossmann, sin embargo dada la limitada similitud entre sus diferentes miembros, las S-AdoMet MT podrían haber evolucionado a partir de un dominio genérico **GxGxG** de unión a nucleótidos¹²⁸ (localizado en la región N-terminal, mencionada anteriormente).

o En 1998 se planteó una hipótesis en la que se resaltaba la importancia de la modificación de ribonucleósidos en el Mundo del RNA, la hipótesis consistía de tres puntos principales: 1) que las modificaciones se ubicaban principalmente en el centro catalítico de los ribosomas primitivos, 2) que tales modificaciones restringían la capacidad autocatalítica del RNA, pero al mismo tiempo le conferían la estabilidad suficiente para la expresión del código genético primitivo, 3) La auto-organización de la materia orgánica se basó en la interacción de estas macromoléculas (proteína y RNA)¹²⁹. Para el año 2007 quedaba establecido que las modificaciones del RNA no sólo modulan las interacciones RNA-RNA o RNA-proteína durante la biogénesis del ribosoma o durante el proceso de traducción¹⁰⁰, sino que también expanden notablemente el repertorio de nucleótidos, rebasando incluso el número de

aminoácidos distintos¹³⁰, lo que implica una ventaja selectiva en la transición Mundo del RNA al Mundo de las RNPs. Podemos observar que las proteínas de modificación de nucleótidos que se identifican como ancestrales actúan sobre el 16SrRNA en la forma libre de proteínas, *i.e.* se añaden las modificaciones conforme se sintetiza¹⁰⁰; además el hecho de que contengan regiones codificadas por tripletes RNY podría indicar que el mecanismo ha sido el mismo desde la aparición de tales enzimas. Se han sintetizado ribosomas totalmente funcionales que no contienen modificaciones en sus nucleósidos, aunque tienen aproximadamente la mitad de eficiencia en la síntesis de proteínas, ya que se desensamblan con mayor facilidad y los organismos con este tipo de ribosomas son menos resistentes a antibióticos, todo eso ha llevado a sugerir que las modificaciones post-transcripcionales son importantes para estabilizar y mantener la estructura del aparato traduccional, lo que sin duda era importante en un ambiente primitivo donde la presión por selección era significativa¹⁰⁰.

o Las diferentes MT han recibido varios nombres para una misma enzima, dependiendo de los estudios realizados acerca de su función o estructura; por lo cual sería preferible instaurar sólo el nombre que describa de manera más adecuada el mecanismo de determinada proteína, de acuerdo con los últimos avances científicos en aras de la universalización del conocimiento.

- **NOL1/NOP2/sun/fmu/RrmB/RsmB** cataliza específicamente la metilación m⁵C de un nucleósido del 16SrRNA, en la forma libre de proteínas¹³¹; de hecho la RsmB tiene actividad incluso en un pequeño bucle de 15 nt, derivado del 16SrRNA¹³². Dado que RsmB y sus homólogos están ampliamente distribuidos en los 3 grandes dominios, es probable que la secuencia ancestral a partir de la cual derivó este gen haya sido muy antigua¹³³, incluso podemos decir ahora que dicha secuencia se remonta al inicio del Mundo de las RNPs. Algunos análisis teóricos y experimentales

revelaron que las RNA: m⁵C MT comparten varias características tanto con MT RNAm: ⁵U MT como con DNA: m⁵C MT, aunque entre estas no hay características en común, lo que sugiere no sólo que hay una relación filogenética entre las tres 5-metilpirimidina MTs^{134,135}, sino que dado que RsmB contiene regiones codificadas por RNY, es probable que sea la más ancestral, inclusive es probable que la última en aparecer fue la DNA MT dado que es evidente que el DNA fue posterior al RNA.

- **GidB/RsmG** es una proteína esencial¹³⁶ que cataliza la metilación m⁷G de un nucleósido del 16SrRNA¹³⁷, depende directamente de la concentración de Mg²⁺ para su actividad y actúa sobre el 16SrRNA libre de proteínas¹³⁸.
- **HemK/PrmC** es una enzima esencial que cataliza la metilación de la amida plana del residuo Gln del motivo **GGQ** de los factores 1 y 2 de liberación del ribosoma (RRF1 y RRF2, por sus siglas en inglés) para formar N⁵-metilglutamina (Q_M), ambos factores promueven la liberación del peptidil-tRNA, el motivo **GGQ** mimetiza el extremo CCA de un tRNA, pero dado que no es posible realizar un enlace peptidil, se induce la terminación abortiva de la traducción; probablemente el motivo **GGQ_M** juega un papel importante en la activación de una molécula de agua involucrada en la hidrólisis del peptidil-tRNA^{139,140}. Homólogos de PrmC se han observado de manera omnipresente en la mayoría de las especies de los tres dominios de la vida, lo que indica su origen ancestral¹⁴¹, además que el motivo **GGQ** es universal a todos los RFs¹⁴². Dado que ninguno de los dos RRFs mencionados aparecen en los resultados de esta tesis, es posible que PrmC fuese cooptada cuando surgieron ambos factores, posteriormente más dominios fueron añadidos a PrmC y le confirieron especificidad por RRF1 y RRF2.
- **Metiltransferasa**, no se indica la función específica de esta proteína pero es posible que dado que contiene regiones codificadas por tripletes RNY, su

actividad tenga impacto directamente sobre RNA o fragmentos de péptidos que participan en el proceso de traducción.

o Con base en un estudio de comparación exhaustiva de secuencias y estructuras, incluso remotamente relacionadas y de los patrones filéticos generados, se encontró que las proteínas que unen S-AdoMet forman 15 grandes superfamilias, de las cuales al menos 5 se encontraban presentes en el LUCA, como las MT de modificación del rRNA y las de modificación de factores de traducción¹⁴³, 4 de las cuales contienen además regiones codificadas por tripletes RNY; adicionalmente se llegó a la conclusión que los ancestros moleculares de las proteínas que unen S-AdoMet pudieron tener una especificidad más laxa que la que actualmente tienen¹⁴³.

o Podemos observar que, tanto si las S-AdoMet son derivaciones evolutivas del plegamiento Rossmann o si provienen de un dominio **GxGxG**, son de cualquier modo descendientes moleculares de regiones que unen ácidos nucleicos, lo que es consistente con el hecho de contener regiones codificadas por tripletes derivados del Mundo del RNA y particularmente del motivo mencionado es abundante en aa codificados por tripletes RNY.

o **PARTE DEL APARATO DE TRADUCCIÓN: Factor de Elongación G y Factor de Reciclaje del Ribosoma.**

o El ribosoma es una molécula muy antigua así como algunos factores asociados al ribosoma durante el proceso de traducción, como lo indica el hecho de que contengan regiones codificadas por tripletes RNY. Uno de los mencionados elementos es el factor G de elongación de la traducción (EF-G), presente en los tres dominios de la vida, el cual puede interactuar con el factor de reciclaje del ribosoma (RRF) para que, una vez terminada la traducción, el complejo traduccional se desensamble⁹⁵.

- **EF-G** trascuela el ribosoma con relación al mRNA utilizando la energía liberada por hidrólisis de GTP y también participa en el desensamblaje del complejo traduccional. Puesto que comparte características comunes con el EF-Tu, se postula que ambos factores surgieron por duplicación/fusión de un ancestro común¹⁴⁴; sin embargo el gen que codifica para EF-Tu no fue identificado utilizando la metodología presentada en esta tesis, lo que me sugiere que EF-G es de las dos la proteína más antigua, siendo el motivo de unión o al RNA o partes del GTP la región ancestral^{cfr.145}. Se propone que el complejo EF-G-GTP mimetiza un aminoacil-tRNA¹⁴⁶ y que la hidrólisis no se requiere para la trascolación sino para liberar el EF-G posteriormente; la trascolación ocurre por el re-arreglo de las dos subunidades ribosomales, cambio inducido por la unión del EF-G^{147,148,149}. El EF-G es una GTPasa universalmente conservada en los tres dominios de la vida¹⁵⁰, lo que significa que muy probablemente estaba presente en el LUCA y el hecho de que el gen que la codifica contiene tripletes RNY, es indicador que al menos esa región es tan antigua como el inicio del Mundo de las RNPs, de hecho sería la GTPasa más ancestral y localizada en el aparato traduccional.
- **RRF** es una proteína esencial para la vida bacteriana y, en coordinación con EF-G, cataliza el desensamblaje del complejo ribosomal impidiendo que se reinicie la síntesis proteica, lo que también podría ocurrir si la concentración de Mg^{2+} es baja; una segunda función de RRF es la prevención de errores durante la traducción, *i.e.* está involucrado en la elongación del péptido¹⁵¹. El RRF semeja la forma y funciones del tRNA, sin embargo es más flexible, lo que permite inducir cambios en la conformación estructural que favorecen el desensamblaje del ribosoma^{152,153,154}. El RRF contiene algunos residuos altamente conservados a lo largo de la evolución¹⁵⁵ y alguno de ellos está probablemente codificado por tripletes RNY. Se ha mostrado que

el RRF puede participar en el rescate de ribosomas que hayan quedado inactivados durante alguna fase estacionaria o estrés ambiental¹⁵⁵, todo lo anterior significa que las funciones de esta enzima son tan amplias que representaron una ventaja selectiva en un mundo ancestral correspondiente a la transición entre Mundo del RNA y Mundo de las RNPs.

o Las acciones del EF-G, tanto en la traslación como en el reciclaje del ribosoma, son controladas por el péptido que se encuentra en el tRNA en el sitio P, en ambos procesos el EF-G-GTP empuja el peptidil-tRNA del sitio A/A al sitio P/P (con el consecuente movimiento del tRNA desacetilado del sitio P/P a la salida). Durante la terminación, el RRF ocupa la posición A/A, lo que provoca cambios estructurales y la consecuente disociación de las subunidades ribosomales^{156,157}; RRF es una proteína esencial que induce cambios estructurales en el ribosoma, lo que estimula directamente el reciclaje del complejo ribosomal en coordinación con EF-G, probablemente mediante una interacción directa a través de los dominios que mimetizan el brazo anticodón del tRNA en ambas moléculas^{158,159}. Se ha mostrado experimentalmente que hay ribozimas que catalizan la síntesis del enlace peptidil^{23,24} y esto probablemente era funcional durante el Mundo del RNA, sin embargo la transición hacia el Mundo de las RNPs posiblemente fue encabezada, o incluso llevada a cabo únicamente, por EF-G y RRF pues son las únicas moléculas de este mecanismo moderno que contienen regiones codificadas por tripletes RNY.

o **AMINOACIL-tRNA SINTETASAS.**

o Las aminoacil-tRNA sintetetasas (aaRS) son enzimas que acoplan aminoácidos con sus tRNA correspondientes, utilizando la energía derivada de la hidrólisis del ATP. Debido a la importancia medular de las aaRS en el mantenimiento de la precisión del código genético, es muy probable que sean de las proteínas más

antiguas, pues significan la conexión entre el Mundo del RNA y el Mundo de las proteínas, es decir que dan la pauta para el comienzo del Mundo de las RNPs¹⁶⁰.

o Las aaRS se han dividido en dos clases, que a su vez se dividen cada una en tres subclases. Ambas clases se han distinguido entre sí con base en diversos factores como: modo de unión al ATP, reconocimiento del tRNA, modo de activación del tRNA (mediante la adición del aa) y organización por módulos y cooperatividad entre ellos, con base en todo lo cual se ha teorizado acerca del origen y evolución de las aaRS¹⁶¹. Habiéndose establecido que el surgimiento de las aaRS es previo a la divergencia de procariontes y eucariontes, el árbol filogenético canónico de especies se trató de enraizar utilizando únicamente 3 aaRS de la clase Ia, demostrándose claramente que Arqueas y Eucariontes son dominios hermanos¹⁶². Recientemente se han llevado a cabo análisis filogenéticos de las aaRS, por ejemplo con base en el tipo de tripletes que codifican para los aminoácidos que corresponden a los tRNA que reconocen determinados grupos de aaRS, la conclusión de tal análisis es que el árbol filogenético canónico no puede reconstruirse con base en cada una de ellas¹⁶³, posiblemente porque se ignoró la naturaleza modular de estas enzimas y que sus módulos pueden transferirse horizontalmente (ver más adelante).

o El reconocimiento del tRNA y el acoplamiento del aa son llevados a cabo por diferentes dominios, lo que puede reflejar la evolución gradual modular de la familia de las aaRS, de las pudieron haber dos diferentes ancestros desde el origen correspondientes a la resolución de cada uno de los problemas de aminoacilación de los tRNA, ya **sea en 2'OH o en 3'OH**. La complejidad incrementó paulatinamente mediante la adición de dominios que cooperaban recíprocamente con los ya existentes; después del primer dominio de unión a aminoácidos, surge el dominio capaz de unir el anticodón del tRNA, posteriormente dominios capaces de proveer energía (hidrólisis de ATP), especificidad y regulación. El mecanismo más simple para obtener dos clases de aaRS derivadas de un mundo ancestral, es a través de

un código binario que necesita únicamente dos sitios diferentes en un RNA, dando lugar a dos diferentes centros quirales para la aminoacilación e imponiendo entonces dos diferentes centros de reacción^{161,164}.

o Aparentemente el sitio activo y el mecanismo de aminoacilación es lo único químicamente similar en todas las aaRS sin importar la clase a la que pertenecen¹⁶⁵, por lo que probablemente es la región más antigua. Ambas clases de aaRS llevan a cabo el mismo mecanismo de aminoacilación y requieren de un ion Mg^{2+} para llevar a cabo la catálisis, el cual es aparentemente necesario para estabilizar la estructura del tRNA, se ha probado que algunos oligorribonucleótidos sintetizados químicamente y conformados en mini-hélices, son reconocidos por las aminoacil-tRNA sintetasas de igual modo que los tRNA correspondientes¹⁶⁶, todo lo anterior podría indicar que en el origen del proceso de aminoacilación, tanto las aaRS como los tRNAs eran menos complejos y los iones metálicos eran fundamentales, máxime si en verdad que el origen del módulo de energía (hidrólisis del ATP) es evolutivamente más reciente que el sitio catalítico.

o Las aaRS contienen algunas zonas, denominadas determinantes o anti-determinantes, que sirven para asegurar unívocamente la unión del tRNA cognado (afín). La mayoría de las reglas de identidad están conservadas evolutivamente y los principales determinantes y anti-determinantes sirven para identificar ***grosso modo*** los tRNAs de cada clase¹⁶⁷; sin embargo el análisis de la comparación de los módulos de las aaRS ha arrojado que la mayoría de las especificidades han emergido a partir de sinapomorfias recientes (*i.e.* caracteres que se originaron en el ancestro común de un grupo y por lo tanto son por lo tanto compartidos por todos los miembros incluido ese ancestro común), por lo que el origen de cada módulo puede ubicarse en distintos periodos evolutivos y la mayoría pueden transferirse horizontalmente aún entre algunas especies¹⁶⁸; más aún el hallazgo mencionado puede indicar que el reconocimiento preciso de los tRNA pudo no

ocurrir tan eficientemente como hoy día durante el incipiente Mundo de las RNPs, por lo que posiblemente más de un tRNA era cargado por más de una aaRS.

o Actualmente queda claro que existen diferentes rutas de aminoacilación, en las que la relación aaRS-tRNA no es necesariamente unívoca y se corrige mediante mecanismos de edición, ya sea en la misma enzima o con la ayuda de proteínas adicionales¹⁶⁹, dichos mecanismos pudieron ser los que existían de manera ancestral debido a que es posible que en un inicio las aaRS fuesen menos complejas y específicas que en la actualidad^{cf.160} y sólo reconociesen estructuras generales en los tRNAs; de hecho es común en las bacterias que los aminoácidos de estructura y/o características fisicoquímicas similares sean cargados en el mismo tRNA por las aaRS y posteriormente sean modificados^{170,171}; posteriormente las sintetasas adquirieron otros módulos y se incorporaron tanto aminoácidos diferentes como enzimas que se volvieron más precisas^{169,172,173}. La falta inicial de especificidad y las mutaciones de las aaRS pudieron contribuir a la redundancia del código genético, permitiendo diversas adaptaciones a nichos ecológicos particulares; las coadaptaciones de los contactos aaRS-tRNA debieron haber ocurrido a través de la evolución, respetando el incipiente código genético. Actualmente existen grupos de reconocimiento tanto en el tRNA como en las aaRS mediante los cuales se reconocen con precisión ambas moléculas entre sí^{174,175}.

o Con base en el análisis de alineamientos estructurales de las aaRS, se ha descrito lo que podría constituir la región catalítica primordial de cada una de las dos clases; además se han definido dos nuevas subclases para la clase I, esto al parecer agrupa de manera más adecuada estas enzimas y ayuda a reflejar de mejor manera el origen y evolución de la familia completa de las aminoacil-tRNA sintetasas¹⁷⁶.

o Las aaRS para valina (ValRS), treonina (ThrRS) y asparagina (AsnRS) son las únicas que fueron identificadas como ancestrales, lo que concuerda con que los aminoácidos V, T y N son codificados por tripletes del tipo RNY.

- **ThrRS.** Es actualmente una aaRS homodimérica perteneciente a la clase IIa y es la única sintetasa que controla la expresión de su propio gen mediante retroalimentación negativa^{177,178}. La ThrRS se une a su propio mRNA¹⁷⁹ mediante las mismas reglas de identidad que el tRNA cognado¹⁸⁰ y compete con el ribosoma por impedimento estérico¹⁸¹, pero cuando el tRNA aparece, la enzima se une a este último permitiendo la traducción del mensajero¹⁸². El doble control represión/des-represión permite que la tasa de síntesis de la ThrRS se ajuste de manera precisa al nivel del tRNA correspondiente en la célula. Por otro lado, los niveles de mRNA, tRNA y ribosoma dependen a su vez de la tasa de crecimiento del organismo¹⁸³. Se puede ver entonces que la traducción del mRNA del gen *thrS* es regulada por dos diferentes mecanismos de competencia que actúan de manera coordinada provocando diferentes cambios en su conformación¹⁸⁴. Al momento de la activación del tRNA^{Thr} por el aa correspondiente, la ThrRS moderna es capaz de discriminar treonina sobre valina, que es químicamente similar y sobre serina, que es de tamaño equivalente¹⁸⁵; contra valina se aplica una selección negativa mediada por un ion Zn^{2+} ^{186,187} y para evitar la incorporación de serina al mensajero, el tRNA con carga incorrecta es hidrolizado¹⁸⁸ o bien su tasa de transferencia disminuye¹⁸⁹. Lo anterior concuerda con el hecho que mientras valina era un aa que podía ser reconocido por su propia sintetasa en el mundo ancestral, serina era un aa codificado por un triplete RNY para el que sin embargo no había una sintetasa cognado específica, de modo que es posible que una ThrRS ancestral fuese capaz de cargar los tres aa (Ser, Val y Thr), ya que poseen

características físicas y químicas similares. Al igual que todas al aaRS, la ThrRS puede ser vista como consecuencia de la adquisición de diferentes módulos funcionales¹⁹⁰, donde el dominio catalítico es la parte considerada como ancestral; incluso se propone que las regiones regulatoria y de edición en la enzima actual han sido una adición evolutiva reciente, sin perder el segmento más ancestral que era únicamente aquel del reconocimiento del tRNA^{Thr}¹⁹¹, reconocimiento que incluso pudo haber sido más laxo que en la actualidad.

- **ValRS.** Es un monómero perteneciente a la clase Ia codificado por el gen *valS*^{192,193}, cuya unión con ATP provoca cambios conformacionales en la enzima¹⁹⁴. El suero que contiene ValRS puede provocar reacción cruzada con IleRS¹⁹⁵ y ambas sintetetas presentan alta homología incluso entre especies lejanamente relacionadas¹⁹⁶, lo que apoya de manera substancial la hipótesis que ambas enzimas (ValRS e IleRS), en conjunto con LeuRS y MetRS, provienen de un ancestro común que no presentaba el dominio de unión a iones metálicos como Zn²⁺¹⁹⁷, que usualmente se asocia al reconocimiento específico de una aaRS con su aa correspondiente. A nivel de codones, los tripletes que codifican para Val e Ile siguen la estructura RNY y se ha descubierto que si solamente la primera y/o tercera base son mutadas los codones pueden codificar para Met o Phe¹⁹⁸. Podemos observar que tanto Val como Ile eran aminoácidos codificados por tripletes RNY y una ValRS primitiva podía activar tRNAs con ambos aminoácidos e incluso puedo plantear que existían más aminoácidos en un ambiente primitivo, que si bien no tenían correspondencia en el código genético primitivo sí podían ser incorporados a los mensajeros por aaRSs con laxo reconocimiento por su molécula cognada; como refuerzo experimental de la propuesta anterior, se ha encontrado que las aaRS para Val e Ile pueden

activar el tRNA correspondiente a la otra, en ausencia del cognado¹⁹⁹. Adicionalmente se ha encontrado que la secuencia CCA terminal del tRNA^{Val} previene que sea activado con Thr²⁰⁰, aunque quizá los tRNA primitivos no poseían estrictamente sólo esta secuencia²⁰¹ y eran cargados con otros aminoácidos, de modo que aun cuando había una enzima para Thr en los albores del Mundo de las RNPs (ver aaRS anterior), la relación no era unívoca y el aa podía ser cargado por otras sintetetasas, quizá reflejando su importancia en el contexto primitivo mencionado. Se ha encontrado incluso que la ValRS es capaz de activar tRNAs, no solamente con Ile y Thr, sino también con α -aminobutirato y con cisteína sin diferencias en la tasa de trascolación de la enzima²⁰², lo que podría reflejar el proceso de aminoacilación en los albores del Mundo de las RNPs. Actualmente existen determinantes y anti-determinantes²⁰³, tanto en el tRNA^{Val} (además del contexto en que se encuentra la secuencia CCA)²⁰⁴ como en la ValRS²⁰⁵ que permiten el doble filtrado de valina por sobre isoleucina y treonina²⁰⁶, así como sitios de edición en la sintetetasa que se localizan lejos del centro catalítico²⁰⁷ y comparten homología con los dominios de edición de IleRS y LeuRS²⁰⁸, lo que puede indicar que las tres sintetetasas comparten un ancestro común cuyo origen fue posiblemente la ValRS ancestral que podía unir tanto Val, Ile y Thr, probablemente incluso al mismo tiempo, ya que se ha mostrado que la ValRS contiene dos sitios funcionales de unión a tRNA²⁰⁹. La reacción de edición de la ValRS comprende dos vías distintas denominadas pre-transferencia y post-transferencia²¹⁰, que posiblemente no son ancestrales.

- **AsnRS.** Es una sintetetasa codificada por el gen *asnS* que pertenece a la clase IIb²¹¹. En algunos lactobacilos, el complejo Asn:tRNA es formado al sustituir directamente el aspartato cargado en el tRNA^{212,213}; en algunos

otros organismos, que carecen tanto de AsnRS como de GlnRS, una sola glutamina amidotransferasa dual es responsable de la formación de Asn:tRNA y Gln:tRNA a partir de Asp:tRNA y Glu:tRNA, respectivamente^{214,215}; algunos extremófilos obtienen la asparagina del complejo Asn:tRNA formado a partir de la transamidación del complejo Asp:tRNA, aparentemente numerosos organismos obtienen asparagina sólo de esta manera²¹⁶; para algunas otras bacterias la biosíntesis del Gln:tRNA es de forma directa, mientras que la formación de Asn:tRNA se lleva a cabo de manera indirecta²¹⁷; se ha determinado que existen pares de bases específicos en el tRNA^{Asn} activado con aspartato (y en el tRNA^{Gln} activado con glutamato) involucrados directamente en la transamidación del aa²¹⁸; como podemos observar existen diversas vías para la síntesis y utilización del complejo Asn:tRNA^{Asn}, lo cual parece indicar que la asparagina era de fundamental importancia además de ser un aa codificado por un triplete RNY, posiblemente la AsnRS era capaz de cargar otros aminoácidos además de asparagina. Desde 1950 se determinó que asparagina=aspartato+NH₄⁺²¹⁹; de manera que aunque no existiese una aaRS específica para Asp, dado que es un aa codificado por tripletes RNY, éste pudo haber sido incorporado a las proteínas mediante una aaRS ancestral, como una AsnRS con baja especificidad, que pudiera reconocerlo, pues **D** y **N** son aa químicamente similares y de este modo no se estarían violando las hipótesis más aceptadas en las que ambos aminoácidos son prebióticos y de los primeros en ser incorporados a oligopéptidos funcionales^{74,75}. De hecho se ha mostrado que, a nivel de secuencia, la AsnRS comparte homología significativa con AspRS aún entre especies muy divergentes, lo que indica una relación evolutiva entre ambas enzimas²²⁰; también se ha mostrado que AsnRS comparte homología a nivel de secuencia con LysRS²²¹, por lo

que es posible que la AsnRS ancestral pudiese cargar Asn, Asp y Lys (aún si este último aa no es codificado por tripletes RNY); es posible incluso que dicha AsnRS primitiva diese origen a las tres enzimas de la clase IIb, de hecho la AsnRS y la AspRS son estructuralmente idénticas, excepto por tres diferencias clave que les proveen del modo preferencial de interacción con su aa respectivo; también se ha observado que la unión del ATP provoca cambios muy similares en la configuración estructural de ambas sintetasas, todo lo cual apunta nuevamente a un origen evolutivo común²²². Se ha encontrado que dos moléculas de agua podrían estar involucradas directamente en el estricto reconocimiento de la asparagina y discriminación contra el aspartato por parte de la AsnRS, proceso que no ocurre en la AspRS²²³; posiblemente durante los orígenes del Mundo de las RNPs, el agua si rodeaba las incipientes enzimas, sin embargo éstas no proveían el soporte estructural necesario y por tanto en el caso específico de la AsnRS ancestral el reconocimiento del aa era más laxo que en la actualidad, reconociendo tanto Asn como Asp y posiblemente incluso Lys.

o Podría parecer inesperado el hecho que las aaRS para los otros aminoácidos codificados por tripletes RNY no se identificaron como ancestrales, sin embargo anteriormente vimos varios ejemplos con base en los cuales podemos inferir que cuando las sintetasas no son del todo específicas, como pudo haber ocurrido durante los inicios de la vida, son capaces de reconocer aa con atributos similares. Varios escenarios evolutivos se han propuesto con respecto al origen y evolución de las aaRS; se ha observado experimentalmente que algunas ribozimas son capaces de acoplar aminoácidos a moléculas de tRNAs y se postula que durante el Mundo del RNA, el proceso de aminoacilación se siguió de este modo. Posterior a la riboaminoacilación, en los albores del Mundo de las RNPs, surgieron dos plegamientos de oligopéptidos distintos, derivados posiblemente de la decodificación

de las dos hebras de una molécula de RNA bicatenario; de los dos plegamientos surgidos se formaron dos clases distintas de aaRS, cada vez más especializadas mediante la adición paulatina de módulos funcionales, concomitantemente las aaRS se diversificaron mediante duplicaciones y mutaciones^{Cfr.160,161,224}; actualmente se ha observado que el reconocimiento y activación del tRNA, así como el sitio catalítico de las moléculas de cada clase de aaRS, son imágenes especulares y por tanto las estructuras de las aaRS de cada clase serían el opuesto complementario una de la otra^{véase188}.

o De acuerdo a varias propuestas teóricas, la región de reconocimiento del aa es la más ancestral, sin embargo me es posible conjeturar que el reconocimiento era laxo y una misma enzima era capaz de reconocer varios aminoácidos, en particular las tres aaRS existentes posiblemente podían cargar los ocho aa codificados por tripletes RNY y quizá otros aminoácidos que, si bien no eran codificados por tripletes derivados del Mundo del RNA, si se encontraban presentes en el medio. En resumen, posiblemente los aa incorporados a las proteínas durante los albores del Mundo de las RNPs fueron: **G** y **A**, que pudieran haberse polimerizado mediante ribozimas^{cfr.24}; **V**, **T** y **N**, para los que existían las aaRS rudimentarias; algunos otros aminoácidos pudieron ser cargados por las aaRS primitivas de baja especificidad, como Ile **I**, Asp **D** y Ser **S** que además son codificados por tripletes RNY, además de E y K que quizá son prebióticos.

o **PROTEÍNA DE LA FAMILIA DE PROCESAMIENTO appr-1-p.**

o La proteína appr-1"-pasa es una enzima eucarionte involucrada en el procesamiento del tRNA, la appr-1"-pasa cataliza la conversión de ADP-ribosa-1"-monofosfato (Appr-1'-p) a ADP-ribosa; la molécula posee una cavidad que semeja un plegamiento Rossmann, de unión a nucleótidos, lo que no es de sorprender ya que es una molécula que interacciona con ADP-ribosa, además de dos iones sulfato,

un ion sodio y una molécula de etilenglicol, así como 4 moléculas de agua que son desplazadas cuando se une la molécula de ADP-ribosa.

- o En bacterias y arqueas existen proteínas ubicuas que se incluyen en la misma familia de la appr-1"-pasa pues contienen el dominio característico funcionalmente relevante de dicha molécula²²⁵, así como en algunos virus de RNA de cadena positiva^{226,227} y aunque parece ser una enzima dispensable, se sugiere que pudiera estar involucrada en la regulación de la replicación viral o de las interacciones del virus con su hospedero²²⁸.

- o El dominio *macro* es un módulo común de todas las proteínas que unen ADP-ribosa, lo que sugiere entonces que dicho dominio está involucrado en la homeostasis del NAD⁺ y sus metabolitos celulares importantes, así como diferentes metabolitos de la ADP-ribosa^{229,230}; el dominio *macro* es *per se* de origen antiguo, codificado en el genoma de muchos procariontes, de donde se supone que los virus lo secuestraron ya que este dominio es capaz de unir nucleótidos²³⁰.

- o Parte de las moléculas NAD⁺ y NADP⁺ son sintetizadas prebióticamente y si además en el metabolismo primigenio, éstas se utilizaban como moléculas energéticas fundamentales, es de esperarse que una proteína involucrada en la homeostasia o sencillamente en la unión de alguno de esos cofactores contenga regiones codificadas por tripletes RNY^{cfr.231}, por otro lado la propuesta del secuestro del dominio *macro* por parte de los virus, se ajusta diversas propuestas que se han hecho con respecto a la convivencia simultánea de partículas virales y de entidades de tipo celular en un mismo tiempo evolutivo.

- o **POLIMERASA POLI-A.**

- o La polimerasa de poliadeninas (PoliAPol) es una nucleotidiltransferasa (NT) que añade adeninas (poliAs), sin un templado, **en el extremo 3' de una cadena de RNA**, usando ATP como sustrato. PoliAPol pertenece a la familia 1 del grupo 3 de la

superfamilia de la DNA polimerasa β (Pol β), dado que contiene el dominio mínimo de las NT (MNT), el cual se propone como el sitio activo común y el más antiguo; dicho MNT contiene una región rica en **G** y otra con un motivo **DxD**; la distribución filética aparente de la superfamilia Pol β sugiere que esas NT desempeñaron un papel muy importante en la utilización de los nucleótidos en la evolución temprana y fueron reclutadas para participar en diversos procesos a través de la acreción de dominios adicionales que les confirieron diversas arquitecturas y distintas actividades enzimáticas y regulatorias, entre estas la polimerización de nucleótidos²³²; por todo lo anterior, es probable que el mencionado MNT sea codificado por tripletes RNY.

- o PoliAPol contiene cinco **Asp** altamente conservados en el sitio activo putativo, así como una **Gly** estrictamente conservada cerca de la región N-terminal; esta región N-terminal es homóloga a la enzima que añade CCA al extremo 3' de los tRNAs, lo que no sorprende porque que la actividad de ambas enzimas consiste en **polimerizar extensiones de nucleótidos en el extremo 3' de moléculas de RNA**. El dominio C-terminal es el sitio de unión a RNA y a otros factores, este sitio se encuentra en una región no conservada entre las NT, lo que sugiere que la función de las NT se determina por los elementos de especificidad al sustrato, localizados fuera del sitio catalítico conservado²³³.

- o Se ha identificado que PoliAPol puede localizarse tanto en membrana como en citosol, acoplándose de este modo al proceso de la síntesis proteica y la degradación del mRNA²³⁴. PoliAPol desempeña un papel importante en la regulación post-transcripcional del metabolismo del RNA pues la adición de adeninas en el extremo **3' de un mRNA lo marca para su degradación por nucleasas 3', al mismo tiempo que esa cola de PoliAs evita que las nucleasas 3' degraden la parte informacional antes de que se lleve a cabo la traducción**; adicionalmente parece ser que la poliadenilación del RNA justamente regula la expresión de genes de dichas ribonucleasas. También se ha encontrado que PoliAPol interactúa con helicasas de la

familia **DEAD**^{233,235} (mismas que también contiene regiones codificadas por tripletes RNY, ver más arriba).

o En los procariontes, sólo una pequeña fracción de las moléculas ribonucleicas tienen colas de poli(A) y por tanto no son marcadas para su degradación por este medio, pues la actividad de la enzima es inhibida por una estructura tipo horquilla **que se forma en el extremo 3' de la mayoría de los RNA, sin embargo una pequeña adición de nucleótidos en el extremo 3' de esta estructura permite la poliadenilación** de manera eficiente²³⁵; el mecanismo regulatorio de la poliadenilación mencionado fue probablemente eficiente en un mundo ancestral en el que existían sólo algunas enzimas (Mundo de las RNPs), ya que con unas cuantas modificaciones el RNA era degradado o protegido.

o Es importante señalar que aunque PoliAPol tiene mayor preferencia por ATP como sustrato, pudiera ser capaz de polimerizar otros NTPs en el orden ATP>CTP>UTP>>GTP²³⁶, lo que pudo implicar ventaja selectiva en un ambiente primitivo donde los recursos eran limitados, aunque en la práctica PoliAPol sólo forma ribo-homopolímeros poli(A), poli(C) y poli(U), aunque no poli(G)²³⁴.

o Se ha observado que la especificidad al sustrato de las RNA polimerasas independientes de templado, queda restringida cuando ocurren cambios estructurales en el dominio catalítico y/o en regiones externas²³⁷, lo que podría dar sustento a la propuesta de que al dominio mínimo de las nucleotidiltransferasas (MNT) se le añadieron dominios que le confirieron especificidades y diversas actividades catalíticas.

o Una posibilidad es que el mecanismo de adición de adeninas fuese introducido viralmente para proteger el material genético de las nucleasas de la célula y dar tiempo para reclutar la maquinaria traduccional del nuevo hospedero^{cfr.25,26}. El mecanismo de la poliadenilación es además independiente de la transcripción y no requiere de un templado para su funcionamiento, por tanto es factible que su origen

se remonte a los inicios del Mundo de las RNPs. Adicionalmente, las investigaciones en química prebiótica indican que la adenina fue la primera base nitrogenada que apareció y, dado que la síntesis biológica de esa base púrica tiene un bajo costo energético^{cfr.23}, el añadir adeninas supone una gran ventaja frente al riesgo de perder información. Por otro lado es posible que, dado que la enzima de adición de CCA no fue identificada como ancestral, la región primordial de poliApol (de acuerdo con la metodología presentada en esta tesis) corresponda a la región homóloga entre ambas proteínas que además corresponde con el dominio MNT, pues esa zona es abundante en aa codificados por tripletes RNY; ese primordio habría surgido entonces en los albores del Mundo de las RNPs, e inicialmente pudo evolucionar hasta una enzima que añadía adeninas sin especificidad a diversas hebras de RNA incluyendo proto-tRNAs, más adelante generaría la poliApol moderna al combinarse con otros módulos.

- o **DNA PRIMASA.**

- o La clasificación COGs coloca la DNA primasa entre las proteínas relacionadas con el procesamiento del DNA, porque su función enzimática es importante para el inicio de la replicación del mismo; sin embargo he decidido incluirla dentro del grupo del procesamiento del RNA, ya que de acuerdo con sus funciones es una RNA polimerasa y de hecho es posible que sea la proteína más antigua de su tipo (seguir leyendo).

- o La DNA primasa es una NT que inicia la síntesis nucleotídica en DNA monocatenario²³⁸, que pudiera considerarse análogo de la molécula de RNA que había en el mundo primigenio. **La DNA primasa sintetiza cebadores o "primers" de oligorribonucleótidos**, requeridos tanto para iniciar repetidamente los fragmentos de Okasaki para la hebra rezagada, como para iniciar una sola vez la síntesis de la

hebra líder, donde en ambos casos el paso limitante en la reacción es la formación del primer enlace fosfodiéster²³⁸.

o Los motivos conservados en las DNA primasas contienen **Gly** y **Asp** principalmente^{238,239}, los motivos conservados son²⁴⁰: 1) al menos un motivo de unión a Zn^{2+} , que es la zona más altamente conservada en la región N-terminal; 2) al menos tres motivos de unión a Mg^{2+} , uno de los cuales semeja el motivo para **"magnesio activado" encontrado en todas las RNA y DNA polimerasas²³⁸**, lo que parece apoyar la propuesta que la antigüedad de la DNA primasa se remonta a los inicios del Mundo de las RNPs; **el motivo de "magnesio activado" puede unirse por especificidad de secuencia al ssDNA en el triplete d(CTG) la mayoría de las veces, aunque otros tripletes como d(TTG) son también reconocidos²³⁸** (podemos observar que el reverso complementario de cada uno de ellos es del tipo RNY); 3) el motivo RNAP está localizado en la región central y es característico de todas las RNA polimerasas, no así de las DNA polimerasas²³⁸; parte de esta región contiene un **plegamiento denominado "toprim"²⁴⁰**, que es estructuralmente similar al dominio central de las DNA topoisomerasas²⁴¹, lo que podría indicar que el dominio toprim en la DNA primasa es el dominio ancestral común a ambas proteínas y por tanto dio origen por un lado a la DNA primasa y por otro lado a la DNA topoisomerasa (ambas contienen una región codificada por tripletes RNY, ver DNA topoisomerasa en 'DNA, estructura y modificaciones').

o **En el complejo denominado "replisoma"**, la DNA primasa y la DNA helicasa interaccionan íntimamente. Recientemente se ha propuesto que ambos genes provienen de la duplicación de un gen ancestral; dicho gen ancestral habría codificado para una proteína bifuncional, con la DNA primasa en el N-terminal y la DNA helicasa en la parte C-terminal, así como una región de enlace entre ambas; la duplicación de esta región de unión, seguida por una separación y divergencia, dio como resultado dos proteínas separadas²⁴². De acuerdo con los presentes

resultados, es posible que la región más antigua de tal gen ancestral, *i.e.* que puede ser trazada hasta el inicio del Mundo de las RNPs, haya sido la correspondiente a la DNA primasa, en particular el dominio toprim²⁴¹, posteriormente hubo una duplicación sin separación, con lo cual surge la función de DNA helicasa hacia la parte C-terminal y posteriormente quizá prosiguió el escenario planteado.

- o Las DNA primasas de procariontes, bacteriovirus y plásmidos comparten el motivo característico toprim²⁴⁰, lo que podría indicar que, no sólo que es quizá la parte más antigua de la DNA primasa, sino que ello sustenta los planteamientos que mencionan que ambos sistemas coexistieron ancestralmente^{25,26}, aunque sin clarificar aún cuál de ellos es más antiguo.

- o Los cebadores tienen usualmente una longitud promedio de 11 ± 1 nt, sin embargo la DNA primasa es capaz de sintetizar cadenas desde 2nt y posteriormente múltiplos de 10nt hasta 60nt *in vitro*²⁴³, lo que podría indicar que existían cadenas nucleotídicas de diversa longitud al inicio del Mundo de las RNPs.

- o Uno de los puntos a favor de la existencia de la DNA primasa como la primera RNA polimerasa es que las concentraciones de NTPs son más altas que las de dNTPs en un ambiente fisiológico²⁴³, es de esperarse entonces que la DNA primasa haya funcionado como la primera proteína capaz de polimerizar los únicos nucleósidos existentes en el Mundo del RNA.

- o La DNA primasa sintetiza el cebador y posteriormente los transfiere a la DNA polimerasa²⁴³, posiblemente reflejando la cronología evolutiva del proceso, *i.e.* la primasa surge en los albores del Mundo de las RNPs sintetizando oligorribonucleótidos, sin embargo es desplazada como molécula principal, por la DNA polimerasa, para la síntesis de las nuevas cadenas nucleotídicas en el mundo DNA+proteínas.

- o La DNA primasa es la polimerasa más lenta y con la mayor propensión a introducir errores (error prone)^{238,243}, sin embargo discrimina notablemente las

bases nucleotídicas no-naturales, de hecho los únicos NTPs que las primasas pueden polimerizar son los que forman puentes de hidrógeno tipo Watson-Crick con las bases del templado, incluso estando en las formas tautoméricas²⁴³. Todas esas características hacen de la DNA primasa una polimerasa eficiente con respecto a la inclusión sólo de NTPs adecuados para la transmisión de la información genética; al mismo tiempo la introducción de errores es reminiscente de un ambiente ancestral, en que la baja fidelidad en los procesos de polimerización favorecía la aparición de nuevos caracteres que servirían como materia prima a la selección natural.

- o **AUSENCIA DE RNA POLIMERASA.**

- o El hecho que la RNA polimerasa no contenga aparentemente una región codificada por tripletes RNY podría parecer ilógico en primera instancia, sin embargo puedo plantear algunas posibilidades: la primera posibilidad es que el RNA dominaba *per se* como molécula autocatalítica y de almacenaje de la información, de modo que habría sido redundante una enzima que lo replicase, de hecho las incipientes proteínas pudieron no llevar a cabo reacciones catalíticas sino fungir como elementos estabilizadores de moléculas prebióticas; la segunda posibilidad es que una DNA primasa semejante a la moderna pudo haber sido la primera RNA polimerasa, que comenzó a coexistir con el RNA como molécula catalítica de la replicación; la tercera posibilidad, apuntaría hacia la veracidad del Mundo Viral^{cfr. 25, 26} en que una DNA polimerasa viral comenzó replicando las moléculas con muy baja fidelidad, teniendo como sustrato lo mismo el DNA de los virus que el RNA de los organismos celulares en los que fue introducida; la cuarta posibilidad apuntaría a que la metodología presentada en esta tesis no es lo suficientemente sensitiva como para captar la región codificada por tripletes RNY de cada una de las proteínas.

T5. Características plausibles de la probable región codificada por tripletes RNY en las proteínas identificadas del procesamiento del RNA.

Conjunto	Proteína (nombre traducido y adaptado)	Características plausibles de la probable región codificada por tripletes RNY.
Proteínas ribosomales	Proteína ribosomal S18	Interacción con RNA, posiblemente mRNA o 16SrRNA
	Proteína ribosomal L6	Interacción con RNA, posiblemente 23SrRNA
Helicasas de RNA	Helicasa de RNA, dependiente de ATP, familia DEAD/DEAH	Interacción con ATP, o Mg2+, o ATP-Mg2+
Proteínas modificadoras del RNA	Metilasa putativa de RNA, familia NOL1/NOP2/sun	No precisada, pero posiblemente unión de S-AdoMet
	Proteína B de la división inhibida por glucosa (GidB)	No precisada, posiblemente unión de S-AdoMet o de 16SrRNA
	Metilasa de modification, familia HemK	No precisada, posiblemente unión de S-AdoMet o del motivo GGQ
	Dominio de una proteína metiltransferasa	No precisada, posiblemente unión de S-AdoMet
Parte del aparato de traducción	Factor G de elongacion de la traducción (EF-G)	Interacción con GTP o unión de RNA.
	Factor de reciclaje del ribosoma (RRF)	Interacción con RNA
Aminoacil-tRNA sintetetas (aaRS)	Treonil-tRNA sintetasa	Reconocimiento de tRNA y/o unión de S/V/T
	Valil-tRNA sintetasa	Reconocimiento de tRNA y/o unión de V/I/T
	Asparaquinil-tRNA sintetasa	Reconocimiento de tRNA y/o unión de N/D/E/K
	Proteína de la familia appr-1-p	Interacción con NAD+
	Polimerasa políA	Unión de NTPs, o interacción con RNA o con helicasas DEAD
	DNA primasa	Unión de cationes Mg2+ y/o Zn2+, o unión de nt

6.3.1.2 DNA, estructura y modificaciones.

Varias de las proteínas encontradas corresponden a enzimas involucradas en la modificación topológica, polimerización, reparación o modificación del DNA o bien funcionan como reguladores transcripcionales (en T6 y T7). Encontrar este tipo de proteínas en un escenario donde los ribonucleótidos fungían como las principales moléculas informacionales parece totalmente contradictorio, sin embargo es posible que algunas de las moléculas que interactuaban inicialmente con el RNA fuesen cooptadas hacia el DNA cuando éste se volvió la molécula dominante; probablemente en cada una de las proteínas del procesamiento del DNA, la región codificada por tripletes RNY se encarga de unir genéricamente moléculas nucleotídicas o incluso de algún otro tipo, de manera que aun cuando esas regiones forman parte de proteínas que actualmente pertenecen al procesamiento del DNA, bien pudieron haber surgido en los albores del Mundo de las RNPs.

La propuesta alternativa del Mundo Viral, que defienden Eugene Koonin y Patrick Forterre, sugiere que fue en los virus donde se originó la molécula de DNA y las proteínas relacionadas con su replicación y topología. Según la propuesta de la Virósfera, el RNA en los virus se modificó, confiriendo a estas entidades de la **ventaja selectiva de "burlar" las** ribonucleasas de la célula infectada, protegiendo así el material genético viral; posteriormente cuando los provirus se escinden, parte del material genético y las proteínas asociadas a la nueva molécula de DNA, se quedan en el huésped formando parte de su repertorio, poco a poco el DNA se convierte en la molécula dominante incluso en los organismos celulares^{cfr. 25,26}.

Order	Sc.	E-v	ID	%	PROTEIN / FAMILY / DOMAIN
1	32	7.00E-08	32/32	100%	PRD domain protein (70.88 %)
2	30	1.00E-06	30/30	100%	DNA topoisomerase I (90.14 %)
3	28	2.00E-05	28/28	100%	DNA-binding response regulator (68.66 %)
4	25	0.001	25/25	100%	site-specific recombinase, phage integrase family (4.42 %)
5	25	0.001	25/25	100%	DNA polymerase III, delta subunit (23.92 %)
7	24	0.004	24/24	100%	plasmid recombination enzyme (35.43 %)
8	24	0.004	24/24	100%	DNA ligase, NAD-dependent (39.68 %)
9	24	0.004	24/24	100%	FtsK/SpoIIIE family protein (41.71 %)
11	23	0.015	23/23	100%	prophage LambdaSa03, structural protein, putative (16.52 %)
12	23	0.015	23/23	100%	prophage LambdaSa03, tail component, putative (43.15 %)
13	22	0.06	22/22	100%	exodeoxyribonuclease VII, large subunit (85.97 %)
14	22	0.06	22/22	100%	prophage LambdaSa04, tail tape measure protein, TP901 family (52.58 %)
15	22	0.06	22/22	100%	DNA-entry nuclease, putative (60.76 %)
16	22	0.06	22/22	100%	catabolite control protein A (1.00 %)
18	22	0.06	25/26	96%	transcriptional regulator, AraC family (41.01 %)
19	22	0.06	22/22	100%	transcriptional regulator, DeoR family (87.21 %)
20	22	0.06	22/22	100%	UvrABC system, B protein (13.31 %)
21	22	0.06	33/36	91%	transcriptional regulator GlnR (38.54 %)
22	22	0.06	22/22	100%	PRD domain protein (86.99 %)

T6. Proteínas del procesamiento del DNA que contienen regiones codificadas por tripletes RNY, nomenclatura y código de colores igual que T1 y T2.

o REPRESIÓN CATABÓLICA POR CARBONO.

o Tanto la **proteína con PRD** (**PTS** regulatory **d**omain) como la **Proteína A de control de catabolitos** (CcpA) son parte del mecanismo denominado represión catabólica por carbono (CCR), mientras que las primeras son reguladores específicos de la transcripción, la CcpA es un regulador global. A través del mecanismo CCR se reprime, mediante cascadas de fosforilaciones, la expresión de genes requeridos para la utilización de fuentes alternativas de energía, lo que ocurre si la bacteria se encuentra expuesta a una mezcla de carbohidratos donde la fuente preferida está presente, glucosa generalmente^{244,245}.

o Las proteínas con PRDs son reguladores transcripcionales que pueden unirse tanto a RNA (antiterminadores) en bacterias G⁺ y G⁻, como a DNA (activadores) sólo en G⁺^{246,247}. Un estudio sobre la filogenia de los reguladores transcripcionales del PTS, sugiere que los dominios de unión a RNA y los de unión a DNA de los PRDs no son homólogos²⁴⁸; es posible suponer entonces que dado que en este tipo de reguladores hay parte codificada por tripletes RNY, muy posiblemente se trata del segmento correspondiente al antiterminador porque es la región que interactúa con RNA y por ende la más antigua.

- o La CcpA es parte del mecanismo CCR al unirse, junto con otras proteínas, a secuencias específicas del DNA, la CcpA es característica de bacterias G⁺ y es similar en estructura a proteínas hélice-vuelta-hélice (HTH) que también se unen al DNA^{249,250}, además es posible que la CcpA tenga relación evolutiva con el metabolismo del nitrógeno en bacterias G⁻²⁴⁹; por lo que resulta en el presente trabajo, es posible que los orígenes evolutivos de la CcpA se encuentren en algún punto del Mundo del RNA, ya que alguna parte de esta proteína se encuentra codificada por tripletes del tipo RNY.

- o El hecho que una región de cada una de las dos proteínas antes mencionadas (proteína con PRD y la Proteína A de control de catabolitos, CcpA) estén codificadas por tripletes RNY, pudiese indicar que existían precursores de la regulación transcripcional del metabolismo de la glucosa al inicio del Mundo de las RNPs; no obstante lo más probable es que tal región ancestral en cada una corresponda en realidad a las regiones de unión a hebras helicoidales de nucleótidos, semejantes a las formas que puede adoptar no sólo el DNA sino el RNA al plegarse sobre sí mismo.

- o **TOPOISOMERASA TIPO I.**

- o La topoisomerasa tipo IA (TopA) es una proteína monomérica que produce una rotura en una de las hebras de la doble hélice del DNA a la cual se acopla formando **un enlace 5'fosfotirosina, permitiendo que la** hebra intacta pase a través y liga nuevamente la hebra previamente fragmentada, de esta manera cataliza la isomerización topológica de la molécula del DNA al relajar negativamente la estructura súper-enrollada. El Mg²⁺ incrementa la tasa de ruptura del DNA y es requerido para la relajación del súper-enrollamiento, aunque no es requerido para la escisión misma. Más aún, la TopA no requiere de ningún cofactor energético para realizar su función enzimática, ya que la fuerza del súper-enrollamiento es suficiente

por sí misma para proveer de impulso mecánico a la enzima^{251,252,253}. TopA tiene un característico plegamiento toroidal formado por cuatro dominios²⁵²; de los cuales los más significativos son: el dominio 5Y-CAP encontrado en un gran número de proteínas que **unen DNA y el dominio "toprim" que está estructurado en un plegamiento Rossmann**, el cual también se encuentra en primasas y algunas nucleasas, dado que es característico de proteínas que se unen a DNA o RNA²⁵¹.

- o La propuesta de Forterre y colaboradores se basa en el origen viral del DNA y proteínas relacionadas, todo lo cual fue introducido en células de RNA mediante ciclos infecciosos, paulatinamente surgieron organismos celulares con DNA como principal molécula informacional, lo que conllevó problemas topológicos y el surgimiento de la DNA topoisomerasas²⁵⁴. Forterre y sus colaboradores encontraron que TopA es ubicua en los 3 dominios celulares, pero según su propuesta es posible que sólo se encontrara en el Último Ancestro Común Bacteriano (LBCA) y después fuera distribuida hacia otros organismos por transferencia horizontal, mediante virus o plásmidos de DNA²⁵⁵.

- o La propuesta mencionada de Forterre es plausible, pero con las técnicas actuales no es posible decidir si es correcta. No obstante, por lo que respecta a la presente investigación, TopA no necesita ningún cofactor energético y quizá únicamente Mg^{2+} para su óptimo funcionamiento, lo que puede hacerme suponer que la región que une este ion metálico es antigua, pues se ha mencionado que en un ambiente primitivo las moléculas energéticas eran más escasas que ahora y eran utilizadas preferentemente en el metabolismo de los organismos primigenios, además los metales son donadores de electrones eficientes y aún funcionan de este modo en las proteínas actuales. Sin embargo TopA contiene dos dominios que unen nt; uno de ellos es característico de las proteínas de unión a DNA, molécula que aún no existía en el Mundo de las RNPs aunque no es descartable que esté codificado por tripletes RNY; el otro motivo de unión a nucleótidos es el motivo toprim, que

también está presente en la DNA primasa (ver en 'RNA, estructura y modificaciones') y del que sabemos contiene varios aa codificados por tripletes RNY, como **G** y **D**, de manera que es probable que dicho sitio haya sido el reconocido como ancestral por la metodología descrita en la presente tesis.

- o **REGULADOR DE RESPUESTA DE UNIÓN AL DNA.**

- o El sistema COGs coloca esta proteína en la categoría "transducción" pues es parte del sistema de dos componentes, pero he decidido registrarla dentro de las proteínas relacionadas con el DNA porque éste es el módulo que se une al material genético, regulando la expresión de genes como respuesta a diversos factores ambientales. Este tipo de reguladores contienen el dominio hélice-vuelta-hélice (HTH), característico de todas las proteínas que interaccionan directamente con el DNA²⁵⁶. Es relevante que parte de esta proteína sea ancestral, porque si bien el código RNY surge en el Mundo del RNA, es posible que ciertos módulos que reconozcan nucleótidos genéricos sean antiguos, mismos que pudieron acoplarse posteriormente a otros módulos que estaban en contacto con el medio externo para generar un mecanismo similar al de los sistemas actuales de dos componentes.

- o **PROTEÍNAS QUE FAVORECEN LA INSERCIÓN DE ELEMENTOS EXTRACROMOSOMALES.**

- o Tanto la **recombinasa de la familia de la integrasa del fago (Int)** como la **enzima de recombinación de plásmidos (Pre)** son proteínas que favorecen la inserción de elementos extracromosomales en el genoma de manera sitio-específica.

- o Por su parte las recombinasas de la familia Int no requieren de cofactores energéticos para su mecanismo, sino que conserva la energía de la rotura del enlace fosfodiéster del DNA para volver a unirlo, el mecanismo es semejante al realizado

por la DNA topoisomerasa IB²⁵⁷. La recombinación específica de sitio incluye la rotura, el intercambio y la religación de las cuatro hebras del DNA mediante "Holliday junction", ya sea en *cis* o en *trans*²⁵⁸. Se han identificado 3 regiones conservadas en recombinasas de la familia Int²⁵⁹, posteriormente se identificaron 4 residuos totalmente conservados sólo en el dominio activo (CAT), así como una región medular de unión a DNA (CB)²⁶⁰, dicho sitio activo está conformado en un bucle flexible que tiene la capacidad de unir tanto DNA como RNA²⁶¹.

- o Se ha propuesto que en un mundo ancestral la transferencia horizontal de genes era mucho más frecuente que en la actualidad e incluso era un proceso esencial para la sobrevivencia de los organismos primitivos, de tal modo que no es de sorprender que parte el primordio de las proteínas que favorecen tal proceso haya surgido en el Mundo de las RNPs. Ciertamente tanto los plásmidos como un fago tipo λ (virus asociado a *SagA909*) son elementos de DNA de doble cadena (dsDNA), molécula que no surgió sino hasta el LUCA; sin embargo como se indicó el dominio CAT une ambos tipos de ácidos nucleicos (DNA y RNA) y por tanto es probable que sea la región codificada por tripletes RNY, pues inicialmente dicho bucle habría reconocido únicamente ribonucleótidos y posteriormente cadenas de DNA cuando esta se volvió la molécula hereditaria.

- o Si la hipótesis del Mundo Viral^{cfr. 25,26} es cierta se simplifica la explicación del origen evolutivo de la recombinasa de la familia Int y de la enzima Pre, pues podríamos suponer que la recombinasa surgió en virus de DNA y permitió la recombinación de su material genético en un genoma proto-celular de RNA, la enzima se habría transferido entonces al organismo celular por escisión defectuosa del provirus, posteriormente la transición hacia el Mundo de DNA en la célula comenzó por plásmidos o provirus antiguos que codificaban la enzima que les permitía recombinarse en los genomas.

o **DNA POLIMERASA III, SUBUNIDAD DELTA (δ).**

o La subunidad δ , también conocida como HoloA, es parte del subcomplejo DnaX sobre el cual se ensambla, mediante la unión de ATP, la subunidad β de la holoenzima DNA polimerasa III (DNApolIII) que replicará el material genético²⁶². De manera singular, la subunidad δ no acopla por sí misma ningún cofactor energético para unirse al DNA de cadena sencilla (ssDNA) cebado^{263,264} y su secuencia está pobremente conservada en diversas bacterias, sin embargo esta subunidad (δ) tiene relación estructural con las otras subunidades del complejo DnaX que sí unen ATP, incluso es posible englobarla dentro de la familia AAA+ (ATPasas Asociadas con diversas Actividades celulares)²⁶⁵.

o Es posible que dado que la subunidad δ se une ssDNA, haya podido unir ancestralmente cadenas de RNA y entonces fue cooptada en la transición hacia la utilización de DNA como principal molécula informacional, de manera que la región codificada por triplete RNY en HoloA tiene la capacidad de unir ribonucleótidos y desoxirribonucleótidos.

o **DNA LIGASA DEPENDIENTE DE NAD.**

o La DNA ligasa bacteriana cataliza la formación de un enlace fosfodiéster entre dos fragmentos de DNA, utilizando NAD⁺ como cofactor²⁶⁶. La ligasa dependiente de NAD⁺ consiste de dos dominios independientes, el dominio N-terminal une el NAD⁺ y se autoadenila, mientras que el dominio C-terminal une DNA y Zn²⁺²⁶⁷.

o Es posible que ambos dominios de la DNA ligasa hayan surgido de manera separada. Por un lado la propuesta de Koonin e Iyer, acerca del reclutamiento de enzimas del Mundo del RNA hacia procesos involucrados en DNA, *i.e.* que una RNA ligasa haya evolucionado hasta la DNA ligasa de hoy en día, podría sugerir el dominio de unión a DNA es el más antiguo. No obstante la antigüedad del dominio

de adenilación, si es que es la parte codificada por tripletes RNY, encuadra mejor en los modelos de mundos primigenios cuando los cofactores energéticos que existían eran derivados de la flavina o contenían un anillo de adenina; esto aunado al hecho de que las DNA ligasas bacteriana y las de eucariontes comparten una pequeña homología justamente en este dominio²⁶⁸.

- o **FAMILIA FtsK/SpoIIIE.**

- o Las proteínas de la familia FtsK/SpoIIIE son de las más simples que existen en cuanto a maquinaria de transporte del DNA asociada a la membrana^{269,270}. Los miembros de tal familia participan en la traslocación de moléculas de DNA hacia una célula progenie o una espora, durante la división celular o la esporulación respectivamente; dicho proceso es dependiente de ATP, que sin embargo sólo se utiliza para bombear el ssDNA hacia fuera de la célula progenitora²⁷¹.

- o Los transportadores de la familia FtsK/SpoIIIE están conservados entre las bacterias y sus miembros se han englobado en una superfamilia de ATPasas de bucle-P denominada FtsK-HerA, donde se encuentran también las ATPasas que sirven para el empaquetamiento del DNA en virus, así como nucleasas que contienen un plegamiento tipo Rossmann (implicado en el reconocimiento de nucleótidos); según los resultados de dicho estudio, el origen evolutivo de la familia FtsK-HerA no puede ser rastreado hasta LUCA, pero se piensa que una nucleasa ancestral pudo ser el punto de partida²⁷². Dicho razonamiento validaría el hecho de que parte de algún miembro de la familia FtsK/SpoIIIE está codificada por tripletes RNY, esa parte justamente podría corresponder precisamente al plegamiento Rossmann característico de proteínas que unen nucleótidos y que pudo unir antiguamente RNA tal como actualmente une DNA o algún NTP, constituyendo el primordio de la familia FtsK/SpoIIIE que surgió en al iniciar el Mundo de las RNPs.

o **PROTEÍNAS DE LA ARQUITECTURA DE UN PROFAGO TIPO LAMBDA.**

o En el genoma de *SagA909* existen 3 proteínas provirales integradas que contienen regiones codificadas por tripletes RNY, todas ellas están relacionadas con la estructura viral y al menos 2 están involucradas en la constitución de la cola del virus.

o Los genes de la cola y la cabeza de algunos fagos tipo λ (lambdoides) tienden a conservar altos niveles de identidad, pero en forma de mosaico unos con respecto de otros. En algunos casos los genes de la cola de los fagos lambdoides pueden intercambiarse o recombinarse mediante pequeñas regiones homólogas y conferir ventaja selectiva al virus, ya que porque le permite ampliar su repertorio de hospederos^{273,274}; esas regiones de micro-homología pudieran ser las más antiguas de los genes estructurales²⁷³ y por ende ser codificadas por tripletes RNY. Por otro lado se ha propuesto que las proteínas provirales pueden servir de reservorios genéticos para los fagos, ya que es más probable haya intercambios genéticos entre fagos residentes que entre fagos que se encuentran en el ambiente²⁷².

o Según las hipótesis de Forterre y de Koonin los virus son al menos tan antiguos como el Mundo del RNA y pudieron ser los elementos responsables de introducir el DNA en las células y por tanto que éste comenzara a desplazar al RNA como el material hereditario²⁷⁵, de manera que es posible que los primordios de las proteínas provirales identificadas correspondan a los albores de dicha transición, misma que habría ocurrido en el Mundo de las RNPs.

o **SUBUNIDAD GRANDE DE LA EXODESOXIRRIBONUCLEASA VII.**

o La exodesoxirribonucleasa VII (ExoVII) degrada ssDNA (DNA de cadena sencilla) bidireccionalmente (tanto del extremo 5' como del 3'), procesivamente (comienza generando oligonucleótidos y subsecuentemente los degrada) y no requiere ningún cofactor energético para llevar a cabo su función²⁷⁶.

- o La subunidad XseA corresponde al dominio estructural de la ExoVII y es la más grande de las subunidades que contiene esa proteína. XseA tiene un dominio OB que une genéricamente ácidos nucleicos genéricos en el extremo N-terminal^{277,278}, por lo que es posible que contenga varios aa codificados por tripletes RNY.

- o Una hipótesis alternativa es que la ExoVII, al menos la parte que fue reconocida como ancestral, haya servido desde el principio para la degradación de DNA, en este caso del material genético de un virus de DNA que infectara células de RNA existentes, bajo la hipótesis del Mundo Viral defendida por Forterre y Koonin.

- o **NUCLEASA PARA ENTRADA DE DNA.**

- o EndA es una nucleasa, localizada en la parte externa de la membrana, comúnmente asociada con la degradación de DNA aunque también degrada RNA²⁷⁹. La enzima EndA funciona actualmente para facilitar la entrada del microorganismo en un hospedero al degradar el DNA de las defensas de este último, aunque EndA también funciona durante el proceso de competencia para la captación de DNA extracelular²⁸⁰. La proteína EndA forma parte de las nucleasas de **DRGH**, nombradas así debido a que son los aminoácidos del sitio catalítico.

- o En un ambiente primitivo la nucleasa EndA fue utilizada posiblemente en la degradación del material genético de los virus de DNA que para el momento del Mundo del RNA eran recientes, o para la degradación del RNA de otros riboorganismos; no obstante, en el Mundo de las RNPs es más probable que la pequeña región codificada por tripletes RNY de la nucleasa mencionada, sencillamente unía y por ende confería estabilidad a los ácidos nucleicos, en particular al RNA.

o **REGULADORES TRANSCRIPCIONALES.**

o De los reguladores transcripcionales que contiene SagA909, sólo se identificaron 3 que contienen tripletes del tipo RNY en su secuencia:

- El regulador **GlnR** reprime la expresión del gen de la glutamina sintetasa²⁸¹; es un regulador local de la transcripción.
- El regulador **DeoR** reprime el operón Deo, que codifica enzimas catabólicas de nucleótidos y desoxirribonucleótidos²⁸²; es un regulador global de la transcripción.
- El regulador **AraC** regula positiva y negativamente la expresión de las proteínas necesarias para la absorción y el catabolismo del azúcar L-arabinosa²⁸³; es un regulador global de la transcripción.

o Todos los reguladores transcripcionales contienen generalmente en su estructura motivos hélice-vuelta-hélice (HTH, en inglés). El motivo HTH y sus variaciones estructurales son característicos de las proteínas de unión a DNA en procariontes, en particular de factores de transcripción. Recientemente se ha mostrado que proteínas virales del tipo HTH parecen estar involucradas en el empaquetamiento de las hebras virales de DNA dentro de la cápside proteica y probablemente los actuales reguladores transcripcionales con motivos HTH hayan evolucionado a partir de proteínas virales con esos patrones²⁸⁴. Es posible que la parte de los reguladores transcripcionales que está codificada por tripletes RNY sea justamente el dominio HTH en caso que la hipótesis del Mundo viral sea cierta, ya que dicho motivo habría unido el material genético viral que ya era DNA cuando la genética de los proto-organismos celulares estaba basada en RNA.

o Por otro lado pudiera ser posible, aunque poco probable, que la parte codificada por tripletes RNY en cada uno de los reguladores identificados no tenga relación con el motivo HTH sino que más bien se identificó en cada uno la región involucrada en el reconocimiento del metabolito correspondiente, pues las moléculas

asociadas a los reguladores mencionados fueron posiblemente fundamentales y abundantes en un mundo ancestral: la glutamina (para GlnR), aunque no forma parte de los aminoácidos codificados por tripletes RNY, es posible que haya estado presente en un mundo primitivo, pudo servir para la formación de otros aminoácidos, está involucrada en el metabolismo del nitrógeno y constituye parte fundamental de las proteínas ancestrales; los ácidos nucleicos (para DeoR) son evidentemente primordiales en el marco biológico de un mundo ancestral; la L-arabinosa (para AraC) es un carbohidrato que, tal como en la actualidad, pudo utilizarse como fuente alternativa de energía desde los albores de la vida.

o Como podemos observar, con la información recabada hasta el momento no es posible discernir aun, cuál sitio de los reguladores transcripcionales mencionados es codificado por tripletes RNY, aunque existen algunos candidatos.

o **PROTEÍNA B DEL SISTEMA UvrABC.**

o La exonucleasa UvrABC está involucrada en la reparación por escisión de nucleótidos (NER) del material genético de los procariontes. La subunidad B (UvrB) es el componente central en dicho proceso, ya que interactúa con todos los elementos²⁸⁵. Actualmente la proteína UvrB contiene pequeños motivos claramente definidos, comunes a las helicasas de DNA y RNA²⁸⁶, como los motivos Walker A y B que forman un sitio de unión a nt²⁸⁷ y una horquilla- β que se inserta en el DNA²⁸⁸.

o El mecanismo actual de NER es el siguiente²⁸⁹: La proteína UvrA carga a la UvrB en el DNA dañado, utilizando ATP para lograrlo y liberando la RNA polimerasa que se ha atascado; una vez formado el complejo UvrB-DNA, UvrA se desensambla usando la energía liberada de la hidrólisis del ATP, lo que activa la proteína UvrB al cambiar su conformación; se acopla UvrC a UvrB-DNA y se realiza la escisión de nucleótidos en los extremos 3' y 5'. UvrD remueve el oligonucleótido escindido

liberando UvrC; la DNA polimerasa I rellena el espacio sintetizando los nucleótidos correspondientes y libera finalmente a UvrB. La DNA ligasa sella la hebra de DNA.

- o En ausencia de UvrA, por ende también en ausencia de ATP que pueda cambiar la conformación estructural, la UvrB puede sin más unir el ssDNA en los motivos de tipo helicasa, mismos que unen DNA o RNA²⁸⁹ y ello que sugiere que ésta puede ser la región codificada por tripletes RNY, que en un inicio reconociera cadenas genéricas de ácidos nucleicos, pues quizá un pequeño fragmento de ssDNA pudiera equivaler estructuralmente a una cadena de RNA de los inicios del Mundo de las RNPs.

- o **AUSENCIA DE, AL MENOS, TOPOISOMERASA TIPO II.**

- o Pese que existe una DNA topoisomerasa tipo II en *SagA909*, ésta no fue identificada como ancestral por la metodología presentada en esta tesis. Una de las posibilidades es que dicha TopoII fue introducida independientemente en el linaje bacteriano posterior a LUCA; otra posibilidad es que, pese que contiene un motivo topim que pudo ser codificado por triplete RNY, el contexto de la secuencia de dicho motivo no contiene tripletes de ese tipo; esto aunado a que es incluso posible que la metodología presentada no es lo suficientemente sensitiva para captar la región codificada por tripletes RNY de cada una de las proteínas.

T7. Características plausibles de la probable región codificada por tripletes RNY en las proteínas identificadas del procesamiento de DNA.

Conjunto	Proteína (nombre traducido y adaptado)	Características plausibles de la probable región codificada por tripletes RNY.
Represión catabólica por carbono (CCR)	Proteína con dominio PRD	Unión de RNA
	Proteína A de control de catabolitos	Unión a DNA (quizá RNA)
	Topoisomerasa tipo I	Unión de Mq2+ o de ssDNA ¿posiblemente viral? (o quizá RNA)
	Regulador de respuesta de unión al DNA	Interacción con DNA (quizá RNA)
Proteínas que favorecen la inserción de elementos extracromosomales	Recombinasa de la familia integrasa del faqo; Int	Unión de RNA y de DNA.
	Enzima de recombinación de plásmidos	Unión de dsDNA (quizá un RNA plegado sobre sí mismo)
	DNA polimerasa III, subunidad delta	Unión de ssDNA ¿posiblemente viral? (o quizá RNA)
	DNA ligasa, dependiente de NAD	Unión de NAD (o quizá de DNA o de Zn2+)
Proteínas de la arquitectura de un profago tipo lambda	Proteína de la familia FtsK/SpoIIIE	Unión de nucleótidos: DNA, o RNA, o algún NTP
	Sa03, proteína estructural, putativa	No precisada, quizá regiones de micro-homología
	Sa03, componente de la cola, putativa	No precisada, quizá regiones de micro-homología
	Sa04, proteína de la medición de la cola, familia TP901	No precisada, quizá regiones de micro-homología
	Exodeoxirribonucleasa VII, subunidad grande	Unión de ácidos nucleicos genéricos
Reguladores transcripcionales	Nucleasa para la entrada de DNA, putativa; EndA	Interacción con DNA y RNA
	Familia AraC	Unión a DNA (quizá RNA) o a L- arabinosa
	Familia DeoR	Unión a DNA (quizá RNA) o ácidos nucleicos en general
	GlnR	Unión a DNA (quizá RNA) o a Q
	Proteína B del sistema UvrABC	Unión de ATP o de ssDNA ¿posiblemente viral? (o quizá RNA)

6.3.1.3 Metabolismo.

El metabolismo es una característica esencial de los sistemas vivos, ya que permite a los organismos obtener energía mediante ciclos de asimilación de moléculas orgánicas *i.e.* anabolismo y catabolismo. En particular el ciclo de los ácidos tricarboxílicos (TCA) se considera el metabolismo central de todos los seres vivos actuales, ya que es posible sintetizar diversos compuestos a partir de sus metabolitos, desde lípidos para la membrana celular, aminoácidos para proteínas, ácidos nucleicos para usarlos como moléculas de información genética y azúcares para sintetizar pared celular y obtener energía^{93,94}. Adicionalmente, se ha mostrado que el ciclo TCA que puede funcionar como un sistema de transducción de señales, ya que cualquier señal ambiental altera las actividades del ciclo TCA, cambiando entonces el estado metabólico de una bacteria y por ende su estado redox²⁹⁰. Pese a todo ello, aunque el ciclo TCA es universal y sus metabolitos se regeneran por retroalimentación interna, éste se lleva a cabo sólo en organismos aerobios^{93,94} y por tanto es descartable como ciclo metabólico ancestral, ya que la atmósfera oxidante fue posterior al LUCA. Dado que la glucosa se puede metabolizar tanto en aerobiosis como en anaerobiosis^{93,94}, es más probable que su el ciclo metabólico correspondiente sea el más antiguo; mientras que el ciclo TCA se fue generando por convergencia de distintos procesos metabólicos.

El ciclo TCA puede presentarse también en reversa (ciclo rTCA) y es posible que ése si sea antiguo, ya que con base en minería de datos de la base de datos KEGG y comparación de rutas metabólicas se ha propuesto que el LUCA era un organismo quimiolitautótrofo *i.e.* habría utilizado el ciclo rTCA como ruta metabólica y por tanto obtenía energía de la oxidación de compuestos inorgánicos, y del carbono a partir la fijación de CO₂²⁹¹.

Se han propuesto al menos dos teorías para explicar la evolución de las rutas metabólicas. Un escenario plantea mediante la hipótesis retrógrada, según la cual las primeras enzimas metabólicas que surgieron son actualmente las últimas que intervienen en cualquier ruta metabólica; dicho escenario no es aplicable a gran parte del metabolismo y existen argumentos de carácter fisicoquímico y bioquímico y estructurales que contradicen firmemente la hipótesis retrógrada, no obstante tal planteamiento no puede descartarse en algunos casos de evolución metabólica. Mucho más ampliamente probada y aceptada es **la teoría de parches o 'patchwork'** que explica la evolución del metabolismo mediante el reclutamiento de enzimas relativamente pequeñas e ineficientes, de amplia especificidad, que reaccionaban con una amplia variedad de sustratos; dicha teoría es la que presenta mayor sustento bioquímico, fisicoquímico y por comparación de secuencias, en este caso la evidencia muestra que los eventos de duplicación génica, seguidos por la acumulación gradual de mutaciones y otros cambios genéticos, derivó en la especialización y modificación de las funciones enzimáticas, al menos en las rutas metabólicas primordiales²⁹².

En general las enzimas metabólicas están altamente conservadas y, con base en un análisis de co-conservación de las enzimas desplegadas en KEGG, se ha mostrado que existe un núcleo enzimático altamente conservado aunque flexible de enzimas involucradas en múltiples reacciones a través de distintas rutas metabólicas, mientras que las enzimas en la periferia de ese núcleo están menos conservadas y asociadas con innovaciones específicas de los taxones. Dicho núcleo enzimático está comprendido por los metabolismos de: nucleótidos, aminoácidos, carbohidratos, energía y lípidos, lo que implica que estos procesos proveen las funciones básicas vitales y se muestra además que el metabolismo efectivamente ha podido evolucionar por modularidad²⁹³.

Recientemente se ha tratado de dilucidar cual pudo ser el metabolismo de organismos previos al LUCA; asumiendo el Mundo del RNA como ambiente ancestral, se ha realizado un estudio de reconstrucción filogenética para indagar cuáles habrían las rutas metabólicas subsecuentes. Se ha señalado que el metabolismo se originó en enzimas con un plegamiento denominado bucle-P de hidrolasa ('P-loop hydrolase'), por lo que el primer conjunto enzimático en emerger habría estado relacionado con la síntesis de nucleótidos en el Mundo del RNA *i.e.* enzimas de las subredes metabólicas de purinas y de pirimidinas que contienen dicho plegamiento, lo que en realidad no es de sorprender asumiendo que el RNA es una molécula prebiótica; también son ancestrales, aunque habrían surgido poco después del metabolismo de nucleótidos, están los metabolismos de aa, carbohidratos y lípidos, todo lo cual habría sido parte de un estallido enzimático que daría origen a los primeros proto-organismos²⁹⁴.

Se ha establecido entonces, mediante distintas metodologías, que los procesos metabólicos primigenios están relacionados con: energía, aminoácidos, nucleótidos, carbohidratos y lípidos, de los cuales existen proteínas cuyo origen es tan ancestral que contienen regiones codificadas por tripletes RNY, aunque también hay enzimas con la misma antigüedad que no necesariamente están involucradas un metabolismo específico (en T8 y T9). A continuación se muestran argumentos plausibles que puedan explicar el hecho de que algunas enzimas metabólicas de **SagA909** fueron identificadas como ancestrales, éstas se encuentran categorizadas con base en las clasificaciones COG y KEGG.

Order	Sc.	E-v	ID	%	SubCateg.	PROTEIN / FAMILY / DOMAIN
1	43	2.00E-14	54/57	94%	Ener.	ATP synthase F1, alpha subunit (67.77 %)
2	30	1.00E-06	30/30	100%	Carb.	D-mannonate oxidoreductase UxuB, putative (4.89 %)
3	30	1.00E-06	30/30	100%	Aa.	glyoxylase family protein (48.91 %)
4	30	1.00E-06	30/30	100%	Ion	sulfatase family protein (91.93 %)
5	28	2.00E-05	28/28	100%	Nucl.	GMP synthase (95.39 %)
6	28	2.00E-05	28/28	100%	Carb.	phosphopentomutase (93.64 %)
7	28	2.00E-05	39/42	92%	Aa.	aminotransferase, classes I and II (50.21 %)
8	27	6.00E-05	27/27	100%	Nucl.	hypoxanthine phosphoribosyltransferase (29.70 %)
9	27	6.00E-05	27/27	100%	Lip.	enoyl-CoA hydratase/isomerase family protein (65.87 %)
10	27	6.00E-05	27/27	100%	Carb.	glycogen synthase (14.62 %)
11	27	6.00E-05	27/27	100%	Ener.	NADH oxidase, water-forming (67.23 %)
12	26	2.00E-04	37/40	92%	Carb.	HAD-superfamily hydrolase, subfamily IIA (95.71 %)
13	25	0.001	34/37	91%	Lip.	acetyl-CoA carboxylase, biotin carboxylase (24.53 %)
14	25	0.001	25/25	100%	Ener.	glycerol-3-phosphate dehydrogenase (NAD(P)+) (29.63 %)
15	25	0.001	25/25	100%	Other	4-oxalocrotonate tautomerase family enzyme (0.00 %)
16	24	0.004	24/24	100%	Carb.	glucose-1-phosphate adenyltransferase (73.40 %)
17	24	0.004	24/24	100%	Ener.	ATP synthase F1, alpha subunit (87.11 %)
18	24	0.004	24/24	100%	Coe.	oxygen-independent coproporphyrinogen III oxidase, putative (33.89 %)
19	24	0.004	24/24	100%	Carb.	inositol monophosphatase family protein (98.36 %)
20	24	0.004	24/24	100%	Mult./Gral.	oxidoreductase, short chain dehydrogenase/reductase family (81.47 %)
21	24	0.004	24/24	100%	Ener.	manganese-dependent inorganic pyrophosphatase (79.04 %)
16	24	0.004	24/24	100%	Coe.	prenyltransferase, UbiA family (97.66 %)
2	23	0.015	23/23	100%	Mult./Gral.	glutamine amidotransferase (84.10 %)
23	23	0.015	23/23	100%	Carb.	glucuronate isomerase (71.50 %)
24	23	0.015	23/23	100%	Aa.	3-dehydroquininate synthase (1.31 %) chorismate synthase (0.60 %)
27	22	0.06	22/22	100%	Nucl.	inosine-uridine preferring nucleoside hydrolase family protein (29.50 %)
28	22	0.06	25/26	96%	Ener.	ATP synthase F0, C subunit (61.00 %)
29	22	0.06	22/22	100%	Ener.	NADH oxidase, water-forming (18.25 %)
30	22	0.06	25/26	96%	Aa.	glyoxylase family protein (48.43 %)
31	22	0.06	25/26	96%	Nucl.	purine nucleoside phosphorylase I, inosine and guanosine-specific (71.94 %)
32	22	0.06	22/22	100%	Mult./Gral.	FMN-binding protein (93.58 %)

T8. Proteínas del metabolismo que contienen regiones codificadas por tripletes RNY, nomenclatura y código de colores igual que T1 y T2.

o **METABOLISMO ENERGÉTICO.** Primeramente debemos decir que las moléculas más importantes para la bioenergética de los organismos son los cofactores ATP, NADP⁺ y FMN/FAD, que participan de uno u otro modo en el metabolismo de todos los organismos conocidos, fungiendo como acarreadores de electrones (e⁻). El origen químico de esas moléculas es relativamente sencillo: primeramente tenemos el ATP que es probablemente uno de los compuestos más antiguos (obsérvese que la adenina es una purina, que resulta ser el primer tipo de base en el código RNY), pudiendo sintetizarse completamente de manera prebiótica bajo una atmósfera reductora y posiblemente llegando a ser uno de los compuestos más abundantes; bioquímicamente el ATP es aquella adenosina del RNA unida a dos fosfatos más, la coenzima NADP⁺ es nuevamente esa adenosina unida a una estructura similar con nicotinamida, para sintetizar el FMN/FAD se utiliza NADP⁺ y nuevamente ribosa^(cfr.23). Es de esperarse entonces que la mayoría de las reacciones

bioquímicas que ocurrían en organismos antecesores de LUCA, utilizasen al menos una de estas moléculas como donadores/aceptores de electrones

o La fuente energética sintética prebiótica primaria debió ser la energía lumínica, sin embargo posteriormente algún tipo de energía pre-bioquímica derivó en procesos sintéticos más eficientes y con ello emergieron formas incipientes de vida, es posible que tal forma de energía pueda encontrarse aún en la red metabólica de organismos actuales. Dichas fuentes energéticas pre-bioquímicas pudieron ser: agentes condensantes de aminoácidos, purinas y carbohidratos simples o incluso monómeros activados de nucleótidos o aminoácidos que se polimerizaban mediante la inclusión de cationes divalentes; moléculas pigmentadas que funcionaban como acarreadores de electrones; energía liberada de la hidrólisis del pirofosfato (PPi); gliceraldehído cuya producción puede ser continua a partir de compuestos prebióticos; piritita que favorecería interacciones electrostáticas que polimerizarían otros compuestos como el fosfato de gliceraldehído y el fosfato de dihidroxiacetona (DHAP), tal como propuso Wächterhäuser o bien energía quimiosmótica que provenía de gradientes de pH por potenciales iónicos derivados de la diferencia de carga entre ambos lados de una capa lipídica, misma que también puede formarse a partir de compuestos prebióticos. En cualquier caso, se ha señalado que el catalizador de un sistema replicante en condiciones prebióticas debía incorporar tanto información genética como actividad polimérica, utilizando las fuentes de energía disponibles²⁹⁵.

o Evidencia molecular sugiere que la respiración aerobia pudo aparecer antes que la fotosíntesis, puesto que el denominado LUCA pudo haber tenido reductasa de óxido nítrico y citocromo oxidasa²⁹⁶, sin embargo tales enzimas no fueron identificadas como ancestrales, por lo que es poco probable que un organismo con código genético derivado directamente del Mundo del RNA fuese aerobio.

o Después de las reacciones fermentativas, la actividad bioenergética más ampliamente distribuida es la respiración aerobia y ambas rutas pueden ser trazadas al menos hasta el LUCA; de modo que aparentemente el LUCA era un organismo principalmente aerobio que sin embargo era bioenergéticamente flexible, ya que contenía varias enzimas bioenergéticas distintas, lo que le confería estrategias de supervivencia en diversos ambientes. Posteriormente hubo una fuerte reducción evolutiva como resultado de la adaptación a nichos particulares especializados; el modo de distribución de las enzimas energéticas sugiere entonces que éstas han evolucionado siguiendo la teoría por parches²⁹⁷.

o Sólo recientemente se ha tratado de describir un origen plausible del grupo de enzimas bioenergéticas que existió antes de LUCA. La semejanza de muchas de esas enzimas entre sí, parece indicar tuvieron su origen en pocos bloques estructurales, que habrían constituido el botiquín esencial redox para la construcción de enzimas diversas²⁹⁸. Otra metodología enfocada a descubrir las enzimas bioenergéticas ancestrales se describe en la presente tesis y a continuación se describe la información obtenida.

- ATP SINTASA.

- El ATP sintasa es una enzima crucial para la vida, está notablemente bien conservada a través de la evolución y es una de las proteínas más antiguas, cuyo origen precede incluso al de la maquinaria respiratoria y la fotosintética²⁹⁹. El ATP sintasa es un canal iónico que acopla el movimiento de protones y la síntesis/hidrólisis de un nucleósido trifosfato (NTP), generalmente ATP; dicho mecanismo es completamente reversible dependiendo de las condiciones de medio interno con respecto a las condiciones ambientales. En condiciones aerobias la ATP sintasa utiliza la fuerza H⁺-motriz (ΔH^+) o (Na⁺-motriz) para generar ATP, mientras que en

condiciones anaerobias, presentes en varias especies de bacterias, el ATP se hidroliza para trascolar H^+ fuera de la célula y mantener el gradiente de protones, por lo que funcionaría como una ATP hidrolasa³⁰⁰, liberando $\sim 7\text{kcal/mol}$ ²⁹⁹.

- Debido al carácter reversible de la ATP sintasa, se ha preferido la denominación ATPasa F_0F_1 , nombrando así además a los 2 componentes fundamentales de las ATPasas: el módulo de membrana F_0 y el módulo citoplásmico F_1 , tal nombre además la distingue de otros tipos de canales-ATPasas, de los que existen al menos otros 4 tipos³⁰¹.
- El módulo F_0 , que está embebido en la membrana, está compuesto de tres subunidades en estequiometría $a_3b_3c_{9-12}$ de las cuales, las c forman un anillo por el que los protones atraviesan la membrana. El módulo F_1 es citoplásmico está formado básicamente por subunidades designadas por letras griegas en estequiometría $\alpha_3\beta_3\gamma_1\delta_1\varepsilon_1$, de las cuales α y β unen ATP, pero sólo β lo hidroliza, mientras que las otras subunidades sirven de conexión física o mecánica con el módulo en la membrana^{302,303,304,305}. Todas las subunidades de la ATPasa se encuentran codificadas usualmente en un mismo operón^{306,307,308}, cuya transcripción se encuentra actualmente regulada por la tasa de crecimiento³⁰⁹.
- Con base en análisis comparativos, se ha propuesto que la ATPasa ancestral era una hidrolasa del tipo anaerobio, dicha ATPasa ancestral se componía de 6 sitios para la hidrólisis del ATP (cat) y 12 para formar el canal (ch) para los protones, por lo que la tasa de H^+ extraídos/ATP hidrolizado era de 2, posteriormente diversos eventos de mutación que consistieron en **fusiones e inversión de función (que bien podría denominarse 'hipótesis fus-inv')** llevaron a: 1) fusiones de los sitios cat e inversión de función resultó

en 3 sitios cat y 12 ch con una tasa aumentada de 4 H⁺ recolectados/ATP sintetizado en organismos aerobios; 2) una fusión de canales y una inversión de función produjo 6 ch y 3 cat con una tasa de 2 H⁺ extraído/ATP sintetizado en vacuolas; 3) una duplicación final de la subunidad **a**, cuya tasa había sido 1, derivó en una tasa de 4 H⁺ extraído/ATP sintetizado para algunas arqueas^{310,311}; ello implica que la ATP sintasa de bacterias actuales tiene una tasa 3 ATP hidrolizados por cada 12 protones que son trascolados³¹².

- Se ha propuesto que las subunidades **α** y **β** del módulo análogo a F₁ en cada una de las ATP sintasas existentes se originaron por la duplicación de un gen ancestral³¹² y se ha sugerido que todas las enzimas de este tipo pudieron haber estado presentes en el LUCA; sin embargo según estudios basados en análisis de las relaciones filogenéticas de las diversas ATP sintasas, la transferencia horizontal de genes (HGT) puede explicar mejor los datos³¹³; la HGT no es descartable y la hipótesis fus-inv como origen de los distintos tipos de ATPasas es plausible, de hecho posiblemente hubo una combinación de ambos procesos.
- Observamos entonces que ambos módulos de ATPasas se componen de subunidades que llevan a cabo funciones distintas y existen varios modelos para explicar el origen y evolución de esas enzimas. Por otra parte, los dos módulos de una ATPasa F₀F₁ pueden estar disociados *in vitro*, de modo que F₁ es sólo una ATPasa y F₀ un acarreador pasivo de H⁺, aunque dichas funciones no tienen relevancia funcional *in vivo* actualmente^{304,314}; adicionalmente, dentro del operón de la ATPasa F₀F₁ algunos genes de las distintas subunidades se traducen de modo acoplado, donde el vínculo más estrecho es entre los genes denominados *atpH/atpE* (subunidad **c**, de ~8.5

kDa actualmente), y *atpA* (subunidad α , de ~55 kDa actualmente)^{315,316}; ambos aspectos son relevantes para la presente investigación, ya que solamente la subunidad c de F_0 que podría conformar el protonóforo y la α de F_1 que une ATP aunque no lo hidrolice, fueron identificadas como ancestrales al contener regiones codificadas por tripletes RNY.

- **Subunidad α de F_1 .** El módulo F_1 es un complejo esferoide en que 3 subunidades α y 3 β se encuentran intercaladas³¹⁷. La subunidad α es la proteína más grande de la ATPasa F_0F_1 , ya que comprende ~1539 nt, ~513 aa y ~55kDa, de los cuales son 46% de hélices- α y 19% de hebras- β ; cuando se observa el uso de codones y la composición de aa de la subunidad α ^{318,319}, es posible notar que son numerosos los codificados por tripletes RNY. Tanto la subunidad α como la β contienen sitios de unión a ATP, sin embargo sólo la subunidad β hidroliza el NTP y por tanto sólo ésta se considera relevante para la catálisis de la ATPasa F_0F_1 , se ha señalado entonces que α sirve actualmente sólo como subunidad regulatoria de la hidrólisis del ATP^{320,321}. De acuerdo con análisis comparativos, las subunidades α y β pudieron haber surgido por duplicación de un gen ancestral, ya que son sumamente homólogas tanto en secuencia como en estructura; ambas subunidades comparten una estructura común compuesta por un dominio-N pequeño y un dominio-C helicoidal, entre estos existe un dominio central de unión a nucleótidos que es la región con mayor semejanza entre ambas subunidades (aproximadamente residuos 60 a 130 de α), denominado bucle-P que corresponde al motivo Walker A y que junto con Walker B lineales forman el plegamiento Rossmann en 3D; dicha estructura la comparten todas las enzimas que unen nt. Al comparar subunidades α de diferentes organismos se ha observado que son idénticas

en 59% entre sí, pero sólo 20% idénticas con respecto a las subunidades β , siendo más ordenado el bucle-P de la α que de la β , además el residuo Q²⁰⁸ presente en β , más no en α , se ha señalado como responsable de la hidrólisis del ATP^{322,323,324,325,326}. El consenso de los motivos Walker A y B son **G**_{x4}**GKT**_{x6}**[I/V]** y **[R/K]**_{x3}**G**_{x3}Lh₄**D** (h es aa hidrófobo), respectivamente³²²; mientras que la forma ancestral del bucle-P se considera que es h**GxGGxGK****[T/S]**³²³. Otros residuos involucrados en la hidrólisis del NTP son: el K¹⁷⁵, localizado en la subunidad α e involucrado en la unión al nt³²⁷; los residuos P²⁸¹, **A**²⁸⁵, R³⁷⁶, **S**³⁷³ son importantes y en algunos casos esenciales para la regulación de la hidrólisis del ATP³²⁸; el **D**²⁶¹ es un residuo clave en la unión no catalítica del NTP cuando forma un complejo con el ion Mg²⁺ y el residuo R³⁶⁵ une la adenina del ATP³²⁹; finalmente, de la secuencia conservada **VISITDG**, **S**³⁴⁷ está involucrado en la unión del fosfato γ del ATP mientras que se desconoce la función específica de **G**³⁵¹ aunque está marcadamente conservado³³⁰. Generalmente se considera que la proteína ancestral catalizaba la hidrólisis del ATP y semejaba más a la β , mientras la α perdió la función catalítica en la duplicación quedando como reguladora; no obstante sólo la subunidad α contiene una región codificada por tripletes RNY y no así la β , lo que sugiere fuertemente que es la unión del ATP y no su utilización la característica ancestral de esta proteína; en particular es relevante señalar que el aa que se identificó como principal responsable de la hidrólisis del ATP en la subunidad β , no está presente en α y tampoco es codificado por tripletes RNY; mientras que varios de los aa de los motivos Walker A y B involucrados en la unión del ATP son prebióticos y están codificados por tripletes RNY.

- Subunidad c de F₀.** El módulo F₀ se encuentra embebido en la membrana mediante la subunidad **c** que, a pesar de ser la subunidad más pequeña de la ATPasa F₀F₁ con ~80 aa y ~8.1kDa, se homo-oligomeriza hasta formar en un anillo de 9 a 12 proteínas a través del cual pueden pasar los protones, por lo cual se puede denominar como protonófero, el cual tiene forma de reloj de arena; cada subunidad **c** es de forma oval y es altamente hidrófoba con 68% aa no polares, en particular hay dos segmentos de 20 a 25 aa hidrófobos separados por (al menos) los aa L³¹**GGKFLEGAARQ**⁴², todo ello conforma una topología hélice-bucle-hélice que se inserta en la membrana como una horquilla mientras el bucle queda expuesto al citoplasma^{331,332,333,334,335}. De cada subunidad **c**, la hélice-C es altamente hidrófoba mientras que la hélice-N es rica en **A** y **G**, de modo que en el homo-oligómero los segmentos helicoidales forman dos círculos concéntricos de 10 Å y 55 Å de diámetro, con los dominios-N al interior y los dominios-C al exterior, mientras 4 o 5 bucles forman parte del enlace físico con el módulo F₁; existe un alto grado de similitud entre las subunidades **c** de diferentes organismos, pero el único aa invariante es el aa **D**⁶¹ localizado en la región transmembranal de la proteína, aproximadamente a la mitad de la hélice-C que lo contiene, e involucrado directamente en la trascolación de H⁺/Na⁺ ^{336,337,338,339,340}. Alrededor del **D**⁶¹ invariante existe un motivo estructural, conformado por los aa conservados **V**⁷⁸, **A**⁷⁷, M⁷⁵, **G**⁷¹, **V**⁶⁸, C⁶⁷ e **I**⁶³, que provee una superficie para el tránsito de los protones³⁴¹; mientras que **A**²⁴, **I**²⁸ y **A**⁶² están ubicados espacialmente en el sitio de unión al protón^{337,342,343}. Otros aa invariantes son **G**²³, **G**²⁷ y **G**⁵⁸ que se consideran críticos para la función de **c** como protonófero³⁴⁴. Una substitución en L⁶⁴ de **c** causa la pérdida de la fosforilación oxidativa, debido a que el complejo de la ATPasa F₀F₁ no puede

ensamblarse adecuadamente³⁴⁵. Como se puede observar la subunidad **c** no posee ninguna actividad catalítica, sino que es un ensamblaje directamente involucrado en la conducción de H^+/Na^+ , a pesar de lo cual la filogenia basada en subunidades **c** (y sus homólogos) de diferentes organismos concuerda en lo general con la filogenia basada en el 16SrRNA³⁴⁶, lo que indica que probablemente dicha subunidad era parte de no sólo del LUCA sino de un organismo anterior. De la subunidad **c**, la región que más probablemente se identificó como ancestral es la que corresponde al **D**⁶¹ invariante y el motivo estructural alrededor, ya que los aa de esa zona son en su mayoría codificados por tripletes RNY, mismos que fácilmente hubiesen interaccionado con moléculas sencillas, totalmente prebióticas y vitales como son los protones.

- Si la atmósfera terrestre primigenia era reductora, debió resultar altamente favorable para los primeros organismos poseer un complejo molecular que al mismo tiempo evitara el exceso de acidez al interior de esas proto-células y mantuviese un gradiente de H^+ que podía utilizarse para diversas reacciones, aún si dicho proceso fuera en detrimento de una molécula energética como el ATP, que pudo haberse sintetizado prebióticamente^{cfr.23}; las proteínas de ese conjunto se fijarían entonces como esenciales y sus genes formarían parte de los llamados 'de mantenimiento' o "housekeeping". Con base en los presentes resultados, el complejo mencionado tiene sus orígenes en el Mundo de las RNPs en forma de primordios de las subunidades **α** y **c**, mismos que corresponden respectivamente: por un lado un péptido que unía ATP que como se ha mencionado es posible que existiera prebióticamente, dicho péptido formaría la subunidad **α** que mediante duplicación y mutación originaría la

subunidad β que hidrolizaba la molécula; por otro lado un péptido que unía H^+/Na^+ que, posiblemente mediante copias de sí mismo, habría formado un poro en la proto-membrana a través del cual pasaran esos cationes prebióticos. Posteriormente se habrían conformado otras subunidades y finalmente se ensamblaría una ATPasa F_0F_1 que se integraría en el repertorio molecular de LUCA.

- NADH OXIDASA FORMADORA DE AGUA.
- Las proteínas relacionadas con la transferencia de electrones mediante moléculas nucleotídicas son probablemente de las primeras clases que se formaron; en particular la enzima NADH oxidasa formadora de agua (NOXasa), que une NADH y puede unir FAD, contiene una región codificada por tripletes RNY.
- La NOXasa, cuya reacción catalizada es: $2NADH + 2H^+ + 2O_2 \rightarrow 2NAD^+ + 2H_2O$, pertenece a la familia de las flavoenzimas; aparentemente dicha reacción no es eficiente para la preservación de energía, sin embargo el organismo puede obtenerla mediante la utilización de cadenas de fosforilación de una amplia variedad de sustratos. Además de las características mencionadas, existen particularidades bioquímicas y estructurales que permiten distinguir las NOXasas de distintas fuentes; se han encontrado NOXasas en bacterias del tipo anaerobias fermentativas, aerobias facultativas y aerobias obligadas, así como en arqueas que son extremófilas anaerobias; algunas NOXasas requieren la molécula FAD como cofactor en cada subunidad, mientras que para otras NOXasas la adición o restricción de FAD no parece afectar su funcionamiento; típicamente la NOXasa une NADH y libera agua, pero existen NOXasas que pueden, tanto unir NAD^+ o NADPH, como liberar además H_2O_2 ; estructuralmente las NOXasas son monómeros de ~50kDa,

pero también existen dímeros, tetrámeros y hexámeros, inclusive en un mismo organismo donde aparentemente cada conformación cataliza reacciones distintas aunque interrelacionadas^{347,348,349,350,351,352,353,354,355,356,357,358,359,360,361,362,363,364,365,366,367,368,369,370,371,372,373,374,375,376}.

- Aunque la NOXasa y la NADH peroxidasa (NPXasa,) están estructural y funcionalmente relacionadas, la NOXasa es más interesante bioquímicamente hablando³⁷⁵; pues mientras la NPXasa genera H₂O₂ que produce estrés oxidativo a la célula y para evitarlo deben entrar en juego otras enzimas, la NOXasa reduce el potencial tóxico del oxígeno, ya que cataliza su oxidación completa hasta H₂O, permitiendo así que cualquier organismo que la posea sea capaz de resistir ambientes oxidantes y aún crecer en estos^{349,353,355,357,359,360,364,366-368,370-373} y simultáneamente resistir ambientes con bajo pH³⁷⁴; debido a dicha aerotolerancia, se ha encontrado que gran cantidad de microorganismos utilizan la NOXasa durante la competencia y la patogenicidad^{358,361-363}, incluyendo el grupo GBS³⁶⁹ al que pertenece **SagA909**; la NOXasa además regenera los niveles de NAD⁺/NADH que pueden reingresar al catabolismo de los carbohidratos, estimulando así la glucólisis^{353,356,357,359,365,367,368,373-376}.
- El sitio catalítico de las NOXasas es un centro redox invariablemente conservado que corresponde a la Cys⁴² de la NOXasa de **Enterococcus faecalis 10C1**³⁷⁷; sin embargo, dado que C no es codificado por tripletes RNY y tampoco es un aa prebiótico, es poco probable que dicho centro redox haya sido identificado como ancestral. Adicionalmente, a pesar de las diferencias mencionadas entre NPXasa y NOXasa, ambas NADH oxidasas comparten gran similitud en 3 regiones, dos involucradas en la unión de las moléculas flavonoides FAD y NAD, así como la región central antes mencionada que contiene el centro cisteinil redox, localizada entre ambos

dominios en NOXasas de distintas fuentes; también se encontró una región conservada que es esencial tanto para la unión de FAD como para el correcto plegamiento de la NOXasa; otras flavoproteínas que unen FAD y NAD(P) también contienen regiones similares a las mencionadas^{349,350-354,366,370,372,373,375,376}. Conjuntamente, debido al alto grado de similitud interna del plegamiento de unión para FAD y NAD(P)H (y posiblemente también NADH) en las flavoenzimas dependientes de nucleótidos del tipo piridina, se ha sugerido que tales enzimas surgieron por la duplicación interna de una proteína ancestral, conformada en un dominio $\beta\alpha\beta$ que actualmente interacciona con la parte adenosil-fosforilo de la coenzima, dicha región contiene una secuencia homóloga al consenso **GxGxxG** presente en varias proteínas de unión a nucleótidos^{378,379}. Por otra parte la diferenciación de especies de cierto género de bacterias con base en el gen de la NOXasa concuerda con la clasificación basada en el 16SrRNA³⁸⁰, lo que no sorprende dado que se prueba, según esta tesis, que es un gen muy antiguo cuya región ancestral era ya codificada en el Mundo de las RNPs.

- Dado que aparentemente no todas las NOXasas unen FAD, pero su característica intrínseca es unir NADH, es probable que la región identificada como ancestral corresponde a los vestigios del péptido antes de la duplicación antes mencionada, aunque semejando más a la región que une NAD. Posteriormente dicho fragmento peptídico evolucionó en complejidad hasta transformarse en una enzima, de modo que si la NOXasa es tan ancestral que un organismo que existiese en una atmósfera reductora como la de la Tierra primitiva pudo tenerla, el medio interno de tal microorganismo no habría sido tan ácido, generando además suficiente diferencia de potencial, entre ambos lados de la incipiente membrana, para provocar un gradiente de protones.

- GLICEROL-3-FOSFATO DESHIDROGENASA.
- El glicerol (1,2,3-propanotriol) es una pequeña molécula simple pero esencial cuya síntesis prebiótica es relativamente sencilla y eficiente y por tanto es de las primeras moléculas en aparecer; el glicerol actualmente resulta del catabolismo de la glucosa, de las proteínas, del piruvato, del triaciglicerol y de otros glicerolípidos, así como de la liberación de grasas provenientes de la dieta; el glicerol es una molécula de alto valor adaptativo en varias especies, los procariontes por ejemplo utilizan el glicerol como fuente de carbono, para la producción de energía, osmorregulación y crioprotección, la importancia metabólica del glicerol es particularmente notoria en bacterias que crecen en condiciones anaeróbicas o aeróbicas carentes de glucosa; el metabolismo del glicerol conlleva a una conversión de exceso de NAD(P)H en NAD(P)⁺, siendo entonces esencial para balancear el potencial redox de las células, además el glicerol está involucrado en el reciclaje del fosfato inorgánico utilizado en la glicólisis; finalmente, el glicerol es precursor, entre muchos otros compuestos, del glicerol 3-fosfato (G3P), una molécula versátil y potencialmente energética. Dado que los procesos metabólicos en que está involucrado el glicerol son muy similares en una amplia diversidad de organismos, es altamente probable que tales procesos compartan un origen común y antiguo³⁸¹; dada la importancia del G3P, existen tres glicerol 3-fosfato deshidrogenasas (G3PDH) que catalizan su síntesis a partir del fosfato de dihidroxiacetona (DHAP), una aeróbica y otra anaeróbica, ambas dependientes de FAD, así como una oxidorreductasa basada en NAD(P)⁺ (G3PDH-NAD(P)⁺)³⁸², identificada como ancestral por el método aquí descrito, que no depende de la cantidad de O₂. La proteína G3PDH-NAD(P)⁺, contiene una región **codificada por tripletes RNY y se le considera parte del grupo "producción y**

conversión de energía” porque está involucrada en el metabolismo del glicerol y en el metabolismo del piruvato, molécula a partir de la cual es posible obtener energía, independientemente de la acetil-coenzima A (AcCoA), mediante metabolismo fermentativo.

- El glicerol 3-fosfato (G3P) es el precursor requerido para la síntesis de todas las especies de fosfolípidos y la enzima G3PDH-NAD(P)⁺, codificada por el gen *gpsA* o *gldA*, cataliza la reducción reversible de DHAP a G3P, utilizando NAD(P)⁺ como aceptor de electrones que se transforma en NAD(P)H, dicha enzima también cataliza la oxidación del propano-1,2-diol y con mucha menor afinidad la reducción del sulfato de glicerona. La mencionada G3PDH-NAD(P)⁺ consta de un peso molecular de ~40 kDa, aunque inicialmente se determinó que constaba de 65 kDa a 80 kDa y que posiblemente estaba conformada por subunidades^{383,384,385}. Especialmente la G3PDH-NAD(P)⁺ se localiza en la superficie interna de la membrana celular en procariontes ya que el G3P, además de ser fundamental para sintetizar los fosfolípidos, también estimula el transporte de otras moléculas pequeñas y esenciales, como succinato y aa³⁸⁶.
- Dada la importancia del G3P *per se*, dado que su síntesis requiere muchos recursos energéticos y está acoplada a la síntesis de la membrana, es necesario evitar la sobreproducción de G3P; de tal modo que la síntesis de dicha molécula se encuentra firmemente regulada, ya sea a nivel enzimático o a nivel genético, dependiendo del tipo de organismo y en ambos casos cuando hay un excedente del compuesto³⁸¹⁻³⁸⁶; en algunas especies de procariontes el G3P se une a la G3PDH-NAD(P)⁺ en un sitio alostérico, actuando como inhibidor competitivo de la DHAP^{387,388,389}; en otros microorganismo el G3P parece ser el factor de regulación de la expresión de su propio gen³⁹⁰.

- Una de las principales diferencias estructurales entre bacterias y arqueas se asumía que era la membrana, pues mientras que las primeras utilizan G3P las segundas utilizan G1P, debido principalmente a la diferencia de sus hábitats y por tanto sus GPDH son distintas y aparentemente sin relación entre sí; sin embargo el arquea hipertermófila *Archaeoglobus fulgidus* posee una G3PDH que, a diferencia de los ortólogos de fuentes bacterianas, es termoestable a 85°C y posiblemente es resultado de un proceso de HGT, aunque bien pudo ocurrir que la gran mayoría de las arqueas perdieron su G3PDH³⁹¹.
- La secuencia de todas las G3PDH, incluyendo la de arqueas, se componen de dos segmentos: el de unión a (NAD(P))⁺ en el dominio-N y el de unión a glicerol en el dominio-C, por un lado el dominio **GxGxxG** que une el cofactor es homólogo al de otras enzimas que unen moléculas nucleotídicas, mientras que el dominio que une el glicerol es catalítico y parece ser homólogo al sitio que une el gliceraldehído 3-fosfato (GA3P) o el lactato, en la deshidrogenasa de gliceraldehído 3-fosfato (GA3PDH) y en la deshidrogenasa de lactato respectivamente; otros sitios conservados de la G3PDH-NAD(P)⁺ son K¹⁰⁵, K¹⁹² y D²⁵⁰ en el sitio de unión al G3P y N¹⁹³, T²⁵⁴ y N²⁶⁰ alrededor del 3-fosfato^{392,391}.
- Se ha descrito un modelo de evolución prebiótica en el que una cadena de G3P y GA3P se autorreplicaría por complementariedad y sería resistente a errores de copiado, de forma análoga a una cadena de nucleótidos pero cuyos enlaces serían covalentes y no mediante puentes de hidrógeno³⁹³.
- La autorreplicación del polímero prebiótico de G3P-GA3P, generado por el método descrito anteriormente, no requiere una molécula adicional que le confiera estabilidad, además no habría ofrecido materia prima para la evolución dado que es resistente a mutaciones, pero si habría constituido

una plataforma estructural con estabilidad química a largo plazo, a partir de la cual o podrían generarse moléculas más complejas, o bien dicho polímero podría haber fungido como un reservorio de materia prima para procesos metabólicos que aparecerían ulteriormente en la evolución. Por todo lo anterior, es posible que de los dos dominios de la enzima G3PDH, el identificado como ancestral sea el que une NAD(P)^+ , ya que además de la estabilidad *per se* de polímeros de G3-GA3P, el consenso del sitio de unión a nucleótidos está constituido por aa codificados por tripletes RNY.

▪ PIROFOSFATASA DEPENDIENTE DE MANGANESO.

- La pirofosfatasa inorgánica dependiente de manganeso (Mn^{2+} -iPPasa), identificada como ancestral, cataliza la hidrólisis de la molécula química no-orgánica más pequeña y más importante para las células: el pirofosfato inorgánico (PPi), una sencilla molécula esencial para la vida capaz de sustituir al ATP bajo condiciones de estrés ambiental y del cual se considera su precursor en la evolución de la bioenergética; el Pi obtenido provee además el impulso termodinámico para numerosas reacciones biosintéticas y puede servir para diversas reacciones de fosforilación. La clasificación original de las iPPasas se basaba en su fuente biológica, por lo que existía la clase A de eucariontes y la clase B de procariontes, aunque ambas clases contienen Mg^{2+} como cofactor; posteriormente se describió otra clase totalmente distinta de iPPasa, denominada entonces clase C, que sólo existe en procariontes pero que depende de Mn^{2+} para su correcto funcionamiento. Más adelante se cambió la clasificación con base en el catión metálico de las iPPasas, por lo que describieron dos clases aparentemente sin ninguna relación filogenética entre sí: la clase I dependiente de Mg^{2+} presente en eucariontes y procariontes y la clase II

dependiente de Mn^{2+} que sólo se ha descubierto en procariontes, tanto en bacterias como en arqueas^{394,395,396}.

- Los primeros 39 aa de la Mn^{2+} -iPPasa corresponden a la secuencia ¹MEKILIFGHQNPDTDTIXAIAYADLKNKLGFNAPVR³⁹L, mientras que la enzima completa consta de 310aa y ~35 kDa por subunidad, siendo menos activo el dímero que la forma el trimérica, cuya formación depende de la concentración del catión metálico y del pH. La especificidad de la Mn^{2+} -iPPasa es más alta que la de las Mg^{2+} -iPPasas y su actividad catalítica de un orden mayor de magnitud, con las que sólo comparte 2/3 de similitud a lo largo de la secuencia; de hecho de los 13 sitios funcionalmente activos, solamente dos están conservados en ambos tipos de iPPasas. La estructura tridimensional de ambas familias de iPPasas es también totalmente distinta, con excepción del arreglo espacial idéntico del sitio activo, que contiene invariablemente **9D**, 3E, 4K, 1R, **2N**, 3Q, 4H, **4T** y **5S**, lo que sugiere que ambos mecanismos son homólogos, aun cuando las dos familias iPPasas no están filogenéticamente relacionadas; adicionalmente existen varias regiones funcionales conservadas, destacando la región ²⁹⁴**SRKKQVV**³⁰¹P de 8 aa cerca del extremo-C, lo que llevó a sugerir que dicha secuencia puede fungir como identificador de las Mn^{2+} -iPPasas^{394,395,397}.
- La estructura de los primeros 189 aa del dominio-N de las Mn^{2+} -iPPasas es un plegamiento α/β , mientras que los últimos 114 del dominio-C tienen una conformación de hojas- β mixtas. En la región entre estos dos dominios está el sitio activo, formado por aa polares de ambos, donde **5D**, 2H, 1E y **2S** participan directamente en la unión del Mn^{2+} , del substrato, un sulfato y agua, mientras que la participación de 1K y 1R es indirecta. El dominio-N en particular activa los cationes metálicos mediante sus **5D**, estos iones al unirse favorecen la oligomerización de la enzima e incluso pueden cambiar

la geometría estructural de su sitio activo, lo que permite una catálisis más eficiente tanto de la hidrólisis del PPI, unido directamente por el dominio-C, como de la liberación de los productos, la enzima es además altamente sensible a las mutaciones de estos residuos, por lo que se considera que dichos **Asp** conforman la región fundamental del sitio activo. Adicionalmente, las iPPasas de la familia II están relacionadas estructuralmente con la familia de fosfoesterasas DHH^{396,398,399,400}.

- La región entre el dominio-N y el dominio-C de las iPPasas de la familia II posee alta flexibilidad estructural, por lo que funge como bisagra, lo que conlleva a que la Mn²⁺-iPPasa pueda existir tanto en conformación abierta como cerrada en distintas posiciones; es posible además que las Mn²⁺-iPPasas sean alostéricas, ya que las posibles configuraciones espaciales confieren muy alta especificidad por el PPI con respecto a otros polifosfatos³⁹⁶; dicha bisagra estructural contiene los aa **AGTN**, de los cuales **A**¹⁸⁹, **G**¹⁹⁰, [**A/T**]¹⁹¹ y **K**¹⁹⁶ están altamente conservados en las iPPasa de la familia II de diversas bacterias y tanto la longitud como la composición de esta región afecta dramáticamente la actividad enzimática de la enzima⁴⁰¹.
- Varios de los aa invariantes de la estructura de la Mn²⁺-iPPasa están codificados por tripletes RNY, sin embargo éstos se encuentran muy alejados uno del otro a lo largo de la secuencia y no es posible dilucidar cuál de ellos proviene del Mundo de las RNPs, además el substrato y los distintos cofactores son muy posiblemente de origen prebiótico; por otro lado, la flexibilidad mecánica de la bisagra es esencial para los pasos catalíticos subsecuentes y todos los aa de esa zona son codificados por tripletes RNY, es probable entonces que ésta haya sido la región identificada como ancestral o bien los **Asp** conservados del sitio activo.

o **METABOLISMO DE CARBOHIDRATOS.** Se han realizado numerosas propuestas evolutivas en que las enzimas del ciclo reverso de los ácidos tricarbónicos (rTCA) y de la gluconeogénesis son proteínas primordiales^{402,403,404,405}; sin embargo es notable que ninguna de ellas fue identificada como ancestral por el método descrito en esta tesis y por otro lado las que contienen regiones codificadas por tripletes RNY no han sido jamás incluidas como enzimas glicolíticas antiguas, por lo tanto no se describen los mencionados trabajos. Pese a lo anterior, todas las enzimas incluidas en este grupo unen moléculas glicosiladas o bien tienen alguna relación con la glucólisis/gluconeogénesis; además el anabolismo de la glucosa se conecta con el metabolismo de otros compuestos glicosilados como los aminoazúcares, el uronato o el glucógeno, que a su vez desembocan en otras rutas metabólicas distintas como el metabolismo de las purinas, o la síntesis de glicolípidos o glicoproteínas.

- ANABOLISMO DEL HEXURONATO.
- Los ácidos urónicos pueden servir como única fuente de energía en las bacterias, para lo cual una reacción de isomerización siempre precede a una reducción⁴⁰⁶, orden en el cual se encuentran los genes que codifican para las enzimas que catalizan dichas reacciones, aun tratándose de organismos distintos, mientras que la región regulatoria respectiva puede tener un ordenamiento distinto^{407,408,409}; las enzimas involucradas en los dos procesos antes mencionados (isomerización y reducción) contienen regiones codificadas por tripletes RNY. La proteína UxaC cataliza la isomerización de glucuronato a fructuronato y de galacturonato a tagaturonato, utilizando iones metálicos como cofactor; mientras que la proteína UxuB cataliza la reducción de fructuronato para formar manonato y de tagaturonato para formar altronato, utilizando cofactores

nucleotídicos^{410,411}, no obstante en algunos organismos UxaC sólo pertenece al grupo de la utilización del galacturonato y UxuB solamente al grupo de utilización del glucuronato⁴¹².

- **UxaC** o isomerasa de glucuronato tiene un peso molecular de ~50kDa, se sintetiza a un nivel basal relativamente alto y es aparentemente el primer gen transcrito en el operón *exu*. UxaC cataliza el primer paso en los metabolismos de los ácidos glucurónico y galacturónico; es decir cataliza la isomerización del ácido glucurónico y del galacturónico a sus análogos ceto correspondientes fructuronato y tagaturonato, utilizando Zn^{2+} , Co^{2+} o Mn^{2+} como cofactores^{413,414,415}. El monómero consiste de una cadena polipeptídica de 451 aa, compuesta de 25 hélices y 9 hebras, dividida en dos dominios. El sitio activo se forma por los aa H³⁰, H³², W³⁶⁶ y D³⁹⁷, localizados en ambos dominios, los cuales coordinan un ion metálico y sólo D³⁹⁷ participa en la reacción de isomerización; el dominio A, cuyo homólogo estructural más cercano (70%) es una fosfotriesterasa, contiene dos segmentos no-contiguos (0 al 42 y 135 al 450), de hecho se considera que la enzima UxaC es un miembro no-hidrolítico de la superfamilia de las amidohidrolasas; el dominio B comprende la región intermedia de la enzima (43 al 134) y carece de homólogo conocido^{416,417,418}.
- **UxuB**, u oxidoreductasa del D-manonato, cataliza la respectiva reducción del ácido fructurónico y del ácido tagaturónico a D-manonato y D-altronato, utilizando NADH y/o NADPH como cofactores, los cuales pueden unirse de manera reversible a la enzima, siendo ésta 5% más activa con NADPH que con NADH; después de la unión de los cofactores se cataliza la reducción de los uronatos, reacción que es reversible, inducible por los substratos y reprimible por la presencia de glucosa^{419,420,421,422}. Posiblemente exista un

ancestro común de UxuB y una enzima denominada UxaB, específica sólo para catalizar la conversión de tagaturonato a D-altronato⁴²³.

- Dado que los uronatos y sus derivados son complejos, es poco probable que sean moléculas prebióticas y por tanto que la región que actualmente los une esté codificada por tripletes RNY. De modo que es posible que la región identificada como ancestral de la enzima UxuC sea la que contiene el **D**³⁹⁷, ya que éste contribuye a la unión del ion metálico Zn^{2+} que si es prebiótico; por otro lado, es probable que la región codificada por tripletes RNY de la enzima UxaB sea la que une el cofactor nucleotídico, ya que se ha mostrado que ese tipo de moléculas son tan ancestrales que pudieron estar presentes antes incluso de los albores del Mundo de las RNPs.

- ANABOLISMO DEL GLUCÓGENO.

- Se considera que el glucógeno es un reservorio de carbono y energía, cuya síntesis se lleva a cabo bajo condiciones limitadas de crecimiento bacteriano, en presencia de exceso de fuentes de carbono y normalmente hasta la fase logarítmica tardía o estacionaria temprana, dicha reserva sirve además para hacer frente a condiciones temporales de restricción nutricional o estrés hiperosmótico. La expresión de los genes para la biosíntesis del α -1,4 glucano está acoplada con la expresión de proteínas requeridas para la utilización de otras fuentes nutritivas; dos de los genes del anabolismo del glucógeno, GlgA y GlgC cuya regulación es co-dependiente, son absolutamente necesarios para su biosíntesis, éstos se encuentran adyacentes en su operón, cada uno comprende ~1.3 kpb y ~50kDa, ambos son transcritos en sentido inverso a las manecillas del reloj^{424,425,426,427,428,429,430,431} y además parte de ellos es codificado por tripletes RNY.

- La **GlgA** o sintetasa de glucógeno es codificada por el gen *glgA* que consta de poco más de 1.4 kpb y está localizado cerca del origen del cromosoma; la enzima, de cerca de 500 aa con peso molecular de ~50kDa, cataliza la elongación del glucógeno, mediante la adición una molécula de glucosa (Glc) a una molécula cebada (primer) de glucano utilizando ADP como cofactor^{432,433,434}; la fuente de ADP-azúcar puede ser la enzima GlgC u otras fuentes enzimáticas, dependiendo de la fuente de carbono disponible⁴³⁵. La región crítica de la enzima GlgA contiene tres Cys, de las cuales sólo la C³⁷⁹ está involucrada directamente en la catálisis⁴³⁶; por otro lado el aa K¹⁵ (numeración basada en *Escherichia coli*) se encuentra también en el sitio activo de la enzima GlgA en una secuencia Kx**GG**, dado que la secuencia **Gx₄GK** está conservada en varias proteínas que unen ATP y GTP, es posible que la K¹⁵ corresponda a esa Lys, de lo cual se desprenden entonces dos posibilidades: 1, que la Lys interactúe directamente con la parte fosfato del substrato ADP-Glc y las Gly sólo provean flexibilidad estructural a esta región o 2, que las **Gly** interactúen directamente con la parte del polifosfato⁴³⁷; adicionalmente otra Lys en el sitio activo es K²⁷⁷ que está situada directamente en el sitio catalítico⁴³⁸. La GlgA es una glicosiltransferasa de la familia 5 (GT-5) que transfiere la glucosa del complejo ADP-glucosa al glucano reteniendo la configuración del carbón anomérico, por lo que se considera una GT retentiva (GT-B) (más detalle sobre las GT en 'Proteínas, procesamiento y modificación'); al comparar la GlgA con otras GT-5 (de plantas y bacterias) y con GT-3 (de mamíferos y levaduras) se observan 5 regiones con características similares que funcionan como reguladores, además de un motivo ⁵¹⁰E-x₇-⁵¹⁸E conservado del que se desconoce su función aunque los Glu son invariantes; cuando se evalúa la similitud en secuencia sólo de la familia GT-5, se encuentra que los aa H¹⁶¹,

D¹³⁷, **R**³⁰⁰ y **K**³⁰⁵ también están conservados y son esenciales para la catálisis^{439,440}. Estructuralmente la GlgA, como otras GB-T tanto de bacterias como de arqueas, se forma por dos plegamientos Rossmann, siendo la hendidura entre ellos donde se localiza el sitio catalítico, lo cual sugiere que la unión del sustrato induce cambios conformacionales y la enzima puede encontrarse en posición abierta o cerrada; mientras que la glucosa del ADP-Glc interactúa directamente sólo con **N**²⁴⁶, el bolsillo de unión al ADP se encuentra en la región C-terminal del inter-dominio donde el heterociclo de adenina es estabilizado por **Y**³⁵⁴, **S**²⁹⁸ y **S**³⁵⁹, mientras que el átomo O2 de la ribosa se une mediante puentes de hidrógeno a **I**²⁹⁷, el OH⁻ a **T**³⁸¹ y los grupos fosfato a **R**²⁹⁹ y **K**³⁰⁴/**K**³⁰⁵ (dependiendo del organismo). Cuando al ADP se une el sitio catalítico es activado y la enzima pasa a la conformación cerrada, hacia dicho sitio se orientan **H**¹⁶¹, **R**³⁰⁰, **K**³⁰⁵ que son residuos críticos para la transglicosilación, así como **E**³⁷⁷ que juega un papel electrostático para orientar adecuadamente el anillo glucosídico; los cambios conformacionales incluyen el acercamiento del dominio-N donde se encuentra el motivo conservado **KxGGL** del que las Gly hacen contacto con los grupos nucleótidos-fosfato y la Lys con la ribosa, aunque sólo las glicinas y un aspartato parecen ser esenciales para la catálisis, en particular la **G**¹⁸ interacciona con el fosfato distal y el **D**²¹ con el átomo O2 de la ribosa^{441,442,443,444,445}.

- La **GlgC** o pirofosforilasa de ADP-glucosa cataliza la formación reversible del oligosacárido ADP-glucosa (mismo que será donado a la GlgA y la reacción procede entonces hacia la biosíntesis) a partir de ATP y alfa-glucosa-1-fosfato (α Glc1P) liberando PPI, por lo que también se le conoce como adenililtransferasa de glucosa-1-fosfato; dado que al incrementar la síntesis de ADP-glucosa incrementa la síntesis del producto, podemos considerar

que éste es el paso limitante en la elongación de glucógeno. La enzima GlgC, cuya estructura es α_4 , consta de cuatro subunidades de ~50kDa cada una, codificadas por ~1.3 kpb y poco más de 400 aa; cada una de las subunidades está conformada por tres regiones: los dominios -N (~10 a 20 aa) y -C (~los últimos 100 aa), que contienen sitios regulatorios para la unión específica de activadores e inhibidores, así como la parte central (~20 a 300 aa) que comprende el módulo esencial catalítico de la enzima, el cual semeja al de otras enzimas que unen nucleósidos-di-fosfatos (NDP). GlgC es regulada alostéricamente por activadores (act) e inhibidores (inh), cuyos sitios de unión se encuentran en las dos regiones terminales de la enzima. Por lo general los activadores (act) son metabolitos glicolíticos de alto peso molecular y alto contenido de carbono (3-fostoglicerato (3PG), hexosas-6-fosfato (Fru,Man,Rib,Glc6P), fructosa-1,6-bisfosfato (FruBP), pentosas-5-P (como Rib5P), GA3P, piruvato y fosfoenolpiruvato (PEP); mientras que los inhibidores (inh) son de bajo peso molecular y por lo **general se trata de los metabolitos finales de todo proceso metabólico (5'-AMP, ADP, PPI o Pi)**. Existen 8 diferentes clases de GlgC, dependiendo del tipo de metabolismo (met) del microorganismo que la posee, lo que también influye en el tipo de reguladores: 1) met: Embden-Meyerhof; act: FruBP; inh: AMP; 2) met: Embden-Meyerhof; act: FruBP y Fru6P, inh: ADP y AMP; 3) met: Embden-Meyerhof; act: no parece ser sensible a activadores; inh: AMP; 4) met: Entner-Doudoroff; act: FruBP y piruvato, teniendo el primero mayor influencia; inh: ADP, AMP y Pi; 5) ambas rutas antes mencionadas; act: FruBP, Fru6P y piruvato; inh: no se han descrito; 6), met: anaerobiosis heterótrofa y fotosíntesis anoxigénica autótrofa; act: piruvato; inh: no se conocen; 7) esporulación bacteriana; aparentemente la GlgC correspondiente no es regulada; 8) fotosíntesis oxigénica; act: 3PG en

proporción al inh: Pi. Usualmente los activadores se unen en el dominio-N y aunque algunas pocas GlgC contienen sitios de unión a inhibidores en el dominio-C, la regulación alostérica resulta en realidad de la interacción de ambos dominios, lo que provoca cambios en la conformación estructural de la enzima^{446,447,448,449,450,451,452,453,454,455,456,457}. El dominio-N de la enzima GlgC contiene una considerable cantidad de aa hidrófobos como Leu, **Val** y Pro; en la activación alostérica participa un ion Mg^{2+} y la mayoría de los aa que se encuentran en la interface entre los dominios -N y -C. Los aa R⁴⁰, R⁵², R⁶⁷, Q⁷⁴, W¹¹³, K²⁹⁶, **G**³³⁶, P²⁹⁵ actúan conjuntamente para regular las propiedades alostéricas del activador FruBP y el inhibidor AMP; K³⁹ une el activador FruBP; los aa **S**²¹², **G**¹⁷⁹, **D**²³⁹, F²⁴⁰ y el totalmente conservado K¹⁹⁵ están involucrados en la unión del substrato α Glc1P y además E¹⁹⁴, **D**²³⁹ y **D**²⁷⁶ tienen contactos directos con el ligando ADP-Glc, más no participan en la catálisis; los aa Q¹⁰⁵, **G**¹¹⁶ interaccionan con ATP y Y¹¹⁴ une ATP, AMP y el activador FruBP; la **A**⁴⁴, conservada invariablemente en la secuencia KR**A**K**P**A**V**, también está involucrada en la unión a los activadores alostéricos y al grupo piridoxal-fosfato también es unido por una Lys que se encuentra la secuencia ¹**AVDENDKIIIEFVEK**P**ANPPSS**^{21458,459,460,461,462,463,464,465,466,467,468}. Posiblemente los sitios regulatorios que permiten clasificar las ADP-glucógeno pirofosforilasas surgieron evolutivamente después que el sitio catalítico, el cual presenta un plegamiento tipo Rossmann de unión a nucleótidos y es probablemente el dominio común ancestral de las GlgC y otras enzimas que unen NDP^{449,456,457}.

- Como se puede observar, GlgA y GlgC contienen el plegamiento tipo Rossmann característico de unión a nucleótidos, estas moléculas pueden ser sintetizadas prebióticamente y el sitio mencionado contiene secuencias específicas con varios aa codificados por tripletes RNY; por lo anterior es

plausible que el origen de las ambas enzimas se remonte a los albores del Mundo de las RNPs y la región mencionada haya sido la identificada como ancestral mediante la metodología descrita en la presente tesis.

▪ FOSFOPENTOMUTASA.

- La fosfopentomutasa, PPM o DeoB de casi 400 aa y ~45kDa codificada por el gen *ppm* de cerca de 1200 nt, está involucrada en el reciclaje/biosíntesis de nucleósidos. DeoB cataliza la transferencia reversible de un fosfato al C1 de la ribosa o de la desoxirribosa mediante la formación de un intermediario (o de la di-desoxirribosa o de la arabinosa), utilizando dos iones Mn^{2+} , Co^{2+} o Ni^{2+} como cofactor, finalmente se libera Rib1P o dRib1P (o didRib1P o Ara1P), nucleósido que puede ser utilizado como fuente de energía o bien fungir como el enlace entre el metabolismo de la glucosa y el de los nucleótidos^{469,470}. Estructuralmente DeoB pertenece a la superfamilia de la fosfatasa alcalina (AP, por sus siglas en inglés), ya que uno de sus dominios es característico de ese grupo mientras el otro es característico de la PPM; no obstante el mecanismo de transferasa intermolecular que se lleva a cabo es análogo al de las superfamilias de la α -phosphohexomutasa (PHM) y de la deshalogenasa haloácida (HAD). El sitio activo de la enzima se localiza en medio de los dos dominios donde se localizan los aa **T**⁸⁵ y **D**¹⁵⁶ que son esenciales para la catálisis, mientras que M⁸⁶ y R¹⁹³ influyen en las distintas conformaciones estructurales; la PPM es catalíticamente activa cuando el nucleófilo **T**⁸⁵ del sitio activo es fosforilado al interactuar con **D**¹⁵⁶, actividad incrementada por la presencia de GluBP^{471,472}. Así mismo la PPM comparte dos motivos con la 2-desoxirribosa 5-fosfato aldolasa (DERA), con lo cual existiría un enlace metabólico entre el metabolismo central de los carbohidratos y el de las pentosas; una **Ser** también localizada en el sitio

activo se encuentra en el motivo **TxSHNP**, mientras que el motivo **DxDxDR** se considera de unión a cationes metálicos⁴⁷³. Con base en el tipo de substrato, la PPM pertenece también a una superfamilia de metaloenzimas que, aunque llevan a cabo distintas actividades catalíticas, actúan en fosfo-/sulfo-carbohidratos. En particular los aa **D**³²⁷, **H**³³¹ e **H**⁴¹² unen uno de los cationes, mientras que **D**⁵¹, **D**³⁶⁹ e **H**³⁷⁰ unen el otro ion metálico; por otro lado los aa **T**³⁶⁷, **T**⁴¹³ y **Q**⁴¹⁰ podrían participar en la unión de la parte de carbohidrato del substrato, mientras que **D**³⁴⁶ y **D**⁴³⁷ podrían estar involucrados en el mantenimiento estructural de la PPM, estos aa junto con **T**⁴⁴¹ están altamente conservados en la PPM y la superfamilia mencionada, aunque no tienen actividad catalítica⁴⁷⁴.

- Es posible que la región identificada como ancestral de la PPM corresponda a alguna de las que permiten identificarla como miembro de alguna determinada superfamilia, ya que varios de los aa conservados son codificados por tripletes RNY. Una de esas regiones podría ser el sitio activo que contiene una Ser, sin embargo el substrato de la enzima actual es un azúcar compleja que es poco probable existiese en un momento evolutivo cercano al origen de la vida; es posible que la región de aa estructurales estrictamente conservados haya sido la identificada como ancestral, no obstante los cationes metálicos ya existían en los albores del Mundo de las RNPs y su sitio de unión es abundante en aa codificados por tripletes RNY, éste habría adquirido otros dominios y divergido en enzimas diversas.

- HIDROLASA DE LA SUPERFAMILIA HAD.
- La subfamilia IIA de las deshalogenasas halo-ácidas (HAD) está involucrada en el procesamiento de azúcares o de compuestos glicosilados, por lo que se considera parte del metabolismo de los carbohidratos.

- Las deshalogenasas haloácidas (HAD) son una superfamilia ubicua de enzimas pequeñas (de 200 a 250 aa) que utilizan el catión metálico Mg^{2+} como cofactor y cuyos diversos substratos contienen grupos fosfato, estas enzimas incluyen fosfoesterasas, ATPasas tipo-P, fosfonatasas, deshalogenasas y fosfomutatas de azúcar; todas ellas consisten fundamentalmente del dominio HAD aunque algunas pueden poseer pequeños dominios accesorios. El dominio HAD es un plegamiento de tipo Rossmann que adicionalmente contiene dos estructuras clave: una vuelta helicoidal ('squiggle') y una horquilla- β ('flap'), localizadas inmediatamente debajo del módulo Rossmann, que facilitan la formación de un intermediario covalente acilo-fosfato; adicionalmente algunas HAD contienen insertos estructurales denominados 'tapas'^{475,476}.
- El proceso catalítico involucra cuatro sitios conservados y ocurre mediante dos mecanismos distintos: el primero involucra una catálisis general de base y el segundo incluye la formación del intermediario covalente mediante el grupo fosfato, sin embargo el primer substrato que se une a la enzima es el nucleótido^{475,476}.
- Los aa de los motivos conservados de las HAD convergen alrededor de una hendidura de la porción-C de una hoja- β central para formar el sitio activo, éstos son: motivo I: **DxDx[T/V][L/V]** donde el primer Asp, localizado al final de una hebra- β hidrófoba, es el aa nucleófilo que forma un enlace acilo-substrato; motivo II: **[S/T]** que une el grupo fosforil mediante puentes de hidrógeno; motivo III: **K-x_n-[G/S]-[D/S]-x_n-[D/N]** que orientan adecuadamente el aa nucleófilo y el grupo fosfato para su interacción, además unen el ion Mg^{2+} ; los aa conservados en los motivos II y III proveen de estabilidad al intermediario catalítico durante la hidrólisis

enzimática y en ambos casos los aa se localizan en una estructura hebra- β -
-bucle--hélice- α ; finalmente el motivo IV contiene también un aa ácido
presente en uno de tres posibles motivos: **DD** o **GD_{x3}D** o **GD_{x4}D**, que unen
también el cofactor Mg²⁺. Algunas HAD contienen dos dominios, uno de
ellos es sólo un subdominio y el otro se considera el dominio esencial de la
enzima, en cuya interface se localiza el nucleófilo activo **D**, así como los aa
adicionales ⁴¹R y ¹⁵⁷Y que lo rodean^{475,476,477}.

- Estructuralmente las HAD se clasifican en tres clases con base en la
ausencia o presencia de una '**tapa**' sobre el sitio activo: **clase I**: contiene
una tapa pequeña entre los motivos I y II; **clase II** (a la que pertenece
IIA): contiene una tapa grande, cuya estructura predicha es un plegamiento
mixto α/β , entre los motivos II y III; **clase III**: no contiene tapa^{476,478}.
- Algunas HAD pertenecen también a una superfamilia de fosfohidrolasas
asociadas a membrana denominada DDDD por la presencia de residuos Asp
estrictamente conservados en los motivos [F/**I/L/M/A/T/V**]-**D**-
[**I/L/F/R/M/V/Y**]-**D**-[**G/S/N/D/E**]-[**T/V**]-[**I/L/V/A/M**]-[**A/T/S/V/I/L/M/C**]-x-
{Y/F/W/H/K/R}-x-{Y/F/W/H/**N/Q**}-x₁₀₂₋₁₉₁-{K/R/H/**N/Q**}-**G-D**-{F/Y/W/H/**I**/
L/**V/M/C**}-{**Q/N/H**}-{F/W/Y/**G/P**}-**D**-{P/**S/N/Q/Y/W**} a lo largo de la
secuencia, del mismo modo algunos miembros de DDDD son HAD al mismo
tiempo, aunque ninguna forma parte de la otra en su totalidad^{479,480}.
- Actualmente las '**tapas**' funcionan como determinantes de la especificidad al
substrato y facilitan la hidrólisis, pero es posible que en un inicio las
inserciones permitieron la diversificación de la superfamilia a partir del
módulo de la fosfatasa, por tanto el estado más ancestral corresponde a las
HAD sin '**tapa**'. Asimismo, las pequeñas estructuras '**squiggle**' y '**flap**', que

actualmente son importantes en la catálisis, pudieron haber emergido como un mecanismo rudimentario de exclusión de sustratos^{476,481}.

- Se ha encontrado también que los componentes receptores de los sistemas denominados de dos componentes, esenciales para la transducción celular, comparten características catalíticas y estructurales con las HAD; por otro lado se propone que la ATPasa tipo-P surgió por la fusión del dominio HAD con una proteína transmembranal; además de las demás familias que se han descrito pertenecen a este grupo y que pudiesen haber surgido del mismo modo por la fusión del dominio HAD con otro tipo de estructuras^{482,483,476}. Como se puede advertir, el dominio HAD fungió como el andamiaje o materia prima evolutiva para el surgimiento de varias enzimas, tal dominio posiblemente provenía de una clase de plegamiento Rossmann que poseía actividad fosfoesterasa y/o quelante de Mg^{2+} caracterizado por un aa ácido estrictamente conservado en la primera unidad- β del módulo^{476,484}.
- Dado que los aa de los 4 motivos conservados en las HAD son codificados por tripletes RNY y todos ellos están involucrados en la unión de los grupos fosfato y/o del Mg^{2+} que pueden ser prebióticos, es posible el origen de cualquiera de esos dominios se remonte a los albores del Mundo de las RNPs y esa de manera constituiría la raíz de la superfamilia HAD.
- INOSITOL MONOFOSFATASA.
- La enzima inositol monofosfatasa (IMPasa) fue aislada de mamíferos, aunque actualmente se ha descrito en otros eucariontes y en varios procariontes; la familia de dicha enzima se considera parte del COG de los carbohidratos porque el inositol (in) es un carbohidrato simple y la IMPasa comparte características con enzimas del llamado metabolismo central⁴⁸⁵.

- La familia de la IMPasa está relacionada con la síntesis o degradación de moléculas mensajeras fosforiladas, en particular la IMPasa es un oligómero que cataliza la hidrólisis de fosfatos de inositol utilizando Mg^{2+} como cofactor. La IMPasa puede estar presente como homodímeros u homotetrámeros, cuyas subunidades son de ~20kDa a ~30kDa cada una y poseen la estructura $\alpha\beta\beta\alpha$, en que cada sitio activo se localiza en una caverna hidrofílica en la base de las dos hélices centrales hacia las que convergen varios de los elementos estructurales. Con base en alineamiento de secuencias de diversas fosfatasas, se ha sugerido que éstas conforman una familia estructural, pues se encontró que comparten un sitio, con el mismo patrón de hidrofobicidad, involucrado en la unión a grupos fosfato y a iones metálicos, aunque existen algunas variaciones dramáticas en la secuencia de una enzima a otra; en particular se conservan 3 regiones (tomando la IMPasa humana como referencia) que comprenden los aa: 69 al 72 la región A, 87 al 100 la región B y 219 al 233 la región C, esos aa son: K³⁶, E⁷⁰, **D**⁹⁰, I⁹², **D**⁹³, **G**⁹⁴, **T**⁹⁵, **A**¹⁹⁶, H²¹⁷, C²¹⁸ y **D**²²⁰, que son cruciales para la unión al substrato, al catión metálico y a la molécula de agua. Las cationes metálicas que pueden actuar como activadores no-competitivos en la enzima pues estabilizan la transición de estado son, en orden decreciente de activación: $Mg^{2+} > Mn^{2+} > Zn^{2+} > Co^{2+} > Ca^{2+}$, mientras que el Li^{2+} y Gd^{2+} actúan como inhibidores; por otro lado la molécula de agua estabiliza el oxianión del inositol actuando como activador nucleófilo; finalmente se une el fosfato de inositol a la enzima^{486,487,488,489}.
- En procariontes no es destacada la función del inositol como molécula de señalización celular e incluso se sugiere que esta molécula es innecesaria en dichos organismos, sin embargo se ha observado que en algunas bacterias varios de los componentes de su pared celular están basados en

la estructura del fosfatidilinositol e inclusive la carencia de IMPasa conlleva a una alteración en la permeabilidad, lo que provoca que la bacteria sea más vulnerable a cambios en el medio ambiente, en esos microorganismos la IMPasa parece ser más estable en condiciones de pH alcalino y se observa un aumento directamente proporcional de la actividad enzimática con respecto al incremento en la temperatura; no obstante lo anterior, en el dominio bacteriano el inositol y sus derivados no son comunes y la síntesis de esos compuestos no depende directamente de la IMPasa, por lo que por mucho tiempo no hubo pruebas fehacientes su papel catalítico en procariontes^{490,491,492,493}.

- Se han descrito IMPasas en diversos procariontes, desde bacterias como *Escherichia coli* y de los géneros *Mycobacterium* y *Staphylococcus*, en bacterias hipertermófilas como *Thermothoga maritima* y en arqueas como *Mehtanocaldococcus janashii* y *Archaeoglobus fulgidus*. Sólo algunas IMPasas de procariontes son inhibidas por Li^{2+} , pero todas son activadas por al menos tres Mg^{2+} que promueven la dimerización o tetramerización de la enzima, ya que la IMPasa no es constitutiva en procariontes, al unirse en sitios homólogos a los de la IMPasa humana; de hecho el sitio activo de la IMPasa contiene un bucle con los aa F²²**GRKDKSYVVGTS****PSGDETEI**_{X₂}**D**⁴⁴, algunos de los cuales (subrayados) están muy próximos a los sitios de unión de los tres cationes y de los productos in y Pi. Empero, la característica más relevante de las IMPasas de procariontes descritas hasta ahora es que éstas llevan a cabo funciones duales, es decir las IMPasas de procariontes catalizan tanto la hidrólisis de fosfatos de inositol (mono-, di-, tri- o poli-fosfato) como de moléculas estructural y/o químicamente semejantes, como FruBP, glucitol 6-fosfato (Glu6P), glicerol 2-fosfato (G2P), 2'-AMP, 2'PAP, 3'PAP y/o NADPH; lo cual es posible debido a que

estas IMPasas poseen un sitio activo más amplio que permite acomodar una mayor diversidad de sustratos, hecho notable pues antes que se describiesen este tipo de IMPasas se asumía que no existían varias de las enzimas involucradas en el llamado metabolismo central como la FruBPasa o la NADP(H) fosfatasa, lo que implicaba rutas metabólicas distintas a las descritas en la mayoría de los organismos^{490-493,494,495,496,497,498,499,500}.

- Por otro lado, la IMPasa humana contiene una región altamente similar en secuencia y estructura a una región de la enzima fructosa-1,6-bisfosfatasa (FruBPasa) del metabolismo central, el monómero de ambas enzimas posee la misma estructura $\alpha\beta\alpha\beta$ y su sitio activo está localizado en una hendidura hidrofílica cerca del inter-dominio en el segmento entre los aa 68 y 125; a partir de esas estructuras y los aa conservados en cada una, se ha deducido que las regiones conservadas constituyen un dominio de unión a cationes metálicos y el sitio activo une específicamente ligandos de ésteres de fosfato⁵⁰¹. De hecho la IMPasa pertenece a una gran superfamilia de hidrolasas de fosfatos de azúcares, incluyendo la FruBPasa y otras involucradas en las reacciones individuales que son capaces de catalizar las IMPasas de función doble que poseen algunos organismos. Las hidrolasas referidas contienen en su secuencia el motivo **DP[I/L]D[G/S][T/S]** que participa directamente en la unión a cationes metálicos y se localiza cerca del sitio catalítico, mientras que las estructuras de las enzimas son superponibles emergiendo el plegamiento común $\alpha\beta\alpha\beta$; si bien los análisis filogenéticos sugieren entonces que todas las proteínas de la superfamilia estructural de la IMPasa comparten un ancestro común que existió bastante tiempo atrás, a partir del cual las secuencias divergieron extensamente al incorporarse elementos estructurales secundarios, de todos los sitios

conservados un motivo **DPxD**, que une directamente cationes metálicos e interacciona con el in, es totalmente invariante^{492,495,501,502,503}.

- Dada la plasticidad de la IMPasa en el reconocimiento de los substratos, ésta puede ser un determinante importante en la fisiología de los organismos, ya su función que no está restringida a la hidrolización de moléculas específicas; en particular, de todas las hidrolasas de azúcares fosfatadas, la FruBPasa semeja a la IMPasa más que a la Fru1Pasa que es su vecina metabólica en la glucólisis; dado que la gluconeogénesis se considera una ruta metabólica más ancestral que la glucólisis, las IMPasas de procariontes establecen el enlace entre el metabolismo del inositol, la respuesta al estrés y la gluconeogénesis, por tanto es posible que constituya el eslabón evolutivo de esa superfamilia cuyo ancestro común semejaba probablemente a las IMPasas de función dual del tipo que poseen los organismos hipertermófilos, pues además que se considera que son de los más antiguos, la IMPasa de bacterias hipertermófilas es más semejante a la de arqueas que a la de otras bacterias. Dado que además el motivo **DPxD** es invariante en la superfamilia de la IMPasa y ya que ésta prueba incidentalmente que las enzimas más antiguas eran menos especializadas con respecto a su substrato⁴⁸⁶⁻⁵⁰³, es probable que la región que contiene dicho motivo sea codificada por tripletes RNY, ya que además los cationes metálicos existían indudablemente en los albores del Mundo de las RNPs.

○ **METABOLISMO DE AMINOÁCIDOS.** Dado que en el ambiente prebiótico existían algunos aa libres, los ciclos químicos de algunos de ellos y su interacción con otros compuestos prebióticos, delinearon de las primeras rutas metabólicas^(cfr.504) y parte de ciertas proteínas involucradas en el metabolismo de aa está codificada por tripletes RNY.

- PROTEÍNA DE LA FAMILIA DE LA GLIOXILASA I.
- El metiglixal (MG) es un metabolito tóxico derivado del fosfato de dihidroxiacetona (DHAP) durante la glucólisis, éste se puede transformar en lactato y/o piruvato mediante 4 sistemas distintos, uno de los cuales es el sistema de la glioxilasa; dicho sistema consiste en las subsecuentes transformaciones del MG en **S**-lactoilglutación (**S**-LG) y de éste a lactato, reacciones catalizadas respectivamente por la glioxilasa I (GloI) y la glioxilasa II (GloII), cada una las cuales pertenece a una familia enzimática distinta⁵⁰⁵, siendo una proteína de la familia GloI la que contiene una región codificada por tripletes RNY.
- GloI es una metaloenzima, aparentemente regulada por fosforilaciones/desfosforilaciones, que cataliza la isomerización del **S**-LG a partir del hemitioacetal de MG (en general tiohemiacetales), formado por la reacción no-enzimática de MG con glutación (γ Glu-Cys-Gly) reducido (GSH); dicho proceso se lleva a cabo mediante la abstracción de un protón del C1 del substrato por un Glu conservado, resultando un intermediario enolato de alta energía, cuya estabilidad es mediada por un catión metálico, que coordinan el mencionado E y otros 3 aa y una molécula de agua^{506,507,508}.
- La estructura de las metaloenzimas 'quelantes del oxígeno vecinal' (VOC, por sus siglas en inglés) que incluye a GloI, está conformada por dos plegamientos $\beta\alpha\beta\beta$ dispuestos pseudo-simétricamente y que forman los dominios -C y -N; se sabe que los sitios de unión a metales son catalíticamente muy versátiles, ya que proveen la atmósfera electrófila necesaria para diversas reacciones redox, de modo que el plegamiento pseudo-simétrico sólo maximiza dicha versatilidad. Se ha propuesto que un sólo plegamiento $\beta\alpha\beta\beta$ era codificado por un mini-gen, que posteriormente

se duplicó originando la metaloproteína VOC pseudo-simétrica progenitora con un dominio-N (aa 3 al 60) unido por un bucle al dominio-C (aa 72 al 126). Existen dos clases de Glol, una activada por Zn^{2+} (de ~180aa), encontrada en eucariontes y una activada por Ni^{2+}/Co^{2+} (de ~130aa), encontrada en procariontes; ambas clases preservan la estructura $\beta\alpha\beta\beta$ típica de isomerasas y el sitio catalítico de unión a metales. El sitio de unión a metales de la Glol de procariontes está formado por H^5 , E^{52}/E^{56} , H^{74}/H^{75} y E^{122}/E^{124} , mientras el bolsillo catalítico se forma por L^{41} , F^{43} y V^{54} del dominio-N y A^{76} , V^{103} , I^{110} , F^{112} e I^{124} del dominio-C y el hemitioacetal es unido por R^9 , N^{60} , R^{98} y K^{104} ; finalmente, las interacciones internas en el dominio-C (E^{26} - A^{129} , D^{128} - R^{131}) y de éste dominio con el bucle que lo une al dominio-N (T^{107} - L^{133} , T^{107} - G^{134} y T^{108} - G^{132}) mantienen el ambiente hidrófobo dentro del sitio catalítico^{509,510,511,512}.

- Dado que actualmente aún no existen estudios que determinen los motivos conservados dentro de la familia de la Glol, no es posible dilucidar la región identificada como ancestral por la metodología descrita en la presente tesis, sin embargo varios de los aa que conforman el bolsillo catalítico, de unión a cationes metálicos, son codificados por tripletes RNY, por lo que es posible que sea ésta la región que existía en los albores del Mundo de las RNPs.
- AMINOTRANFERASA, CLASES I Y II.
- Las aminotransferasas (ATs) juegan un papel importante en el metabolismo de todas las especies de aminoácidos (aa), pues catalizan la transferencia del grupo amino- α de un aa- α a un cetoácido- α , estas enzimas también se denominan transaminasas, pues canalizan los grupos amino de una amplia variedad de aa al α -cetoglutarato (α KG) para su conversión a NH_4^+ en una

reacción simple y básicamente formada por dos medias reacciones, en una se acepta un aa como sustrato y sustrae de éste su grupo amino, mientras la segunda transfiere el grupo amino al oxo-ácido de un segundo aa. Todas las AT pertenecen a la súper-familia de enzimas PLP que utilizan el piridoxal fosfato o vitamina B₆ como grupo prostético, el cual se compone de un anillo piridina ligeramente básico y un grupo hidroxil fenólico ligeramente ácido, en consecuencia los derivados del PLP pueden constituir formas tautoméricas estables, en las cuales el átomo de nitrógeno de piridina es protonado y por tanto positivamente cargado mientras que el grupo hidroxilo es desprotonado formando un fenolato; no obstante el grupo más importante de las PLP es el aldehído, ya que es el grupo que interacciona con los aa mediante una base de Schiff^{513,514}.

- Las AT catalizan las reacciones más ubicuas con mayor repertorio de sustrato, con base en lo cual se han definido 3 grupos: el tipo I de AT interconvierte pares de ceto- α y aa- α , son AT con extrema estéreo-selectividad entre D-aa y L-aa, aunque sólo regular estéreo-especificidad por aminoácidos con cadena lateral similar; el tipo II de AT transfieren el grupo amino de los aa terminales dejando un grupo aldehído en su lugar, dependiendo del tipo de aa se sub-clasifican en: AT-IIA que usan aa básicos, AT-IIB que usan aa- ω y AT-IIC que usan diaminas; finalmente el tipo III de AT son las enzimas más específicas de la familia. Dado que las AT catalizan reacciones naturalmente reversibles, su potencial debe ser suficientemente alto para promover direccionalidad *in vivo*. La amplia especificidad de las enzimas ancestrales maximiza la versatilidad catalítica de las células ancestrales en los tiempos evolutivos remotos cuando existían pocos compuestos disponibles para las reacciones metabólicas. La marcada pluriespecificidad de la mayoría de las AT ejemplifica la

ambigüedad de las enzimas ancestrales por su substrato, además es probable que el empalme de especificidad comúnmente observado en las AT, esté dado porque el origen de tales enzimas fue la duplicación de un gen primario y la evolución de la regulación diferencial derivó en enzimas más especializadas⁵¹⁵. Las secuencias de AT de eucariontes y procariontes, se alinearon con base en similitud, patrones de hidrofobicidad y estructuras secundarias predichas, con base en lo cual se han identificado cuatro aa totalmente invariantes: **G**³¹⁴, **D/E**³⁴⁰, **K**³⁸⁵ y **R**⁵⁶², de un total de 400aa. Las secuencias alineadas pueden además ser divididas en cuatro subgrupos, de los cuales se evaluó la homología intra-grupo e inter-grupos que indica que los substratos en cada subgrupo son bioquímicamente similares entre sí: el I contiene AspAT, AlaAT, TyrAT, PheAT y AT de fosfato de histidinol; el subgrupo II contiene AT de los aa metabólicos, no-codificados y compuestos relacionados, acetil ornitina, ornitina, aa- ω , 4-aminobutirato y diaminopelargonato; el subgrupo III contiene AT para D-Ala y para aa de cadena ramificada; finalmente el subgrupo IV contiene AT de fosfoserina y SerAT. De estos subgrupos, el III es el más distante evolutivamente, aunque en realidad todas las AT constituyen un solo grupo de proteínas homólogas que aparentemente divergió en los cuatro subgrupos antes del LUCA, con algunas excepciones de enzimas individuales específicas⁵¹⁶.

- La transferencia específica de grupos amino de los aa a oxo-ácidos es sólo una de las numerosas transformaciones llevadas a cabo por enzimas dependientes de piridoxal-5'-fosfato o vitamina B₆ (enzimas PLP), aunque en realidad estas enzimas constituyen un extenso grupo de enzimas que en lo general catalizan transformaciones de aminoácidos; de hecho las reacciones de las enzimas PLP proceden a través de la formación de un intermediario aniónico base de Schiff entre el substrato y la coenzima. La

identidad que comparten todas las enzimas PLP es de >15%, pues de un total de 400aa, los aa invariantes son sólo 12, están agrupados en regiones muy específicas y sus funciones se empalman al interactuar con los sustratos y con el cofactor PLP que es unido por 6 de esos aa, o bien son esenciales para la conformación estructural de la enzima, 8 de esos aa invariantes están localizados en un segmento de 74 aa que comprende de Asn194 a Gly268, dicho segmento conforma un módulo esencial de dos hélices- α adyacentes y dos y media hebras- β en el interior del gran dominio de unión a la coenzima. Los aa invariantes son: Y⁷⁰, P¹³⁸, **N**¹⁹⁴, P¹⁹⁵, **G**¹⁹⁷, **D**²²², Y²²⁵, K²⁵⁸, R²⁶⁶, **G**²⁶⁸, **A**³⁰³ y R³⁸⁶, con base en la AspAT de *Escherichia coli*, cuyo sitio esencial es prototipo de todas las AT pues comparten alta similitud; a pesar de todo lo anterior, es posible que las características estructurales y mecánicas comunes estén dictadas por las propiedades químicas del PLP, más que debido a un ancestro común de las enzimas que lo utilizan^{516,517,518}. La superfamilia de enzimas PLP se subdivide en 3 familias con base en criterios estructurales que correlacionan con su regio-especificidad, es decir las familias α , β y γ se clasifican con base en el átomo de carbono del sustrato que lleva el grupo amino y forma un enlace imina con la coenzima, por lo que su covalencia cambia o es eliminado durante la catálisis; de éstas, la familia α es la más grande e incluye a casi todas las AT, con excepción de algunas del subgrupo III; pues las AT de la familia I contienen 11 aa altamente conservados que interactúan en su mayoría con el PLP, 4 de estos están presentes en todas las AT y 2 de ellos son invariantes en la división α de las PLP^{519,520,521}. Todas las enzimas PLP son homodímeros compactos o bien grandes oligómeros con simetría dihedral. Sus subunidades tienen en general dos dominios aunque en ocasiones tres,

el dominio central es una estructura abierta α/β que contiene al aa esencial K del sitio activo, situado en la interface entre ambos dominios y al cual contribuyen además aa de los otros dominios. Con base en las características estructurales de las AT, tanto la familia α como la γ tienen el mismo tipo de plegamiento denominado I, mientras que la AspAT conforma el plegamiento tipo IV; los plegamientos no mencionados pertenecen a otras enzimas PLP, modificadoras de aa pero que no son AT, lo que conforma un total de cinco plegamientos no relacionados entre sí⁵²². Las enzimas PLP comparten 7 segmentos comunes que conforman una organización estructural común que incluye algunos sitios clave y contiene un módulo fundamental hidrófobo mínimo, formado por cuatro aa que unen el cofactor; a pesar de ello, se ha propuesto que en lo general las enzimas PLP no tienen un origen común, pues en realidad los aa que unen los substratos, rara vez se encuentran en posiciones equivalentes y en realidad distintos grupos surgieron al en menos 5 ocasiones durante la evolución, pues la orientación del anillo piridoxal, así como las gran mayoría de los aa que unen el cofactor y por tanto la estrategia catalítica son distintas en cada una^{523,524}; inclusive las tasas de evolución de las enzimas PLP son totalmente distintas entre sí, por lo que se deduce evolucionaron por distintos constreñimientos funcionales y por tanto provienen de múltiples orígenes evolutivos; sin embargo la especialización funcional, primero por reacción y después por substrato, de la mayoría de estas proteínas ya era completa en el LUCA, de donde evolucionaron finalmente por especiación^{525,526}.

- Es posible que la especificidad de las enzimas primitivas fuese más laxa que en la actualidad y utilizaban familias de moléculas como substrato, esas enzimas ganaron paulatinamente mayor especificidad mediante la

duplicación de genes y el reclutamiento de distintos dominios polipeptídicos⁵¹⁶, en particular las enzimas PLP primordiales eran aparentemente catalizadores regio-específicos generales, que posteriormente se especializaron por sustrato; dado que todos los miembros de la superfamilia PLP utilizan la misma coenzima y un andamiaje peptídico general para ello, aun cuando llevan a cabo funciones distintas, es posible que las restricciones catalíticas de los distintos grupos hayan estado presentes en el LUCA o quizá desde antes, es decir que el ancestro común de las PLP, fungía como una unidad multi-propósito en los sistemas primitivos y catalizaba reacciones de amplia especificidad⁵²⁰. No obstante, pese a la evolución convergente, todas las enzimas PLP están evolutivamente relacionadas con la AspAT⁵²²; de hecho las clasificaciones que se han hecho de las AT con base en secuencia, estructura y estereoquímica del hidrógeno transferido coinciden, en particular esta última característica refleja además la evolución molecular de dichas enzimas a partir de una única proteína o un péptido común ancestral, posiblemente con excepción de la AspAT del subgrupo III que pudo haber surgido por convergencia evolutiva⁵²¹. Dado que el PLP es un cofactor nucleotídico que surgió prebióticamente y desde el inicio catalizó reacciones metabólicas, es muy probable incluso que sea la coenzima más versátil de modo que hubo varios grupos de enzimas dependientes de tal molécula, que surgieron paralogamente en varias ocasiones y dieron origen a linajes independientes⁵²⁶.

- Por un lado la superfamilia de las enzimas PLP surgieron por evolución paróloga varias veces en la evolución, aunque su repertorio estaba prácticamente concluido en el LUCA; una de esas familias de enzimas PLP son las AT, las cuales comparten un solo ancestro común; por otro lado una

de las AT I/II se identificó como ancestral por la metodología descrita en la presente tesis, no se define a cuál de las dos familias pertenece la enzima identificada, pero dado que la subfamilia III está evolutivamente distante de las otras dos, podrían agruparse las subfamilias I y la II en una sola entidad; adicionalmente, tenemos que el PLP es un cofactor nucleotídico que puede sintetizarse en condiciones prebióticas, hecho importante ya que es posible deducir que desde la Tierra primitiva pudieron haber coexistido los nt y los aa. Por todo lo anterior, es posible que la región de la AT ancestral que une la coenzima sea el primordio de todas las enzimas PLP, al menos del linaje que más tarde originó a las AT, cuya antigüedad se remontaría entonces a los albores del Mundo de las RNPs.

- AA AROMÁTICOS, DESHIDROQUINATO SINTASA / CORISMATO SINTASA.
- La ruta del ácido shikímico conecta el anabolismo de los aminoácidos aromáticos (aro aa) con el metabolismo central únicamente en procariontes y plantas, *i.e.* no en animales. El shikimato es un ácido carboxílico a partir del cual se deriva el corismato, que es precursor de una amplia variedad de compuestos aromáticos como los aa Phe, Tyr y Trp; la biosíntesis aromática consiste de siete pasos metabólicos en total, ***grosso modo***: fosfoenolpiruvato (PEP) + enol-4-fosfato (E4P) → 3-desoxi-D-arabinoheptulosonato-7-fosfato (DAHP) → deshidroquinato (DHQ) → shikimato (Shk) → 3-enolpiruvilshikimato-5-fosfato (EPSP) → Corismato (Cori) → aro aa (F, Y, W). La mayoría de los genes involucrados en la biosíntesis de los compuestos mencionados, se localizan dispersos en el cromosoma, excepto dos de ellos que, pese a que sus productos catalizan funciones metabólicamente separadas entre sí, **se encuentran a menos de 1' de separación** y por tanto muchas veces no pueden distinguirse en secuencia

sino por función, incluso se considera que pertenecen al mismo operón o aún conforman una misma entidad genética; dichos genes codifican para las enzimas AroB y AroC, que utilizan una molécula nucleotídica para llevar a cabo su proceso catalítico respectivo liberando Pi. Las enzimas mencionadas se consideran redox neutrales, ya que la molécula nucleotídica actúa como grupo prostético, es decir se une covalentemente a la enzima, guardando relación estequiométrica con ésta y no con el sustrato e interviene directamente en la catálisis como un donador interno de e-, mismo que se regenera durante el proceso catalítico y por tanto no es consumido en la reacción final *i.e.* no conlleva una pérdida^{527,528,529}. Tanto AroB como AroC son las únicas enzimas de la ruta del shikimato que contienen una región codificada por tripletes RNY.

- La **AroB** o **DHQS (deshidroquinato sintasa)** cataliza el segundo paso de la ruta de shikimato, la ciclización DAHP → DHQ; para su actividad máxima AroB utiliza NAD⁺, en equivalencia estequiometría no con el sustrato sino con la enzima y con un mol de Zn²⁺ o Co²⁺; esta reacción involucra a su vez cinco transformaciones sucesivas que en su conjunto provocan que la catálisis sea irreversible. Varios análogos, que actúan como inhibidores competitivos, pueden unirse al sitio de unión al sustrato con mayor afinidad que DHAP, sin embargo la enzima sólo cataliza la ciclización de este último^{529,530,531,532,533,534,535,536,537,538,539}. Aun cuando la DHQS utiliza una molécula nucleotídica como grupo prostético, su secuencia contiene el motivo característico de las proteínas que unen nucleótidos NAD⁺ o FAD⁺ como cofactor, los aa 96 al 126 presentan la estructura predicha βαβ con el motivo **GxGxxG**₁₈₋₂₀Q, donde la Gln (aa cargado negativamente) es específica de la DHQS y varios de los demás aa son hidrófobos, dicho segmento se considera además el módulo esencial de la proteína; otros aa

que se consideran conservados son dos Met que se consideran críticas para la unión del sustrato, pues también están presentes en la DHQ cinasa^{540,541}. La metaloenzima AroB es un homodímero que consta de poco más de 350aa y ~40kDa cada subunidad, donde el dominio-N tiene una estructura α/β característica del plegamiento Rossmann y contiene el sitio de unión al NAD⁺; el dominio-C está compuesto de una hélice- α y está encargado de unir DHAP, que evita que la enzima pierda el NAD⁺ rápidamente, y el catión metálico, Zn²⁺ que estimula la función o Co²⁺ que preserva la estabilidad de la enzima; finalmente, el sitio activo se localiza en la hendidura inter-dominio y contiene varios aa conservados, muchos de los cuales están involucrados directa o indirectamente o en la función catalítica de la proteína y/o en la preservación de la estructura y algunos de los cuales también unen la parte fosfato del NAD⁺, de hecho a la posición del sitio activo se le conoce también como media-media, porque se localiza en la región medial intra-dominio que también corresponde con la porción medial inter-dominio del dímero^{542,543,544,545,546,547,548,549}. La DHQS de procariontes es usualmente monofuncional, aunque en algunos organismos extremófilos la AroB es trifuncional, mientras que la DHQS de eucariontes es usualmente pentafuncional, catalizando sola varios pasos de la vía del shikimato; a pesar de ello las AroB de diversas fuentes comparten ~30% de identidad en secuencia y, pese a las diferencias sutiles en la reacción debido a las secuencias particulares, la estructura general se preserva así como los aa clave que interaccionan con las moléculas involucradas en la reacción enzimática^{550,551,552,553}.

- La **AroC** o **CS (corismato sintasa)** está involucrada en el séptimo y último paso común de la biosíntesis aromática, catalizando la reacción EPSP → Cori en anaerobiosis, o bien en condiciones aerobias utilizando FMN neutral

preferentemente, aunque también puede utilizarse FMN reducido (FMN^{red} o FMNH₂) o FAD; ese grupo prostético es un donador-aceptor de e- que sirve para estabilizar las cargas de los intermediarios durante la reacción y estimula la fosforólisis y la desprotonación, de tal modo que la molécula nucleotídica se libera en el mismo estado redox al finalizar el proceso catalítico; el paso limitante en la reacción es posterior a la unión del FMN y posiblemente sea la liberación del fosfato o la transformación del grupo prostético en distintos intermediarios, finalmente se da la liberación del EPSP. La CS de procariontes es usualmente monofuncional, mientras que la CS de varios eucariontes es bifuncional, pues utiliza NADPH como cofactor además del grupo prostético FMN, a pesar de lo cual los mecanismos son muy similares^{529,554,555,556,557,558,559,560,561,562,563,564,565,566,567}. AroC es un homotetrámero, o bien un dímero de dímeros estructurados en topología $\beta\alpha\beta$, donde cada subunidad es de ~350aa, ~40kDa y disposición pseudo-simétrica; exceptuando los últimos aa de la CS que son innecesarios y posiblemente sólo favorecen el adecuado plegamiento o bien están involucrados indirectamente en la catálisis y son los que determinan el tamaño mayor de la CS en eucariontes. La CS contiene grupos bien definidos de aa conservados involucrados en la unión tanto del FMN como del EPSP, en particular la hélice 7 es abundante en aa pequeños no cargados y la hélice 4 contiene un aa ácido y cinco básicos, los sitios de unión a cada molécula son distintos aunque interdependientes e incluso existe un pequeño puente de hidrógeno entre ellos, en esta composición el grupo prostético se localiza muy profundamente en la parte media-media, mientras el sustrato se une a modo de tapón impidiendo la salida del primero. En particular la región que une el anillo de isoaloxazina del FMN es hidrófoba aunque tiene pocos contactos y en su mayoría con aa pequeños

como Ala, en contraste con la cadena ribitol; por otro lado los aa que interaccionan con el EPSP son Arg en su mayoría^{556,568,569,570,571,572,573,574,575,576,577,578}. Dos secuencias permiten identificar una CS en cualquier organismo, éstas son: RP[**G/S**]H[**A/G**]D_{x5}K y Rx**S[A/G]**R[E/V][**S/T**]_{x3}**V**_{x2}**G**_{x6}L, que se han propuesto como el sitio activo de la enzima; posteriormente se describieron otros 3 motivos conservados: 1. **G[D/E/S]SH[G/C]**_{x2}[L/**I/V/M**][**G/T/I/V**]_x[L/**I/V/T**][L/**I/V**][**D/E/S/T**]**G**_x[P/**V**], 2. [**G/E**]_{x2}**S[A/G]**Rx[**S/T**]_{x3}[**V/T**]_{x2}[**G/A**][**S/T/A/V/Y**][L/**I/V/M/F**] y 3. R[**S/H**]D[P/**S/V**][C/**S/A/V**]_{x4}[**S/G/A/I/V**]_x[**I/V/G/S/T/A/P**][L/**I/V/M**]_xE[**S/T/A/H**][L/**I/V/M/A**], que contienen aa que unen directamente el FMN. La homología de las CS va desde >50% hasta <95% y con base en análisis evolutivos se ha concluido su origen es monofilético, posiblemente la enzima ancestral era monofuncional e incapaz de reducir FMN, por lo que unía FMNH₂^{556,579,580,581, 582,583}.

- Como podemos observar, tanto AroB como AroC acoplan una molécula nucleotídica, que si bien es distinta para cada enzima, varios de los aa conservados en ambos sitios de unión son codificados por tripletes RNY, dado que los substratos de cada enzima son distintos y muy complejos para haber sido sintetizados prebióticamente, es posible que el origen evolutivo de dicha región de unión a nt se remonte a inicios del Mundo de las RNPs.
- AA AROMÁTICOS, AMIDOTRANSFERASA DE GLUTAMINA.
- Las amidotransferasas dependientes de glutamina (GlnAT) son enzimas que catalizan la generación de amoniaco, a partir de la transferencia del nitrógeno de la amida de la glutamina, a siete diferentes clases de aceptores, dependiendo del dominio con el que estén asociadas en su extremo-C que le confiere además nombre a la enzima. Estructuralmente,

las GlnAT tipo I (GlnAT-I) se caracterizan por contener una tríada catalítica CHE, mientras que las GlnAT-II por poseer una topología nucleófila. Las GlnAT poseen especificidad dual por Gln generando Glu y por NH_3 , ya que pueden utilizar ambas moléculas como fuente de nitrógeno, la GlnAT del Glu-tRNA^{Gln} posee además especificidad por Asn; por otro lado el dominio-C asociado contiene el sitio donde ocurre la síntesis de la molécula objetivo mediante la transferencia del grupo amino; ambos sitios se conectan mediante una estructura flexible que funge como túnel hidrófobo interno para mediar la transferencia de la molécula extraída, dicho conector permite al menos tres cambios conformacionales durante el ciclo catalítico de la enzima y aunque su delección no implica la pérdida de función de ninguno de los dominios, ya que éstos pueden llevar a cabo sus reacciones catalíticas de manera independiente, al no existir un enlace tampoco hay regulación alostérica y en realidad la enzima *per se* tampoco se forma^{584.585}.

La GlnAT identificada en **SagA909** está asociada con un dominio involucrado en la síntesis de triptófano, por lo que se designa como sintetasa II de antranilato, o TrpD/TrpE. Como todas las amidotransferasas, TrpD es una enzima modular en que el dominio-N, de acuerdo a lo esperado lleva a cabo la actividad GlnAT-I, que a diferencia de otras enzimas de la familia sólo une amoníaco y no el aa Gln; mientras el dominio-C corresponde a una transferasa de fosforibosilantranilato (PRAT). TrpD cataliza entonces la primera reacción en la síntesis de Trp, sin mediar la utilización de cofactores o coenzimas: corismato + $\text{NH}_3 \rightarrow$ antranilato + piruvato; en bacterias, TrpD se asocia usualmente con la sintetasa I de antranilato o TrpE/TrpG, que sí cataliza la hidrólisis $\text{Gln} \rightarrow \text{Glu} + \text{NH}_3$. Estructuralmente ambos dominios de la TrpD están claramente definidos y a lo largo de la secuencia cada uno ocupa regiones específicas, pues la GlnAT abarca el

primer tercio de la proteína y la región PRAT abarca los dos últimos tercios de la enzima, misma que consta de casi 1,600 nt y ~530aa; por lo anterior se ha propuesto, similar a otras GlnAT, que posiblemente la TrpD sea resultado de la fusión de un gen que codificaba una GlnAT-I específica para NH₃ y un gen distinto que codificaba una PRAT^{586,587,588,589,590,591,592,593}. Hasta la fecha no hemos encontrado que se sugieran los motivos de aa conservados en el dominio GlnAT de la TrpD fuera de la tríada CHE misma que no contiene aa codificados por tripletes RNY, por lo tanto no ha sido posible definir con precisión la región identificada como ancestral por el método descrito en la presente tesis; no obstante es posible inferir, que dado que el corismato es una molécula compleja y de formación poco probable sin catálisis enzimática y por otro lado el amoniaco se considera esencial para la química prebiótica, es probable que la región que une esta sencilla molécula haya surgido en los albores del Mundo de las RNPs.

o **METABOLISMO DE NUCLEÓTIDOS.** Los nucleótidos son biosintetizados por dos vías distintas. En la vía *de novo* los nucleótidos son ensamblados a partir de compuestos más simples; en el caso de las pirimidinas la base se ensambla primero pieza por pieza y después se une a la estructura de la ribosa, mientras que para el caso de la purinas la base se ensambla directamente en la ribosa, ambas rutas comprenden un pequeño número de reacciones elementales repetidas iterativamente para generar diferentes nucleótidos y se plantea que dichas rutas aparecieron muy tempranamente en la evolución; en las rutas de reciclaje las bases y demás componentes son recuperados y reconectados a la unidad de ribosa⁵⁹⁴ (de allí 25.1 a 25.6). Sin embargo sólo algunas de las enzimas involucradas en el metabolismo de ribonucleótidos, existían en organismos predecesores de LUCA y las que se identificaron por contener regiones codificadas por tripletes RNY se consideran

enzimas del metabolismo de las purinas, relacionadas particularmente con la transformación de ribonucleótidos y ribonucleósidos.

- SINTASA DE GMP.
- La sintasa de GMP (GuaA), formalmente llamada **xantosina 5'-P** ligasa de amoniaco, cataliza la reacción reversible $\text{XMP} + \text{ATP} + \text{Gln/NH}_3$ (como fuente de nitrógeno) \leftrightarrow $\text{GMP} + \text{AMP} + \text{PPi} + \text{Glu}$ (sólo si Gln fue donador) y Mg^{2+} como cofactor. Algunas peculiaridades de esta reacción son: existe preferencia para utilizar Gln, más que NH_3 , como fuente de N; se pueden utilizar análogos de XMP aunque éste es el substrato más eficiente; el PPi actúa como inhibidor competitivo del ATP y de hecho la unión de cualquiera de estas dos moléculas determina la dirección del proceso catalítico; finalmente el exceso de GuaA es letal pues conlleva a una depleción de substratos en cualquier sentido de la reacción^{595,596,597,598,599,600}.
- GuaA es una enzima dimérica, de casi 1600 nt, poco más de 500 aa y ~60kDa por subunidad, que pese a la amplia diversidad de organismos procariontes y eucariontes que la poseen, es altamente significativo el grado homología; esta enzima pertenece a la clase I de la familia 'amidotransferasas de glutamina' (GlnATs) cuya característica es poseer especificidad dual por el aa Gln y por NH_3 y en que ambos sitios de unión, a uno u otro substrato, se conectan con el sitio activo mediante un túnel estructural interno^{585,601,602}.
- Se considera que los primeros 200 aa o dominio-N corresponden al sitio de unión a la glutamina, donde es particularmente crítico el residuo Cys⁸⁶ que forma un intermediario covalente para llevar a cabo la transferencia de la amida, lo que provoca inactivación selectiva de la enzima para utilizar solamente Gln como molécula donadora de nitrógeno, este residuo es parte

de la triada conservada CHE que promueve la hidrólisis enzimática; a pesar de ello, el dominio-N no posee actividad glutaminasa *per se* en ausencia de los otros substratos, aunque si el dominio-N se elimina, aumenta la hidrolización de ATP, por lo que se considera que el dominio-N regula dicha actividad. El dominio-C contiene el característico motivo-PP o bucle-P cuya secuencia es **SGGVDSS** y al que se une la adenina del ATP, tal sitio funge como una ATP-PPasa que hidroliza el ATP y donde se lleva a cabo la condensación de éste con el XMP, adicionalmente el dominio-C contiene el sitio de unión al NH₃; se considera que el dominio-C es la porción ancestral de la enzima, la cual se fusionó con un dominio de glutaminasa en el extremo-N, lo que permitió que la enzima funcionase aun cuando la concentración de amoniaco fuera baja^{585,603,604,605,606,607,608,609,610,611}.

- La glutamina no es un aa del Mundo del RNA y los aa de este sitio catalítico tampoco son aa codificados por tripletes RNY, por otro lado el nitrógeno es una molécula presente en la tierra primitiva y tanto el ATP como el XMP pueden ser sintetizados prebióticamente, se ha planteado además que el dominio-C es la porción ancestral de la GMP sintetasa y dado que todos los aa del bucle-P, que unen esas moléculas prebióticas, son codificados por tripletes RNY, es probable que su origen se remonte a los albores del Mundo de las RNPs.
- FOSFORRIBOSILTRANSFERASA DE HIPOXANTINA.
- La fosforribosiltransferasa de hipoxantina (HPRT) cataliza la adición del **5'fosforibosil- α -1-pirofosfato**, considerado como el metabolito central en la síntesis y reciclaje de nucleótidos, a las 6-oxo-purinas hipoxantina, guanina o xantina (en orden de eficiencia) para producir inosina-, guanosina- o xantosina-monofosfato respectivamente, en la reacción: PRPP + Hx/Gu/Xa

→ PPI + IMP/GMP/XMP, utilizando dos cationes iguales Mg^{2+} o Mn^{2+} como cofactores, siendo el primero más eficiente. La HPRT de procariontes es un homotetrámero, u homohexámero en algunos casos, de ~200 aa y ~21 kDa por subunidad, de estos aproximadamente los primeros 100 aa contienen el sitio de unión a nucleótidos, mientras que el substrato PRPP se une en el dominio-C. Se ha propuesto que varias PRT y algunas nucleosidasas, pese a tener distinta especificidad, comparten un ancestro común con base en el alineamiento de secuencias, pues existe un considerable grado de conservación en el sitio de unión a fosfato. A su vez la HPRT pertenece a las PRT de purina (PRT-R) y éstas a la familia I de las PRT (PRT-I), en ellas la estructura medular consiste de 4 o 5 hojas- β paralelas rodeadas por 3 o 4 hélices- α , donde los extremos-C de las hojas- β forman el sitio activo ubicado en la base de la estructura. En el sitio activo donde se une el PRPP hay un motivo conservado de 13 aa, que contiene por lo general 4 hidrófobos, 2 ácidos y 7 variables, usualmente G o T; de estos se identifica una porción de 11 aa donde los 2 primeros son ácidos y los 9 restantes conforman un bucle flexible que une la ribosa-5-P del PRPP o bien un bucle de 4 aa que une el PPI y donde el último aa es **[G/A/S]** que es esencial para la unión del catión metálico. La secuencia del sitio de unión al PRPP es **V[V/I/L][V/I/L][V/T]DDG[V/I/A/S]A[T/M]G** **[A/G/T/Y/S][S/T]**, rodeada por los aa hidrófilos Arg³², Arg⁵⁵, Lys⁵⁶, Glu⁵⁴, Arg¹⁰⁵ y Glu¹⁹³. Las PRT-I presentan en particular un plegamiento tipo Rossmann cuya región medular consiste de 4 bucles que despliegan funciones importantes durante el proceso catalítico: el bucle I contiene el tripéptido invariante **A**³⁴**R**³⁵**G**³⁶ que sirve para el reconocimiento del mononucleótido; el bucle II, que contiene K⁵⁹X₁E⁶⁹, es altamente flexible y se dobla hacia el sitio activo protegiéndolo; el bucle III contiene un sitio de

unión a fosfato con el motivo $\mathbf{V}^{86}\mathbf{X}_{10}\mathbf{T}^{95}$, que sirve para el fosfato del nucleótido o para el **5'P** del PRPP, además de un aa conservado \mathbf{D}^{137} que participa activamente en la catálisis actuando como base general y una \mathbf{K}^{165} que posiciona la purina; del bucle IV se considera que no contiene motivos de aa característicos, salvo un aa aromático (usualmente \mathbf{F}^{186}) que une la purina, un \mathbf{D}^{197} que une un \mathbf{Mg}^{2+} y finalmente una \mathbf{R}^{199} que participa en la unión de un grupo fosfato que puede ser del PRPP o del PPI, estos últimos sirven para mantener la actividad catalítica de la HPRT; adicionalmente, dependiendo de la fuente, existen tres grupos (G) de HPRTs que contienen aa distintos en posiciones específicas: las del G1 contienen \mathbf{L}^{67} en el bucle I y $[\mathbf{S/R/F/L}]^{104}\mathbf{Y}^{105}$ en el bucle II, las del G2 contienen \mathbf{M}^{67} en el bucle I y $\mathbf{R}^{104}\mathbf{Y}^{105}$ en el bucle II, finalmente las del G3 contienen $[\mathbf{A/S}]^{67}$ en el bucle I y $\mathbf{F}^{104}\mathbf{Y}^{105}$ en el bucle II. Excluyendo el motivo conservado de unión al PRPP y los otros motivos conservados, las PRT sólo comparten homología de $<15\%$ ^{612,613,614,615,616,617,618,619,620,621,622}.

- Varios de los aa de algunos bucles conservados en las PRTs son codificados por tripletes RNY; sin embargo la región conservada más extensa que contiene aa de este tipo es el sitio que une el PRPP, dicha molécula es fundamental y se considera universal para la producción de nucleósidos; si además de ello el PRPP o partes de éste son prebióticos, es probable que el motivo de aminoácidos que le une constituya la región identificada como ancestral por la metodología en la presente tesis *i.e.* ya existía en los albores del Mundo de las RNPs.

- PROTEÍNA DE LA FAMILIA DE LA NUCLEÓSIDO-HIDROLASA PREFERENTE PARA INOSINA-URIDINA.
- Las hidrolasas o ribohidrolasas de nucleósidos (NHs) son glicosidasas que hidrolizan el enlace N-glucosídico de los ribonucleósidos- β , en la reacción reversible: nucleósido + H₂O \leftrightarrow ribosa + nucleótido, utilizando Ca²⁺ como cofactor; estas enzimas son altamente específicas para la ribosa, no así para las bases nitrogenadas, ya que no están restringidas estrictamente a algún tipo en particular, haciendo más eficiente el reciclaje de ambos componentes nucleotídicos. No obstante la amplia especificidad del substrato, las NH presentan sutiles preferencias por un tipo de nucleótido, con base en lo cual se han identificado tres subgrupos, uno de los cuales comprende las hidrolasas de nucleósido que muestran preferencia por inosina y uridina (NHIU) que a su vez pertenece al grupo I de las NH (NH-I), que es el más importante en el reciclaje de purinas, ya que es eficiente aún con baja concentración de substratos, que pese a su preferencia éstos pueden ser en realidad todos los nucleósidos de purina y pirimidina comunes; adicionalmente se ha propuesto, con base en alineamiento de secuencias, que algunas nucleosidasas comparten un ancestro común con varias PRT pese a tener distinta especificidad, pues existe un considerable grado de conservación en el sitio de unión a fosfato. Estructuralmente, las NH comprenden una familia de metaloproteínas relacionadas, cuya topología de hojas- β se considera única. Las NHIUs mejor caracterizadas provienen de eucariontes, al parecer no exhiben propiedades alostéricas y se presentan como un homotetrámero de 34kDa por subunidad, cuya estructura consiste en ocho hojas- β rodeadas por 6 hélices- α y un lóbulo pequeño adicional de 4 hélices- α , en las NHIUs el sitio activo se localiza en

la porción terminal-C de las hojas- β . El sitio activo es responsable de la estricta especificidad por la ribosa y la actividad promiscua hacia las nucleobases, las cuales se unen en un bolsillo hidrófobo; en particular 7aa conservados en todas las NHs son responsables de ello y esenciales para la catálisis; por un lado la triada HYY funge como donador de protones para la purina, lo que facilita su liberación de la enzima; por otro lado 4Asp, del motivo conservado **DxDxxxDD** que identifica a todas las NH, interaccionan con la ribosa del nucleósido, uno de éstos interacciona además con la molécula de H₂O que a su vez une el Ca²⁺ en una región profunda de la enzima^{620,623,624,625,626,627}.

- De los dos sitios conservados en la NIUH, aquel que contiene aa codificados por triplete RNY es el sitio de unión a ribosa, dado que éste funciona como **una 'firma' para** detectar NHs, es posible que tal región haya sido la identificada como ancestral por la metodología descrita en la presente tesis; si además se considera la ribosa como prebiótica, es posible entonces que el motivo de aa que une dicha azúcar sea, no sólo aquel cuya antigüedad se remonte a los albores del Mundo de las RNPs, sino que sobre este se ensamblaron otros dominios hasta conformar las NHs específicas actuales.
- FOSFORILASA I DE NUCLEÓSIDOS DE PURINA, ESPECÍFICA PARA GUANOSINA E INOSINA.
- Las fosforilasas de nucleósidos (NP) son enzimas ubicuas involucradas en la fosforólisis reversible de enlaces N-ribosídicos, estas enzimas se clasifican en dos tipos (I y II) dependiendo de su topología; en particular las fosforilasas de nucleósidos de purina (tipo I, PNPs-I) son enzimas inducibles por sustrato que catalizan la hidrólisis reversible de los enlaces de los (desoxi)nucleósidos de purina, liberando sus componentes y revirtiendo la

orientación anomérica del nucleósido en la reacción: β -nucleósido de (desoxi)purina + $\text{PO}_4^{3-} \rightleftharpoons$ purina + α -(desoxi)ribosa-1-P, de este modo las PNP-I son esenciales en el reciclaje de purinas, mismas que pueden interconvertirse mediante otras reacciones, implicando además el reciclaje de energía. Existen dos tipos de PNP-I dependiendo de la especificidad por determinada nucleobase y su arreglo oligomérico, lo que refleja parcialmente la fuente biológica: las PNP triméricas, de ~90kDa con ~31kDa por monómero, son específicas para las 6-oxopurinas (6oxoR) guanosina (Guo) e inosina (Ino) y son características de mamíferos; mientras que las PNP hexaméricas (trímeros de dímeros), de 110 a 150 kDa con ~26 kDa por subunidad, específicas para 6oxoRs y adenosina (Ado), una 6-aminopurina (6NR), son típicas de bacterias; no obstante existen PNP-I triméricas en microorganismos e inclusive organismos con PNPs de ambas configuraciones. Cabe señalar que no obstante las PNPs-I mencionadas comparten sólo ~11% de identidad en secuencia, las estructuras de sus monómeros son notablemente similares, en particular su módulo esencial α/β , del cual sobresalen 3 sitios conservados esenciales para la unión de los substratos/productos (dependiendo de la direccionalidad de la reacción), el sitio activo y el sitio de contacto inter-subunidades (numeración basada en *E. coli*); algunos de esos sitios comparten aa en común y la unión del sustrato induce cambios conformacionales de estructura abierta a cerrada durante la catálisis, lo que conlleva a la convergencia de todas las regiones. Las interacciones inter-subunidades se dan mediante dos tipos de bucles: uno rico en Gly **G⁶³MGIPS⁶⁸**, donde la M⁶⁴ junto con V⁴² y R⁴³ están involucrados además parcialmente en la unión al fosfato, el otro bucle corresponde al di-

aminoácido **I**⁷¹**Y**⁷²; el sitio de unión al fosfato se compone a su vez por tres Arg (R²⁴, R⁴³ y R⁸⁷) que le confieren una fuerte carga positiva, dos moléculas de H₂O y **G**²⁰ y **S**⁹⁰, este último junto con C⁹¹, **G**⁹² y **A**⁹³ conforma además el sitio activo; tanto el sitio que une la ribosa como el que une la purina son generalmente hidrófobos y comparten los aa **V**¹⁷⁸ y M¹⁸⁰, que mientras con E¹⁷⁹ forman enlaces van der Waals con la pentosa y con una His un puente de hidrógeno, junto con F¹⁵⁹, F¹⁶⁷, **S**²⁰³, **D**²⁰⁴ y **I**²⁰⁶ forman el sitio de unión a la base nucleotídica, aunque en realidad sólo el Asp y someramente la Phe interaccionan directamente, lo que pudiese explicar la versatilidad en el reconocimiento de purinas y que no obstante se considera que el sitio catalítico de los hexámeros es más accesible que el de los trímeros, se ha mostrado que cambios de únicamente 4aa permiten ampliar o restringir la especificidad de la enzima por determinados tipos de purinas, manteniendo intactas las demás regiones^{628,629,630,631,632,633,634,635,636,637,638,639,640,641,642}.

- Se ha propuesto que varias PRTs y algunas nucleosidasas, pese a tener distinta especificidad, comparten un ancestro común con base en el alineamiento de secuencias, pues existe un considerable grado de conservación en el sitio de unión a fosfato, en particular la PNP-I pertenece a la familia denominada I que incluye hidrolasas y fosforilasas de nucleósidos en procariontes y algunas enzimas inducibles por estrés en plantas, dicha familia se caracteriza por un motivo [h/C][h/C]_{x2}[**G/N**]_{x2}[**G/A/S**]_{ax5}[a/**A**]_{x1-2}**D**hx[a/C] donde 'h' es un aa hidrófobo [**I/L/V/M/F/Y/W**] y 'a' uno aa alifático [**I/L/V/M**], este sitio corresponde al sitio de unión al nucleósido en las PNPs-I⁶²¹. Como se ha señalado previamente, existen dos familias de NP (I y II), que pudieron haber evolucionado de manera independiente, caracterizadas por topologías distintas y especificidad por

nucleósidos de distintos tipos, cada uno de estos plegamientos se considera entonces el ancestro común de todas las enzimas de esa superfamilia, de tal modo que la familia de NPs-I es específica para purinas y la de NPs-II para pirimidinas, cada una formando su propia clade; *i.e.* las PNPs-I triméricas y las hexaméricas comparten un ancestro común que era capaz de reconocer purinas o sus nucleósidos de manera genérica, además se ha especulado que dado que ambos tipos de NPs tienen un sitio similar de unión al fosfato y el mecanismo catalítico es el mismo, es posible que todas las NPs compartan algún ancestro molecular lejano^{643,644}.

- Los 5 sitios significativos de las PNPs contienen aa codificados por triplete RNY, por ende no es posible determinar cuál de ellos es el más ancestral con base en esa única característica y aunque se especula que el sitio de unión al fosfato es primordial, los análisis filogenéticos sugieren que el sitio de unión al nucleósido es el módulo más antiguo de las PNPs-I además que ésta es la región más extensa cuyos aminoácidos unen una molécula de origen prebiótico, aunque en realidad los 3 componentes de la purina (base nitrogenada, ribosa y el grupo fosfato) surgieron posiblemente antes del Mundo de las RNPs.

○ **METABOLISMO DE LÍPIDOS.** Se ha intentado dilucidar la naturaleza de la frontera biológica del ancestro común, si ésta existía, planteando diversos escenarios evolutivos para su origen; uno de los más plausibles sugiere que LUCA poseía una membrana plenamente definida, pues contenía la mayoría de las enzimas del metabolismo de los lípidos, varios de los cuales pudieron sintetizarse prebióticamente; no obstante se trataba de una barrera heteroquiral, ya que solamente existía la materia prima enzimática de la glicerol fosfato deshidrogenasa pero no la proteína *per se*⁶⁴⁵. Como en todos los ciclos metabólicos existe el

catabolismo y el anabolismo de lípidos, pero únicamente el espiral biosintético de los ácidos grasos cuenta con 2 proteínas que contienen regiones codificadas por tripletes RNY; aunado a lo anterior, se considera que las proteínas identificadas inciden en diversos procesos metabólicos actuales y no sólo en el anabolismo de los lípidos, de modo que su presencia como enzimas ancestrales, o al menos parte de ellas, resulta relevante; como punto de interés adicional, ambas enzimas son entidades independientes en bacterias y no así en eucariontes, donde forman parte de complejos enzimáticos, lo que podría indicar que a través de la historia evolutiva, las unidades funcionales respectivas se fusionaron.

- PROTEÍNA DE LA FAMILIA DE LA HIDRATASA / ISOMERASA DE ENOIL-CoA.
- La crotonasa (ECH o HCH) cataliza el segundo paso clave de la espiral de oxidación- β de los ácidos grasos, éste consiste en la reacción reversible o de hidratación *syn* del 2-trans-enoil CoA o crotonil-CoA (hidratasa de enoil-CoA, ECH), o de hidrólisis del S/L-3-hidroxiacil CoA (hidrolasa de hidroxiacil-CoA, HCH) donde el protón- α del 3-hidroxiacil regenera el H₂O, o bien derivados de éstos con 4 y hasta 20 átomos carbono de longitud. La primera fuente biológica de la crotonasa fue una mitocondria de eucarionte, sin embargo los residuos clave y el proceso catalítico son los mismos para la crotonasa de procariontes; con la diferencia que en eucariontes la crotonasa es una enzima individual, mientras en bacterias es parte de un complejo denominado FOM, que consta de dos subunidades en que la crotonasa se localiza en la porción-N de la subunidad α , en particular los primeros 200aa que son aproximadamente idénticos en ambas proteínas. La crotonasa es un dímero de trímeros (hexámero) que consta de 10 a 12 hojas- β y ~11 hélices- α , siendo el dominio-N abundante en hojas- β y el

dominio-C en hélices- α ; la topología de cada monómero es una espiral de cuatro vueltas orientada hacia la derecha, en que cada vuelta consiste de dos hebras- β y una hélice- α , semejante a un plegamiento Rossmann, seguido por dos pequeños dominios involucrados en la trimerización. La porción acetilo del substrato correspondiente se encuentra localizado en la parte profunda de la enzima hacia el espacio inter-trimérico, mientras las partes ADP-3'-fosfato y ácido-pantoténico se encuentran expuestas al solvente; finalmente la parte del ácido graso se sitúa en un túnel específico que modifica su configuración dependiendo de la longitud de este grupo, que puede curvarse asistido por un bucle flexible **GAGG** y otros dos residuos hidrófobos, de modo que las cadenas más largas de ácidos grasos entran de vuelta al espacio inter-trimérico donde está el extremo acilo de la AcCoA; cerca de dicho espacio se localiza también el sitio catalítico cuya arquitectura es rígida y estable, lo que confiere eficiencia catalítica a la enzima. La huella molecular de la crotonasa se encuentra del aa 131 al 150, donde se localiza el sitio catalítico mencionado, donde hay una cavidad para el oxianión, compuesta por los **A**⁹⁸ y **G**¹⁴¹ (en mitocondria) críticos para la actividad enzimática, así como los residuos clave E¹⁴⁴ y E¹⁶⁴ (E¹²⁰ y E¹⁴⁰ en bacteria), el primero fungiendo como un ácido/base para el adecuado posicionamiento y posterior activación de la molécula de agua durante la catálisis y el segundo involucrado directamente en la protonación/desprotonación el carbono- α del grupo tioéster de la CoA durante la isomerización del 3-enoil CoA; adicionalmente, se ha mostrado que una isomerasa de la superfamilia mencionada puede adquirir actividad como crotonasa, mediante un único cambio de Lys²⁴² en el extremo-C, mismo que no se considera fundamental para ninguna de las dos proteínas,

pues incluso está situado lejos del sitio activo^{646,647,648,649,650,651,652,653,654,655,656,657,658,659,660,661,662,663,664,665,666}.

- La crotonasa forma a su vez una superfamilia de hidratatas/isomerasas funcional y mecanísticamente similares, ya que se consideran parte del metabolismo de lípidos y la catálisis requiere la estabilización de un intermediario enolato derivado de un substrato-AcCoA; pese a que estas enzimas comparten poca identidad en la secuencia general, contienen una región conservada a lo largo del dominio-N, involucrada en la unión del grupo carbonilo del extremo acilo de la AcCoA, la cual corresponde a la huella molecular de la crotonasa que incluye los aa conservados Glu¹⁶⁴ y Gly¹⁴¹ y se caracteriza por la secuencia de aa: F[C/A/T/S][A/S/T] G[F/A/I/H/W/G/L][D/Y]_{x30}[H/L/Q/R/F/E/Y]_{x22}G/A/G/G/AX[G/E/D/V/L]_{x21} [W/E/P/G]_{x6}[N/P/A/Y/D][D/G/N/P] (subrayados los más conservados). Por lo anterior se considera que los miembros de esta superfamilia están evolutivamente relacionados y la diversificación se dio mediante la localización estratégica de aa polares que llevaron a cabo distintas funciones catalíticas^{667,668,669}.
- De acuerdo con algunas investigaciones enfocadas a la indagación del genoma mínimo, la AcCoA podría ser una molécula clave del metabolismo general e indispensable en organismos de metabolismo fermentativo, es posible entonces que LUCA y organismos antecesores tuviesen un metabolismo de este tipo, sin olvidar que probablemente también utilizaban el CO₂ como fuente de energía (ver la sección "transporte y metabolismo de aminoácidos"). Adicionalmente, puede plantearse que varios de los módulos de la AcCoA pudieron haber surgido prebióticamente y como se puede observar, varios de los aa que se encuentran en la región considerada la huella de la crotonasa, que une dicho cofactor, son codificados por

tripletes RNY y aunque no es posible decidir hasta el momento cuál de estos fue identificado como ancestral por la metodología descrita en la presente tesis, no es de sorprender que un motivo de aa que une la AcCoA, o parte de esta molécula universal, haya surgido en los albores del Mundo de las RNPs.

- CARBOXILASA DE BIOTINA DE LA CARBOXILASA DE ACETIL-CoA (AcCoA).
- La carboxilasa de acetil CoA (AcCoA) cataliza el primer paso en el anabolismo de lípidos / ácidos grasos, mediante dos medias reacciones, cada una de las cuales es catalizada por un péptido distinto, ya sea que provengan de dominios de una misma proteína como en eucariontes o bien de proteínas individuales como en el caso de bacterias, con excepción de las cianobacterias donde ambas proteínas son entidades separadas como en eucariontes. En cualquier caso, la carboxilasa de biotina (AccC) es un homodímero estable, cuya interface es principalmente hidrófila y consta de ~450aa y ~50 kDa por subunidad, ésta cataliza la primera media reacción de la carboxilación de la AcCoA y utiliza Mg^{2+} o Mn^{2+} como cofactor. El mecanismo consiste en la fosforilación del bicarbonato por ATP para formar un intermediario carboxifosfato, que será transferido al grupo carboxilo del cofactor biotina para formar carboxibiotina, ya sea que se encuentre libre o unido a su acarreador; para dicha reacción, la presencia del catión metálico divalente incrementa la unión de la enzima al ATP, mismo que reacciona con el bicarbonato sin la necesidad de la biotina, la unión del complejo ATP- Mg^{2+} promueve a su vez un cambio en la conformación de la enzima de estructura abierta a cerrada y la hidrólisis de esta molécula es estimulada únicamente por la interacción con la biotina. Existen dos sitios activos completos por homodímero y el ciclo catalítico descrito ocurre

alternadamente en cada subunidad y aunque la dimerización no es esencial para la catálisis, ésta disminuye considerablemente si el dímero no se forma, *i.e.* son subunidades interdependientes^{670,671,672,673,674,675,676,677,678,679,680,681,682,683,684,685,686}.

- Estructuralmente se considera que cada subunidad de la AccC consta de tres dominios: el A de M¹ a **I**¹⁰³, un pequeño dominio HTH, el B de **V**¹³¹ a Y²⁰³ y el C de R²⁰⁸ hasta el último aa (numeración basada en *E. coli*). De estos, sólo el dominio central contiene el sitio activo, conformado o rodeado por los aa: K¹¹⁶, K¹⁵⁹, H²⁰⁹, E²¹¹, H²³⁶, K²³⁸, E²⁴¹, E²⁷⁶, **I**²⁸⁷, E/K²⁸⁸, **N**²⁹⁰, R²⁹², Q²⁹⁴, **V**²⁹⁵, E²⁹⁶ y R³³⁸, algunos de los cuales se conservan estructuralmente en otras enzimas de unión a ATP; de los aa mencionados la secuencia E²⁸⁸**MNPR**²⁹² corresponde al sitio catalítico; los aa involucrados en la unión de nucleótidos son K¹¹⁶, E²⁰¹, K²⁰², L²⁰⁴, H²⁰⁹, H²³⁶ y E/K²⁸⁸, de los cuales K¹¹⁶, K²⁸⁸ y H²³⁶ interactúan con los grupos fosfato, mientras que E²⁰¹, K²⁰² y L²⁰⁴ interactúan con el anillo de adenina del ATP; el motivo **G**¹⁶²**GGGGRG**¹⁶⁸ rico en Gly y que confiere flexibilidad a la enzima, también es un sitio de unión a ATP, en particular las **Gly**¹⁶⁵ y **Gly**¹⁶⁶ unen el complejo ATP-Mg²⁺ mediante el fosfato-β. Por otro lado, de los aa mencionados E²¹¹, E²⁸⁸, **N**²⁹⁰ y R²⁹² conforman además el sitio de unión a biotina junto con C²³⁰, C³³⁶, K²³⁸ y un motivo MK[M/L] que están conservados en todas las carboxilasas dependientes de biotina^{674,679,680,682,687,688,689,690,691,692}.
- La secuencia y mecanismo catalítico de la AccC presenta homología considerable con otras carboxilasas dependientes de biotina, que utilizan esta molécula como sustrato o como ligando alostérico, dichas enzimas se encuentran altamente conservadas en los 3 dominios de la vida y por tanto es posible que hayan emergido tempranamente en la evolución y

probablemente compartan un ancestro común. Inicialmente se consideraba que una de las regiones que contenía ese ancestro era precisamente el sitio de unión a la biotina, sin embargo no existe un motivo universal de unión a tal cofactor. Por otro lado el plegamiento ATP-grasp, cuyo motivo es el característico **GGGGGRG**, que semeja a su vez al sitio **GxGxxG** de enzimas que unen nucleótidos, se considera ubicuo en la naturaleza, en particular es parte fundamental de todas las carboxilasas dependientes de biotina descritas hasta ahora y en general se encuentra en numerosas enzimas que transforman ATP en ADP durante una reacción de ligación^{674, 679,680,682,693,694,695,696}.

- Como se puede observar, algunos de los aa del sitio activo como **Val**, **Ile** y **Asn** son codificados por tripletes RNY, no obstante la región extensa que contiene aa de este tipo es el bucle de unión al ATP-Mg²⁺ abundante en **Gly** y dado que los componentes que conforman dicho complejo pudieron existir en un ambiente prebiótico, es probable que el sitio que lo reconoce haya surgido probablemente en los albores del Mundo de las RNPs y por tanto constituiría el primordio de la enzima.

- **METABOLISMO DE COENZIMAS.**

- Las coenzimas son compuestos nucleotídicos de origen prebiótico que funcionan como donadores/aceptores de electrones y dado que varios aa son también de síntesis prebiótica, no es de sorprender en realidad que ambos tipos de moléculas comenzaran a interactuar en ciclos proto-metabólicos con cierto grado de afinidad. Posteriormente las coenzimas escasearon debido a su utilización masiva y posiblemente los péptidos que inicialmente sólo las unían participaron en su biosíntesis, la síntesis intracelular de coenzimas habría asegurado entonces cierta independencia del medio externo, lo que promovería más adelante el desarrollo de

organismos más complejos como el LUCA^{697,698}. De las diversas proteínas consideradas actualmente del metabolismo de coenzimas, las únicas que contienen regiones codificadas por tripletes RNY, cuyo valor-E es $0.05 < E < 0.1$, son la oxidasa de coproporfirinógeno III, independiente de oxígeno y una preniltransferasa de la familia UbiA.

- OXIDASA DE COPROPORFIRINÓGENO III, INDEPENDIENTE DE OXÍGENO.
- La oxidasa, independiente de oxígeno, de coproporfirinógeno III (HemN), cataliza el 8º paso de los 10 necesarios para la síntesis específica de heme, a partir del cual se puede producir también siroheme y cobalamina, o el segundo paso en la síntesis general de tetrapirroles, que son cofactores ubicuos esenciales para la transferencia de electrones en los 3 dominios de la vida. El nombre sistemático de HemN es oxidoreductasa/d Descarboxilasa de coproporfirinógeno-III (**copro'gen**) : S-adenosil-L-metionina (S-AdoMet), o bien deshidrogenasa de protoporfirinógeno IX (**proto'gen**), cuya reacción catalizada sigue: **copro'gen** + 2(S-AdoMet) → **proto'gen** + CO₂ + 2(L-Met) + 2(**5'desoxyadenosina**, **5'dAdo**), utilizando NADH⁺, NADPH⁺ y Mg²⁺ como cofactores; de tal modo que los grupos propionato de los anillos A y B del **copro'gen** son descarboxilados, quedando en su lugar grupos vinilo y formándose el **proto'gen**, mientras que los 4e⁻ liberados son captados por S-AdoMet que se descompone en L-Met y 5'dAdo. Dado que HemN es independiente de oxígeno, utiliza la molécula flavonoide, NADH o NADPH, como aceptor de electrones, proveyendo mayor actividad la primera que la segunda, por lo que puede considerarse también una flavoproteína; por otro lado, a diferencia de otras enzimas que utilizan S-AdoMet, esta molécula se transforma durante el proceso catalítico de HemN en lugar de utilizarse para una metilación, por lo que se considera un co-substrato y no

un cofactor y, aunque HemN puede utilizar L-Met y ATP en lugar de S-AdoMet, se considera miembro de la superfamilia de enzimas dependientes de S-AdoMet. Algunos procariontes poseen enzimas adicionales denominadas HemZ y HemY, que funcionan en condiciones aerobias o en condiciones de estrés respectivamente, sin embargo sólo HemN es esencial para las bacterias porque puede funcionar bajo cualquier condición ambiental, aunque en la anaerobiosis ésta depende de la activación de otras enzimas. HemN está codificada por casi 1,500 nt que dan ~500 aa, conformados en un monómero de 52 ± 5 kDa que contiene dos motivos distintivos, esenciales para el mecanismo catalítico: por un lado un bloque rico en Cys CxxxCxxC, del que Cys⁶², Cys⁶⁶ y Cys⁶⁹ son parte y que es esencial para la formación de un grupo FeS, al que Y⁵⁶ e H⁵⁸ proveen de integridad estructural, por otro lado un bloque rico en Gly **GGGTP**, que contiene **Gly**¹¹¹ y **Gly**¹¹³, es esencial para la unión de S-AdoMet y para el mecanismo radical. Ambos motivos mencionados son característicos de las enzimas dependientes de S-AdoMet^{699,700,701,702,703,704,705,706,707,708,709}.

- Existen dos tipos de enzimas dependientes de S-AdoMet, las metiltransferasas (MT) y las radicales, a éste último pertenece HemN, cuya estructura contiene dos sitios altamente similares de unión a S-AdoMet consecutivos en la estructura, con lo que se llevan a cabo las descarboxilaciones de los propionatos simultáneamente; este tipo de S-AdoMet radicales contienen los bloques mencionados, el rico en Cys para la unión del grupo FeS y el rico en Gly que une S-AdoMet; adicional a este último, existen algunos otros aa que conforman el bolsillo de unión a la adenina de la S-AdoMet, como el bucle **I**²¹¹**IGL**²¹⁴ y uno de F²⁴⁰ a **A**²⁴³; finalmente, otros bloques característicos e importantes para el mecanismo catalítico son: un motivo P²¹**RYTSYPTA**²⁹, crítico para la unión al sustrato y

dos motivos K⁷³**IVT**⁷⁶ y T²⁴⁷**IFAAQ**²⁵², que pudiesen servir como donadores de e- para la flavodoxina; todos los motivos mencionados están conservados en el dominio-N, pues sólo este contiene el sitio catalítico de HemN. Dado que HemN es la única enzima de su tipo que funciona bajo estrés oxigénico, ésta se ha considerado un fósil molecular cuyo origen se remonta a la atmósfera anoxigénica primitiva^{710,711,712,713,714,715}.

- La HemN puede considerarse como un fósil molecular, cuyo origen se remonta a la atmósfera anoxigénica primitiva, no obstante en la actualidad su función catalítica bajo condiciones anaerobias depende de otras enzimas, de tal modo que posiblemente el dominio más antiguo no sea el involucrado directamente en la catálisis **y puesto que copro'gen y proto'gen** poseen cierto grado de complejidad que no habría existido en las moléculas prebióticas, lo más probable es que la región identificada como ancestral corresponda al sitio de unión al co-substrato S-AdoMet o bien al sitio de unión a la coenzima nucleotídica, pues ambos contienen motivos altamente conservados de aa codificados por tripletes RNY.
- PRENYLTRASNFERASA DE LA FAMILIA DE LA UbiA.
- Las quinonas son moléculas lipófilas e hidrófobas, biológicamente relevantes pues fungen como acarreadores redox móviles en varias cadenas de transporte de electrones. La síntesis de las quinonas procede a partir del corismato, sirviendo de precursores el 2-succinilbenzoato y subsiguientemente el 1,4-dihidroxi-2-naftoato y su síntesis es necesaria para la síntesis de orotato bajo condiciones anaerobias, además que su deficiencia conduce a una estricta dependencia de una fuente externa de uracilo en anaerobiosis. En bacterias existen dos quinonas isoprenoides, ubiquinona y menaquinona, ambas poseen una cadena lateral compuesta

de un número variable de residuos isoprenoides insaturados designados K-n donde 'n' especifica el número de isoprenoides, generalmente ocho y que por tanto corresponden a un octaprenilo; la adición de dicha cadena a la estructura básica de las quinonas es catalizada por octapreniltransferasas (OPT), diferenciadas en OPT de menaquinona (MenA) u OPT de ubiquinona (UbiA); pese a lo cual ambas OPTs son altamente semejantes en estructura y mecanismo catalítico, pues ambas muestran dependencia por el catión metálico divalente Mg^{2+} , comparten el reservorio común de octaprenildifosfato y tanto los sustratos como las enzimas están asociados a la membrana^{716,717,718,719,720,721}.

- En particular el 4-hidroxibenzoato (4-HB) es un intermediario clave en la biosíntesis de la ubiquinona (UQ) y la enzima UbiA, que consta de ~300aa y ~30 kDa, cataliza la transferencia del octaprenilo a la posición 3 del anillo del 4-HB, reacción que se considera el segundo paso o paso limitante de la biosíntesis de UQ pues la transcripción del gen que codifica para UbiA es directamente proporcional al nivel de oxígeno e inversamente proporcional al suministro de glucosa; empero la reacción mencionada, UbiA posee amplia especificidad por poliprenilos, ya que ésta no radica en la longitud de la cadena lateral de isoprenoides. Las preniltransferasas (PT) poseen en general dos dominios típicos (I y II) que poseen un sitio de unión a poliprenildifosfato (PPDP) cada uno; estas enzimas catalizan las reacciones de formación de enlaces C-C entre un residuo aromático y un isoprenildifosfato (IPDP) como electrófilo. Existen tres clases de preniltransferasas, siendo UbiA de la clase I de sintetisas de IPDP, éstas se caracterizan porque la parte del carbohidrato del sustrato se une mediante interacciones hidrófobas y los aa de su sitio activo, que permiten la liberación del pirofosfato e interaccionan con el ion metálico, son diversos

Asp, Glu y Lys. Los dos sitios de unión al PPDP de las PT aromáticas se caracterizan por un motivo conservado **[N/D]xxD**, aunque UbiA sólo presenta el segundo cercano al dominio-N, que en particular se caracteriza por el motivo conservado **DDxxD** y del que sus aa son esenciales en la catálisis; adicionalmente, la enzima UbiA contiene una región hidrófila de 22 aa (183 a 205), que podría tratarse del sitio de unión al 4HB. Por otro lado el motivo mencionado de unión al PPDP también está presente en sintetasas de prenildifosfato (PDP) de elongación de cadena, al comparar MenA con UbiA se encuentra que la región de unión al PPDP contiene el motivo conservado **LANDYGD**. Estructuralmente, las PT pertenecen a una amplia clase de proteínas cuya conformación estructural es una hélice- α , aunque en realidad conforman su propio grupo, ya que todas la PT presentan un barril particular cuya conformación se basa en repeticiones de un módulo ancestral de plegamiento $\alpha\beta\beta\alpha$ ^{722,723,724,725,726,727,728,729}.

- Por secuencia, mecanismo y estructura, se ha indicado que todas las PT comparten un ancestro común que posiblemente contenía el sitio de unión al PPDP cuyos grupos fosfato pueden ser de origen prebiótico, hacia dicho sitio convergen varios motivos que contienen aa codificados por tripletes RNY y es probable que alguno de ellos haya sido el identificado como ancestral por la metodología descrita en la presente tesis, *i.e.* posiblemente surgió al inicio del Mundo de las RNPs.

- **METABOLISMO DE IONES INORGÁNICOS.** Son numerosas las proteínas que contienen sitios de unión a iones inorgánicos⁷³⁰, sin embargo la única particularmente considerada del 'metabolismo de iones inorgánicos' pertenece a la familia de las sulfatasas, posiblemente debido a que en la clasificación de las otras enzimas tiene más relevancia la reacción catalizada que la unión de iones.

- PROTEÍNA DE LA FAMILIA DE LA SULFATASA.
- Las enzimas de la familia de las sulfatasas catalizan la hidrólisis de enlaces ésteres de sulfato de una amplia variedad de sustratos sulfatados. Pese a que el azufre no es un metal las sulfatasas se agruparon inicialmente en una superfamilia de metaloenzimas, cuyas características comunes son sitios de unión a iones metálicos con un grupo de **Asp** altamente conservados, **D**⁵¹, **D**³²⁷, **D**³⁴⁶, **D**³⁶⁹ y **D**⁴³⁷ (numeración basada en una metaloenzima bacteriana que no es sulfatasa) y cuya mutación conlleva a la supresión de la actividad enzimática, pues están involucrados el mantenimiento de la integridad estructural y son indispensables para la unión del grupo sulfato en las sulfatasas así como en la unión de grupos fosfato en otras metaloenzimas. Al describirse mayor número de sulfatasas se encontró que éstas son enzimas conservadas en los 3 dominios de la vida, cuya homología en secuencia varía entre el 20% y el 60% y cuyas estructuras son totalmente superponibles, en particular en los residuos del sitio activo que conforman la región más conservada evolutivamente, dicho sitio incluye un catión metálico divalente, que puede ser Ca²⁺ o Mg²⁺, como cofactor localizado en el mismo sitio en que se unen los sustratos. Los aa críticos se localizan en la primera mitad de la cadena polipeptídica, estructuralmente en una cavidad catalítica de difícil acceso y con una disposición espacial que indica que llevan a cabo funciones idénticas en el proceso catalítico que genera y resuelve el intermediario éster sulfato-enzima, *i.e.* los residuos críticos involucrados en el proceso catalítico están conservados en secuencia y estructura de las sulfatasas, lo que indica un mecanismo catalítico común sin importar los distintos sustratos que pueden ser hidrolizados. Uno de los sitios esenciales, característicos y únicos de las sulfatasas es una hidroxiformilglicina (fG_{OH}), que se puede

formar a partir de la modificación de una cisteína conservada en eucariontes y algunos procariontes (C→fG_{OH}) o de una serina conservada en procariontes (Ser→fG_{OH}), lo que permite la distinción sulfatasas-Cys y sulfatasas-Ser; tal modificación se lleva a cabo mediante 3 procesos mecanísticamente distintos, uno propio de eucariontes y los otros dos presentes en procariontes, dependiendo si las condiciones ambientales son aerobias o anaerobias. Los aa conservados en las sulfatasas (numeración basada en una sulfatasa humana) son entonces: **D**³⁵, **D**³⁶, fG_{OH}⁷⁵, R⁷⁹, K¹³⁴, H¹³⁶, H²⁹⁰, **D**³⁴² y K³⁶⁸. Por lo anterior se considera que todas las sulfatasas provienen de un ancestro común; de hecho se ha mostrado que tan sólo un péptido pequeño que contenga una Cys/**Ser** y los aa circundantes correspondientes, es suficiente y necesario para la modificación del primero a fG_{OH}, **dicho péptido se conoce como 'firma de la sulfatasa'** y su secuencia es [C/**S**]x[P/**A**]xRx₄[**T/S**]**G**. De acuerdo con análisis filogenéticos las sulfatasas-Ser son las más ancestrales de ambas y de éstas las que contienen **Ala** (en P/**A**) están presentes en varias bacterias; adicionalmente se ha mostrado que las sulfatasas y enzimas denominadas pirofosfatasa de ectonucleótidos comparten homología en el sitio activo, lo que sugiere que posiblemente también lleven a cabo mecanismos catalíticos similares y compartan un origen evolutivo común^{474,731,732,733}.

- Las sulfatasas contienen varios sitios que posiblemente pudieron haberse identificado como la región ancestral por el método descrito en la presente tesis: por un lado pudiera ser antigua la secuencia común que comparten las pirofosftatasas mencionadas y las sulfatasas, por otro lado la **denominada 'firma de las sulfatasas'** es un bloque que contiene varios aa codificados por tripletes RNY y finalmente varios de los aa conservados a lo largo de la secuencia de las sulfatasas son Asp, en cualquier caso es

probable que la región que surgió en los albores del Mundo de las RNPs haya evolucionado hacia una Ser-sulfatasa que contenía Ala en su 'firma'.

o **NO INCLUIDAS EN COGs.** Las proteínas incluidas en algún COG representan proteínas presentes en al menos tres organismos con relación filogenética distante *i.e.* son codificadas por genes ortólogos y por tanto poseen dominios o motivos comunes; sin embargo algunas proteínas no están incluidas en algún COG porque los programas de alineamiento no los reconocen como homólogos y no se les ha sometido a una curación manual; algunas de ellas se incluyen en una categoría nombrada 'otras', mientras que aquellas cuyo mecanismo metabólico incide en diversos metabolismos se les denomina 'multifuncionales' o 'generales'^{734,735}. Tres enzimas consideradas metabólicas, que contienen regiones codificadas por triplete RNY, no están incluidas dentro de ningún COG, éstas son: una enzima de la familia de la 4-oxalocrotonato tautomerasa, involucrada en la biodegradación de xenobióticos según la base de datos KEGG; una oxidoreductasa de la familia denominada SDR (oxidoreductasa/deshidrogenasa de cadena corta, por sus siglas en inglés) y una proteína de unión a FMN, éstas últimas pudieran ser agrupadas dentro del COG "conversión y producción de energía", pues el FMN de la primera y el NAD de la segunda son importantes acarreadores de electrones que intervienen en múltiples vías metabólicas.

- TAUTOMERASA DE 4-OXALOCROTONATO (4OT).
- La tautomerasa de 4-oxalocrotonato (4OT) cataliza la isomerización reversible de α -cetoácidos insaturados, que mediante un intermediario dienólico se transforman en cetonas insaturadas, las que después de una descarboxilación entran al ciclo TCA. El mecanismo catalítico de la 4OT involucra una Pro conservada en el extremo-N, ése residuo evita las

reacciones nucleófilas de amina en el sitio activo de la enzima y en su lugar se promueven reacciones imina/enamina con varios aldehídos y compuestos cetónicos, de modo que la 4OT muestra además catálisis promiscua hacia diversos sustratos y productos, adicionalmente un sólo cambio en este aa de P¹A o P¹G provoca que no sólo se catalice la tautomerización del oxalocrotonato, sino además la descarboxilación de oxaloacetato. El mecanismo de isomerización involucra la transferencia de un H⁺, mediada por P¹ que actúa como base general y R³⁹ que actúa como ácido e interacciona con el sustrato junto con R¹¹, proceso en que también está involucrado R⁶¹ que contribuye a la estabilización del intermediario, posiblemente I⁷ y L⁸, así como M⁴⁵ y F⁵⁰ que contribuyen a generar un ambiente hidrófobo alrededor de la Pro favoreciendo una catálisis altamente eficiente, del orden de 10⁸. Estructuralmente la 4OT es un homohexámero de entre 30kDa y 40kDa, en que cada subunidad consta de sólo ~60 aa, cuya conformación se presenta como un trímero de dímeros en que la interacción de las subunidades no está mediada por enlaces covalentes. El hexámero contiene 6 sitios activos conformados por los sitios de interacción de las subunidades, de los que únicamente 3 se encuentran ocupados por cada ciclo catalítico, debido al impedimento estérico y electrostático de las moléculas que unen; el plegamiento de cada monómero es $\alpha\beta\alpha$, cuya conformación es una hélice- α , dos hebras- β , una horquilla- β , dos bucles y dos vueltas, mientras los aa remanentes se enroscan sobre sí mismos, cada subunidad está interconectada por dos bucles y tres vueltas, mientras que el sitio activo está conformado por la segunda hebra- β , el bucle que la precede y la horquilla- β , donde se encuentran los aa clave Arg³⁹ y Pro¹; un par de monómeros forman un dímero, que consta de 32 simetrías y

consiste entonces de 4 hebras- β con dos hélices- α y dos hebras- β pequeñas en cada extremo; los dímeros son ensamblados alrededor de un eje de rotación triple para formar el hexámero mediante pequeñas hebras- β mediante las cuales cada dímero contacta los dímeros vecinos. La secuencia de cada monómero de la 4OT de *E. coli*, que es **PIAQIHILEGRSDEQKETLIREVSEAISRSLDAPLTSVRVIITEMAKGHFGIGGELASKVRR**, no muestra homología con secuencias existentes en las bases de datos ni con enzimas mecanísticamente similares; no obstante la Pro¹ está conservada en isomerasas mecanísticamente relacionadas por lo que es posible que ésta sea la base catalítica general de las isomerasas^{736,737,738,739,740,741,742,743,744,745,746,747,748,749,750,751,752,753,754,755,756,757}.

- Diversos procariontes extremófilos poseen enzimas cuya secuencia o es semejante a la de la 4OT aunque no poseen Pro¹, o son proteínas que poseen la Pro¹ y son similares a la 4OT pero no están anotadas como homólogas. Por otro lado existen enzimas cuyo plegamiento es $\beta\alpha\beta$, contienen Pro¹ como residuo crítico y, aunque catalicen reacciones distintas pues dicho aa funciona en algunas como base y en otras como ácido, se considera que todas ellas conforman la familia de la 4OT. Dicha familia se subdivide en 5 subfamilias con base en la conservación de los residuos catalíticos y estructurales a lo largo de la secuencia, de éstas la subfamilia 5 es la más distante, pues sólo comparte el plegamiento característico y la Pro inicial; las demás familias contienen la P¹ y un residuo aromático o hidrófobo en la posición 50, aro⁵⁰ [F/Y/W] o hfb⁵⁰ [**V/M/I**], seguido del motivo **G**⁵¹x**GG**⁵⁴, que en la familia 5 es substituido por **A**⁵¹**VAG**⁵⁴ y provee estabilidad al oligómero, en particular la última **Gly** es esencial para la formación de una vuelta- β que servirá de anclaje para ensamblar una

subunidad con otra; adicionalmente la familia 1 contiene R¹¹ y R³⁹, la familia 2 R¹¹ y [H/V]³⁹, la familia con 3 L¹¹, L³⁹ y K³⁶, finalmente la familia 4 contiene V¹¹ y V³⁹. Dado que el plegamiento βαβ y el motivo conservado con Gly⁵⁴ son fundamentales para mantener la integridad estructural de las proteínas de la familia de la 4OT y ésta en particular es una enzima verdaderamente pequeña y se ha propuesto que nuevas enzimas pudiesen surgir a partir de modificaciones de sus características esenciales, *i.e.* la 4OT es materia prima con aún mucho potencial evolutivo^{758,759,760,761,762,763}.

- La 4OT es una proteína pequeña que parece no presentar homología con otras descritas previamente, por tanto no es posible establecer cual región es ancestral, aunque evidentemente contiene varios aa codificados por tripletes RNY a lo largo de su secuencia; en particular los residuos 50 al 54 proveen de estabilidad estructural a la proteína y esa característica fue probablemente muy importante para los primeros péptidos sintetizados en los albores del Mundo de las RNPs.
- DESHIDROGENASA/OXIDOREDUCTASA DE CADENA CORTA.
- Las deshidrogenasas/reductasas (DR) de alcoholes, dependientes de NAD, son enzimas diméricas o triméricas que catalizan los procesos redox de dichas moléculas. Con base en la longitud de su cadena pueden ser clasificadas en DR de cadena corta (SDR, DR de cadena mediana (MDR) y DR cadena larga, con menos del 20% de identidad entre los miembros dentro de cada grupo; una de las SDRs contiene una región codificada por tripletes RNY. Las SDRs constan de ~250aa de longitud y su identidad intra-grupo es de 15% a 30%, mientras que los residuos con baja identidad pudieran estar relacionados con eventos de paralogía. No obstante el bajo nivel de identidad en secuencia, es posible identificar algunos motivos clave

que se consideran la firma de las SDRs: uno de ellos es Y¹⁵²xxxK¹⁵⁶ que junto con **N**¹¹¹ y **S**¹³⁹ conforman la tétrada catalítica; otro motivo relevante es la secuencia que estructuralmente conforma un plegamiento tipo Rossmann cuyas mitades son estructuralmente semejantes entre sí, éste plegamiento interacciona con la coenzima, específicamente con NAD(H) mediante el motivo **T**¹²**G**xxx**G**x**G**¹⁹ y **D**³⁷, o con NAD(P)H mediante la modificación **T**¹³**G**xx**K****G****G**¹⁹ y R³⁹; los aa **D**⁶⁰ y el motivo **N**⁸⁶**NAG**⁸⁹ son también esenciales para la interacción con el NAD; la **N**¹⁷⁹ permite la conexión del sitio de unión al sustrato con el sitio activo y finalmente otras Gly adicionales conservadas, **G**¹³¹ y **G**¹⁸³, son importantes para el correcto plegamiento de las proteínas SDR. Las SDRs identificadas comparten una identidad del 15% al 60% con un promedio de 25% que, con base en la **similitud de sus regiones conservadas y utilizando un método de 'cadenas de Markov escondidas', se han** dividido en 7 subfamilias que comparten 3 grandes segmentos conservados, con un total de 40 posiciones que abarcan del sitio de unión a la coenzima o dominio A al sitio activo o dominio D; los aa conservados en las SDR canónicas o en las de configuración extendida son (a=aromáticos, h=hidrófobos, p=polares, c=cargados y x=cualquiera): **TG**xxx**G**h**G**/**TG**xx**G**h**aG**, con un papel estructural en la región de unión a la coenzima; **D**hx[c/p]/**D**hx**D**, que une el anillo adenina de la coenzima; **G**xh**D**hhh**NNAG**h/[**D**/E]xhhHx**AA**, que estabiliza la estructura; tres motivos: h**N**hx**G**/h**N**hh**GT**xxh**hc**, **G**xh**hx****SS**h/hhhx**SS**xxh**aG** y Yx[**A**/**S**][**S**/**T**]K/PYxx[**A**/**S**]Kxxh[**D**/E], que conforman el sitio activo; h[K/R]h[**N**/**S**]xhx**PG**xxx**T**/h[K/R]xx**NGP**, cuya función es estructural y ayuda a dirigir al reacción. De los aa indicados, la región que presenta mayor grado de conservación es la correspondiente a la unión de la coenzima, pues seis de sus aa están estrictamente conservados y siete más

se encuentran en la mayoría de las SDRs, siendo **Gly** el aa más abundante y los aa hidrófobos en segundo lugar; lo anterior indica que las regiones de homología detectable reflejan patrones comunes de plegamiento y por tanto la importancia de restricciones conformacionales; por otro lado la región más divergente de las SDRs se encuentra en el último tercio de la secuencia, que define en realidad la especificidad al sustrato. De la conservación de secuencias se infiere que la configuración estructural monomérica de las SDRs es similar, no obstante que las reacciones particulares que catalizan y las propiedades fisicoquímicas de sus sustratos, como hidrofobicidad, tamaño o rigidez, son muy distintas. De lo anterior se sugiere que primeramente evolucionó la conformación estructural fundamental de las SDRs, posteriormente la capacidad de unir una molécula nucleotídica, luego la actividad deshidrogenasa/reductasa y finalmente la especificidad por un sustrato determinado; en particular el origen de las SDRs habría sido un plegamiento Rossmann simple del tipo α/β que unía moléculas presentes en el Mundo del RNA, tal plegamiento, mediante duplicaciones internas, habría dado origen a una SDR, que siguió divergiendo y adquiriendo otros tipos de plegamientos y dominios dando origen a las MDR y posteriormente a las MDR dependientes de Zn^{2+} . Filogenéticamente se ha mostrado que la familia SDR es de las más antiguas o quizá la más antigua surgida probablemente antes del LUCA; actualmente es la familia más extensa con representantes en todos los taxa, incluyendo virus e incluso posiblemente presente en organismos aún por descubrir^{764,765,766,767,768,769,770,771,772,773,774,775,776,777,778,779,780}.

- Varios de los aa conservados en las SDR están codificados por tripletes RNY, en particular los que preservan la estructura de la enzima y los involucrados en la unión de la coenzima; estos últimos, como se ha

mencionado previamente con base en la información filogenética estructural de las SDRs, fueron posiblemente los primeros en surgir como un plegamiento tipo-Rossmann rudimentario que sólo unía la molécula nucleotídica sin poseer actividad catalítica, por tanto es esta la región que más probablemente fue identificada como ancestral por la metodología seguida en la presente tesis y es posible entonces situar su origen evolutivo en los albores del Mundo de las RNPs.

- PROTEÍNA DE UNIÓN A FMN.
- El **piridoxal 5'-fosfato (PLP)** es una molécula ubicua que participa en el metabolismo de aminoácidos, el PLP se deriva de la oxidación de piridoxina/**piridoxaminamina 5'-fosfato (PNP/PMP)** que a su vez proviene de la vitamina B₆ o piridoxina (PN). Por su parte, la mayoría de las proteínas de unión a flavina mononucleótido (FMN), además de estar involucradas en numerosos procesos redox, corresponden precisamente a oxidasas de PNP (PNPOx), también llamadas enzimas dependientes de PLP, siendo de este tipo la enzima de unión a FMN identificada como ancestral en *SagA909*. La PNPOx es un homodímero de ~50kDa, que depende de dos moléculas de FMN unidas no covalentemente, una por subunidad; dicha enzima puede funcionar en condiciones aerobias, utilizando el PLP como sustrato y produciendo H₂O₂ con el FMN como intermediario para el acarreo de electrones, o anaerobias, en cuyo caso funge únicamente como una flavoproteína oxidasa monofuncional^{781, 782, 783, 784, 785, 786, 787}.
- La PNPOx conforma a su vez una familia de enzimas que comparten homología significativa en secuencia y estructura, que consiste de 8 hojas-β rodeadas por 5 hélices-α. A lo largo de la secuencia existen dos grandes regiones o parches hidrófobos altamente conservados, en la parte medular

de la enzima (con base en la PNPOx de *E. coli*), de los que el parche largo consta de 40 aa hidrófobos de los cuáles 11 son aromáticos mientras que el parche corto consta de 23 aa hidrófobos de los que 9 son aromáticos. Cada anillo isoaloxacina de los FMN, parte fundamental de esas coenzimas, se une en su oquedad respectiva en la parte más profunda de la interfaz del dímero, contactando a ambas subunidades mediante los aa: **V**⁸⁰**FYTN**⁸⁴, **R**⁸⁷**IVLLK**⁹⁰, **E**¹⁸⁹**FWQG**¹⁹³, **R**¹⁹⁷**LHDR**²⁰¹ y los aa **Q**¹⁴⁶ y **S**¹⁴⁷, todos ellos altamente conservados en la familia de la PNPOx, por su parte el PLP queda orientado paralelamente al FMN en la enzima mediante **N**⁸⁴ y los aa **F**¹⁷⁷ y **K**¹⁴⁵; por otro lado, la unión a los grupos fosfato ocurre en la parte externa de la enzima y es mediada por varias Arg (**R**²³, **R**²⁴, **R**²⁵ y **R**¹⁵³) y **G**¹⁵⁴; finalmente 5 aa que son **K**⁷², **Y**¹²⁹, **R**¹³³, **S**¹³⁷ y **R**¹⁹⁷ contactan con el PNP y también se conservan totalmente en las PNPOx. Con base en la estructura medular y las regiones conservadas antes mencionadas, se han identificado algunas PNPOx, o al menos enzimas de unión al FMN en diversos organismos^{788,789,790,791,792,793,794,795,796,797,798,799,800}.

- Las PNPOx comparten a su vez, estructura y parcialmente secuencia con otras enzimas de unión a FMN, con las que comparten un ancestro común que posiblemente existió antes que el LUCA, el cual ya contenía muchas de las especializaciones distintas presentes actualmente los organismos⁷⁸⁸. Adicionalmente se considera que el PLP es de origen prebiótico, con base en trabajos que mencionan que la cinética del metabolismo primitivo estaba controlada por péptidos y por lo que actualmente se conoce como cofactores nucleotídicos como PLP⁸⁰¹, PMP, PNP, FMN, etc. Por lo anterior, es probable que la región que une a alguna de las dos moléculas nucleotídicas, haya sido la identificada como ancestral por el método descrito en la presente tesis, ya que ambos sitios contienen varios aa

conservados que son codificados por tripletes RNY, siendo el sitio de unión al FMN el que contiene más aa de este tipo.

Era de esperarse que varias de las enzimas identificadas como antiguas contengan regiones codificadas por tripletes RNY, como la ATP sintasa o las relacionadas con el metabolismo de nucleótidos, algunas otras como las dos involucradas en el metabolismo de lípidos o algunas del metabolismo de carbohidratos como la G3PDH tampoco son una verdadera sorpresa; no obstante se esperaba identificar como antiguas muchas más enzimas del ciclo TCA/rTCA que se ha propuesto como medular para todos los organismos toda vez que provee los metabolitos para otras rutas metabólicas, o bien de la glucólisis/gluconeogénesis como la triosa fosfato isomerasa (TIM/TPI) que por un tiempo se consideró universal aunque en realidad es propia de organismos aerobios y por ende la probabilidad de que haya sido una ruta metabólica antigua es baja, aunque sí era esperado encontrar una ruta metabólica como la fermentación casi completa sí es plausible como el primer proceso metabólico; otras enzimas identificadas fueron del todo inesperadas como la 4OT, UxuB y UxaC, una preniltransferasa o una IMP.

De acuerdo con un modelo de evolución del metabolismo denominado **'desplazamiento de genes no ortólogos'** (NOGD, por sus siglas en inglés), que indica que algunos módulos originales de las proteínas fueron posiblemente desplazados por otros y éstos a su vez desplazaron a los que surgirían posteriormente, se destaca que la gran mayoría de las enzimas indicadas como ancestrales coinciden con varias de las identificadas por la metodología descrita en la presente tesis; no obstante con dicho modelo cualitativo se indican como antiguas proteínas que no se identifican como tal bajo nuestra metodología, en particular las del metabolismo de carbohidratos aunque coinciden casi en su totalidad las del metabolismo de nucleótidos y de coenzimas⁸⁰², posiblemente debido a que el modelo NOGD pudiese

no reflejar con precisión la antigüedad en algunos casos al basarse solamente en la probabilidad de patrones filéticos.

Se identificaron primordios de enzimas fundamentales de distintas rutas metabólicas que pudieran conformar el repertorio de un organismo quimioheterótrofo. Dada la heterogeneidad de las enzimas metabólicas identificadas como antiguas bajo la metodología descrita en la presente tesis, se vuelve necesario analizar sus características comunes desde un enfoque que no involucre las propiedades bioquímicas actuales, de tal forma es posible observar que todas las enzimas metabólicas encontradas son dependientes de algún cofactor nucleotídico o un catión metálico, ambos probablemente de origen prebiótico^(cfr. 4,23), dichos cofactores se unen en los motivos conservados respectivos en cada proteína y éstos se conforman en su mayoría por aa codificados por tripletes RNY, lo que refuerza la propuesta de que el metabolismo evolucionó **'por parches'** *i.e.* que no todas las rutas metabólicas estaban completas al inicio y algunos de esos módulos surgieron en los albores del Mundo de las RNPs.

T9. Características plausibles de la probable región codificada por tripletes RNY en las proteínas identificadas del metabolismo.

Conjunto	Proteína (nombre traducido y adaptado)	Características plausibles de la probable región codificada por tripletes RNY.
Metabolismo energético	ATP sintasa F1, subunidad alfa ATP sintasa F0, subunidad C NADH oxidasa, formadora de agua Glicerol-3-fosfato deshidrogenasa, dependiente de NAD(P)+ Pirofosfatasa inorgánica, dependiente de manganeso	Unión de ATP Unión de H+/Na+ Unión de NAD(H) o quizá de FAD Unión de NAD(P)+ Unión de Mn2+, o de PPI, o de un sulfato, o de agua; o quizá la bisagra estructural
Metabolismo de carbohidratos		
Anabolismo del hexuronato	Isomerasa de glucuronato; UxaC Oxidoreductasa del D-manonato, putativa; UxuB	Interacción con Zn2+, o Co2+ o Mn2+ Unión de NADH y/o NADPH
Anabolismo del glucógeno	Sintetasa de glucógeno; GlqA Pirofosforilasa de ADP-glucosa; GlqC	Unión de grupos nucleótidos-fosfato Unión de nucleótidos, del tipo NDPs
	Fosfopentomutasa	Unión a cationes metálicos divalentes
	Hidrolasa de la superfamilia HAD, subfamilia IIA	Unión de grupos fosfato y/o del catión metálico Mg2+
	Proteína de la familia de la inositol-monofosfatasa; IMP	Unión de tres cationes Mg2+
Metabolismo de aminoácidos, aa		
	Proteína de la familia de la glioxilasa I	Quizá unión de cationes metálicos (Ni2+/Co2+)
	Aminotransferasa, clases I y II	Unión de PLP
Aminoácidos aromáticos	Deshidroquinato sintasa; AroB Amidotransferasa de glutamina	Unión de cofactor nucleotídico, NAD+ y FMN respectivamente No precisada, posiblemente unión de amoniaco
Metabolismo de nucleótidos, nt		
	Sintasa de GMP; GuaA Fosforribosiltransferasa de hipoxantina; HPRT Nucleosido hidrolasa preferente para inosina-uridina; NH-IU PNP-I, específica para inosina y guanosina	Unión de nitrógeno y/o o de ATP y/o de XMP Unión de PRPP (o partes de éste) Unión de ribosa (de un nucleósido) Unión de nucleótido (o partes de éste, como fosfato y/o base nitrogenada y/o ribosa)
Metabolismo de lípidos		
	Proteína de la familia ECH; crotonasa Carboxilasa de biotina de la carboxilasa de AcCoA	Unión de AcCoA Unión de ATP, o Mg2+, o ATP-Mg2+
Metabolismo de coenzimas		
	Oxidasa de copro'gen, independiente de oxígeno; HemN Preniltrasnferasa de la familia UbiA	Unión de S-AdoMet, o de NADH+/NAD(P)H+ o de Mg2+ Unión de PDP, quizá de los grupos fosfato
Metabolismo de iones inorgánicos		
	Proteína de la familia de la sulfatasa	Unión de iones metálicos, o quizá la firma de la sulfatasa
NO incluidas en COGs		
	Enzima de la familia de la tautomerasa de 4OT	Quizá estabilidad de la configuración estructural
¿Metabolismo energético?	Oxidoreductasa, familia SDR Proteína de unión al FMN	Unión de NAD(H)+/NAD(P)(H)+ Unión de FMN, o de PLP

6.3.1.4 Proteínas, procesamiento y modificaciones.

Entre el 25% y 30% de las proteínas que sintetizan las bacterias tienen funciones membranales o justo afuera de la célula en su pared externa⁸⁰³; algunas de las proteínas actualmente esenciales para la secreción de dichas proteínas contienen regiones codificadas por tripletes RNY (en T10 y T11), incluyendo subunidades del complejo encargado de trascolar las proteínas a través de la membrana celular, la sortasa que las corta en regiones específicas y la proteína que las ancla a la membrana. También se identificaron como antiguas, una peptidasa y una proteína rica en repeticiones de leucina.

Order	Sc.	E-v	ID	%	PROTEIN / FAMILY / DOMAIN
2	37	7.00E-11	40/41	97%	preprotein translocase, SecY subunit, putative (31.41 %)
24	27	6.00E-05	27/27	100%	membrane protein oxaA, putative (68.59 %)
73	24	0.004	24/24	100%	sortase family protein (29.61 %)
98	23	0.015	23/23	100%	leucine-rich repeat protein (0.28 %)
109	22	0.06	22/22	100%	preprotein translocase, SecY subunit (16.41 %)
120	22	0.06	22/22	100%	peptidase, S8 (subtilisin) family (39.22 %)
139	22	0.06	22/22	100%	preprotein translocase, SecA subunit (52.33 %)

T10. Proteínas que a su vez procesan o modifican proteínas y que contienen regiones codificadas por tripletes RNY, nomenclatura y código de colores igual que T1 y T2.

o **COMPLEJO DE SECRECIÓN DE PROTEÍNAS.**

o La secreción de proteínas se lleva a cabo por varios sistemas, sin embargo el complejo denominado Sec está encargado de la secreción de la mayoría y la mayoría de sus subunidades están conservadas en los tres dominios de la vida, *i.e.* bacterias, arqueas, la membrana del tilacoide del cloroplasto y el retículo endoplásmico de los eucariontes; el complejo Sec está integrado por un motor molecular llamado SecA y un poro conductor formado principalmente por SecY, aunque hay proteínas auxiliares como OxaA (homólogo de YidC) para que la trascolación sea más eficiente; el proceso está guiado tanto por hidrólisis de ATP como por fuerza protón-motriz (ΔH^+)^{803,804,805,806,807,808,809,810,811,812,813}.

- **SecA** es un homodímero hidrófilo, de 92kDa a 94kDa y 901 aa^{814,815,816}. SecA pertenece al grupo de motores moleculares, que catalizan ciclos de reacciones utilizando la energía derivada de la hidrólisis de ATP, en este caso SecA acopla la hidrólisis de ATP a la secreción de pre-proteínas⁸¹⁷, liberándolas en la porción asociada a la membrana del complejo de traslocación, aunque si SecA es eliminada de la reacción, la ΔH^+ puede mantener el proceso de traslocación hasta que haya finalizado; en determinados organismos el Na^+ provee la ΔH^+ ^{818,819} (Para mayor información consulte 803-813,820,821). Como toda ATPasa, cada dominio del homodímero SecA contiene los dos motivos Walker, o dominios de unión a nucleótidos (NBD) característicos; el NBD1 se compone a su vez de los motivos I, Ia, II y III característicos de la superfamilia de helicasas de DNA y RNA (ver más adelante), está localizado en el extremo N-terminal y tiene alta especificidad por el ATP, además sirve como región de interacción con SecY; el dominio NBD2 comprende los motivos IV, V y VI de las helicasas, se localiza en el extremo C-terminal y tiene baja especificidad por la molécula energética, sirve para la asociación con la membrana celular mediando la interacción con los fosfolípidos; ambos NBD son esenciales para el funcionamiento de la proteína y por si mismos podrían ser suficientes; finalmente el sitio de unión a la pre-proteína se localiza entre NBD1 y NBD2^{*822,823*,824,825,826}. Adicionalmente, al hacer una comparación de la secuencia de SecA de *Escherichia coli*, *Bacillus subtilis* y una bacteria corniforme, se observa que en la porción N-terminal se encuentran la mayoría de los dominios compartidos, por lo que en el trabajo se indica que la región N-terminal contiene las funciones básicas comunes de la proteína SecA, mientras que la región C-terminal determina la especificidad entre especies⁸²⁷. Se ha especulado que la unión e hidrólisis de ATP provoca

varios cambios conformacionales en SecA, pues mientras ATP está unido, la proteína se encuentra en su forma relajada, pero cuando sólo queda ADP en la enzima, ésta pasa a la forma compacta o globular, dicha tasa de recambio de ATP aumenta cuando SecA se une a SecY; tales cambios en la conformación de la proteína provocan ciclos procesivos de inserción/desinserción de SecA en la membrana conjuntamente con la pre-proteína y el complejo Sec embebido en la membrana^{822,823,828,829,830}. Durante un ciclo de procesividad se trasuelan ~5kDa de proteína (20 a 30 aa), en dos pasos consecutivos de aproximadamente 2 a 2.5 kDa cada uno, el primero corresponde a la inserción acoplada de SecA+pre-proteína y el segundo corresponde a la unión ATP+SecA+pre-proteína^{831,832,833}. Se ha encontrado que la regulación de la actividad enzimática de SecA reside en un regulador intramolecular de ATP (IRA), cuyo sitio de contacto es la región entre NBD1 y NBD2 donde SecA interacciona con la pre-proteína; se propone que IRA restringe los movimientos de SecA durante los ciclos de inserción/desinserción dirigidos por la hidrólisis de ATP⁸³⁴. Se ha propuesto un segundo mecanismo de regulación independiente de la regulación de IRA, en dicho mecanismo el sitio NBD2 se disociaría de NBD1, con lo cual disminuye la eficiencia enzimática de SecA; si se activan ambos mecanismos de regulación dicha eficiencia disminuye considerablemente a menos del 50%⁸³⁵. La eficiencia enzimática de SecA depende además de la presencia de lípidos cargados negativamente en la membrana, los cuales interaccionaban con el dominio C-terminal cargado positivamente, de tal modo que las cadenas de acil-fosfolípidos evitan que SecA pueda asociarse a la membrana y únicamente podría interactuar con SecY y de esta forma insertarse en la membrana durante el proceso de trascolación⁸³⁶. La eficiencia enzimática de SecA también depende de la utilización de los

cationes divalentes Mg^{2+} y Ca^{2+} ^{837,838,839}. Los primeros datos del proceso de trascolación de la proteína, sugieren el dímero SecA se disocia y sólo un monómero entra en la membrana llevando la pre-proteína consigo ^{840,841}. Dicha dimerización ocurriría de manera antiparalela, de tal manera que los extremos C-terminal interactúan con los extremos N-terminal ⁸⁴²; no obstante el análisis de la estructura cristalina de SecA de *Thermus thermophilus* revela que ésta se dimeriza de manera paralela ⁸⁴³. Se ha indicado que la estructura homodimérica de SecA puede formar un anillo que insertado en la membrana produce un poro capaz de trascolar polipéptidos *per se* ⁸⁴⁴. La conformación dimérica de SecA parece ser esencial durante la trascolación de las proteínas, aunque no se ha descartado la participación de la forma monomérica en la reacción ^{845,846,847,848}. De manera contrastante, un estudio estructural de SecA parece mostrar que la forma activa de la enzima es el monómero y que la forma abierta de la enzima, observada en otros trabajos, no se debe a una dimerización de la misma, sino a cambios conformacionales provocados por la unión de la pre-proteína ⁸⁴⁹; incluso algunas investigaciones señalan que SecA funciona como monómero durante la trascolación de la proteína *in vivo* ⁸⁵⁰. Una explicación posible a la estequiometría variable de SecA, es que los diferentes grupos de investigación observan diferentes fases del mismo proceso de trascolación ^{851,852}. Se ha mostrado que la proteína SecA regula su propia traducción uniéndose a una secuencia **río arriba ('upstream')** del gen que la codifica; dicha autorregulación reside en los siete motivos conservados de la superfamilia de helicasas de DNA y RNA presentes en SecA, en particular de la familia DEAD; adicionalmente SecA presenta un sitio de unión a ATP, presente en numerosas enzimas de unión a nucleótidos. SecA compete

alostéricamente con el ribosoma por el sitio de unión al mRNA, mientras el dominio helicasa relaja las regiones de horquillas y bucles de su propio mRNA. Se sugiere entonces que SecA es una ATPasa bifuncional con actividades distintas e independientes: por un lado es un motor molecular para la trascolación de las pre-proteínas, por otro lado también posee baja actividad como helicasa de RNA no específica de la secuencia; se ha propuesto que la actividad helicasa es un remanente evolutivo de una proteína ancestral cuya actividad como helicasa era más robusta. De los siete motivos de la superfamilia de helicasas, aparentemente sólo el motivo IV está implicado directamente en la actividad de trascolación de SecA; por otro lado, los motivos III y V son requeridos tanto para la actividad helicasa como para la de motor molecular en la trascolación^{853,854,855,856,857,858}. Adicionalmente a su función como ATPasa y helicasa, se ha propuesto que SecA tiene actividad como chaperona de las pre-proteínas que serán secretadas⁸⁵⁹. Incluso se ha mostrado que SecA puede unirse *per se* al ribosoma para acarrear las proteínas recién sintetizadas al complejo de trascolación asociado a la membrana⁸⁶⁰.

- **SecY** es la proteína más larga del complejo de trascolación de las proteínas y constituye la estructura básica del canal a través del cual transita la gran mayoría de las proteínas secretadas (Para mayor información consulte 803-813,861,862). SecY es un polipéptido de 438 aa y 47.5 kDa, cuya expresión depende de la fase de crecimiento, ya que no se produce SecA (anteriormente descrita) ni SecY durante la fase estacionaria y a su vez SecY es esencial para el crecimiento⁸⁶³. SecY interacciona directamente *in vivo* con SecA y la pre-proteína que será secretada⁸⁶⁴; en el complejo de trascolación actual, tanto SecY, SecE y SecG son unidades importantes para la secreción de polipéptidos, aunque sólo SecY y SecE son esenciales para ello, y son

suficientes para activar la actividad ATPasa de SecA⁸⁶⁵. De todas las subunidades propias del complejo de trascolación (sin tomar en cuenta el motor molecular), únicamente SecY fue identificada como ancestral, por lo que en la presente tesis se denominará complejo de SecY al complejo Sec asociado a la membrana. A pesar de las diferencias estructurales de la pared citoplásmica, los segmentos de SecY comparten ~50% de identidad entre organismos G- y G+^{866,867}, así como 20% de identidad y 52% de similitud compartida con arqueobacterias⁸⁶⁸, inclusive el cloroplasto posee un homólogo de SecY⁸⁶⁹. SecY contiene 10 segmentos transmembranales, cinco partes expuestas al periplasma y seis regiones citoplasmáticas que incluyen los dominios N-terminal y C-terminal, debido a la topología es posible decir que SecY forma un canal para los polipéptidos que serán secretados⁸⁷⁰. Seis de los 10 segmentos transmembranales (STM) de SecY son funcionalmente los más importantes, de estos es necesario que la tasa de mutación del 3 y del 4 sea lo más baja posible⁸⁷¹ por lo que pudiera tratarse de los segmentos más antiguos. El complejo de SecY actual no sólo se encarga de la trascolación de las proteínas, sino también de su integración en la membrana, aunque ambas actividades estriban en dominios distintos de SecY y por lo tanto son procesos independientes⁸⁷², ya que mientras la trascolación radica en el dominio C-terminal, la delección de éste no parece afectar la integración de las proteínas a la membrana⁸⁷³. El análisis estructural del complejo Sec asociado a la membrana muestra que los segmentos transmembranales de SecY están dispuestos en forma de dos conos encontrados por el vértice, compuestos por aproximadamente la mitad de segmentos cada uno, formando una estructura similar a un reloj de arena, a través de cuyo centro o poro pasa el polipéptido secretado en forma extendida o de α -hélice cuando mucho. Una porción del STM2 de

SecY funge como tapón para el poro, obstruyendo el paso de una molécula cualquiera y manteniendo así el gradiente electroquímico; dicho tapón es desplazado cuando la pre-proteína entra en contacto con SecY, lo que también provoca desplazamiento del STM7, provocando que el poro del canal se expanda. El tapón del canal mantiene además la estructura del complejo de SecY y confiere selectividad a la reacción de trascolación y sensibilidad al efecto de ΔH^+ . Las paredes del poro del canal están conformadas principalmente por aa hidrófilos y algunos con carga, aunque en la constricción del centro, adicionalmente al tapón del STM2, hay un anillo de residuos hidrófobos que provocan fricción con el polipéptido lo que previene que éste se deslice cuando es liberado por SecA durante cada ciclo de inserción/desinserción al hidrolizarse el ATP; de modo que, tanto el tapón, como la constricción central y la inserción misma del polipéptido, bloquearían el paso de moléculas como iones y agua^{874,875,876,877,878,879}. Algunos trabajos mencionan que la estequiometría de SecY no es de 1:1 con SecA, sino de al menos 2:1, pues mientras una SecY funciona como anclaje a la SecA que lleva la pre-proteína, la otra recibe el polipéptido y permite su trascolación, se menciona que incluso pueden existir 4 complejos de SecY asociados^{880,881,882,883}, aunque la gran mayoría de las evidencias señalan que únicamente una subunidad SecY forma el canal de trascolación de proteínas^{884,874-879}. SecY se encuentra en el mismo operón que 10 proteínas ribosomales⁸⁸⁵, descubrimiento que toma relevancia recientemente, ya que en algunos organismos esta enzima interacciona directamente con el ribosoma a través algunas proteínas ribosomales y, sin mediación de SecA, se acopla al polipéptido naciente^{886,887}.

- o Se ha sugerido que SecA y SecY pueden constituir por si solas un canal para las pre-proteínas, aunque la coordinación entre SecY y SecA puede verse

afectada por la interacción con ligandos como el ATP, el polipéptido secretado o el ribosoma^{888,889,890,891,892,893,894,895}. Recientes resultados apuntan que no existen ciclos de inserción/desinserción de SecA en la membrana y lo que en realidad ocurre es que el ATP es hidrolizado por SecA para mantener abierto el poro de SecY, de este modo la secreción de la proteína es resultado de movimiento Browniano y la tasa de trascolación es incrementada por ΔH^+ ⁸⁹⁶.

▪ **Proteína de la familia OxaA.** Los miembros de la familia OxaA/Alb3/YidC son, en sentido el más amplio, chaperonas insertasas de membrana, *i.e.* catalizan el plegamiento e inserción de las proteínas que serán integradas en la membrana celular. Las proteínas de la familia OxaA/Alb3/YidC están presentes en los tres grandes dominios de la vida: YidC se encuentra en la membrana interna procariontes tanto G+ como G- con algunas diferencias, OxaA u Oxa1 se localiza en la membrana interna de la mitocondria y en la membrana de *Saccharomyces cerevisiae*, mientras que Alb3 se encuentra en la membrana del tilacoide de cloroplastos. Cualquiera de los miembros de la familia OxaA/Alb3/YidC puede actuar de manera dependiente o independiente del complejo de trascolación Sec (Para más información : 803,807,809,813, 897,898,899,900,901). Debido a que YidC es el nombre por el que se conoce a la chaperona insertasa bacteriana, se le mencionará de este modo a la proteína ancestral putativa identificada como OxaA. La proteína YidC co-purifica parcialmente con el complejo de SecY, por lo que se deduce que al menos temporalmente están asociados^{902,903}. Mediante liposomas reconstituidos que contienen ribosomas, una proteína hidrófoba (FtsQ) y el complejo de SecY-YidC, se muestra que YidC asiste en la inserción de las proteínas secretadas por el complejo Sec, acomodando los segmentos transmembranales de éstas en la fase lipídica de la membrana citoplasmática, pausando la secreción para estabilizar la proteína secretada

en posiciones específicas⁹⁰⁴; según algunos investigadores, la interacción de YidC con el complejo de SecY se da mediante el complejo no esencial SecDFyajC⁹⁰⁵. De acuerdo con otros trabajos, la interacción entre YidC puede ser directamente con SecY e incluso con el ribosoma⁹⁰⁶. Se ha demostrado que la enzima YidC juega un papel fundamental en el correcto plegamiento de proteínas politópicas de membrana (como LacY)⁹⁰⁷. Por todo lo anteriormente mencionado, los miembros de la familia OxaA/Alb3/YidC, se describen como chaperonas insertasas esenciales para el organismo correspondiente⁹⁰⁸. La depleción de YidC interfiere con la inserción de proteínas dependiente de Sec⁹⁰⁹, aunque con menor efecto que durante el proceso de inserción independiente del complejo de SecY, pues en este caso YidC se requiere de manera estricta, como en el caso de la inserción de las subunidades F₀F₁ de la ATP sintasa e incluso de algunas proteínas del profago M13 que se insertan en la membrana^{910,911,912}. Se ha mostrado que YidC es de hecho suficiente para las proteínas que se integran de modo independiente del complejo de SecY, no sólo promoviendo su correcto plegamiento, sino también su inserción en la membrana, ya que provee de una superficie anfífila a través de la cual se transfieren las regiones polares de las proteínas⁹¹³. Tanto la topología, como el proceso de inserción independiente de YidC, son características conservadas por homólogo mitocondrial OxaA, además que ambas enzimas son esenciales para el crecimiento celular del organismo respectivo⁹¹⁴. YidC es una proteína politópica de 61 kDa, ubicada en la membrana interna e involucrada en la inserción y adecuado plegamiento de las regiones hidrófobas de proteínas membranales; YidC atraviesa seis veces la membrana bacteriana, mientras que en eucariontes sólo existen 5 dominios, mismos que corresponden a las regiones conservadas en la

familia OxaA/Alb3/YidC (TMS2 al TMS6); en particular los residuos 322-342 de YidC parecen ser esenciales para la viabilidad de la célula y la actividad insertasa de la enzima, tanto en el proceso dependiente como en el independiente de Sec⁹¹⁵. De acuerdo con algunos trabajos, la inserción de proteínas mediante YidC procede co-traduccionalmente, tanto en el proceso dependiente como en el independiente de Sec. Para ambos tipos de proceso, YidC puede presentarse como monómeros o formar dímeros; se propone, de acuerdo al análisis de la proyección de la imagen cristalográfica de la enzima, que un monómero puede yuxtaponerse al complejo de SecY, formando una salida lateral para las proteínas que se integren a la membrana, o bien acoplarse directamente al ribosoma en forma de dímeros, constituyendo en ambos casos una especie de canal hidrófobo para la proteína secretada⁹¹⁶. YidC es esencial para la viabilidad de bacterias, mitocondrias y cloroplastos, aunque no parece haber ningún homólogo en la membrana del retículo endoplásmico⁹¹⁷. Con base en análisis filogenéticos, se distinguen seis subgrupos en la familia OxaA/Alb3/YidC, los cuales probablemente surgieron a partir de tres duplicaciones genéticas independientes, seguidas de especializaciones funcionales en cada grupo de organismos, dichas especializaciones consistieron en la adición de regiones o dominios a la estructura conservada de 5 dominios transmembranales⁹¹⁸.

o Un análisis filogenético y comparativo acerca de la maquinaria de trascolación, menciona que su evolución no ha sido continua y más bien es resultado de una serie de eventos discretos, en que los componentes adicionales a la maquinaria canónica, agregados independientemente en cada división bacteriana, contribuyen sinérgicamente a la capacidad trascolación, formando una ruta cada vez más precisa para la proteína trascolada. De acuerdo con el trabajo mencionado, la

maquinaria canónica de trascolación incluye el complejo SecYEG, el motor molecular SecA, la chaperona insertasa YidC y el complejo SecDFYajC; dicha maquinaria canónica se encuentra en las bacterias G+, mientras que las G- poseen un sistema más complejo⁹¹⁹ (Para mayor información consulte 920,921).

o De todas las proteínas de la maquinaria canónica de trascolación, únicamente SecA, SecY y YidC contienen regiones codificadas por tripletes RNY; estas 3 proteínas pueden interaccionar entre sí, llevando a cabo el proceso de trascolación de polipéptidos aunque de manera ineficiente; dichas proteínas también interaccionan con moléculas nucleotídicas, lípidos y RNA, que pueden ser sintetizadas prebióticamente. Por todo lo anterior, podemos suponer que la trascolación de proteínas ya se llevaba a cabo en organismos ribonucleoproteicos, o al menos los dominios polipeptídicos que darían origen a la maquinaria de trascolación actual son muy antiguos; dicho de otro modo, a pesar que la maquinaria canónica de trascolación de proteínas posiblemente ya formaba parte del LUCA, únicamente los primordios de SecA, SecY y YidC pueden ser trazados hasta los albores del Mundo de las RNPs.

o **ANCLAJE DE PROTEÍNAS EXTRAMEMBRANALES.**

o Las proteínas de superficie celular sirven como interfaz entre el organismo que las posee y su ambiente, existen varios tipos de éstas proteínas y cada uno es reconocido y anclado por enzimas específicas, que son miembros de la familia denominada "Anclajes de superficie de pared celular"⁸⁰³. Dos enzimas de la familia mencionada fueron identificadas como ancestrales, una de estas sólo es mencionada como proteína de esa familia, la otra es una sortasa.

▪ **Proteína de la familia "Anclajes de superficie de pared celular".**

Aparentemente la característica general de las enzimas de esta familia es reconocer motivos específicos en las proteínas que serán de superficie y

catalizar su anclaje mediante mecanismos particulares. Dado que la familia es muy diversa, podría tratarse cualquiera de los miembros⁹²², de manera que sólo quedará mencionada sin especular de cual enzima podría tratarse.

- **Sortasa.** Muchas de las proteínas de la superficie celular de los microorganismos están covalentemente asociadas a ésta mediante transpeptidasas denominadas sortasas; existen diferentes tipos de sortasas que anclan las proteínas en distintas posiciones, además algunas sortasas y proteínas tipo sortasa tienen más especificidad por su sustrato que otras, más de una clase de los distintos tipos de sortasas y proteínas semejantes se puede encontrar en un mismo organismo; a pesar de poseer arquitecturas celulares diferentes, existen sortasas y proteínas relacionadas en todas las bacterias G+, en algunas G- e incluso en arqueas^{923,924}. Las proteínas que serán ancladas covalentemente a la superficie celular contienen un motivo conservado de aproximadamente 5 aa, hidrófobos la mayoría, en el extremo C-terminal; este dominio C-terminal es reconocido por la sortasa (señal para la sortasa, SS), que cataliza de modo consecutivo la hidrólisis específica de la SS y su enlace con un componente de la superficie celular⁹²⁵. Con base en diversos análisis de la secuencia primaria, topología en la membrana, posición genómica y especificidad por los motivos de las SS, se han descrito 5 subfamilias de sortasas (Srt) en bacterias G+: A, B, 3, 4 y 5; así como una subfamilia distinta en organismos G-⁹²⁶. Con base únicamente en el alineamiento de 61 secuencias de sortasas de organismos G+, se han describen 4 clases: A, B, C, y D, cada una corresponde a las subfamilias descritas previamente, así la clase A corresponde a SrtA, la clase B corresponde a SrtB, la clase C es la subfamilia Srt3 y la clase D incluye Srt4 y Srt5⁹²⁷. Los miembros de la subfamilia SrtA se encuentran distribuidos en un amplio rango de géneros

bacterianos, aunque principalmente en bacterias G+ con bajo %GC; los substratos de esta subfamilia son numerosos y diversos y los genes que los codifican no están agrupados junto a los genes que codifican las proteínas; las enzimas del tipo SrtA contienen sólo una región de aa hidrófobos en la región N-terminal, por lo que su topología corresponde al tipo I; las SrtA son específicas para la SS LP[K/**E/N/Q/A**][**T/A**]**G**[E/**D/S**]^{926,927}. La clase B es el grupo más pequeño, presente únicamente en algunos cocos y bacilos con bajo %GC; la mayoría de sus miembros de la familia SrtB están involucrados en la adquisición de heme-Fe y los genes que los codifican están adyacentes a sus substratos, mismos que contienen la inusual SS **NP**[Q/K][**T/S**][**N/G/S**][**D/A**]^{926,927}. La subfamilia Srt3 o clase C es el grupo más extenso y heterogéneo, está presente tanto en bacterias de bajo %GC como de alto %GC; aunque las enzimas de esta clase son más especializadas en sus substratos, pueden reconocer algunos de la clase A; la expresión de los genes que codifican las enzimas de clase C y sus respectivos substratos está co-regulada, pues éstos se encuentran siempre adyacentes en el genoma; las proteínas tipo Srt3 son las únicas que contienen aa hidrófobos en ambos terminales (C- y N-), por lo que su topología corresponde al tipo II; se ha propuesto que las sortasas de clase C contienen una región propensa a transferencia lateral que puede requerirse en la adaptación a diferentes nichos; la mayoría de los substratos de las Srt3 están involucrados en la adhesión celular y contienen la SS [L/**I**]P[K/**S**/E/L/**A/N**]**TGG**^{926,927}. La clase D, que agrupa Sr54 y Srt5, incluye algunas sortasas de bajo %GC y alto %GC que pueden dividirse en tres subgrupos que reflejan la filogenia (bacilos, clostridia y actinomicetales)⁹²⁷. Las enzimas de la subfamilia Srt4 se encuentran únicamente en bacilos, los genes que las codifican se encuentran

usualmente adyacentes a los genes de sus substratos, que en su mayoría son otras enzimas que contienen la SS LP[**N/D/E/K**]**TA**[**T/S**]^{926,927}. La subfamilia Srt5 es característica de bacterias G+ con alto %GC; los genes que codifican estas enzimas, al igual que en el caso de SrtA, no se encuentran adyacentes a los genes de sus substratos, que contienen la SS **LA**[**E/A/S/H**]**TG**[**A/S/G/V**]^{926,927}. Las sortasas de organismos G- están muy relacionadas entre sí, su función es mucho más especializada que las sortasas de organismos G+ y los substratos que reconocen este tipo de sortasas contienen la SS [L/M/Y]P[Q/K]**T**[**A/S**][**T/L**]⁹²⁶. SrtA y en general las sortasas de su clase se utilizan como modelo experimental, a partir de lo cual se deduce que la estructura, mecanismo de catálisis y evolución es similar en todos los tipos de sortasas. Las sortasas llevan a cabo dos reacciones consecutivas en lo que se considera una transpeptidación, entre el substrato y un componente de la superficie celular, sin utilizar ATP; primero cataliza la hidrólisis del enlace peptídico entre dos de los aminoácidos de la SS y posteriormente cataliza la de un enlace tipo amida entre el carboxilo que queda libre en el substrato y el grupo amino de un componente denominado lípido II, que es un intermediario en la formación de la pared celular **bacteriana (ver más adelante en 'Estructuras membranales y extramembranales')**; de este modo las proteínas quedan literalmente incorporadas a la superficie celular del organismo^{928,929}. La caracterización de los substratos que reconocen las proteínas SrtA, indica que han evolucionado de manera modular, por un lado surge el motivo de la SS, por otro lado la región que reconoce el ligando de esas proteínas de superficie y finalmente el cuerpo de la proteína que ha evolucionado a través de la adquisición de diferentes dominios, algunos de los cuales se han expandido por duplicación y recombinación homóloga⁹³⁰. La enzima

SrtA posee una estructura de barril- β flanqueada por dos hélices- α cortas; su sitio catalítico está formado por C¹⁸⁴ e H¹²⁰, el residuo R¹⁸⁷ también está conservado y es necesario para la catálisis eficiente de la SrtA^{931,932}; la catálisis es estimulada por el ion Ca²⁺⁹³³, pues facilita el reconocimiento adaptativo entre la sortasa y la SS del sustrato⁹³⁴. Una comparación estructural entre los sitios activos de las enzimas SrtA y SrtB muestra que posiblemente el mecanismo de anclaje de las proteínas de superficie sea universal; la topología de ambas sortasas también muestra una idéntica disposición de la estructura de barril- β y la disposición espacial de los aa catalíticos también está conservada; además se indica que el sitio de unión a la SS en ambos tipos de sortasa está compuesto y rodeado de aa hidrófobos **A**⁹², **A**¹⁰⁴, P¹⁶³, **I**¹⁸², **I**¹⁹⁹ y W¹⁹⁴ además de Y¹⁸¹, Y¹²⁸, **S**²²¹, E²²⁴ y **N**⁹² que facilitan el secuestro del sustrato⁹³⁵. Se ha observado que la SrtA puede existir como dímero o monómero, los residuos K₁₆₂, K₁₃₇, Y₁₄₃ y K₁₅₂ son esenciales para la dimerización⁹³⁶. Se menciona que alineamiento de las secuencias de las 61 sortasas antes mencionadas, revela la presencia de tres dominios conservados: D1, D2 y D3. Los residuos de aa, que se piensa están involucrados en la formación del sitio catalítico, se encuentran conservados en los dominios D2 y D3 en todas las sortasas analizadas, además cada clase de sortasas posee un patrón específico de aa conservados⁹²⁷. En términos generales las sortasas catalizan la unión de grupos amino, por lo tanto se ha propuesto que dichas enzimas pueden catalizar la unión de cualquier par de moléculas que contengan los grupos químicos antes mencionados, incluyendo lípidos y carbohidratos⁹²⁵; con base en lo anterior el escenario más plausible es que la actividad catalítica de las sortasas primitivas era mucho más general que en la actualidad, pues habrían catalizado la 'adhesión' a diversas moléculas de origen

prebiótico mientras éstas fuesen de característica hidrófoba con grupos amino en su estructura. De la sortasa identificada como ancestral, probablemente la región codificada por tripletes RNY corresponde al bolsillo hidrófobo de unión al substrato (también hidrófobo), de hecho es posible que las sortasas y sus substratos hayan coevolucionado, siguiendo un **modelo de 'llave-cerradura'**. Por todo lo descrito, parecería que las subfamilias SrtA y Srt3 son las más antiguas, tanto por su distribución y diversidad, como por el tipo de aa que componen la SS de sus substratos; con base en el presente análisis sólo se puede mostrar que el origen de las sortasas de clase C se remonta al inicio del Mundo de las RNPs, ya que es el único tipo de sortasa que posee *SagA909*, el siguiente paso es analizar el genoma RNY de una amplia gama de procariontes, para encontrar efectivamente cuales de todas las subfamilias de sortasas que existen actualmente contienen regiones codificadas por tripletes RNY.

o **PROTEÍNA CON REPETICIONES RICAS EN LEUCINA.**

o Las proteínas que contienen repeticiones ricas en leucina (PRRL) representan un grupo diverso de moléculas con distintas funciones y localizaciones celulares, con numerosos representantes de PRRL en los 3 dominios de la vida y en virus; esas repeticiones ricas en leucina (RRL) son motivos de secuencia cortos, encontrados en las estructuras primarias de numerosas proteínas que son parte de diversos procesos de reconocimiento molecular, como transducción de señales, adhesión celular, desarrollo celular, reparación de DNA y procesamiento de RNA; en tan amplia gama de funciones biológicas vitales, el dominio RRL funge como el soporte estructural de las interacciones proteína-proteína e incluso es posible que la estructura repetitiva del motivo RRL pueda generar nuevas variantes si se requiere^{937,938,939,940,941,942,943,944}.

o Las PRRL se presentan usualmente en bloque y el número de motivos RRL varía desde uno a más de 40. La longitud más común de un motivo RRL es de 20 a 30 residuos⁹³⁷⁻⁹⁴⁴. La secuencia consenso para las PRRL contiene Leu o algún otro aa alifático en las posiciones 2, 5, 7, 12, 16, 21 y 24, finalmente **Asn/Cys/Thr** en la posición 10, la mayoría contiene **Asn** en este sitio; adicionalmente, las secuencias consenso derivadas de los motivos RRL, contienen residuos conservados en otras posiciones, la mayoría son aa alifáticos como **Gly** y aromáticos como Pro, aunque en raras ocasiones de algún otro tipo; los residuos de consenso hidrófobo en la parte carboxilo de las repeticiones, generalmente están espaciados por 3, 4 o 7 residuos. Las regiones terminales N- y C- son cortas y muestran una débil homología con el motivo interno repetido. La estructura esencial de las PRRL está constituida por las cadenas laterales de Leu y otros residuos alifáticos en las posiciones 2, 5, 7, 12, 17, 20, 23 y 24, generalmente Cys en la posición 17 y de **Asn/Cys** en la posición 10⁹³⁸.

o Los intrones de las PRRL usualmente se localizan en posiciones similares a las repeticiones, con mayor frecuencia en las posiciones 1 y 6, con menor frecuencia entre las posiciones 11 y 18 y muy rara vez en otros lugares. El análisis de secuencia de las PRRL sugiere que el motivo RRL evolucionó por la duplicación de un sólo exón y el rearrreglo de los resultantes; con base en ello, se propusieron dos posibles escenarios evolutivos: 1) un módulo ancestral RRL evoluciona primero y se duplica dando lugar a miembros individuales de la superfamilia que más tarde desarrollaron sus propias secuencias consenso únicas o 2) las duplicaciones ocurrieron de manera separada en cada familia de proteínas y con base en sus constreñimientos estructurales dieron origen a secuencias consenso similares pero únicas, cabe resaltar que ambos esquemas no son mutuamente excluyentes⁹³⁸.

o Las predicciones de estructura secundaria de las PRRL no mostraban patrones repetitivos claros y fue hasta la cristalización de la primera PRRL descubierta, el inhibidor de ribonucleasa porcina (RI), que se encontró que cada uno de los motivos

RRL corresponden a unidades estructurales $\alpha\beta$, éstas consisten de una hoja- β corta y una hélice- α aproximadamente paralelas entre sí. Todas las repeticiones, incluyendo las regiones terminales, adoptan estructuras similares y todas las repeticiones estructurales consisten de 28 o 29 residuos, excepto las regiones terminales que contienen 25 residuos. Se ha mostrado que repeticiones cortas similares pueden adoptar también estructuras $\alpha\beta$ semejantes⁹³⁸.

- o Las unidades estructurales están dispuestas de manera que todas hojas- β y hélices- α son paralelas a un eje común, resultando en una estructura no globular poco convencional con forma de herradura, en que las hojas- β componen la cara externa y las hélices- α la cara interna de las semicircunferencias⁹³⁸. Esa composición es entrópicamente la menos desfavorable y permite modificar la especificidad de la proteína mediante la sustitución de aa en la superficie, por lo que lejos de ser una plataforma estable el dominio RRL es una armazón flexible que puede ajustarse a los requerimientos de una interacción particular⁹³⁹.

- o La estructura no globular en forma de herradura es responsable de la afinidad por proteínas globulares, en particular la superficie de hojas- β expuestas al solvente, en que hojas- β y hélices- α de otras proteínas se aglutinan; esa misma conformación estructural ofrece también una extensa área de unión y formación de numerosos contactos favorables para una amplia variedad de pequeños ligandos proteicos no relacionados entre sí^{938,939}.

- o La parte variable de las PRRL adopta probablemente una conformación en la cual: los residuos no-polares conservados están dentro del eje hidrófobo, los residuos polares conservados están involucrados en la formación de puentes de hidrógeno dentro de la estructura y los residuos no conservados están expuestos en la superficie fuera de la estructura⁹⁴².

o El análisis estructural de las PRRL arroja que se trata de una superfamilia que se subdivide en al menos siete familias, caracterizadas tanto por diferentes longitudes y secuencias consenso de las repeticiones como por distintas orientaciones, interacciones moleculares y estructuras tridimensionales de las regiones variables; por otro lado las repeticiones de diferentes subfamilias retienen un plegamiento súper-helicoidal similar, pero difieren en las distribuciones tridimensionales de las repeticiones individuales; todo ello provoca que los distintos diseños estructurales no sean compatibles entre sí y por tanto no puedan acoplarse dentro de un mismo tipo de proteína. La transformación de una subfamilia en otra, requeriría numerosas mutaciones simultáneas, lo que sugiere que el escenario evolutivo más probable es que las distintas subfamilias hayan emergido independientemente, *i.e.* que no compartan un ancestro común⁹⁴⁰⁻⁹⁴⁴.

o La secuencia consenso de la región de hojas- β es una característica de la superfamilia PRRL; dicha secuencia consenso comprende únicamente once residuos que son LxxLxLxx[**N/C**]xL, donde x puede representar cualquier aa y la posición alifática L puede ser ocupada por **V/I/F**⁹⁴²⁻⁹⁴⁴. De la secuencia consenso la región mejor conservada es el motivo LxxLxL^{940,941}, que contiene el adecuado espaciamiento entre residuos polares e hidrófobos y es suficiente para generar la estructura de herradura característica⁹⁴⁰⁻⁹⁴⁴. La conformación estructural de la parte variable puede ser desde una hélice tipo II de poliprolinas (pP) hasta una hélice- α ⁹⁴⁰⁻⁹⁴⁴.

o Las PRRL bacterianas poseen repeticiones inusuales, cuya secuencia consenso es PxxLxxLx**V**xx**N**xLxxLP[**D/E**]L y se considera la más corta con tan sólo 20 aa sin algún otro dominio adicional, la región variable es una hélice de pP que incluye un **D** conservado que está expuesto al solvente y probablemente es importante en la formación de interacciones intermoleculares⁹⁴⁰⁻⁹⁴⁴.

o A diferencia de los cambios conformacionales observados en la mayoría de las proteínas, en las PRRL no hay articulaciones evidentes, sino que pequeños cambios se acumulan a lo largo de la estructura, lo que resulta en variaciones en la curvatura y el giro de la herradura⁹³⁹. El modelado molecular sugiere que la curvatura estructural de las PRRL es similar en la mayoría de las familias y en la mayoría de los casos posee un giro hacia la derecha⁹⁴⁴. Las PRRL pertenecen a una clase más general de estructuras proteicas con arreglo de solenoide; esas estructuras solenoidales han sido vistas como superhélices continuas, las cuales han sido caracterizadas cualitativamente y cuantitativamente con respecto a su curvatura, tipo y magnitud de giro⁹⁴³.

o Con base en el análisis de secuencia y estructura de las PRRL, se ha propuesto una rica y compleja historia evolutiva que pudo haber surgido de más de una raíz, de hecho hay buenas razones para creer que escenarios divergentes y convergentes han trabajado en conjunto, llegando hasta la superfamilia de PRRL que existe actualmente⁹³⁷⁻⁹⁴⁴.

o Parecería paradójico que una proteína cuyo aa característico (L) no es codificado por algún triplete RNY sea identificada como ancestral; sin embargo es necesario recordar que 1) Leu puede ser substituida por otros aa alifáticos que si son codificados por tripletes ancestrales y 2) además de los aa alifáticos, las PRRL contienen otras posiciones consenso que pueden corresponder a la región identificada como ancestral; dado que la función de los aa conservados en este tipo de proteínas es únicamente estructural, mientras se mantengan las propiedades químicas en las distintas posiciones, la proteína mantiene su configuración. Es plausible entonces deducir que alguno de los patrones en una RRL pudo existir en los albores del Mundo de las RNPs, puede que incluso haya tenido la función que actualmente desempeña como plataforma estructural para las uniones proteína-

proteína (o péptido-péptido). Como podemos observar, las PRRL pudieran ser el modelo evolutivo que mostraría que la forma sí puede dar origen a la función.

- o **PEPTIDASA DE LA FAMILIA S8 (SUBTILISINA).**

- o Las peptidasas intracelulares juegan un papel crítico en la regulación del metabolismo al modificar algunas enzimas y activarlas, inhibirlas y/o dirigir las a determinados sitios dentro de la célula⁹⁴⁵.

- o Sin importar la complejidad de algún organismo, las peptidasas son esenciales en cada fase de la vida de las células individuales, dado que todas las moléculas peptídicas producidas deben ser proteolíticamente procesadas y eventualmente degradadas. Los genes que codifican para peptidasas ocurren a una frecuencia relativamente alta, con un promedio de 3% del genoma⁹⁴⁶.

- o Las peptidasas han sido clasificadas por familias, que a su vez pueden ser agrupadas en clanes, algunos de estos clanes poseen peptidasas de más de 60 líneas evolutivas diferentes *i.e.* con orígenes evolutivos separados, aunque existen algunos claros ejemplos de convergencia.

- o Las propiedades de las peptidasas de cada familia se han considerado desde dos puntos de vista: la diversidad de su actividad catalítica y el número de convergencias evolutivas⁹⁴⁷. El tipo de péptidas se refiere al grupo químico responsable de la catálisis: **Ser**, Cys, **Asp** o metal. La familia describe el grupo de enzimas en que cada miembro muestra relación evolutiva con al menos algún otro miembro, ya sea a lo largo de toda la secuencia o al menos un fragmento de la secuencia responsable de la catálisis. Un clan comprende grupos de familias cuya relación evolutiva esté indicada por el orden lineal de los residuos catalíticos y por la estructura terciaria⁹⁴⁸.

- o Las enzimas proteolíticas que dependen de serina para la actividad catalítica (Ser-peptidasas) están ampliamente distribuidas y son numerosas. Las Ser-

peptidasas se encuentran en virus, bacterias, arqueas y procariontes e incluyen exopeptidasas, endopeptidasas, oligopeptidasas y ω -peptidasas. Con base en la estructura, las más de 20 familias se pueden agrupar en seis clanes que parecen carecer totalmente de alguna relación evolutiva. Uno de esos clanes es el SB, que contiene la familia S8, cuya enzima representativa es la subtilisina y por lo que también se le conoce como familia de subtilasas⁹⁴⁹.

- o La familia S8 es la segunda más grande de las Ser-peptidasas y sus miembros están dispersos en los organismos de los tres grandes dominios de la vida y en virus; esta familia contiene tanto exopeptidasas como endopeptidasas. La familia S8 contiene varias enzimas con extensiones terminales N- o C- que no muestran relación con otras proteínas. Las proteínas de la familia S8 contienen la triada catalítica **SHD**, donde **Ser** es aa nucleófilo, His aa básico y **Asp** aa electrófilo; la mayoría de estas Ser-peptidasas contiene además **Gly** conservadas en la vecindad del residuo catalítico **S**^{948,949}.

- o Las subtilasas, de la gran mayoría de los organismos, son enzimas extracelulares requeridas para la defensa o el crecimiento en sustratos proteicos. La familia de subtilasas puede ser subdividida en seis subfamilias: A) de subtilisina, que incluye las subtilisinas verdaderas, proteasas de alta alcalinidad y proteasas intracelulares y sólo se encuentra en microorganismos; B) de termitasa, que se encuentra en varios microorganismos incluyendo termófilos y alófilos; C) de proteinasa K, que se encuentra en bacterias G- y hongos; D) peptidasas de lantabióticos, que conforman una familia muy especializada de enzimas; E) de kexina, cuyos miembros se encuentran en eucariontes y F) de pirrolisina, que es el grupo más heterogéneo y con poca conservación de secuencia. El dendrograma de esta familia S8 se muestra como (((A)(B))((E)(F)))(C)(D))⁹⁵⁰.

- o A partir de un análisis estructural y de múltiples alineamientos de secuencias de las subtilasas, los iones Ca^{2+} son esenciales para la estabilidad y actividad de

estas enzimas, en las que los únicos aa absolutamente conservados son la tríada **S**²²¹**H**⁶⁴**D**³² y la **G**²¹⁹ embebidos en un andamiaje estructural $\alpha\beta$. En las subtilasas el sitio de unión al substrato puede ser descrito como un canal de superficie con cuatro compartimentos (**S2'**, **S1**, **S2** y **S4**), capaz de acomodar seis aa del substrato correspondiente (**P2'**, **P1'**, **P1**, **P2**, **P3** y **P4**); los compartimentos **S1** y **S4** son grandes e hidrófobos, lo que explica la amplia especificidad de las subtilasas⁹⁵⁰.

- o La familia S8 es una de las 16 familias de peptidasas esenciales, estas son capaces de cubrir todos los requerimientos proteolíticos de digestión y procesamiento de péptidos y proteínas de la vida celular y viral, por lo que se le ha denominado el degradoma basal, cuya importancia radica en el hecho que se ha conservado a pesar de millones de años de evolución y prueba de ello es el hecho que la familia S8 haya sido descubierta en todos los organismos, lo que sugiere que es un grupo ubicuo que ha existido por largo tiempo y dada la extensa distribución de estas familias es improbable un escenario de HGT^{951,950}.

- o No se encontró evidencia del proteasoma como complejo ancestral pero muchas de las familias modernas de peptidasas son muy antiguas, por lo que es posible deducir que la degradación de proteínas mediante peptidasas era la única vía para regular diversas actividades metabólicas y mecanismos estructurales primitivos; se ha sugerido que la familia S8 existía en el LUCA, sin embargo posiblemente su primordio es mucho más antiguo. La región que inicialmente pudiera ser la más antigua correspondería al sitio catalítico, sin embargo en otras enzimas la región ancestral une moléculas de síntesis prebiótica; dado que el substrato de la subtilasa es un péptido y pequeños péptidos éstos ya existían en los albores del Mundo de las RNPs, es probable que alguno de los sitios que lo une está codificado por tripletes RNY.

T11. Características plausibles de la probable región codificada por tripletes RNY en las proteínas identificadas del procesamiento de proteínas.

Conjunto	Proteína (nombre traducido y adaptado)	Características plausibles de la probable región codificada por tripletes RNY.
Complejo de secreción de proteínas	Transcolasa de preproteínas, subunidad SecA	Unión de ATP, o de RNA; o interacción con un proto-ribosoma, o con péptidos hidrófobos
	Transcolasa de preproteínas, subunidad SecY	Diversos aa que interaccionan con péptidos hidrófobos, o con la membrana lipídica
	Proteína de membrana oxaA, putativa	Interacción con péptidos hidrófobos, o con SecA y SecY, o con la membrana lipídica
Anclaje de proteínas extramembranales	Proteína de la familia 'anclajes de pared celular'	No precisada
	Proteína de la familia sortasa, Srt	Unión de péptidos hidrófobos, o quizá interacción con proteínas ribosomales
	Proteína con repeticiones ricas en leucina	Plataforma estructural para interacciones péptido-péptido
	Peptidasa de la familia S8 (subtilisina)	Unión de péptidos pequeños

6.3.1.5 Estructuras membranales y extramembranales.

Una de las primeras distinciones entre microorganismos se basa en la tinción de sus estructuras extramembranales. Los microorganismos G- están rodeados por una delgada pared de peptidoglicano rodeada a su vez por una pared externa de lipopolisacárido, que bloquea el paso de la tinción Gram quedando los organismos teñidos en color rosa. Los microorganismos G+ carecen de la pared externa, por lo que se tiñen en violeta, sin embargo la pared de peptidoglicano está constituida por varias capas formando una protección mucho más gruesa; a través del peptidoglicano de los organismos G+, sobresalen estructuras poliméricas aniónicas denominadas ácidos teicóicos, que bien pueden estar anclados a la membrana celular (LTA) o a la pared de peptidoglicano (WTA)⁹⁵². Los microorganismos deben sobrevivir en ambientes impredecibles y usualmente hostiles en los que se encuentran y las envolturas extramembranales les proveen de la protección necesaria para enfrentar esas condiciones, estas envolturas también segregan diferentes tipos de moléculas e incluso funcionan como superficies en las cuales diversas reacciones ocurren⁹⁵³. Embebidas en la membrana y en la pared celular existen proteínas y lipoproteínas encargadas tanto de adhesión celular, respuesta a estímulos ambientales, e interacción con otros organismos^{953,954}.

Algunas de las proteínas involucradas en la síntesis de componentes extramembranales y algunas proteínas ubicadas en la membrana o en la pared, contienen regiones codificadas por tripletes RNY (en T12 y T13), entre ellas encontramos enzimas que participan en la síntesis de peptidoglicano y ácido lipoteicóico, polisacáridos capsulares, así como lipoproteínas, las proteínas putativas de membrana que se encontraron, también se han incluido en esta categoría (ver metodología).

Order	Sc.	E-v	ID	%	PROTEIN / FAMILY / DOMAIN
1	28	2.00E-05	28/28	100%	membrane protein, putative (22.01 %)
3	27	6.00E-05	27/27	100%	lipoprotein, putative (68.64 %)
4	27	6.00E-05	27/27	100%	UDP-N-acetylglucosamine 1-carboxyvinyltransferase (43.45 %)
1	27	6.00E-05	27/27	100%	streptococcal histidine triad family protein (89.71 %)
5	26	2.00E-04	26/26	100%	glycosyl transferase family protein (19.74 %)
6	26	2.00E-04	26/26	100%	cna B-type domain protein (49.38 %)
7	25	0.001	25/25	100%	cell wall surface anchor family protein (49.54 %)
8	25	0.001	25/25	100%	membrane protein, putative (42.31 %)
9	25	0.001	25/25	100%	dltB protein (76.60 %)
2	24	0.004	24/24	100%	metallo-beta-lactamase family protein (79.51 %)
10	24	0.004	35/38	92%	membrane protein, putative (61.02 %)
11	24	0.004	24/24	100%	UDP-N-acetylmuramoyl-tripeptide--D-alanyl-D-alanine ligase (42.79 %)
12	24	0.004	24/24	100%	fibrinogen-binding protein (91.64 %)
14	24	0.004	24/24	100%	cell wall surface anchor family protein (93.17 %)
15	24	0.004	24/24	100%	membrane protein, putative (19.28 %)
4	23	0.015	23/23	100%	streptococcal histidine triad family protein (30.47 %)
17	23	0.015	23/23	100%	CHAP domain protein (27.90 %)
18	23	0.015	23/23	100%	D-alanyl carrier protein (2.93 %) dltD protein (1.74 %)
19	22	0.06	28/30	93%	glucosamine--fructose-6-phosphate aminotransferase, isomerizing (19.57 %)
20	22	0.06	22/22	100%	UDP-N-acetylmuramoylalanyl-D-glutamate--2,6-diaminopimelate ligase (41.20 %)

T12. Proteínas que contienen regiones codificadas por tripletes RNY, involucradas en la formación de estructuras membranales y extramembranales; nomenclatura y código de colores igual que T1 y T2.

o **PROTEÍNAS DE MEMBRANA.**

o Cuatro de las proteínas de membrana que contienen regiones codificadas por tripletes RNY están anotadas como proteínas de membrana putativas; dado que las proteínas membranales pueden fungir como receptores, canales, transportadores o traslocasas y al momento no se les ha asignado una función específica, decidí agruparlas dentro de la presente categoría tomando en cuenta únicamente su localización.

o **LIPOPROTEÍNA.**

o Las lipoproteínas son todas aquellas proteínas ancladas a la membrana mediante una región lipídica, cuya región proteica queda dentro de la pared celular o se proyecta hacia fuera de ella con el fin de generar interacciones, positivas o negativas, entre el organismo que las posee y el medio externo.

o Las lipoproteínas contienen inicialmente una secuencia señal N-terminal que les permite ser traslocadas por la maquinaria denominada Sec, cuyos componentes principales también contienen módulos o dominios ancestrales (ver sección

'Proteínas, estructura y modificaciones'), la energía es proveída por la fuerza protón-motriz (ΔH^+); posteriormente la secuencia señal es removida por una señalopeptidasa; finalmente se añade una cadena lipídica, unida covalentemente mediante un enlace acilo, dicha cadena lipídica ancla la lipoproteína a la cara externa de la membrana celular^{953,955}.

- o Las lipoproteínas cumplen diversas funciones, algunas sirven para mediar la adhesión del microorganismo a diferentes substratos; para importación de nutrientes, para secreción de péptidos; algunas lipoproteínas están involucradas en la señalización o en la conjugación bacteriana; otras más están involucradas en las interacciones estructurales con el peptidoglicano; algunas lipoproteínas son liberadas de la célula dependiendo de la fase de crecimiento del microorganismo. La función que desempeñe cada lipoproteína en particular determinará si ésta es esencial o dispensable⁹⁵⁶.

- o La mayoría de las lipoproteínas sólo han sido identificadas a partir de predicciones del análisis de la secuencia primaria, haciendo coincidir el sitio consenso que reconoce la peptidasa⁴⁵⁴. La única lipoproteína que se ha identificado mediante la metodología descrita en la presente tesis se anota como putativa, lo que no impide señalar que contiene una región codificada por tripletes provenientes del Mundo del RNA, *i.e.* que su primordio existía ya en el Mundo de las RNPs.

- o **Mur (UDP-N-) & GlmS.**

- o El peptidoglicano o mureína es un heteropolímero polianiónico que forma una red que envuelve el microorganismo, ésta red mantiene la forma celular a pesar de los cambios en presión osmótica⁹⁵⁷. El peptidoglicano está compuesto de unidades repetidas del disacárido N-acetilglucosamina—ácidoNacetilmurámico (GlcNAc-MurNAc), entrecruzado directamente o mediante cadenas de pentapéptidos; el

grupo carboxilo de cada MurNAc es substituido además por una cadena peptídica, que generalmente contiene alanina⁹⁵⁸.

o La biosíntesis del peptidoglicano es un proceso complejo que involucra reacciones en el citoplasma (síntesis de precursores nucleotídicos), del lado interno de la membrana celular (síntesis de intermediarios ligados a lípidos) y del lado externo de la membrana (reacciones de polimerización)⁹⁵⁹. Debido a que el peptidoglicano es una estructura presente tanto en G+ como de G-, la mayoría de las enzimas que lo sintetizan están presentes en ambos tipos de microorganismos.

o Las reacciones citoplásmicas que resultan en pentapéptido—GlcNAc—MurNAc, pueden ser divididas en cuatro grupos: 1) formación de UDP-GlcNAc a partir de fructosa-6-fosfato, 2) formación de UDP-MurNAc a partir de UDP-GlcNAc, 3) ensamblaje del UDP-N-acetilmuramil-pentapéptido y 4) formación colateral de ácido D-glutámico y el dipéptido D-alanil-D-alanina⁹⁵⁹.

o Cuatro de las proteínas que fueron identificadas por contener regiones codificadas por tripletes RNY catalizan reacciones que ocurren en la fase citoplasmática de biosíntesis del peptidoglicano, lo que significa que aunque ciertos dominios de algunas de las proteínas surgen en el Mundo de las RNPs el peptidoglicano no podía ser sintetizado, tales módulos quedaron fijados y posteriormente interaccionarían con otros hasta generar enzimas modernas que con las demás enzimas que se habrían sintetizado posteriormente, formarían la cadena biosintética completa en estadios más avanzados de evolución. Las cuatro enzimas identificadas son: UDP-N-acetilglucosamina 1-carboxiviniltransferasa (MurA), UDP-N-acetilmuramoil-tripéptido--D-alanil-D-alanina ligasa (MurF), UDP-N-acetilmuramoilalanil-D-glutamato--2,6-diaminopimelato ligasa (MurE) y glucosamina--fructosa-6-fosfato aminotransferasa, isomerizante (GlmS); mientras que el orden en que actualmente interaccionan, aunque de manera no secuencial, es: GlmS, MurA, MurE, MurF⁹⁵⁹. La enzima GlmS participa en el primer paso de la

síntesis de peptidoglicano para la formación de GlcN-6-P (glucosamina-6-fosfato) a partir de Fru6P (fructosa-6-P), otros dos pasos subsecuentes catalizados por GlmM y GlmU se siguen para obtener UDP-GlcNAc (UDP-N-acetil-glucosamina)^{957,959}. La siguiente etapa de la fase citoplasmática se inicia con la catálisis de MurA y consiste en la transferencia del residuo enolpiruvato del PEP (fosfoenolpiruvato) a la posición 3 de la UDP-GlcNAc. El siguiente paso es catalizado por MurB y consiste en la reducción de la porción enolpiruvato a D-lactato resultando en UDP-MurNAc (UDP-N-acetil-muramato). Una serie de ligasas dependientes de ATP (MurC a MurF, nombradas consecutivamente por las reacciones en que están involucradas) catalizan la adición de una cadena pentapeptídica en el nuevo grupo reducido D-lactato, lo que resulta en la formación de UDP-MurNAc.

o En los microorganismos actuales G⁺ y G⁻ todas las enzimas MurA-MuF son ubicuas, poseen alta especificidad por su substrato, son esenciales, están altamente conservadas y pueden identificarse motivos estructurales comunes⁹⁶⁰.

- **MurA**, llamada originalmente MurZ⁹⁶¹ cataliza actualmente la transferencia de la porción enolpiruvil del PEP al grupo 3-OH de la porción GlcN de la UDP-GlcNAc mediante la formación de dos intermediarios, con la concomitante liberación de fosfato inorgánico^{960,962,963,964,965}. La reacción catalizada por MurA es actualmente inhibida por UDP-MurNAc, el producto de MurB⁹⁶⁶; sin embargo en el Mundo de las RNPs esa inhibición por retroalimentación no existía pues ni siquiera una porción de MurB fue identificada como ancestral. MurA es una proteína conformada por dos dominios estructuralmente similares que se repelen entre sí, el sitio activo está localizado entre ambos y su formación es inducida por la unión del substrato al provocar cambios en la conformación^{967,968,969}, es posible además que la unión del substrato, UDP-GlcNAc cargado negativamente, neutralice la fuerza repulsiva entre ambos dominios. La estructura de la

proteína consiste de un plegamiento repetido seis veces; el único elemento repetido en la secuencia de aminoácidos es un motivo pequeño L₃-X₃-[G/A] responsable de la formación de puentes de hidrógeno entre los dominios de la enzima^{970,971}. En los organismos G⁺ existen dos enzimas MurA funcionales que son expresadas bajo diferentes etapas de desarrollo, ambas pudieron haber surgido a partir de una duplicación y MurA1 es posiblemente la enzima ancestral, aunque ambas son menos eficientes que las contrapartes de organismos G⁻⁹⁷²; el parásito clamidia, que se presumía carente de pared de peptidoglicano posee la enzima MurA y algunos otros hallazgos sugieren que de hecho sí posee peptidoglicano funcional⁹⁷³. De acuerdo con varios trabajos, la única otra enzima similar a MurA en secuencia, mecanismo y estructura, que incluso comparte sitios estrictamente conservados, es la 5-enolpiruvilshikimato-3-fosfato sintasa (EPSPS o AroA) que está involucrada en el anabolismo de aminoácidos aromáticos^{962-973,974,975}; también se ha encontrado que en las ciclasas de **RNA, que catalizan la formación de extremos 2'-3'-cíclicos** en el RNA utilizando la hidrólisis de ATP, el dominio grande encargado de la unión múltiples grupos fosfatos es similar a uno de los dominios de AroA y MurA, por lo que es posible que las tres enzimas compartan un origen común⁹⁷⁶. Sin embargo el ancestro de las tres proteínas es posiblemente más cercano a MurA, ya que es la única de ellas identificada por la metodología descrita en esta tesis; inicialmente pudiera pensarse que la secuencia ancestral de MurA corresponde al motivo repetido encargado de la formación de puentes de hidrógeno pues éste contiene aa codificados por tripletes RNY; no obstante se menciona que las tres enzimas son prácticamente idénticas entre sí y dado que las tres acoplan una molécula con grupos fosfato, es más probable que la región que une dicha molécula corresponda al

primordio de todas ellas, o del ancestro de estas, que habría surgido en el Mundo de las RNPs.

- **MurF y MurE.** Ambas son enzimas que catalizan los pasos finales de la fase citoplásmica de la síntesis de peptidoglicano utilizando la energía derivada de la hidrólisis del ATP. La enzima MurE cataliza la adición del tercer residuo aminoacídico a la cadena peptídica del UDP-MurNAc, usualmente este tercer residuo es un diaminoácido que varía entre las diversas especies de microorganismos. La enzima MurF cataliza el paso final de la fase citoplásmica de la síntesis del peptidoglicano catalizando la adición de D-Ala-D-Ala a la creciente cadena UDP-MurNAc-tripéptido⁹⁶⁰. Dado que estas enzimas son altamente específicas a sus substratos, las sintetetasas Mur deben poseer sitios bien definidos para la unión del UDP-MurNAc(-péptido) y para el substrato ya sea péptido o aminoácido. De acuerdo con estudios experimentales realizados con la sintetetasa MurC, es muy probable que el orden de unión de los substratos en todas las proteínas Mur sea: ATP, UDP-MurNAc y Ala. Con base en la homología que existe entre las secuencias de aminoácidos de las sintetetasas MurC a MurF, se ha propuesto que evolucionaron a partir de un ancestro común; posteriormente un estudio más fino reveló que es posible incluso detectar invariantes comunes: siete aminoácidos K¹³⁰, E¹⁷⁴, H¹⁹⁹, **N**²⁹³, **N**²⁹⁶, R³²⁷, **D**³⁵¹ y la secuencia consenso **GXXGKT/S**, todos ellos ocupando la misma posición en los alineamientos, además que la conservación de la longitud entre ciertos invariantes sugiere además que hay motivos estructurales comunes; las sintetetasas MurC-F son entonces una clase bien definida de proteínas evolutivamente relacionadas entre sí. En un trabajo posterior, donde se compararon las secuencias de las ligasas Mur y FolC (una enzima cuyo mecanismo es similar y está involucrada en la síntesis el cofactor folato), se definieron las mismas

regiones de homología⁹⁷⁷. De igual modo que para la mayoría de los genes involucrados en la biosíntesis de una ruta metabólica, los genes *mra* que codifican para las sintetasas MurF y MurE, entre otras, se localizan en un sólo operón. El análisis de las secuencias revela que existen homología, no sólo entre las sintetasas Mur como se estableció previamente, sino con algunas secuencias de otras proteínas involucradas en la pared celular y en la división celular. En ese nuevo trabajo, existen cuatro regiones de homología conservadas entre todas las sintetasas MurC-F: la región **GxxGKT/ST/S**, previamente identificada que corresponde al dominio de unión al ATP que está conservado en las proteínas de unión a nucleótidos; un segundo motivo **GxxNxxNxxAAxA-5x-G-18xR**, compuesta por residuos conservados de **Asn** en un contexto general hidrófobo; otros dos motivos conservados adicionales existen entre los mencionados anteriormente, estos son la región III, que es una secuencia ácida corta, mientras que la región II es caracterizada por una Gln y una His conservadas, adicionalmente un residuo alanina en la región IV está conservado en todas las ligasas Mur⁹⁷⁸. Las ligasas MurC-F también llevan a cabo mecanismos enzimáticos similares, que involucran la activación de un carboxilato libre al transferirse el grupo fosforilo del ATP, formando un intermediario acilfosfato, seguido de la formación de un nuevo enlace amida con un aminoácido entrante. La estructura de las ligasas Mur se compone de dos dominios que se aproximan uno a otro al ocurrir la catálisis; el dominio N-terminal, responsable de la unión del substrato UDP-MurNAc, es semejante al plegamiento Rossmann típico de las enzimas que unen nucleótidos y puede considerarse como la parte fundamental de las ligasas Mur; mientras que el dominio C-terminal está muy probablemente asociado con la unión del aminoácido entrante y presenta el grado más alto de diversidad en

secuencia y estructura; en el sitio activo existen compartimentos específicos, alrededor del centro catalítico, para la unión del UDP-MurNAc, Mg^{2+} -ATP y L-Ala. Con base en evidencias bioinformáticas, bioquímicas y estructurales, se ha determinado que todas las ligasas Mur surgieron a partir de la duplicación de un gen ancestral, además MurC y MurD son más similares entre sí del mismo modo que MurE y MurF lo son entre sí⁹⁷⁹. Es posible que el primordio de las cuatro ligasas Mur constituya a su vez el primordio de MurF y MurE; dado que es el substrato UDP-MurNAc es muy complejo como para haber existido en un ambiente primitivo, mientras que tanto el aa Ala como los componentes del Mg^{2+} -ATP son incluso de origen prebiótico, es probable que el sitio que une alguna de estas últimas dos moléculas haya surgido en los albores del Mundo de las RNPs.

- **GlmS** es conocida formalmente como L-glutamina:D-fructosa-6-fosfato amidotransferasa isomerizante y es una enzima ubicua en todos los organismos y en al menos un virus^{980,981,982}. GlmS es una enzima esencial en los procariontes cuando existe poca o nula disponibilidad de aminoazúcares en el medio, pues está involucrada en el anabolismo del peptidoglicano, los lipopolisacáridos y los ácidos teicóicos al catalizar la formación de GlcN6P a partir de Fru6P de manera irreversible⁹⁸³ y por tanto se considera un punto de control metabólico muy importante; sin embargo la enzima GlmS cataliza la hidrólisis de glutamina a glutamato en ausencia de Fru6P, mientras que en ausencia de glutamina lleva a cabo una actividad fosfoglucosa del tipo isomerasa que lleva al equilibrio la Fru6P y la Glu6P, durante el proceso catalítico se forman diversos intermediarios inestables, aunque ninguna de esas reacciones requieren de ATP o algún otro cofactor⁹⁸¹⁻⁹⁸³. GlmS es una enzima lenta y relativamente ácida, que exhibe alta especificidad por L-Gln como donador del grupo amino, no pudiendo

emplear nitrógeno exógeno en su lugar, los substratos se unen estrictamente en el orden: L-Gln y D-Fru-6-P⁹⁸¹⁻⁹⁸³. GImS pertenece a la familia de las amidotransferasas y aunque por mucho tiempo se consideró una isomerasa, ha recibido innumerables nombres como glucosamina-6-P sintasa, hexofosfato aminotransferasa, GlcN-6-P ceto-isomerasa o D-fructosa-6-P amidotransferasa⁹⁸¹⁻⁹⁸³. Como todas las amidotransferasas, la GImS es una enzima modular y homodimérica en que cada monómero comprende dos dominios: el dominio N-terminal une la glutamina y la hidroliza a glutamato y amoniaco; el dominio C-terminal, también conocido como sitio de isomerización, contiene el sitio de unión a la D-fructosa-6-P y utiliza el amoniaco para la síntesis de GlcN6P; el fragmento que une ambos dominios es flexible y permite al menos tres cambios en la conformación de la enzima durante el ciclo catalítico, dicho conector contiene además un **"canal" hidrófobo que permite la transferencia del** amoniaco de un dominio al otro^{981-983,984,985,986,987,988}. Ambos dominios pueden llevar a cabo las reacciones catalíticas correspondientes de manera independiente uno del otro⁹⁸¹⁻⁹⁸³. Los residuos altamente conservados de la GImS se ubican en ambos dominios y están involucrados en la unión de los substratos y en la catálisis⁹⁸¹⁻⁹⁸³, estos son: C¹, **D**²⁹, H⁸⁶, H⁹⁷, **D**¹²³, C³⁰⁰, K⁴⁸⁵, E⁴⁸⁸, H⁵⁰⁴ y K⁶⁰³; la secuencia **DxPxxLAK(SC)VT**, que está altamente conservada en el dominio C-terminal, también parece ser una señal distintiva de GImS⁹⁸⁹; el dominio C-terminal contiene además dos dominios SIS (por isomerasa de azúcares, en inglés), involucrados en la unión no específica de grupos fosfato-azúcar^{981-983,990}. Mientras que eucariontes la expresión de GImS es inhibida por UDP-GlcNAc⁹⁸¹, en organismos procariontes la expresión de GImS es regulada a nivel post-transcripcional, aunque la regulación es distinta dependiendo si se trata de organismos G- o G+, siendo la

regulación en algunos organismos G- más compleja pues depende de más factores⁹⁸²; dado que el organismo utilizado para la metodología presentada aquí es una bacteria G+, la regulación correspondería a organismos de este tipo. La zona 5'-UTR ("untranslated region" o región no traducida 5') del mRNA de *GlmS*, contiene la ribozima/riboswitch *GlmS* (*rzm/rsw GlmS*), la cual es únicamente activada por la unión del metabolito GlcN6P (producido por *GlmS*) que promueve su auto-escisión^{982,991}, de modo que el GlcNP actúa como coenzima para la *rzm/rsw GlmS*⁹⁹² en un circuito regulatorio de inhibición por retroalimentación. Con base en el análisis de la secuencia de la *rzm/rsw GlmS*, se han identificado cuatro regiones que pueden parearse formando pseudonudos, pero sólo dos de esas regiones están altamente conservadas, por lo que se les considera el consenso mínimo de las *rzm/rsw GlmS*s que es esencial para su funcionamiento⁹⁹³; recientemente se ha mostrado que, aún con algunas variantes, este consenso mínimo se conserva en secuencia y estructura, entre bacterias filogenéticamente distantes⁹⁹⁴. Con base en la estructura terciaria se ha establecido que la unión del metabolito no promueve el plegamiento de la *rzm/rsw GlmS*⁹⁹⁵ y de hecho no promueve ningún cambio en la conformación⁹⁹⁶, sino por el contrario la unión ocurre en un sitio preexistente⁹⁹⁷ expuesto al solvente⁹⁹⁸, es posible incluso encontrar la *rzm/rsw GlmS* completamente plegada en solución⁹⁹⁶; además se ha encontrado que el sitio de unión del ligando se encuentra próximo al sitio de segmentación del mRNA⁹⁹⁹ y el reconocimiento es muy específico¹⁰⁰⁰, pues el grupo amina del metabolito GlcN6P unido a la *rzm/rsw GlmS*, lleva a cabo catálisis general ácido-base y catálisis electrostática⁹⁹⁸ sin importar la presentación estereoquímica del cofactor¹⁰⁰¹; recientemente se ha encontrado que la actividad de la *rzm/rsw GlmS* puede ser modulada *in vivo* por diversos metabolitos, mientras sean

hexosas o aminohexosas, lo cual implica que la *rzm/rsw GlmS* sirve para integrar la información acerca del estado metabólico general de la célula¹⁰⁰². La estructura terciaria también revela los nucleótidos esenciales para la unión del ligando y el mecanismo catalítico⁹⁹⁸. A diferencia del común de las proteínas, los iones metálicos divalentes no son esenciales para la función catalítica de la *rzm/rsw GlmS*¹⁰⁰³, aunque funcionan en la coordinación de los sitios de unión de la GlcN6P con el metabolito¹⁰⁰⁴. Como podemos observar, la *rzm/rsw GlmS* manifiesta varios rasgos remanentes del Mundo del RNA ^{cfr.992,1005,1006,1007}; no obstante, dado que la metodología toma en cuenta únicamente los ORFs de cada proteína, sólo es posible aseverar que GlmS contiene una región proveniente del Mundo de las RNPs; ese primordio pudiera corresponder al motivo distintivo de dicha enzima que contiene aa codificados por tripletes RNY, o más probablemente a los dominios SIS que unen complejos fostato-azúcar cuyos componentes son posiblemente incluso de origen prebiótico.

- o La pared de peptidoglicano es una estructura compleja que comparten tanto procariontes G+ como G-⁹⁸¹⁻⁹⁸³; como podemos observar es posible que en los organismos primitivos ya existían ciertos dominios que conformarían más tarde algunas de las proteínas que interaccionarían entre sí para formar enzimas completas y generar entonces la ruta anabólica del peptidoglicano.

- o **PROTEÍNAS DE ADHESIÓN CELULAR (Proteína de la familia de la tríada de histidina estreptococal, Proteína de dominio Cna B y Proteína de unión a fibrinógeno).**

- o Las tres proteínas mencionadas están actualmente involucradas en la patogénesis del organismo, sin embargo decidí incluirlas dentro de esta categoría

porque todas ellas se encuentran ancladas a la superficie celular; dado que antes de la aparición del LUCA no había algún organismo complejo al cual infectar.

- **Proteína de la familia 'tríada de histidina estreptocócica'**. Las proteínas de esta familia se han denominado PhtA, -B, D y -E parecen ser exclusivas del género *Streptococcus*, aunque hasta la fecha no se ha descrito su función general; las cuatro características que presentan en común son^{1008,1009,1010}: 1) un motivo HxxHxxH de donde toman su nombre y está repetido cinco veces a lo largo de la secuencia¹⁰⁰⁹, al menos 3 de los cuales estarían involucrados directamente en la función catalítica al coordinar un ion Zn^{2+} que es esencial para la actividad biológica de la proteína¹⁰⁰⁹; 2) un motivo de 11 aminoácidos, repetido dos veces a lo largo de la secuencia y cuya función se desconoce totalmente, ya que no existe semejanza alguna con dominios conocidos; 3) la inoculación de estas proteínas en un organismo que pudiera servir de huésped al microbio que las presente, le confiere al primero cierta protección frente a las características patógenas del segundo^{1009,1010} y 4) estas proteínas se encuentran ancladas en la superficie celular de los microorganismos¹⁰¹⁰. A pesar que la región de interés médico es el motivo HxxHxxH, no es plausible que ésta sea la región identificada como ancestral, ya que la histidina no es un aminoácido codificado por tripletes RNY ni un aminoácido que pudiese haber sido ser cargado por una aaRS de baja especificidad; posiblemente la región identificada como ancestral, en este miembro de la familia de la tríada de histidina, corresponda al motivo repetido de los 11 aa cuya función se desconoce en la actualidad, lo que sugeriría que su actividad biológica debe ser relevante y por tanto investigada, ya que la secuencia que lo codifica no ha sido eliminada en más 3.5 millones de años de evolución.

- Dominio tipo B de proteína de Cna.** Las proteínas Cna MSCRAMM (componentes de superficie microbiana que reconocen moléculas adhesivas de matriz, por sus siglas en inglés) se caracterizan por una región no-repetitiva (A), que contiene el dominio de unión a colágena (CBD), seguida de uno o hasta cuatro unidades repetidas (B) dependiendo de la cepa bacteriana que se trate, una tercera región de asociación a la pared celular y finalmente un segmento citoplásmico corto^{1011,1012}. Se ha encontrado que la capacidad de unión a la colágena de la proteína Cna es independiente de variaciones en la o las unidades B que contenga¹⁰¹³, incluso se ha mostrado que la región B no es necesaria en absoluto para la unión a la colágena^{1014,1015}. El análisis particular de los dominios B repetidos reveló que comparten entre sí substancial semejanza en secuencia, al ser diferentes en un sólo aminoácido¹⁰¹⁶. Las unidades repetitivas de la región B de la proteína Cna son de aproximadamente 23 kDa, que a su vez están compuestas de dominios entre los que se preserva homología estructural, estos dominios están arreglados en forma de acordeón y, aunque actualmente se desconoce la función de esta región B de la proteína Cna, la estructura sugiere que podría estirarse y contraerse provocando que la región A se proyecte fuera de la superficie celular¹⁰¹⁷. La estructura cristalina de la unidad B1 muestra dos dominios similares, D1 y D2, que comparten estructuras similares e identidad del 67% en las secuencias **DKDNODGKRPEK** y **DNNNODGKRPT**E; en las unidades B1B2 los "loops" o bucles que conectan D1 con D2 y D3 con D4 tienen la secuencia KY**TPGET** y la que conecta D2 con D3 la secuencia KY**TPET**, por lo que se deduce que la primera es más flexible estructuralmente debido a la **Gly**, aunque ambas comparten un alto grado de homología¹¹¹⁷. Según he investigado, no se cuenta con algún trabajo en que se mencione la secuencia completa de la

región B de la proteína Cna, ya que los análisis de esta región se enfocan en la comparación de estructuras y parte de la secuencia de algunas unidades, por lo tanto no resulta asequible dilucidar cuál podría ser el fragmento ancestral de ésta; no obstante, aun cuando la función de los bucles reportados no está definida, podemos observar que contienen los aminoácidos **G**, **D**, **N** y **T**, mismos que son codificados por tripletes RNY, además de K y E que pudieron haber sido cargados por aaRS de baja especificidad durante los albores del Mundo de las RNPs.

- **Proteína de unión a fibrinógeno.** Existen diversas proteínas presentes en la superficie celular de los microorganismos que les permiten interaccionar con el fibrinógeno de sus hospederos^{1018,1019,1020,1021,1022}. Las proteínas denominadas FbsA y FbsB, que fueron descritas por primera vez en *Streptococcus agalactiae* (*Sag*), han sido las moléculas de unión a fibrinógeno más estudiadas, dichas proteínas que no comparten homología en secuencia ni estructura y por tanto no comparten relación filogenética alguna¹⁰²³. FbsA es la proteína de unión a fibrinógeno más estudiada y es el receptor de dicho ligando en *Sag*, ésta posee similitud estructural con muchas de las proteínas asociadas a la superficie celular, es además una proteína que contiene una secuencia señal para secreción y un dominio compuesto por 3 y hasta 30 unidades repetidas, altamente similares entre sí, cada una con los 16 aa **GNVLERRQRDAENRSQ** (cuyo consenso es **G[N/S/T]VL[A/E/M/Q]RRX[K/R/W][A/D/E/N/O][A/F/I/L/V/Y]XX[K/R]XX**) y del que **-NVL-RR** resultan ser los aa críticos para la unión del fibrinógeno^{1024,1025,1026,1027,1028,1029}. FbsB, por su parte, es una proteína no-repetitiva, secretada en vez de unida covalentemente a la superficie celular de hecho la región N-terminal de FbsB es conservada entre diversas proteínas que se secretan y la región C-terminal de 223 aa es casi idéntica

entre al menos tres variantes de FbsB¹⁰³⁰. Recientemente algunos grupos de trabajo se han enfocado en la búsqueda de proteínas de unión a fibrinógeno en otras especies de *Streptococcus*¹⁰³¹. Como podemos observar, varios de los aa que se encuentran en la región de unión a fibrinógeno son codificados por tripletes RNY o bien son aminoácidos que pudieron haber sido cargados por aaRS de baja especificidad; sin embargo, dado que no se tiene un trabajo en el que se reporte la secuencia completa de FbsA, o un trabajo acerca de la posible filogenia de esta proteína, y además el fibrinógeno es una proteína característica de vertebrados que evidentemente no es antigua, no resulta fácil discernir si una proteína de unión a fibrinógeno es verdaderamente ancestral o su identificación se deriva del hecho que está compuesta por unidades repetitivas en las cuales varios de los aminoácidos sí son codificados por tripletes RNY.

o Las tres proteínas previamente analizadas son estructurales, no enzimáticas, están compuestas de unidades repetitivas y todas ellas funcionan como moléculas de adhesión; sin embargo hasta la fecha no existen trabajos acerca de las relaciones filogenéticas de cada una, por lo que no es posible establecer argumentos sólidos acerca de su identificación como proteínas ancestrales. Es posible que por el hecho de que las 3 proteínas anteriores se componen de unidades repetitivas y estas unidades contienen aminoácidos codificados por tripletes RNY, se confiere un resultado falso positivo y dichas proteínas fueron identificadas como ancestrales sin que realmente surgieran en los albores del Mundo de las RNPs; por otro lado también es posible que los segmentos identificados como ancestrales, en efecto lo sean y posteriormente fueron cooptados por otros dominios proteicos, conformándose la respectiva molécula de adhesión actual.

o **PROTEÍNA DE LA FAMILIA DE LAS GLICOSILTRANSFERASAS.**

o Las cadenas de oligosacáridos unidas a la superficie celular, así como las proteínas y lípidos extracelulares, son conocidas por mediar diversas funciones biológicas importantes. La diversidad en la glicosilación se encuentra en cada nivel de organización biológica, tanto entre- como intra-especies, sin embargo las enzimas involucradas en la glicosilación (glicosiltransferasas o GTs) son altamente conservadas¹⁰³². Las GTs son enzimas que sintetizan oligosacáridos, polisacáridos y glicoconjugados de péptidos o lípidos, catalizando la transferencia de la porción azúcar de moléculas donadoras activadas a moléculas aceptoras específicas mediante enlaces glicosídicos. Las GTs se dividen en dos tipos dependiendo de la estereoquímica de sus substratos y productos: las GTs retentivas transfieren el azúcar reteniendo la configuración del carbono anomérico, mientras que las GTs inversoras transfieren el azúcar con inversión de la configuración del carbono anomérico. La diversidad de donadores de azúcar pueden ser di- o polisacáridos, azúcar 1-fosfatos, o más comúnmente nucleótido-difosfo-azúcares (NDP-azúcares). La especificidad de las GTs puede deberse a diferencias estructurales muy pequeñas y específicas; en varias ocasiones se ha observado que las enzimas actúan en substratos similares y actúan con el mismo mecanismo, aunque se clasifican en diversas familias, ya que existen pocas regiones de homología entre las distintas clases y las similitudes son sólo locales. Inicialmente se propuso que una clasificación con base en la similitud de secuencias, estructuras y mecanismos moleculares conservados, ofrecía una clasificación más precisa, permitiendo identificar varias familias poli-específicas y llegando la propuesta que las diferencias observadas en la especificidad por los substratos posiblemente refleja la evolución divergente de la forma ancestral de glicosiltransferasa, aunque también se ha identificado al menos un ejemplo de evolución convergente^{1033,1034,1035}.

o Aunque no existe un alto grado de homología significativa entre las GTs, es claro que los motivos peptídicos altamente conservados en cada clase desempeñan un papel importante en la función enzimática; la característica común dentro de cada grupo homólogo es generalmente la utilización del mismo nucleótido-azúcar, o alternativamente de la misma porción nucleotídica (**e.g.** UDP); de particular interés es la existencia del motivo **DxD** en familias de GTs que utilizan diferentes aceptores y donadores. En todas las estructuras y modelos, los motivos más conservados en las GTs están involucrados en la actividad catalítica y de unión al sustrato¹⁰³⁴.

o Uno de los dominios conservados en las GTs inversoras es el responsable del reconocimiento de nucleótidos del tipo pirimidina, ya sea UDP o la porción TDP de un donador nucleótido-azúcar (ya sea que contenga glucosa, galactosa, ácido glucurónico, xilosa, ramnosa o daunosamina), ya que el nucleótido es la única estructura común entre los donadores y aceptores de ese tipo de GTs, a este dominio se le ha nombrado NRD1 β (por las siglas en inglés de: dominio de reconocimiento de nucleótidos 1 β); el dominio NRD1 β está localizado en la porción C-terminal de las GTs, la porción N-terminal contiene también una pequeña región homóloga al NRD1 β . Un residuo [**S/T**], potencial de ser fosforilado, es localizado en la parte media del NRD1 β y esta fosforilación podría regular la actividad enzimática; es probable incluso que todas las GTs con NRD1 β compartan el mismo mecanismo de reacción. Podemos encontrar GTs con NRD1 β en eucariontes, arqueas y bacterias, estas enzimas pueden ser divididas en cinco clases con base en el grado de similitud de sus secuencias; todas las GTs bacterianas se agrupan en la clase III, el arquea *Methanococcus jannaschii* también posee proteínas de este tipo y por ello mejor conocidas como GTs microbianas, lo anterior indica que las GTs con NRD1 β evolucionaron antes de la separación actual en tres grandes dominios¹⁰³⁵.

o La porción central del NRD1 β es muy similar a uno de los dominios presentes en una de las familias de GTs retentivas, por lo que en este caso se le denomina familia NRD1 α y probablemente la mencionada porción central homóloga tiene función catalítica. Los miembros de la familia de las GTs NRD1 α pueden aceptar nucleótidos-azúcar que contengan ya sea pirimidina o purina, pues carecen de motivos de reconocimiento de nucleótidos específicos; esta familia puede ser subdividida en GTs de procariontes, de eucariontes-plantas y de eucariontes-animales. En el caso de los microorganismos, la amplia mayoría de las GTs que participan en la formación de exopolisacárido (EPS), lipopolisacárido (LPS) o estructuras del polisacárido capsular (CPS) pertenecen a la familia de las GTs NRD1 α , algunas de las cuales comparten el motivo RxxxxK, un área hidrófoba subsecuente de cinco a siete aminoácidos. La comparación de secuencias de GTs NRD1 β y NRD1 α revela además la presencia de dos residuos Glu conservados, así como los motivos PQ y/o **DQ**¹⁰³⁵.

o Un tercer dominio denominado NRD2 se encuentra exclusivamente en las GTs de membrana y todas son inversoras. Los miembros de las GTs NRD2 pueden aceptar nucleótidos-azúcar que contienen tanto pirimidinas como purinas y por lo general son proteínas asociadas a membrana que pueden llevar a cabo contactos con ésta, según los índices hidropáticos que les caractericen; sin embargo la mayoría de los miembros bacterianos de la familia NRD2 carecen de dominios transmembranales en la porción N-terminal, aunque si contienen motivos hidrófobos a lo largo de la secuencia; es posible además identificar una región conservada en esta familia de GTs, localizada exactamente en la porción N-terminal o justo antes del dominio membranal (cuando existe). Varios miembros de la familia NRD2 participan en la formación de la pared celular y de estructuras extramembranales, en particular el grupo 4 de la familia participa en la biosíntesis del CPS y algunas

proteínas sintetizan ácidos teicóicos (más adelante se habla de este tipo de moléculas), el grupo 5 consiste de enzimas involucradas en la síntesis de LPS. Los miembros de la familia NRD2 también son encontrados en varios eucariontes y procariontes distantemente relacionados, lo que posiblemente indica que la presencia de familia de GTs NRD2 existió también antes de la separación de los tres grandes dominios y aún más que su evolución estuvo relacionada con la evolución de la membrana celular. Se ha especulado también que las GTs de la familia NRD2, necesarias tanto para la biosíntesis del LPS bacteriano y de los glicolípidos de eucariontes, han evolucionado a partir de un predecesor común¹¹³⁵.

o Las GTs frecuentemente presentan una estructura modular, en que el módulo catalítico está rodeado de varios módulos ancestrales que, usualmente aunque no siempre, unen carbohidratos. Las estructuras de GTs pueden adoptar un plegamiento GT-A o un plegamiento GT-B. El plegamiento GT-A comprende dos dominios distintos, uno involucrado en la unión del nucleótido (llamado dominio SGC) y el otro en la unión del aceptor; en este tipo de plegamiento la estructura constriñe la coordinación un ión metálico por dos oxígenos fosfato del nucleótido que se une y los aminoácidos **DxD** de la proteína. El plegamiento GT-B se caracteriza por dos plegamientos tipo Rossmann similares; las proteínas con este tipo de plegamiento no tienen el motivo DxD y ningún ion metálico ha sido identificado claramente en la estructura, aunque se ha reportado que algunas de esas enzimas son dependientes de metal. Se ha observado que aunque el plegamiento esté conservado no implica que la función lo esté y la división en GTs retentivas o GTs inversoras no necesariamente se refleja en el tipo de plegamiento¹⁰³⁶. Recientemente las GTs se han clasificado en subfamilias, familias, clanes o superfamilias y plegamiento común con base en los distintos grados de relación de las proteínas; por un lado la categorización a nivel de subfamilia permite

la predicción funcional, por otro lado el agrupamiento por plegamientos permite la detección de relaciones evolutivas distantes¹⁰³⁷.

o Los tres dominios NRD, mencionados anteriormente, se encuentran presentes en GTs virales y de eucariontes y procariontes, donde son indispensables¹¹³⁵, de hecho las GTs ya se encontraban presentes antes del LUCA y quizá su origen es más antiguo. Una posible explicación al hecho que una GT fue identificada como ancestral, puede deberse a que algún segmento de un motivo NRD está codificado por tripletes RNY; adicionalmente diversos escenarios evolutivos planteados implicarían que en los albores del Mundo de las RNPs ya existían azúcares y nucleótidos, que pudieron interactuar formando complejos nucleótido-azúcar que serían reconocidos por algún dominio polipeptídico del tipo NRD, más específicamente conteniendo el motivo **DxD**.

o **PROTEÍNAS DLT.**

o Los ácidos teicóicos son cadenas poliméricas que contienen residuos glicerol-fosfato o ribitol y son polímeros exclusivos de las bacterias G+, que pueden estar anclados a la membrana celular (LTA) o a la pared celular de peptidoglicano (WTA)¹⁰³⁸. **Los ácidos teicóicos crean una "carga negativa continua" en la superficie bacteriana**, la cual puede fungir como una red para los cationes metálicos mono- y divalentes esenciales para diversas proteínas y en general para la homeostasis iónica de la bacteria; los ácidos teicóicos también sirven como receptores o moléculas de anclaje o adhesión que le permiten a la bacteria interactuar con su medio ambiente y con otros organismos; de hecho los ácidos teicóicos también han sido implicados en el crecimiento celular, la división y la morfogénesis¹⁰³⁹. Muchos de los ácidos teicóicos están decorados con azúcares adicionales y aminoácidos, lo que regula la carga de estos polímeros; las modificaciones usuales en las bacterias G+ con alto contenido de G+C (G+ con alto %GC) son acetil, glutamil y lisil,

mientras que los ácidos teicóicos de las G+ con bajo contenido de G+C (G+ con bajo %GC) son usualmente modificados con D-alanina¹⁰⁴⁰.

o Las enzimas del operón *dlt* (**A-E**) catalizan la D-alanilación de los ácidos teicóicos, lo que neutraliza la carga negativa neta por la introducción de aminoácidos básicos. Al parecer las proteínas Dlt no parecen ser esenciales para el desarrollo del organismo y la carencia de D-alanina únicamente resulta en ácidos teicóicos completamente aniónicos¹⁰⁴¹, aunque se ha mostrado además que los genes *dlt* están involucrados en la acumulación de polisacárido intracelular¹⁰⁴² y la D-alanilación no sólo confiere cargas positivas a los ácidos teicóicos sino también a las proteínas secretadas y las asociadas a la pared celular, lo que incrementa su estabilidad y las protege de la degradación durante el plegamiento post-traducciona¹⁰⁴³; se ha propuesto incluso que la carencia de D-alanina interfiere con los procesos celulares normales de transformación, adherencia y división celular¹⁰⁴⁴. Se ha mostrado que la expresión de proteínas Dlt es dependiente de la fase de crecimiento del organismo, siendo la mayor expresión durante la fase exponencial media y durante condiciones de estrés ambiental^{1144,1045,1046,1047}; de hecho parece ser que la carencia de enzimas Dlt vuelve al organismo vulnerable ante el medio ambiente¹⁰⁴⁸, por ejemplo sensible a ácidos, aunque al parecer la ausencia de ácidos teicóicos decorados con D-alanina también se presenta cuando hay cantidad suficiente de iones divalentes esenciales en el medio^{1047,1049}, lo que genera un entorno suficientemente positivo que vuelve dispensables otro tipo de moléculas cargadas positivamente.

o Las proteínas del operón *dlt* no están filogenéticamente relacionadas, sin embargo sus funciones están profundamente interconectadas en la secuencia C, A, B, D, aunque la E para no se ha descubierto función alguna¹⁰⁵⁰; de esas proteínas, las únicas identificadas como ancestrales son: DltB, DltC y DltD.

- **DltB.** Es una proteína esencialmente hidrófoba según su perfil hidropático, lo que sugiere que es una enzima transmembranal que está involucrada en el transporte de la D-alanina a través de la membrana celular, cuando se une al acarreador específico de dicho aa^{1046,1048,1049,1150}. DltB es similar en secuencia a varios anti-portadores de protones involucrados en el transporte de diversos compuestos. Un estudio posterior mostró que DltB es de hecho parte de una superfamilia denominada MBOAT (por O-acetiltransferasas unidas a membrana), que transfieren distintas moléculas a blancos embebidos en la membrana celular; la característica general de los miembros de esta superfamilia es una porción invariante de residuos hidrófobos que sugieren la presencia de 8 a 10 regiones que se extienden a través de la membrana, además de una región con similitud de secuencia detectable¹⁰⁵¹. Posiblemente los residuos hidrófobos fueron los identificados como ancestrales, ya que algunos de los aminoácidos codificados por tripletes RNY presentan ese tipo de característica química.
- **DltC.** Es una proteína ácida¹⁰⁴⁸ que actúa como un acarreador altamente específico de la D-alanina activada que se une al cofactor fosfopanteteína mediante una ligasa específica (DltA)¹⁰⁴⁹, la D-Ala se une a DltB y de ésta a DltD^{1048,1049,1050,1051}. DltC pertenece a la familia de las proteínas acarreadoras de acilos (ACPs), involucradas en la biosíntesis de ácidos grasos¹⁰⁵² y en que la máxima conservación de secuencia se ubica en el sitio de la unión a la fosfopanteteína; por otro lado las DltC de diversos organismos contienen dos regiones homólogas, que no están presentes en las ACPs, Kxx**VDxLA** y **DxVKxNxD** y que podrían determinar la interacción específica de DltC con DltA¹⁰⁵³. DltC contiene además un sitio de interacción específica con los residuos esenciales de los ácidos teicóicos, lo que concuerda con las propuestas que el complejo D-alanil-DltC es trascolado a

través de la membrana al interactuar con DltB. La DltC es además capaz de efectuar la trans-acilación de los residuos D-alanil y los residuos de los ácidos teicóicos, jugando un papel principal en la distribución de la D-alanina¹⁰⁵⁴. Posteriormente se mostró que sólo 20% de los residuos entre DltC y las ACPs son idénticos, aunque la homología estructural es significativamente alta entre estas proteínas; por otro lado DltC es también estructuralmente homóloga a las proteínas de la biosíntesis de poliquétidos y a proteínas de la síntesis de péptidos no-ribosomales, sobre todo en sitio de unión a la D-alanina¹⁰⁵⁵, aunque la homología en secuencia con estas enzimas es sólo de 2.6% a 6.8%¹⁰⁵⁶. Algunas de las proteínas del anabolismo de ácidos grasos contienen regiones ancestrales (ver sección 'Metabolismo') y si la región catalítica de las ACPs es también ancestral, pues de acuerdo a los resultados es homóloga en secuencia, estructura y muy probablemente en mecanismo a DltC, ello podría indicar que el organismo del cual formaban parte tenía el potencial catalítico para sintetizar su propia membrana celular una vez que las enzimas de la vía metabólica completa hubiesen aparecido. De todo lo anterior es posible deducir que la región codificada por tripletes RNY en DltC puede corresponder o al sitio de unión de la fosfopanteteína como en las ACPs, o al sitio de unión a D-Ala como en proteínas de síntesis de poliquétidos y de péptidos no-ribosomales; en cualquier caso la existencia de cualquiera de las dos moléculas es plausible en el Mundo de las RNPs.

- **DltD.** Es una proteína que, según su perfil hidropático, contiene un péptido-señal en la porción N-terminal, por lo que se sugirió que ésta es una proteína que se ancla a la membrana y está involucrada en la transferencia de la D-alanina de su acarreador (DltC) a los residuos de los ácidos teicóicos^{1048,1049,1050,1051}. Resultados experimentales posteriores revelaron

que DltD está efectivamente asociada a la membrana y promueve la unión correcta de la D-alanina con su acarreador específico, mostrando además una actividad de tioesterasa, al hidrolizar el enlace del acarreador de D-alanina cuando está incorrectamente cargado, lo que facilita la unión de la D-alanina^{1051,1057}.

- o Existen tres observaciones relevantes¹⁰⁵¹ que podrían explicar los resultados analizados: 1) se ha demostrado que en la mayoría de las proteínas que sintetizan péptidos por la vía no-ribosomal (sintetasas no-ribosomales), los **dominios tioesterasa se encuentran usualmente "río abajo" (o "downstream")** del módulo sintetasa, 2) la proteína DltD no comparte homología con otras proteínas conocidas y 3) las enzimas DltC y DltD están acopladas traduccionalmente lo que indica la estrecha relación estructural y funcional entre ambas proteínas; dicho acoplamiento se refleja en que la región codificada por tripletes RNY corresponde¹⁰⁵¹ a un fragmento de DltC y de DltD, *i.e.* un mismo fragmento tuvo coincidencia con dos proteínas distintas pero funcionalmente relacionadas.
- o Los D-aa llevan a cabo funciones reguladoras y de señalización aunque, a diferencia de los L-aa, no pueden ser incorporados a las proteínas mediante tRNAs. Estos D-aa sirven como nutrientes durante el crecimiento bacteriano; funcionan como señales autócrinas y parácrinas, tanto intra-especie como entre especies, regulando la formación de esporas bajo condiciones de estrés ambiental y promoviendo la dispersión de las formaciones conocidas como **"biofilms"**; además forman parte de la pared celular bacteriana¹⁰⁵⁸. Se ha encontrado que los experimentos de química prebiótica producen mezclas racémicas de aminoácidos y aunque actualmente existen racemasas que cambian la configuración¹⁰⁵⁸, éstas no fueron identificadas como ancestrales. De manera que posiblemente los dominios que reconocen D-aa sean muy

ancestrales, y es posible entonces que la región codificada por tripletes RNY en DItC y DItD corresponda al sitio de unión de D-Ala, ya que es el metabolito en común entre ellas; esto implicaría entonces que posiblemente un sistema que unía L-aa (sistema ribosomal) coexistiese con dominios polipeptídicos de unión a D-aa (sistema no-ribosomal) en el Mundo de las RNPs.

o **PROTEÍNA DE LA FAMILIA DE LAS METALO-BETA-LACTAMASAS, M β L.**

o Los antibióticos llamados β -lactámicos son una amplia variedad de compuestos que interfieren con la síntesis de la pared de peptidoglicano; las enzimas denominadas β -lactamasas inactivan dichos antibióticos al catalizar la hidrólisis del anillo β -lactámico mediante dos mecanismos distintos y no relacionados filogenéticamente. El primer mecanismo involucra el aa **Ser** localizado en el sitio activo de la enzima y por lo que se denominan S- β -lactamasas; el segundo mecanismo depende de la activación de una molécula de agua por un ion Zn^{2+} , en modo análogo a la metalo-peptidasas, por lo que se han denominado Zn^{2+} - o metalo- β -lactamasas (M β L) a tales enzimas. Las β -lactamasas han sido clasificadas en cuatro grupos, A, B, C y D con base en sus relaciones filogenéticas; de dichos grupos sólo la clase B corresponde a metalo- β -lactamasas, este grupo B se subdivide a su vez en B1-2 y B3¹⁰⁵⁹.

o La primer estructura resuelta de una M β L proviene del microorganismo *Bacillus cereus*, ésta es una enzima que cataliza la hidrólisis de la gran mayoría de los antibióticos β -lactámicos en dicho organismo. El plegamiento de la enzima presenta simetría molecular interna y el sitio activo está cercano al extremo N-terminal, lo que sugiere una posible duplicación genética, mientras que el ion Zn^{2+} es coordinado por tres His y una molécula de agua. Una comparación de secuencias reveló que sólo nueve aminoácidos estrictamente conservados en las cinco

secuencias disponibles hasta el momento: L, **D**, **G**, L, H, **G**, **N** e H, de los cuales la última H y el **D** participan directamente en la hidrólisis del anillo β -lactámico¹⁰⁶⁰. Un trabajo posterior mostró que en realidad son dos iones Zn^{2+} los que se acoplan al sitio activo y la enzima es completamente funcional con cualquiera de los dos unidos¹⁰⁶¹.

- o A pesar que la mayoría de los trabajos relacionados con β -lactamasas se enfocan en la búsqueda de nuevos fármacos, algunos trabajos muy importantes están orientados a buscar el posible origen y evolución de dichas proteínas.

- o Las M β L comprenden en realidad una familia muy diversificada de enzimas, que a su vez pertenecen a una de las dos superfamilias de hidrolasas dependientes de Zn^{2+} ; esta superfamilia comparte cinco regiones conservadas, de éstas las menos conservadas no están presentes en la totalidad de los miembros del grupo; por otro lado la mayoría de las posiciones conservadas corresponden al sitio activo de la M β L de *Bacillus cereus*. Se plantea que la diversidad catalítica y consiguiente adaptabilidad de la superfamilia de Zn^{2+} -hidrolasas podría explicar cómo ciertos miembros pudieron evolucionar hacia M β L en respuesta a presiones ambientales¹⁰⁶².

- o Con base en la secuencia y estructura de la proteína, se desarrolló una clasificación evolutiva de las M β L y se planteó que la actividad catalítica de las proto- β -lactamasas era mucho más general que en la actualidad, pues se trataba únicamente de hidrolasas dependientes de Zn^{2+} u otros cationes divalentes, esas metalo-hidrolasas fueron especializándose bajo diversas presiones de selección y finalmente hubo eventos de exaptación de las formas inactivas de la enzima actual. La familia de las M β L se compone de miembros filogenéticamente distantes que al menos comparten 4 de los 5 motivos conservados¹⁰⁶³.

- o En un estudio distinto, se clasifican los miembros de la familia de M β L en 16 grupos que contienen el dominio o plegamiento M β L, presente en una gran

diversidad de proteínas sin relación filogenética entre sí. La familia incluye: clase B de β -lactamasas, glioxilasa II, arilsulfatasa, flavoproteína, ciclasa/deshidratasa, una **enzima procesamiento del extremo 3' del RNA**, una **proteína de reparación de DNA** y una proteína de absorción de DNA, una hidroxilasa CMP-N-acetilneuraminato, el producto del gen *romA*, una proteína relacionada con la absorción de alquifosfonato, la alquilsulfatasa e incluso hidrolasas de insecticidas¹⁰⁶⁴.

o Más recientemente, se identificaron 3 sublinajes de la clase B de β -lactamasas, identificados como las subclases B1, B2 y B3, de los que B1 y B2 descienden de un único grupo dentro del cual las secuencias exhiben similitudes significativas, por otro lado B3 comparte similitudes estructurales aunque no de secuencia con B1 y B2. En dicho trabajo la superfamilia de las M β L se ha clasificado únicamente en 8 grupos distintos, los cuales incluyen hidrolasas de diversos enlaces -éster del tipo tiol-, fosfodi-, ácido sulfúrico-, así como distintas oxidorreductasas y hasta el momento de la publicación de dicho trabajo, existían más de 6000 miembros en la superfamilia de las metalo- β -lactamasas, los cuales comparten 5 motivos conservados, siendo el motivo H¹¹⁶xH¹¹⁸x**D**¹²⁰H¹²¹ el más característico; todos los miembros además exhiben la topología **$\alpha\beta\beta\alpha$** de las M β L. De acuerdo con el estudio mencionado, el plegamiento M β L, representa claramente una plataforma que ha sido repetidamente utilizada durante la evolución para catalizar un amplio rango de reacciones químicas; estructuralmente esta diversidad se lleva a cabo al variar la secuencia y longitud de las regiones cerca del sitio activo, ya que al combinar el dominio M β L con dominios auxiliares, substratos, productos y/o unión de cofactores distintos, el espectro de funciones se ha extendido considerablemente desde su surgimiento en la historia evolutiva¹⁰⁶⁵.

o El plegamiento de la M β L representa una de las formas básicas de varias actividades biológicas y está presente tanto en forma activa como estructural en los

tres dominios de la vida, por lo que puede deducirse que dicho plegamiento estaba presente en el LUCA¹⁰⁶³. La ubicuidad en la distribución de la familia de las M β L a lo largo de los 3 dominios de la vida, sugiere un origen antiguo e importancia fundamental¹⁰⁶⁴. En algunos trabajos se ha propuesto que los genes de resistencia a antibióticos, particularmente las β -lactamasas, surgieron a partir de genes constitutivos (housekeeping) que evolucionaron como consecuencia de la exposición constante a esos antibióticos¹⁰⁶⁶; de hecho muchos miembros de la superfamilia de las M β L están involucrados en la maduración del mRNA y reparación del DNA, lo que sugiere que esta superfamilia puede ser resultado de evolución divergente que proviene de una proteína ancestral que muy posiblemente no poseía actividad β -lactamasa¹⁰⁶⁵. Con base en la evolución enzimática y flexibilidad catalítica del dominio M β L, se ha extrapolado el umbral de su estabilidad y se ha encontrado que aún no se agota su espectro funcional *i.e.* aún puede expandirse y posiblemente surjan nuevas proteínas que contengan el dominio mencionado¹⁰⁶⁷. Posteriores análisis filogenéticos extensos de los dos grupos de β -lactamasas que existen, confirman que no están relacionados entre sí e indican que la antigüedad de las M β L es de más de 2,000 millones de años (antes de la divergencia de microorganismos G+ y G-), de hecho las filogenias de β -lactamasas y genes constitutivos son **altamente congruentes. El análisis del denominado "resistoma" y en particular de la superfamilia de las M β L, muestra que sus miembros presentan alto grado de divergencia, están presentes en los tres superdominios de la vida y comparten propiedades comunes con proteínas cuya función no es de resistencia; el resistoma es entonces la colección de genes de resistencia y de genes que tienen el potencial para evolucionar hacia elementos de resistencia, denominándose en ese caso genes de proto-resistencia (en analogía a los proto-oncogenes). Dado el origen antiguo y evolución de las β -lactamasas, no es de sorprender que los genes de resistencia**

fuesen rápidamente tomados de reservorios ambientales por organismos patógenos mediante transferencia horizontal de genes^{1059,1068,1069,1070}.

o Los genes de las β -lactamasas son muy antiguos y se han encontrado en ambientes insospechados, remotos y aislados¹⁰⁷¹. Un estudio realizado en el suelo de Alaska reveló nuevos tipos de β -lactamasas bifuncionales que anteriormente no habían sido descritas; también se mostró que algunas bacterias del suelo pueden subsistir utilizando los antibióticos como única fuente de carbono¹⁰⁶⁹. Otro estudio de microorganismos de muestras de subsuelo de más de 3 millones de años, mostró que existen bacterias filogenéticamente muy diversas que presentan resistencia y multi-resistencia a antibióticos sin haber establecido contacto previo con actividades antropogénicas¹⁰⁷². Recientemente se ha mostrado que algunos bacteriófagos de DNA aislados de muestras ambientales de agua, presentan algunos genes de resistencia a antibióticos β -lactámicos, lo que podría indicar que no solamente microorganismos no patógenos, sino que los virus pueden ser y de hecho son reservorios ambientales de genes de resistencia¹⁰⁷³.

o Una proteína de la familia de las metalo- β -lactamasas fue identificada como ancestral, sin embargo no es un resultado totalmente inesperado y de hecho existen varias explicaciones posibles al respecto: 1) este resultado no implica que la proteína identificada tenga actividad antibiótica, pues en realidad la familia de las M β L agrupa proteínas muy diversas cuya única característica en común es el plegamiento M β L; 2) se ha planteado que el plegamiento M β L se encontraba presente en el LUCA y dado que una proteína de la familia M β L contiene tripletes RNY en su secuencia, es posible que al menos un fragmento del plegamiento sea tan antiguo que probablemente corresponda a la región identificada como ancestral; 3) es posible que dicho fragmento codificado por tripletes RNY corresponda a lo que se ha denominado como proteína de proto-resistencia; 4) es también posible que lo

identificado como ancestral, sea alguno de los cinco motivos de la superfamilia M β L, aunque probablemente no el característico HxHxDH que sólo contiene un aa codificado por tripletes RNY, sino alguno de los otros cuatro que contienen, entre otros, los aa **Asp**, **Gly** y **Asn**, aunque con la información general que se tiene no es posible precisar cuál de ellos es el primordio de la proteína de esta superfamilia M β L.

o **PROTEÍNA CON DOMINIO CHAP.**

o La superfamilia de proteínas CHAP fue descubierta recientemente por dos grupos independientes^{1074,1075}, quienes describen amidohidrolasas y peptidasas sin otras características en común más que 1) un dominio de tres motivos conservados, que contienen residuos Cys, His y **Asp**, los cuales parecen ser esenciales para la catálisis y 2) todas estas proteínas hidrolizan sustratos que contienen una porción γ -glutamil^{1074,1075}. La mayoría de las proteínas CHAP se encuentran en bacterias y bacteriófagos^{1074,1075} y están usualmente involucradas de alguna manera en la remodelación o degradación de la pared de peptidoglicano^{1076,1077,1078}; también existen algunas proteínas de esta superfamilia en arqueas y eucariontes^{1074,1075}. Más trabajos sobre CHAP en bacterias y bacteriófagos en: 1079,1080,1081,1082,1083,1084,1085,1086y1087.

o Se ha encontrado que las proteínas de la superfamilia CHAP asociadas al metabolismo del peptidoglicano podrían pertenecer a su vez a las peptidasas del grupo denominado NlpC/p60, que utilizan cisteína como principal residuo catalítico, sin embargo existe poca similitud a nivel de secuencia entre ambos grupos de peptidasas¹⁰⁷⁶. Se ha propuesto además que el dominio CHAP puede conferir especificidades múltiples a las proteínas que lo contengan¹⁰⁷⁷.

o De acuerdo con un estudio de las proteínas con dominio CHAP, de *Staphylococcus aureus* y sus bacteriófagos asociados, es posible distinguir

claramente dos ramas filogenéticas distintas: una que agrupa las proteínas bacterianas que contienen el dominio CHAP en el extremo C-terminal, otra de las proteínas virales que poseen el dominio CHAP en el N-terminal y además contienen el patrón de secuencia F-**[IV]**-R altamente conservado. En el estudio mencionado también se investigaron secuencias semejantes a CHAP que contienen ligeras variaciones, en todas ellas hay seis residuos altamente conservados: Q, C, **G**, H, **V** y **N**, cinco de los cuales están expuestos a la superficie molecular y agrupados en el sitio activo putativo, mientras que la **Val** está profundamente embebida en la región hidrófoba esencial de la proteína¹⁰⁷⁸.

o Probablemente la proteína identificada como ancestral corresponda a una de las variantes CHAP, ya que sólo los aminoácidos **D**, **G**, **V** y **N**, que podrían ubicarse en el sitio activo junto con el **D** de las proteínas CHAP regulares, son codificados por tripletes RNY y no así los aminoácidos Cys e His; adicionalmente es posible que la proteína identificada esté actualmente asociada al metabolismo del peptidoglicano, ya que la amplia mayoría de las proteínas con dominio CHAP lo están y aunque en un inicio quizá su función fue distinta, ésta fue cooptada hacia aquella ruta metabólica durante la evolución.

Con base en algunas extrapolaciones estructurales, a partir de comparaciones de morfología y algunas enzimas responsables de sintetizar componentes extramembranales, Thomas Cavalier ha planteado que los primeros organismos en existir fueron negibacterias sin membrana externa que posteriormente la adquirieron al sintetizar LPS cuando el ambiente se transformó en oxidante, perdiéndolo después y engrosando su pared celular, misma que posteriormente también se perdió surgiendo los eucariontes¹⁰⁸⁸.

De acuerdo con nuestros resultados, algunas de las proteínas que fueron identificadas como ancestrales están involucradas exclusivamente en la síntesis de

estructuras extramembranales de organismos del tipo G+, ciertamente la bacteria de la que hemos obtenido las coincidencias discutidas en la presente tesis es un organismo G+ y ello podría implicar un sesgo en los resultados, sin embargo si la propuesta de Thomas Cavalier-Smith es cierta y los organismos G+ hubiesen surgido después de la aparición del oxígeno, los fragmentos codificados por triplete RNY de las proteínas identificadas, habrían quedado almacenados para evolucionar ulteriormente a proteínas modernas mediante la combinación de diversos dominios.

Por supuesto es necesario contar con una muestra suficientemente variada de organismos, comenzando con G+ y G-, para poder tener resultados completamente concluyentes y reconstruir verdaderamente la historia evolutiva.

Nuestros resultados pudiesen indicar que la evolución de la vida ha sido en realidad un proceso parsimonioso, en donde los organismos de una población pre-LUCA contenían proteínas que posteriormente formaron estructuras extramembranales del tipo G+ o bien estructuras tanto del tipo G+ como G- dentro de mismo organismo, planteamiento que podrá realmente discernirse al obtener más adelante el genoma RNY de organismos G- y en general de muchos más organismos.

T13. Características plausibles de la probable región codificada por tripletes RNY en proteínas identificadas como parte de la membrana o estructuras extramembranales.

Conjunto	Proteína (nombre traducido y adaptado)	Características plausibles de la probable región codificada por tripletes RNY.
	Proteína de membrana	No precisada
	Lipoproteína	No precisada, quizá interacción con la membrana lipídica
Síntesis de peptidoglicano o mureína	MurA MurF MurE GlmS	Posiblemente unión de grupos fosfato Posiblemente unión de L-Ala, o unión de ATP-Mq2+ Posiblemente unión de L-Ala, o unión de ATP-Mq2+ Unión de grupos fosfato-azúcar, o simplemente motivo distintivo de la enzima
Proteínas de adhesión celular	Proteína de la familia 'tríada de histidina estreptocócica' Dominio tipo B de proteína CNA Proteína de unión a fibrinógeno	No precisada No precisada, pero hay motivos con varios aa codificados por RNY No precisada, pero hay unidades repetidas con varios aa codificados por RNY
	Proteína de la familia de las glicosiltransferasas; GTs	Unión de nucleótidos, o de azúcar, o de complejos nt-azúcar
Proteínas Dlt, para ácidos teicóicos	DltB DltC DltD	Residuos hidrófobos para inserción en una membrana lipídica Unión de D-Ala, o de fosfopanteteína No precisada, quizá unión de D-Ala
	Proteína de la familia de las metalo-beta-lactamasas; MβL	No precisada, pero al menos 5 motivos contienen aa codificados por tripletes RNY
	Proteína de dominio CHAP	No precisada, pero varios aa altamente conservados son codificados por tripletes RNY

6.3.1.6 Transportadores y Canales.

Dado que los organismos más antiguos carecían de las enzimas necesarias para sintetizar varios de los compuestos esenciales, su sobrevivencia dependía en gran medida de la capacidad para captar los nutrientes del medio externo y exportar metabolitos de desecho; se ha mostrado que tales sistemas comparten un origen filogenético común y varias de las proteínas que permiten el intercambio de moléculas entre el exterior y el interior de la célula, son mucho más antiguas que el surgimiento del LUCA, lo cual queda señalado por el hecho que varias de estas contienen regiones codificadas por tripletes RNY que entonces son posiblemente tan ancestrales como la formación de la membrana/pared celular misma. Las proteínas de transporte que median el paso de pequeños solutos a través de las membranas bacterianas muestran una fuerte conservación de secuencia en los 3 dominios de la vida. Los motivos estructurales y patrones filogenéticos de transportadores dependientes de alguna forma de energía, sugieren que éstos contienen un módulo esencial que resulta ser el canal mismo a través del cual el soluto traspasa la membrana, mientras que los mecanismos energéticos correspondientes se pueden considerar como características secundarias que fueron añadidas durante la evolución a esas unidades fundamentales de trascolación. El tema común de diseño de esos transportadores son dominios transmembranales similares que comprenden el canal *per se*, embebido en la membrana. Estructuralmente los transportadores de la membrana externa se componen de hebras- β transmembranales, mientras que las proteínas de la membrana plasmática son hélices- α agrupadas en dos dominios de seis hélices cada uno. La diversidad de los transportadores radica tanto en la especificidad de sus canales, como en los dominios periféricos empleados para acoplar la energía^{1089,1090,1091,1092,1093}.

La membrana externa contiene tres tipos de canales cuyo mecanismo es pasivo: 1) porinas, que son grandes canales que permiten el movimiento pasivo de iones y moléculas hidrófilas pequeñas; 2) facilitadores, que poseen sitios específicos que promueven la difusión facilitada de algunos compuestos; 3) sistemas dependientes de TonB, una proteína anclada en la membrana citoplásmica asociada con un receptor de alta afinidad.

Por otro lado, los acarreadores de la membrana interna se clasifican en activos y pasivos con base en la naturaleza de sus mecanismos de transporte¹⁰⁸⁹⁻¹⁰⁹³:

PASIVOS: Permiten la difusión de los solutos a través de la membrana, por lo que se les denomina **'sistemas de difusión facilitada'**, cuyos únicos ejemplos pertenecen a la familia de la **'proteína intrínseca principal'** (MIP) y son el facilitador de glicerol (GlpF) y acuaporinas de diversas fuentes, estos transportadores se componen de 6 hélices- α transmembranales.

ACTIVOS: Poseen algún mecanismo de energía acoplado al transporte.

Primarios: Son sistemas que dependen directamente de la hidrólisis de ATP; uno de estos sistemas utiliza dos proteínas hidrófobas periplásmicas de unión al substrato (PBP), dos proteínas hidrófilas que contienen sitios de unión a nucleótidos (ABC), lo que les confiere el nombre de **'transportadores de unión a ATP'** (transportadores ABC, por las siglas en inglés **'ATP binding cassette'**) y finalmente dos dominios transmembranales (TMD); otro sistema de este tipo son las ATPasas-P que usualmente bombean iones metálicos y contienen el bucle-P de unión a nt.

Secundarios: Son sistemas que acoplan el transporte de nutrientes o moléculas esenciales con el transporte de iones, por lo que también se les llama electroquímicos y pueden ser uniportadores, antiportadores o simportadores, la mayoría de los cuales pertenecen a la **'superfamilia del facilitador principal'** (MFS), aunque hay algunos transportadores secundarios que no pertenecen a esta familia; dichos transportadores están conformados por 2 dominios de 6 hélices- α cada uno y

comparten una identidad del ~30% entre sí; adicionalmente algunos los estos simportadores dependen de un receptor extracitoplásmico que posee alta afinidad con el soluto.

PTS: Son sistemas de trascolación vectorial que acoplan la transferencia de un grupo PO_4^{3-} , que proviene de la hidrólisis del fosfoenolpiruvato (PEP), a diversos azúcares que son transportados a través del canal, por lo tanto se conocen como **'sistemas de transporte dependientes de PEP'** o PTS por sus siglas en inglés.

Algunos de los componentes de estos sistemas acoplados a fuentes de energía muestran homología con enzimas no involucradas en el transporte, lo que indica se naturaleza de mosaico, algunos de esos sistemas han evolucionado por duplicación y otros por fusión de dominios; aunque finalmente se aprecia que sólo son variaciones para cubrir requerimientos específicos, pues los componentes y el mecanismo elemental de transporte se encuentran altamente conservados¹⁰⁸⁹⁻¹⁰⁹³.

Entre las proteínas transportadoras que contienen regiones codificadas por tripletes RNY encontramos: transportadores primarios, transportadores secundarios y complejos PTS. A continuación proporciono los argumentos que expliquen de manera general el hecho de haber encontrado que algunas de estas proteínas (en **T14** y **T15**) que sirven para el transporte de moléculas contienen regiones ancestrales.

Order	Sc.	E-v	ID	%	PROTEIN / FAMILY / DOMAIN
1	31	3.00E-07	31/31	100%	ABC transporter, substrate-binding protein, putative (32.57 %)
2	30	1.00E-06	30/30	100%	glycine betaine/proline ABC transporter, permease/substrate-binding protein (70.30 %)
3	27	6.00E-05	44/50	88%	ABC transporter, ATP-binding protein CylA (5.06 %) ABC transporter, permease protein CylB (1.03 %)
4	27	6.00E-05	27/27	100%	polar amino acid uptake (PAAT) family ABC transporter, amino acid-binding protein (8.81 %)
5	26	2.00E-04	26/26	100%	major facilitator family transporter (26.42 %)
6	26	2.00E-04	26/26	100%	iron chelate uptake ABC transporter, permease protein (1.14 %) iron chelate uptake ABC transporter, permease protein (1.75 %)
7	25	0.001	25/25	100%	peptide/opine/nickel uptake (PepT) family ABC transporter, substrate-binding protein (30.65 %)
8	25	0.001	25/25	100%	permease, putative (43.78 %)
9	25	0.001	25/25	100%	ABC transporter, permease protein, carbohydrate uptake 2 (CUT2) family, putative (98.64 %)
10	25	0.001	34/37	91%	ABC transporter, ATP-binding protein (39.92 %)
11	24	0.004	24/24	100%	polar amino acid uptake (PAAT) family ABC transporter, permease protein (63.89 %)
12	24	0.004	24/24	100%	BioY family protein (53.06 %) (biotin uptake)
13	24	0.004	24/24	100%	permease, putative (89.58 %)
14	24	0.004	24/24	100%	polar amino acid uptake (PAAT) family ABC transporter, permease protein (26.21 %)
15	24	0.004	24/24	100%	spermidine/putrescine ABC transporter, permease protein PotB (23.05 %)
16	24	0.004	24/24	100%	nickel ABC transporter, permease protein (44.17 %)
17	23	0.015	23/23	100%	permease, putative (66.73 %)
22	23	0.015	23/23	100%	PTS system, IIA component, lactose/cellobiose family (17.65 %)
18	23	0.015	23/23	100%	ABC transporter, substrate-binding protein (72.29 %)
19	23	0.015	23/23	100%	ABC transporter, ATP-binding/permease protein (43.24 %)
20	23	0.015	23/23	100%	K ⁺ transporter (Trk) family protein (1.96 %)
21	23	0.015	23/23	100%	glycerol facilitator/aquaporin (16.32 %)
22	22	0.06	22/22	100%	carbohydrate uptake 1 (CUT1) family ABC transporter, permease protein (8.55 %)
23	22	0.06	28/30	93%	major facilitator family transporter (21.76 %)
24	22	0.06	22/22	100%	quaternary amine uptake (QAT) family ABC transporter, permease protein (34.77 %)
25	22	0.06	22/22	100%	PTS system, IIA component, lactose/cellobiose family (11.15 %)
26	22	0.06	22/22	100%	PTS system, IIC component, mannose/fructose/sorbose family (10.71 %)
25	22	0.06	22/22	100%	calcium-transporting ATPase, P-type (transporting), HAD superfamily, subfamily IC (41.58 %)
26	22	0.06	22/22	100%	ABC transporter, permease protein (14.96 %)
27	22	0.06	22/22	100%	polar amino acid uptake (PAAT) family ABC transporter, permease protein (68.20 %)
28	22	0.06	22/22	100%	glycerol facilitator/aquaporin (44.62 %)

T14. Transportadores y canales que contienen regiones codificadas por tripletes RNY, nomenclatura y código de colores igual que T1 y T2.

o **FACILITADOR DE GLICEROL / ACUAPORINA, SUPERFAMILIA MIP.**

o La entrada de agua a las células es esencial para la supervivencia de los seres vivos y por otro lado la osmorregulación es importante para sobrevivir a varias condiciones estresantes en el medio externo; los miembros de superfamilia de la 'proteína intrínseca principal' (MIP) **fungen** entonces como canales en la membrana que transportan selectivamente agua, iones y moléculas neutras pequeñas intra- e inter-celularmente. Dichas proteínas comparten un plegamiento común en forma de reloj de arena, de 26kDa a 34kDa, conformado en seis hélices- α transmembranales y dos enlaces hidrófilos; dicho plegamiento pudo haber surgido por duplicación interna como lo indica la homología del 20% al 40% y la identidad del 20% y la estructura equivalente que comparten el dominio-N y el dominio-C. Con base en similitud de secuencias y características funcionales, la superfamilia ubicua MIP puede ser dividida en tres grupos principales cuya divergencia es muy antigua: los facilitadores de glicerol y de otras moléculas pequeñas (GlpFs) tienen poca permeabilidad al agua y son monoméricos, mientras que las acuaporinas (AQPs, o canales selectivos para agua) y las acuagliceroporinas (AQGPs) se presentan en

forma de homotetrámeros hidrófilos. La comparación filogenética entre AQPs y GlpFs revela que pese a la paralogía existe amplia divergencia entre ellos, lo que indica que el ancestro común es al menos tan antiguo como el LUCA. Todos los miembros de la familia MIP se caracterizan por un par de motivos **NPA(/NPS/NLA)** donde el primero está contenido en la secuencia **SGxHxNPAVT** y el otro en **NPAR**, estos son invariantes, se encuentran opuestos ~ equidistantemente y forman el canal para el soluto^{1094,1095,1096,1097,1098,1099,1100,1101}.

- o Dado que la proteína identificada como ancestral se menciona como miembro de la familia MIP, es en realidad irrelevante mencionar los motivos conservados característicos de cada grupo (GlpF, AQP o AQGP), puesto que no se indica cuál es la proteína específica de la familia MIP que contiene el organismo *SagA909*; no obstante es probable que la región codificada por tripletes RNY, corresponda a una de las regiones invariantes mencionadas, pues cualquiera de las dos interacciona con solutos que existían antes del Mundo de las RNPs.

- o Indudablemente, tanto el agua como el glicerol son moléculas pequeñas antiguas que evidentemente ya existían en el Mundo de las RNPs y por tanto es posible que el primordio de la proteína identificada de la familia MIP haya unido alguna de esas moléculas; por otro lado, los motivos que forman parte del poro son codificados en su mayoría por tripletes RNY y esta fue probablemente la región codificada por tripletes RNY mediante la metodología descrita en la presente tesis.

- o **TRANSPORTADORES ABC.**

- o La familia de transportadores ABC es una de las más extensas, con representantes en cada uno de los dominios de la vida; dichos sistemas son altamente específicos para una amplia variedad de substratos y funcionan en la importación y exportación de solutos e inclusive hay miembros de la familia ABC que no están involucrados en el transporte. Los transportadores ABC son sistemas

multi-componentes, formados canónicamente por tres módulos principales: dos proteínas hidrófobas de especificidad al sustrato (PBP), dos dominios transmembranales (TMD) que forman el canal *per se* y dos proteínas hidrófilas de unión a ATP (dominios ABC que confieren nombre a la familia); no obstante en ocasiones la PBP está ausente o forma parte de un módulo PBP-TMD, en otros casos existe un sólo módulo ABC-TMD y en varios ejemplos los módulos equivalentes están fusionados de diversas formas. Sin importar el arreglo particular de cada sistema ABC, es notable que los tres módulos canónicos se organizan de esta manera seriada en los operones específicos (PBP+TMD+ABC), lo que indica que los distintos sistemas ABC surgieron por duplicación de un sistema completo aunque rudimentario que debió estar formado posiblemente antes del LUCA. De manera particular los módulos ABC son altamente similares entre sí, por tanto las filogenias se han establecido únicamente con base en la correspondencia encontrada entre PBP y TMD. Con base en una clasificación estructural y filogenética de los sistemas ABC se ha encontrado que: existen miembros de la familia que no funcionan como transportadores pues carecen de TMDs, pertenecen a la clase II (cII) y se encuentran presentes en todos los organismos, además la mayoría de esos casos están involucrados en el procesamiento de DNA y/o RNA; por otro lado existen los transportadores de clase I (cI) que sirven para la exportación de productos de desecho y están ubicuos en todos los organismos; finalmente existen los sistemas ABC de clase III (cIII) que dependen de una proteína que canalice los solutos que serán importados, éstos sistemas sólo se encuentran en procariontes aunque algunos eucariontes los poseen de manera incompleta. Posiblemente la clase III de sistemas ABC sea la más antigua, tomando en cuenta al inicio de la vida era imperioso importar nutrientes cuya concentración era baja y en realidad no era necesario exportar productos de desecho, ya posteriormente el LUCA importaba y exportaba solutos (generalidades en: 1102,1103,1104,1105,1106,1107,1108,1109,1110,1111,1112,1113,1114,1115,

1116,1117,1118,1119,1120,1121,1122,1123,1124,1125,1126; PBP: 1127,1128; TMD: 1129,1130; ABC: 1131,1132,1133,1134,1135

,1136,1137). A continuación se describen las características de cada módulo con base en la información antes mencionada (¹¹⁰²⁻¹¹³⁷); para cada descripción general se mencionan las proteínas específicas identificadas como ancestrales por el método descrito en la presente tesis y la posible región que probablemente esté codificada por tripletes RNY, así como las referencias particulares correspondientes.

- **PBP.** Son los dominios periféricos que confieren especificidad y direccionalidad a los importadores ABC (cIII), pues se encuentran asociados solamente a dichos sistemas; la unión del soluto produce cambios conformacionales en los dominios hidrófilos de las PBP, estas interaccionan entonces con las proteínas de membrana que transportarán el substrato; no obstante se considera que la PBP es una proteína accesoria y en realidad dispensable para el adecuado funcionamiento de un transportador ABC. Las PBP son relativamente pequeñas, de 20 a 60 kDa, estables y poseen alta afinidad por sus ligandos, la unión de los cuales tiene efectos en su propio transporte o en la quimiotaxis del organismo. Todas las PBP exhiben la mayor homología en el extremo-N y la menor hacia el extremo-C, aunque la secuencia general presenta baja homología y a pesar de ello son estructural y mecánicamente muy similares; todas las PBPs presentan un plegamiento α/β en que los dominios están dispuestos en forma de bisagra, en cuya hendidura se une el soluto mediante numerosos puentes de hidrógeno y de manera muy similar a una 'Venus atrapamoscas' se cierran cuando el soluto se une, liberándolo en el módulo TMD al momento de su interacción; la PBP y posee una estructura similar al plegamiento Rossmann e inclusive el soluto se une de manera similar y sin importar el tamaño de la PBP dicha región abarca un pequeño porcentaje del total e impone diversas restricciones a la secuencia y estructura de la proteína. Con base

en la topología de la estructura modular se han identificado PBP-I y PBP-II, de los cuales el tipo I es probablemente el más antiguo. Las características mencionadas sugieren que la especificidad por los solutos surgió por duplicación y divergencia y/o eventos de HGT a partir de un módulo fundamental bilobulado conformado solamente por una hebra- β y una hélice- α , eventos evolutivos que en cualquier caso debieron ocurrir antes del LUCA. Dado que las PBP no contienen aa característicos conservados, no es posible dilucidar cuál fue la región identificada como ancestral, sin embargo puede considerarse que posiblemente aquella esencial para preservar **la estructura Rossmannoide o de 'Venus atrapamoscas'**, pues se ha encontrado que en general los aa que confieren especificidad a las proteínas están poco conservados y son en general resultado de procesos de especialización posteriores al LUCA.

- Proteína de unión al sustrato, putativa. Varias de estas se han propuesto como PBP pues están definidas por un ORF que presenta homología en secuencia y estructura con PBPs plenamente identificadas, además que su contexto genético indica pertenencia a algún un operón específico para la utilización de nutrientes importados; aunque al no existir estudios experimentales específicos para mostrar su función, la PBP anotada aquí se considera putativa^{1138,1139,1140,1141,1142,1143,1144,1145,1146,1147,1148}.
- PAAT. Los transportadores de aminoácidos son bidireccionales y se subdividen en dos clases: la familia de transporte de aa polares (PAAT) y la familia de transporte de aa hidrófobos (HAAT); a su vez la familia PAAT contiene 11 distintos tipos de transportadores clasificados en dos categorías con base en su especificidad, misma que es dictada en primera instancia por la PBP y confirmada por el módulo TMD^{1149,1150}.

- Transportador de péptidos/opinas/níquel, familia PepT. Los tripéptidos son una fuente importante de aa para organismos que crecen anaeróticamente y la metalopeptidasa PepT cataliza su formación a partir de escindir péptidos más largos; en varios organismos la secuencia de PepT se superpone con el operón de un sistema ABC denominado PotABCD; PepT da nombre además a una familia de transportadores ABC a la que pertenece el sistema mencionado Pot y transportadores para níquel y para opinas (compuestos derivados de aa, producidos por plantas infectadas por una bacteria específica y cuya síntesis está promovida por el patógeno mismo que los utiliza como fuente de alimento); por lo anterior pudiera deducirse que la familia ABC de PepT está involucrada en el transporte de moléculas relacionadas al procesamiento de péptidos, incluyendo cofactores metálicos^{1151,1152,1153,1154}.
- **TMD**. Se considera el canal *per se*, también denominado permeasa, a través del cual pasarán los solutos e interacciona con el módulo ABC; pese a ser proteínas con un bajo nivel de similitud en secuencia su estructura es notablemente similar entre sí, ésta consta de 2 pares de 6 hélices- α hidrófobas que atraviesan la membrana, conectadas todas ellas por regiones hidrófilas que interaccionan a su vez con los otros dos módulos del sistema. De manera particular los TMD de ABC cIII contienen una secuencia específica denominada bucle-**EA** por los aa conservados que presenta, posiblemente sea esta la región que fue identificada como ancestral, pues aquél es un motivo totalmente conservado en TMD y Ala es un aa codificado por tripletes RNY; aunque como se verá en las descripciones detalladas de cada proteína identificada, éstas contienen otros sitios con aa de tal tipo.
- Permeasa, putativa. Varias de las enzimas anotadas como permeasas, están definidas por un ORF que codifica una secuencia de aa con secciones

hidrófobas definidas, mientras que su contexto genético indica que es parte de algún un operón metabólico, aunque la función precisa de la proteína aquí reportada no está plenamente identificada^{1155,1156,1157,1158}.

- Transportador de glicina-betaína/prolina, TMB/PBP. El operón proU/opuA codifica para el sistema ABC implicado en la importación de glicina-betaína y prolina, que fungen como estabilizantes internos en procariontes y eucariontes ante condiciones de estrés osmótico o por salinidad o por temperaturas bajas. De dicho operón los productos ProW (~36kDa) y ProX (~33kDa) corresponden a los módulos TMD y PBP respectivamente, aunque en algunos organismos ambos componentes están fusionados en un mismo polipéptido donde el dominio-N corresponde al módulo TMD y el dominio-C a la PBP, en cualquier caso la parte TMD presenta el motivo **EAA** como todo ABC cIII^{1159,1160,1161,1162,1163,1164,1165,1166,1167,1168,1169,1170}.
- CyIB. En *S. agalactiae* existe un locus genético que contiene al menos 8 distintos ORFs, de los cuales sólo 6 son esenciales para la producción de citolisina y otros pigmentos utilizados para degradar células procariontes y eucariontes. El producto de uno de esos genes es la proteína CyIB que contiene ~300aa, 6 TMD y un peso de 32kDa, además dentro del operón el gen de CyIB se superpone con el gen de CylA (módulo ABC, más adelante)^{1171,1172,1173,1174,1175}.
- Quelante de hierro. Los sideróforos son proteínas de bajo peso molecular que actúan como quelantes específicos de hierro para transportarlo en condiciones estresantes con baja concentración de este metal esencial, ya sea en forma de Fe²⁺ en organismos que crecen en condiciones de anaerobiosis o bien en microorganismo marinos que habitan cerca de las costas, o como Fe³⁺ en organismos que crecen en ambientes ácidos o en microorganismos marinos que habitan a mar abierto. FhuB es la permeasa

de tal sistema quelante, pesa ~70kDa y sus dos dominios están claramente definidos [-N (1-133) y -C (334-660)], cada uno de los cuales consiste de 10 hélices- α que forman el canal y contienen la secuencia **EAA** característica de los ABC cIII^{1176,1177,1178,1179,1180,1181,1182,1183,1184,1185}.

- CUT1/CUT2. Las permeasas que median la importación de carbohidratos comprenden dos subfamilias: por un lado CUT1 transporta disacáridos, oligosacáridos, glicerol-fosfato y polioles, por otro lado CUT2 transporta solamente monosacáridos y está involucrado en el reciclaje de ribonucleósidos de purina y pirimidina; dado que las permeasas de CUT1 y CUT2 son ABC c-III, contienen el motivo **EAA₃G₉I₁LP** en su secuencia^{1186,1187,1188}.
- PAAT. Los transportadores de aminoácidos son bidireccionales y se subdividen en dos clases: la familia de transporte de aa polares (PAAT) y la familia de transporte de aa hidrófobos (HAAT); los de amplia especificidad con 9 subtipos y los de reducida especificidad con dos subtipos; la diferencia en la especificidad por determinado soluto está determinada por la PBP y por la TMD, pues si bien el dominio-N está conservado en todas las permeasas de PAAT, las de amplia especificidad son 30% más largas y el canal que forman es también más ancho¹¹⁴⁹⁻¹¹⁵¹.
- BioY. En procariontes la importación de biotina es mediada por el sistema BioMNY, del que BioY, que consta de ~20kDa, es una permeasa capaz de transportar el soluto a concentraciones nanomolares o picomolares e incluso sin la presencia los otros dos componentes de dicho sistema y aun cuando el motivo **Ax₃A**, necesario para la interacción con esos módulos, se encuentra conservado; adicionalmente, la capacidad de BioY para transportar la biotina es inversamente proporcional a su afinidad por el

substrato, misma que es mediada por sus hélices- α abundantes en **Gly** y por **D**¹⁶⁴ y **K**¹⁶⁷ como sitios específicos^{1189,1190,1191,1192,1193,1194,1195,1196,1197}.

- Transportador de espermidina/putrescina, PotB. La espermidina y la putrescina son poliaminas necesarias para el crecimiento celular, los sistemas ABC para su importación son homólogos entre sí e incluso en algunos organismos dichos solutos pueden ser transportados por uno u otro sistema indistintamente; de tal modo que la permeasa para importar espermidina, formada por PotB y PotC, se corresponde con la permeasa para importar putrescina, formada por PotH y PotI, con una homología de ~30% por cada par (B-H y C-I); todas ellas constan de ~30kDa, con 6 hélices- α cada una y todas contienen la secuencia conservada **LEAA**[R/K]**LGAS**, como todos los ABC cIII^{1198,1199,1200,1201,1202}.
- Transportador de níquel. El operón *nik* está involucrado en el transporte altamente específico de Ni²⁺, aún en presencia de trazas de este catión que es esencial para el metabolismo fermentativo de numerosos microorganismos; dicho operón contiene de 2 a 5 genes, de los que NikB de ~300aa y NikC de ~200aa conforman la permeasa, que consta de 6 hélices- α y contienen la secuencia característica **EAA** de los ABC cIII^{1203,1204,1205,1206,1207}.
- Importación de aminas cuaternarias, familia OAT. Las aminas cuaternarias están presentes en osmolitos, como betaína, carnitina y colina, que estabilizan la presión osmótica intra-celular; dichas moléculas son importadas en forma de moléculas de amonio cuaternario totalmente metiladas a través de una familia de permeasas, de 312aa a 343aa que constan de 8 hélices- α , que posiblemente se originó por duplicaciones de un

ancestro común que transportaba compuestos estructuralmente similares^{1208,1209,1210,1211}.

- **ABC**. Este es el módulo donde se lleva a cabo la hidrólisis de dos moléculas de ATP-Mg²⁺, lo que provee de gran cantidad de energía al sistema ABC. Sin importar las PBP o TMD asociados, los dominios ABC presentan la mayor similitud entre sí y se consideran el componente más antiguo de los transportadores ABC. La secuencia del módulo ABC consta de 200 a 300 aa de ~30kDa y, como otras enzimas de unión a nucleótidos, su estructura es un plegamiento tipo Rossmann que contiene los conocidos motivos Walker A y B (W_A y W_B) así como una secuencia adicional denominada 'firma ABC'. Los motivos conservados en el módulo ABC unen diferentes partes de la molécula de ATP y linealmente son: **Gx[S/T]GxGK[S/T][S/T]**, que corresponde a W_A o también llamado bucle-P y une la molécula energética + una Gln conservada que da nombre al bucle-Q o 'región estructuralmente diversa' (SDR) + el conector **LSGGQ[Q/R/K]QR** o 'firma ABC', que transmite los cambios conformacionales derivados de la hidrólisis del ATP al transporte de los substratos + h₅**D** que corresponde a W_B cuya conformación es una hebra-β, este **Asp** se localiza en la parte más profunda de la estructura y es esencial para la hidrólisis del ATP (+ una secuencia **TEVGERV** en los sistemas cl) + una His conservada que da nombre al bucle-H. Previo a la unión del ATP el sistema ABC está orientado 'hacia fuera' y recibe al soluto, al llevarse a cabo la hidrólisis el sistema queda orientado 'hacia dentro' y libera el soluto del otro lado de la membrana. De los transportadores ABC, que se consideran de los más antiguos, es el módulo ABC uno de los primeros en surgir; en dicho módulo hay varias regiones con aa codificados por tripletes RNY, no obstante la más extensa de ellas que contiene aa de este tipo es el motivo característico W_A que es

probable haya surgido en los albores del Mundo de las RNPs, pues además una ATP cuyos componentes pudieron sintetizarse incluso prebióticamente.

- Proteína de unión a ATP, putativa. Es el módulo más fácilmente identificable en los sistemas ABC pues contiene los motivos característicos de unión al ATP y la estructura tipo Rossmann; desafortunadamente esta proteína ABC no está plenamente caracterizada y por ende no se ha asociado con alguna PBP o TMD específica^{1212,1213}.
- CylA. En *Sag* existe un locus genético que contiene al menos 8 distintos ORFs, de los cuales sólo 6 son esenciales para la producción de citolisina y de otros pigmentos utilizados para la degradación de células procariontes y eucariontes. El producto de uno de esos genes es la proteína CylA, que contiene los motivos característicos W_A y W_B dentro de sus ~300aa con peso de 35kDa; dentro del operón, el gen de CylA se superpone con el gen de CylB (módulo TMD, revisado anteriormente)¹¹⁷¹⁻¹¹⁷⁵.

- **Proteína de la familia HAD IC, ATPasa transportadora de Ca^{2+}** .

- Las ATPasas-P son transportadores activos primarios de 70 kDa a 200 kDa, que acoplan el transporte de cationes específicos a la formación de fosfointermediarios de alta energía durante su ciclo catalítico y son de vital importancia en los 3 dominios de la vida. La superfamilia de ATPasas-P se clasifica en -P1 a -P4 con base en su estructura primaria, dominios transmembranales (TMD) potenciales y función metabólica; de éstas, la identificada en *SagA909* que transporta Ca^{2+} pertenece a las ATPasas-P2. Todas las ATPasas-P contienen los motivos: **TGE** que se localiza en un bucle flexible, **DKTG** que se encuentra en el dominio de fosforilación, **TGDN** que es fundamental para la unión al complejo ATP- Mg^{2+} y finalmente el motivo **GDGxND** localizado entre estos dos últimos y que también une ATP; adicionalmente la ATPasa-P2 para Ca^{2+} contiene los aa conservados E³⁰⁹, E⁷⁷¹,

N⁷⁹⁶, **T**⁷⁹⁹, **D**⁸⁰⁰ y **E**⁹⁰⁸ (numeración basada en eucariontes), todos ellos importantes para la unión del catión y su posterior transporte. Las ATPasas-P pertenecen a su vez a la familia de las hidrolasas del tipo deshalogenasas haloácidas (superfamilia HAD) con las que comparten varios aa agrupados en motivos a lo largo de la secuencia, en particular el **D**¹⁰ es invariante y la **T**¹⁴ está altamente conservada en el motivo I, un aa hidroxilo conservado en [**S/T**]¹¹⁸ caracteriza al motivo II y finalmente en el motivo III están conservados los aa **K**¹⁵¹, **S**¹⁷⁶ y **D**¹⁸⁰. Se considera que todas las ATPasas-P surgieron a partir de un ancestro común, que a su vez se habría fusionado con el dominio transmembranal de una proteína HAD ancestral^{1214,1215,1216,1217,1218}.

- o Como podemos observar, varios de los aa que unen porciones del complejo ATP-Mg²⁺, están codificados por tripletes RNY, de modo que cualquiera de ellas pudo haber sido la identificada como ancestral por la metodología descrita en esta tesis; dado que el Mg²⁺ como el ATP, en particular el anillo de adenina, son moléculas que pudieron haber existido prebióticamente, no sorprende que un motivo de aa capaz de unirles tenga su origen en los albores del Mundo de las RNPs.

- o **CANALES IÓNICOS Y FACILITADORES.**

- o Los elementos que se utilizan como cofactores en varias reacciones enzimáticas o que confieren de potencial eléctrico a la membrana de las células se encuentran en su forma ionizada, lo que permite que funcionen como donadores/aceptores de electrones; como hemos mencionado a lo largo del trabajo las reacciones de transferencia de electrones son sumamente ancestrales, es por ello que podemos encontrar que ciertas porciones de algunos facilitadores y canales iónicos son también codificadas por tripletes RNY.

- **Superfamilia del facilitador principal, MFS.** Es una superfamilia altamente diversa de transportadores secundarios, la más grande de este

tipo y en general la más grande después de los sistemas ABC; de hecho al menos el 25% de las MFS se consideran ubicuas y ambos tipos de acarreadores son de las superfamilias más antiguas. Los transportadores MFS son energéticamente frugales y más bien utilizan el potencial electroquímico de las moléculas que transportan a través de la membrana citoplásmica. Dicha superfamilia incluye uniportadores (un único soluto), simportadores (dos solutos, misma dirección) y antiportadores (dos solutos, direcciones opuestas), cuyo número en procariontes correlaciona simultáneamente con el tamaño del genoma de un organismo y la versatilidad nutricional del mismo pues provee de adaptación a diversos nichos ecológicos, mientras que en eucariontes el número de MFS está regulado espacial y temporalmente^{1219,1220,1221}. *Grosso modo* uno de los criterios para clasificar las MFS se basa en cuatro características: donde **W** indica el modo de transporte y mecanismo de acoplamiento de energía, **X** indica la familia o superfamilia, **Y** indica la familia o subfamilia y **Z** indica la especificidad por el sustrato, de tal modo que cada permeasa se identifica con un código **WXYZ**¹²²². Las MFS constan en general de 400 a 600 nt arreglados en 12 hélices- α transmembranales (TMD), aunque hay MFS desde 7 y hasta 24 hélices, conectadas por enlaces hidrófilos formando un poro, mientras los extremos -N y -C quedan en el citoplasma. Tal estructura de los MFS está más conservada que la secuencia, pese a lo cual es posible identificar algunos motivos conservados, la mayoría localizados en la cavidad central de la región alostérica que conecta el sitio de unión al sustrato con el sitio de traslocación de éste^{2019-2022,1223,1224,1225,1226,1227}. Las mitades -N y -C de los MFS son altamente similares entre sí, con motivos repetidos que se conservan equidistantemente a lo de la secuencia, de tal modo tenemos los segmentos **A y A', B y B', C y C', D y D'**. En primera

instancia puede reconocerse el motivo **DRxGRR** entre TMD2 y TMD3, que corresponde al motivo homólogo **G_{x3}[D/E][R/K]xG(x)[R/K][R/K]** entre TMD8 y TMD9, donde el Asp y la Gly interna son importantes para el transporte; adicionalmente está **DKIGRKKIQ** conservado **en B y B'** e **IPETKARYTLEE** **en D y D'**. El motivo conservado entre TMD2 y TMD3 puede considerarse extendido hacia **G[R/K/P/A/T/Y]L[G/A/S][D/N][R/K][F/Y]GR** **[R/K][R/K/P][L/I/V/G/S/T][L/I/M]**, donde los aa entre corchetes están químicamente relacionados y la presencia de uno u otro depende de las familias de MFS analizadas; de hecho la filogenia de los MFS correlaciona con la configuración estructural de las moléculas que transportan y la mayoría de las familias de MFS encuentran su divergencia cerca de la raíz de su árbol filogenético y esto aunado a la alta similitud entre el dominio-N y el dominio-C, sugiere que probablemente los MFS surgieron por varios eventos rápidos de duplicación y fusión previos a la aparición del LUCA; inicialmente habría existido una horquilla primordial de 2 hélices- α , misma que se triplicó y posteriormente la estructura se duplicó hasta formar un MFS genérico de 12 TMD, semejante en estructura a los que conocemos actualmente pero sin la especificidad que muestra cada una de las 74 familias de MFS contemporáneos¹²¹⁹⁻¹²²⁷. Algunos de los aa conservados en los diversos motivos que caracterizan la MFS son codificados por tripletes RNY, sin embargo no puede establecerse con precisión cual se identificó como ancestral por la metodología descrita en la presente tesis, aunque es probable que sea parte de la horquilla que se ha mencionado como primordio de la MFS y siendo así podemos ubicar su origen en los albores del Mundo de las RNPs.

- **Transportador de K⁺, familia Trk.** La acumulación intracelular de potasio en altas concentraciones es ubicua en bacterias y es determinada por las

necesidades osmóticas de la célula, pues mantiene la turgencia intracelular. El sistema Trk es el segundo en importancia para el transporte del potasio, pese a lo cual presenta la mayor especificidad y está expresado constitutivamente; cuatro de las proteínas de la familia Trk son importantes para la importación de K^+ : TrH y TrkG son proteínas transmembranales (TMD) altamente similares entre sí, aunque independientes una de la otra, de casi 5000aa de longitud arreglados en 12 hélices- α , estas proteínas contienen segmentos-P consecutivos que forman motivos llamados MPM repetidos al menos 4 veces, por lo que se considera que surgieron mediante eventos de duplicación, motivos similares se encuentran en otros sistemas de transporte de K^+ y otros cationes; dichas TMD se asocian con dos proteínas periféricas, TrkA que es estrictamente necesaria para el transporte de K^+ , consta de ~50kDa y une NAD^+ mediante el plegamiento tipo Rossmann cuyo motivo conservado es **GxGxxG** y finalmente TrkE que posee los motivos W_A y W_B de unión a ATP, que sin embargo no es hidrolizado y más bien actúa como molécula reguladora del sistema, de hecho la energía que requiere el sistema Trk es proporcionada por la fuerza protón-motriz ($\Delta\mu[H^+]$), por tanto dicha proteína no es requerida más que en presencia de TrkH e incluso puede ser substituida por algún modulo ABC de los sistemas del mismo nombre^{1228, 1229, 1230, 1231, 1232, 1233, 1234, 1235, 1236, 1237, 1238, 1239}. Dado que no se especifica cuál proteína del sistema Trk fue identificada como ancestral, no es posible determinar la porción más antigua de tal sistema en **SagA909**; por otro lado aunque las referencias electrónicas parecen indicar que se trata de TrkH, los trabajos de los que se sepa hasta el momento no permiten precisar cuál es la región codificada por tripletes RNY pues no se provee algún motivo conservado, lo que sí puede asegurarse es que esa región surgió en los albores del Mundo de las RNPs.

o **SISTEMAS PTS.**

o Los sistemas de fosfotransferasa (PTS) son de los complejos moleculares más antiguos que existen, dado que permiten que distintos carbohidratos ($C_xH_{2x}O_x$) sean transportados hacia la célula y éstos pueden utilizarse como fuentes de energía y precursores de compuestos en diversas rutas metabólicas, adicionalmente los PTS participan en la regulación de genes del catabolismo mediante mecanismos de auto-regulación, que no obstante no es una característica antigua. Los PTS comprenden las enzimas llamadas EI y EII, ambas dependientes de Mg^{2+} y la proteína de bajo peso molecular estable al calor (HPr), que funge como intermediario. Por un lado la enzima I (EI), fuera de la membrana, cataliza la transferencia del grupo fosfato del (fosfoenolpiruvato) PEP a la HPr (ahora P-HPr). Por otro lado la enzima II (EII), una proteína integral de membrana, cataliza la transferencia dicho fosfato a grupos de β -glucósidos estructuralmente relacionados, liberando la HPr y trascolando el azúcar fosforilado al interior de la célula; de hecho la EII es necesaria y suficiente para esta transfosforilación vectorial (transformación química de los substratos mientras son transportados de uno al otro lado de la célula) e incluso para fosforilación no vectorial (transformación en que todos los procesos ocurren del mismo lado de la membrana)^{1240,1241,1242,1243,1244,1245,1246,1247,1248,1249,1250}.

o Es notable que los PTS están presentes en la gran mayoría de las bacterias, principalmente en las anaerobias, facultativas u obligadas, pues la utilización de ATP conlleva un alto costo energético con respecto a la energía total generada por dicho metabolismo; de hecho es aparentemente el PEP y no el ATP, la moneda energética principal de tales microorganismos. Por su parte las EII, como se ha mencionado, son el componente asociado a la membrana que determina la especificidad del sistema por determinados tipos de $C_xH_{2x}O_x$ y el transporte de éstos acoplado a la donación energética de un grupo fosforilo; la mayoría de las EII consisten de los dominios EIIA, EIIB y EIIC, que son autónomos y dan un total de ~630aa, éstos

pueden encontrarse libres o fusionados de distintas maneras, en cuyo caso existen conectores flexibles entre cada uno, que pueden ser del tipo repeticiones **PA** (Pro-Ala), del tipo Q que contiene al menos una [Q/**N**/H] entre aa neutrales hidrófilos y neutrales hidrófobos del tipo [**S**/**T**/**A**], o bien del tipo genérico que contiene residuos cargados; el dominio citoplásmico monomérico EIIA consta de 100aa a 160aa, con un residuo His fosforilable por la P-HPr, por lo que se considera el sitio de transferencia de energía; el dominio hidrófilo EIIB de ~100aa, que se puede encontrar de uno u otro lado de la membrana, contiene una Cys conservada a la que puede transferirse el fosfato unido previamente como P-HPr; finalmente el dominio dimérico EIIC de ~350aa, es altamente hidrófobo con 6 a 8 hélices- α y contiene al menos un gran bucle hidrófilo, caracterizado por una His y el motivo [**G**/**N**][**I**/**V**][**T**/**S**/**N**]E que constituye el sitio de unión y transporte del $C_xH_{2x}O_x$ ^{1251,1252,1253,1254,1255,1256,1257}.

o No se han encontrado homólogos de algún módulo de los PTS en eucariontes o arqueas; sin embargo son de los sistemas enzimáticos más abundantes en bacterias, principalmente en las anaerobias como se ha mencionado; dado que además las EII confieren especificidad a los PTS por $C_xH_{2x}O_x$ semejantes, es posible evaluar que en general los sistemas específicos para glucosa (Glc) están presentes casi en la totalidad de las bacterias, mientras que los lugares segundo y tercero en prevalencia lo ocupan los PTS para fructosa/manosa/sorbosa (EII^{Fru/Man/Sor}) y para lactosa/celobiosa (EII^{Lac/Cel}). Al analizar casos específicos se ha encontrado que las bacterias menos complejas pueden no poseer EII^{Glc} pero si contienen EII^{Fru}, por lo que éste se plantea como el PTS antiguo a partir del cual surgieron los demás mediante duplicaciones, fragmentaciones y rearrreglos, lo que explicaría que el grado de homología entre PTS no está determinado por la divergencia filogenética de los organismos que los contienen, sino que coincide con la especificidad a distintos tipos de $C_xH_{2x}O_x$ ^{1258,1259,1260,1261,1262,1263,1264,1265,1266,1267}; incluso se ha

planteado que todas las permeasas pudieron haber surgido a partir de un poro rudimentario, constituido únicamente por segmentos consecutivos de aa no polares insertados en la membrana primitiva mediante interacciones hidrófobas¹²⁶⁸. Un planteamiento evolutivo más reciente sugiere que los PTS surgieron como sistemas completos posterior a la divergencia desde LUCA, precisamente debido a que estos sólo se encuentran en bacterias, aunque si es posible que los diferentes módulos hayan evolucionado previamente de manera independiente^{1269,1270}.

o De las proteínas EII identificadas como ancestrales por la metodología presentada en esta tesis no es posible decidir con precisión cuál fue la región identificada por tripletes RNY en cada una, pues de EIIA^{Lac/Cel} no se ha descrito algún motivo conservado de aa codificados por tripletes RNY, mientras que el bucle hidrófilo de EIIC^{Fru/Man/Sor} contiene aa codificados por tripletes RNY aunque el motivo hidrófobo también pudiera contener aa de este tipo. Ciertamente los PTS para Fru/Man/Sor y Lac/Cel son los más abundantes actualmente y posiblemente de los más antiguos; lo que pudiera implicar que este tipo de azúcares eran los más abundantes en un ambiente primigenio y por ende las fuentes primarias de energía preferidas por las células, así como los precursores de otros varios compuestos para la formación de diversas rutas metabólicas, lo que también podría verse reflejado en que son de los PTS más abundantes en general; podemos sugerir entonces que una de las regiones en cada una de las dos EII identificadas surgieron en los albores del Mundo de las RNPs y por ende son los primordios para la subsecuente formación de los PTS, aunque su ensamblaje como parte de un sistema surgiría hasta que el linaje bacteriano comenzara a definirse.

Como se ha mencionado, con base en análisis filogenéticos de la función y estructura, se propone por un lado que todos los transportadores surgieron a partir de un poro rudimentario compuesto de segmentos polares conformados en hélices- α

insertadas en la primitiva membrana y conectadas por bucles hidrófilos; estos poros evolucionaron hacia distintos tipos de canales facilitadores, posteriormente acarreadores primarios y luego secundarios, mediante duplicaciones y mutaciones que confirieron paulatinamente especificidad a las proteínas mencionadas. Por otro lado los dominios que proveen energía en los distintos tipos de transportadores comparten cierta homología entre sí e inclusive con dominios de unión a moléculas fosforiladas en otro tipo de sistemas como los de transducción, por lo que se deduce que todos ellos comparten un ancestro común; finalmente ocurrieron rearrreglos y fusiones entre esos dominios, permeasas y energizadores, hasta conformar las distintas familias de transportadores que existen en la actualidad. De tal manera que antes del LUCA surgieron posiblemente la superfamilia de MIP, la MFS, transportadores ABC y otros dependientes de moléculas fosforiladas, mientras que los PTS habrían surgido próximos al LUCA, aunque algunos de sus primordios son más antiguos; finalmente los transportadores siguieron evolucionando, mayoritariamente como ortólogos en procariontes y como genes parálogos en eucariontes^{1268,1271,1272,1273,1274,1275,1276}. De manera notable los transportadores identificados como ancestrales corresponden a aquellos cuyo origen se ha indicado previo al LUCA y, con base en los resultados obtenidos mediante la metodología descrita en la presente tesis, podemos indicar ahora que su origen se sitúa en los albores del Mundo de las RNPs.

T15. Características plausibles de la probable región codificada por tripletes RNY en los canales y transportadores identificados.

Conjunto		Proteína (nombre traducido y adaptado)	Características plausibles de la probable región codificada por tripletes RNY.	
PASIVOS	SUPERFAMILIA MIP	Facilitador de glicerol (GlpF) / acuaporina (AQP)	Canal para el soluto	
ACTIVOS	PRIMARIOS			
	ABC	PBP	Proteína de unión al sustrato; putativa Familia PAAT, proteína de unión al aa Family PepT, proteína de unión al sustrato	Posiblemente preservación de estructura 'Venus atrapamoscas' Preservación de estructura 'Venus atrapamoscas', o quizá unión de aa polares Preservación de 'Venus atrapamoscas', o quizá unión de cofactores metálicos
		TMD	Proteína permeasa; putativa	Parte de una hélice hidrófoba, replicada, embebida en la membrana
			Transportador de glicina betaína/prolina; TMD/PBP	Parte de hélice hidrófoba, o quizá contexto con EAA, o quizá unión de Gly-bet/Pro
			Proteína permeasa CylB	Parte de una hélice hidrófoba embebida en la membrana, o quizá unión de citolisina
	Captación quelante de hierro, proteína permeasa		Parte de una hélice hidrófoba, o quizá contexto con EAA, o quizá unión de Fe	
	ABC	Captación de carbohidratos, familia CUT1/CUT2, permeasa	Parte de hélice hidrófoba, o quizá contexto con EAA, o quizá unión de carbohidratos	
		Familia PAAT, proteína permeasa	Parte de una hélice hidrófoba embebida en la membrana, o quizá unión de aa polares	
		Proteína de la familia BioY; captación de biotina	Parte de una hélice hidrófoba embebida en la membrana, o quizá unión de biotina	
		Espermidina/putrescina, proteína permeasa PotB	Parte de hélice hidrófoba, o quizá contexto con EAA, o quizá unión de poliaminas	
ABC	Transportador de níquel, proteína permeasa	Parte de una hélice hidrófoba, o quizá contexto con EAA, o quizá unión de Ni		
	Familia QAT, proteína permeasa	Parte de hélice hidrófoba en la membrana, o quizá unión de aminas cuaternarias		
ATPasa-P	ATPasa-P, transportadora de Ca ²⁺ , HAD IC	Unión de ATP, o Mg ²⁺ , o ATP-Mg ²⁺		
SECUNDARIOS	Transportador MFS Proteína de la familia de transportadores de K ⁺ , Trk	Parte de una hélice- α embebida en la membrana; aa importantes para el transporte Quizá unión de K ⁺ , o quizá parte de hélice hidrófoba, o quizá unión de NAD ⁺		
PTS	Componente IIA, familia de lactosa/celobiosa	Unión de Mg ²⁺ , o quizá aa hidrófilos		
	Componente IIC, familia de manosa/fructosa/sorbosa	Unión de Mg ²⁺ , o quizá bucle hidrófilo, o quizá unión de Man/Fru/Sor		

6.3.1.7 Transducción.

Las redes de transducción celular sirven para monitorear parámetros ambientales e intracelulares y modificar la expresión de genes en respuesta a tales cambios. El flujo de información en tales circuitos va desde el dominio N-terminal de las proteínas sensoras (input), generalmente asociadas o embebidas a la membrana celular, hacia el dominio C-terminal de las proteínas efectoras (output), que se unirán al DNA para regular la transcripción de genes. Las proteínas involucradas en la transducción celular son modulares y sus dominios pueden subdividirse en: sensoriales o input (donde se une el ligando y se asocian generalmente con la membrana), de transducción (donde un grupo fosforilo provoca cambios conformacionales como homo-oligomerización) y de salida de información u output (que usualmente se hetero-oligomerizan y son hélices de unión al DNA). En los circuitos más sencillos de transducción, cada dominio se encuentra en una proteína distinta; mientras que en los circuitos más complejos, una misma proteína contiene todos los dominios (input, transmisor y output) a lo largo de su secuencia¹²⁷⁷.

La transducción celular se basa principalmente en la fosforilación y desfosforilación regulada de proteínas. El fosforilo ($[\text{PO}_4]^-$) proviene de la hidrólisis del grupo γ terminal del ATP; la transferencia de ese grupo es catalizada por cinasas, mientras que la remoción hidrolítica del mismo es catalizada por fosfatasas. Ambos tipos de enzima son específicos para transferir o remover (dependiendo si son cinasas o fosfatasas) el $[\text{PO}_4]^-$ de residuos de Ser, Thr, Tyr o His de las proteínas que serán reguladas, aunque la mayoría de las cinasas de His contienen la función de fosfatasa en la misma enzima^{1278,1279}.

Tres enzimas implicadas en la transducción contienen regiones codificadas por tripletes RNY (en T16 y T17): una enzima de la familia cinasa/fosfatasa de His, así como una cinasa de Ser/Thr y una fosfatasa de Ser/Thr.

Order	Sc.	E-v	ID	%	PROTEIN / FAMILY / DOMAIN
1	24	0.004	24/24	100%	sensor histidine kinase (20.00 %)
2	24	0.004	24/24	100%	serine/threonine protein kinase Stk1 (80.97 %)
3	24	0.004	24/24	100%	Ser/Thr protein phosphatase family protein (94.53 %)

T16. Proteínas, involucradas en la transducción, que contienen regiones codificadas por tripletes RNY, nomenclatura y código de colores igual que T1 y T2.

o **CINASAS.**

o Las cinasas son un grupo ubicuo de enzimas (existen en virus y en los tres grandes dominios de la vida)¹²⁸⁰ que catalizan la reacción de transferencia del grupo $[PO_4]^-$ del ATP a algún sustrato receptor¹²⁸¹.

o El resultado de una inspección en 2002 de todas las cinasas disponibles hasta entonces (>17,000) fue que éstas podían agruparse en 30 familias distintas con base en similitud de secuencias, 19 de esas familias (~98% del total) se pueden agrupar a su vez en siete categorías con base en su plegamiento estructural¹²⁷⁵.

o La relación filogenética entre las cinasas de Ser, Thr y Tyr es muy estrecha, mientras que todas ellas están sólo distantemente relacionadas con las cinasas de His a pesar de las numerosas similitudes en el mecanismo catalítico¹²⁷¹. La comparación de varias estructuras que se han resuelto, ha revelado la existencia de un dominio esencial universal a todas ellas, que consiste únicamente de las regiones requeridas para la unión del ATP y para la reacción de fosfotransferencia. El dominio esencial catalítico de la cinasas consiste de un subdominio N-terminal pequeño de usualmente una hoja- β y un subdominio C-terminal más grande de usualmente una hélice- α , mientras que el sitio de unión al ATP es un bolsillo que se forma entre ambos subdominios que pueden formar una estructura abierta o cerrada, los residuos involucrados en la reacción de fosfotransferencia se localizan en el borde de éste sitio de unión y están altamente conservados¹²⁸². Fuera del dominio esencial universal, las cinasas poseen una variación de secuencias considerable¹²⁷⁴, lo que incide directamente en su especificidad a los sustratos¹²⁷⁶.

- o La **cinasa sensora de histidina** (HK, por sus siglas en inglés), corresponde a uno de los módulos del sistema bacteriano llamado de dos componentes, mismo que se considera la modalidad más sencilla de transducción celular, el sistema usual en bacterias y por tanto es el mejor caracterizado. Las **'proteínas de la familia de la HK'** pueden denominarse **'proteínas HK'**, **'familia HK'** o simplemente **'HKs'** en plural¹²⁸³.
- o En el sistema de dos componentes sólo existe la molécula cinasa que monitorea los parámetros y es fosforilada y la molécula reguladora que recibe el $[\text{PO}_4]^-$ y se une al DNA. El mecanismo general de las proteínas de la familia HK consiste en: 1) percepción del estímulo, 2) homodimerización, 3) unión e hidrólisis del ATP en el sitio catalítico, 4) autofosforilación en el residuo His, 5) transferencia del estímulo al segundo componente del sistema de señalización y 6) desfosforilación del regulador¹²⁸⁴; durante el ciclo catalítico ocurren diversos cambios en la conformación de una HK¹²⁸⁵. El sitio sensor está localizado usualmente en el dominio N-terminal, presenta alta variabilidad y está asociado a la membrana, ya sea embebido o solamente acoplado; el estímulo se transmite entonces a la parte C-terminal de los dominios, donde un monómero es fosforilado, se transfiere el $[\text{PO}_4]^-$ a una His en el otro monómero y finalmente este grupo se transfiere un regulador (o segundo componente), que se unirá al DNA; en la mayoría de los casos el regulador es desfosforilado por la misma HK¹²⁸⁶.
- o Existen diferentes subclases dentro de la familia de la HK que monitorean distintos parámetros ambientales; estas subclases han evolucionado independientemente una de la otra, pero manteniendo un dominio esencial altamente conservado¹²⁸⁷ que define a la familia completa. Dicha región se localiza cerca del dominio C-terminal, dentro del cual hay seis residuos especialmente conservados: **N**...(15 a 45 aa)...**DXGXG**...(20 a 50 aa)...**GXG**

(regiones II y III). Adicionalmente, varias cinasas poseen una His que precede por 100 residuos a la Asn conservada, dicha His es el sitio de autotransfosforilación de la enzima¹²⁸⁸.

- o Los residuos conservados mencionados corresponden a cinco regiones de similitud o cajas que se presentan usualmente de forma continua aunque espaciada a largo de ~200 aa del longitud, definen el sector esencial de las HK y son nombradas por el aa más abundante en cada una: el residuo His, que confiere su nombre a la caja H, se autotransfosforila utilizando Mg^{2+} como cofactor y genera fosfoamidatos, dicha autotransfosforilación ocurre entre homodímeros; siguen las cajas N, G1, F y G2 que conforman la hendidura de unión a la parte nucleotídica del ATP, donde se llevan a cabo las actividades de cinasa/fosfatasa, *i.e.* es tanto el módulo de unión de ATP/ADP como de fosfotransferencia. Varias HKs carecen de la caja H pero contienen las otras 4 y una His no conservada en el dominio N-terminal, por lo que se les considera de la misma familia. Algunos renombraron las cajas **N**, **G1**, F y **G2** como **N**, **D**, F y **G** respectivamente, para evitar confusiones y respetando los aa conservados universalmente en las HKs. El Asp de la caja D se une directamente al anillo de adenina del ATP mediante puentes de hidrógeno, a este anillo se unen también a los aa de la caja N y otros de la caja D mediante dos moléculas de agua con las que forman también puentes de hidrógeno; adicionalmente el Mg^{2+} sirve como enlace entre los fosfatos del ATP y los aa de la caja N. La caja F es parte del dominio que envuelve al ATP, mientras que la caja G forma la bisagra del extremo de ese dominio; la bisagra es tan flexible que funge como unión entre el dominio catalítico y el de dimerización, lo que resulta crítico para los rearrreglos estructurales durante el ciclo catalítico, en el cual la enzima fluctúa entre la actividad como cinasa y la actividad como fosfatasa^{1289,1290,1291,1292,1293,1294,1295,1296,1297}. Según un estudio

llevado a cabo en *Streptococcus pneumoniae* se ha definido la caja E con base en comparación de secuencias, ésta se localiza después de las cinco que se habían caracterizado previamente, no obstante sólo está presente en algunas de las cinasas estudiadas en dicho organismo¹²⁹⁸.

- o Dentro de las proteínas HK existen la clase I y la clase II, cada clase contiene el dominio catalítico de unión a ATP (CA) pero difieren en el dominio que contiene el residuo His que será fosforilado. La clase II contiene el dominio llamado de fosfotransferencia que contiene histidina (HPT) antes de un módulo regulador, luego el módulo CA y al final otro regulador. La clase I contiene primero un módulo sensor y luego el dominio de histidina, dimerización y fosfotransferencia (DHP), que es el mismo CA en su versión más sencilla. Adicionalmente algunas HKs consideradas híbridas (HKH) contienen además un dominio receptor y al final un dominio HPT adicional, cuya His puede o no fosforilarse^{1299,1300}; estas HKH están prácticamente ausentes en procariontes, aunque las arqueas poseen un poco más que las bacterias, no obstante comprenden el tipo más abundante de HK en eucariontes, principalmente en hongos, amibas y plantas terrestres, aunque no en metazoarios¹³⁰¹.
- o El dominio catalítico o módulo transmisor comparte claramente homología en secuencia y estructura con los dominios de unión a ATP de otras enzimas, de hecho las HK se han agrupado junto con las girasas, las chaperonas moleculares Hsp90 y las enzimas del tipo MutL en una superfamilia denominada GHKL, que a pesar de compartir poca homología entre ellas las estructuras de sus respectivos dominios de unión a ATP son esencialmente superponibles. La familia GHKL tiene cuatro dominios conservados: motivo I [uubEuuu**NouDA**], motivo II [uxux**DNGxG**uxbaauxxu], motivo III [u**GxxG**xoux**S**xxxuoxbu**TuxT**] y motivo IV [**T**X_n**GT**] (u=aa voluminosos hidrófobos, o=aa pequeños, b=aa básicos, a=aa ácidos). Los primeros tres

motivos corresponden a las cajas N, G1 y G2 de las HK, adicionalmente se ha propuesto nombrar al motivo IV como G3; en la familia GHKL los cuatro motivos definen el plegamiento de unión a ATP denominado 'Bergerat' en honor de quien lo describió por vez primera¹³⁰².

- o Las proteínas HK, además de las cinco cajas conservadas, contienen un dominio flexible denominado HAMP (por las siglas en inglés de: cinasas de histidina, ciclasas de adenina, proteínasceptoras de metilo y fosfatasas, con 'ph' en vez de 'f'), el cual conecta la región sensorial de la membrana con el dominio esencial y permite los distintos cambios en la conformación de la proteína durante el ciclo catalítico^{1303,1304}.
- o Además del módulo CA, el conector HAMP y el HPT, se han descrito otros dominios en las HK: **PAS**, **GAF**, **GGDEF**, **EAL** y **HD-GYP**, que se nombran por los aa más conservados en cada uno¹³⁰⁵, estos dominios están pobremente caracterizados aunque se han predicho algunas de sus funciones y se sabe que distintas combinaciones de ellos originan distintas redes de regulación¹³⁰⁶.
- o El análisis genómico de las secuencias de 348 HKs de procariontes en 2001, reveló que éstas pueden agruparse en 5 distintos tipos y 11 subtipos cuya distribución depende del tipo de microorganismo y aunque un mismo organismo puede contener los cinco tipos de HK en general se observó que: a) muchas arqueas carecen de HK o contienen de un solo tipo, probablemente adquiridas mediante HGT y b) todas las bacterias contienen HK, exceptuando endosimbiontes y *Mycoplasmas*, los patógenos contienen pocas, los microorganismos de genomas grandes contienen más y los de vida libre son los que mayor número de HKs contienen¹³⁰⁷. En 2006 y 2010 se describió el mismo patrón en la distribución de las HKs con mayor número de genomas secuenciados y se consideró entonces que un mayor número de tales enzimas permite a los microorganismos adaptarse a un rango más amplio de

condiciones ambientales, pues aunque la mayoría de las arqueas son extremófilas su ambiente no cambia significativamente como sí suele ocurrir con bacterias de vida libre^{1308,1309}.

- o De acuerdo a un trabajo sobre la filogenia de los sistemas de dos componentes, ambos módulos (el sensor y el regulador) coevolucionaron primero en bacterias y posteriormente se adquirieron en pares por arqueas y eucariontes¹³¹⁰.
- o **En términos prácticos, cualquiera de las 'cajas' distintivas de las HKs pudo haber tenido su origen en los albores del Mundo de las RNPs, pues varios de los aa conservados en dichas enzimas son codificados por tripletes RNY y la gran mayoría, con excepción del motivo involucrado directamente en la catálisis pues la His no es un aa de ese tipo, convergen en la unión de porciones del ATP/ADP, del que al menos la porción adenosil se considera prebiótica.**
- o La **proteína cinasa Stk1 de serina/treonina** pertenece a la superfamilia de las Ser/Thr cinasas (STKs), cuyo nombre sistemático es ATP: fosfotransferasas de proteínas; dichas cinasas intervienen en la modificación covalente de un sinnúmero de proteínas ubicuas, pues catalizan la transferencia del fosfato γ del ATP a algún residuo Ser o Thr de esas proteínas receptoras creando fosfoésteres; durante el proceso las catalítico STKs utilizan Mg^{2+} como cofactor, Ca^{2+} en algunas ocasiones y liberan ADP^{1274,1275,1276,1285,1311}.
- o Se considera que la fosforilación de proteínas puede ser un mecanismo regulatorio antiguo presente incluso en los organismos celulares primitivos; en bacterias las proteínas fosforiladas por STKs están involucradas en la fosforilación de proteínas del metabolismo central y de mantenimiento o

'housekeeping', así como de proteínas que regulan las respuestas adaptativas al ambiente y la patogenicidad del microorganismo^{1312,1313,1314}.

- o De manera particular, la Ser/Thr cinasa denominada Stk1 está conservada universalmente en bacterias y, de acuerdo con la metodología descrita en la presente tesis, contiene una región codificada por tripletes RNY; dicha STK se encuentra asociada a la membrana celular y fosforila una pirofosfatasa inorgánica, una sintetasa de adenilosuccinato, un regulador de respuesta denominado CovR y al menos otras 7 u 8 proteínas distintas, todas ellas involucradas en el crecimiento y/o morfología y/o segregación celular y/o patogenicidad del microorganismo; algunos estudios sugieren que la Stk1 es de hecho la única Ser/Thr cinasa encontrada en el grupo B de estreptococos (GBS), al cual pertenece **SagA909**^{1315,1316,1317,1318,1319,1320}.
- o De manera general la fosforilación de proteínas por STKs es un factor clave en la transducción de diversas señales intra- e inter-celulares, es importante notar que el número de STKs es directamente proporcional con el tamaño del genoma, complejidad e independencia del organismo, de modo que los simbioses y organismos parásitos no poseen STK alguna, aunque en general hay STKs en eucariontes y arqueas, recientemente se ha caracterizado su presencia en bacterias y se han encontrado en algunos virus. Las STKs son usualmente tetrámeros de ~130kDa, cuyo monómero es generalmente bilobulado a modo de bisagra, que cambia de estado abierto a cerrado dependiendo del proceso catalítico y en cuya hendidura se localiza el sitio activo. En el sitio activo están localizados la gran mayoría de los aa conservados, involucrados en la unión del complejo ATP-Mg²⁺ justo en la base de la hendidura catalítica y en la fosfotransferencia; en particular el bucle **GxGxxGxV**, altamente conservado, ubicado entre la hebra-β1 y la hebra β-2 y que es típico de enzimas que unen nucleótidos, sirve de anclaje a los grupos

fosfato no transferibles, mientras el Mg^{2+} se une entre el fosfato α y γ del ATP para crear el complejo ATP- Mg^{2+} ; algunos aa como L y **V** rodean el bucle de Gly para formar un bolsillo hidrófobo que secuestra el anillo de adenina del ATP; adicional al bucle de Gly también están conservados dos Asp y una Asn que son: **D**¹⁸⁴, **D**¹⁶⁶ y **N**¹⁷¹ (con base en una STK de eucariontes), éstos interaccionan también con partes del complejo ATP- Mg^{2+} , siendo invariante en las STKs el primer **Asp**. Por otro lado cada subgrupo funcional de las STKs es determinado por los dominios con que está asociado el módulo esencial, lo que implica que mientras las STKs comparten de 20% a 30% de identidad con las cinasas generales y éste es un porcentaje mayor que el obtenido en las comparaciones dentro de cada subgrupo, el dominio medular de todas las STKs está marcadamente conservado y consta de ~270aa dispuestos en un dominio-N pequeño compuesto por hebras- β en su mayoría y un dominio-C grande de hélices- α mayoritariamente. Tomando en cuenta los datos obtenidos se deduce que todas las STKs provienen de un ancestro común cuyo origen se remonta al menos hasta LUCA^{1321,1322,1323,1324,1325,1326,1327,1328,1329,1330}.

- o Todos los aa característicos de las STKs unen distintas porciones del complejo ATP- Mg^{2+} , que a su vez contiene varios componentes de origen posiblemente prebiótico; varios de esos aa son codificados por tripletes RNY, aunque la región más extensa con aa ancestrales es el bucle abundante en **Gly** y por ello es probable que su origen de remonte a los albores del Mundo de las RNPs.

- o **FOSFATASAS.**

- o La transducción celular se basa en la fosforilación y des-fosforilación regulada de proteínas, mecanismo regulatorio que se considera tan antiguo que incluso los organismos celulares primitivos pudieron ya poseerlo. Las fosfatasas catalizan la

hidrólisis de los enlaces fosfoésteres de residuos Ser/Thr/Tyr de un sinnúmero de proteínas ubicuas (las fosfatasa de His están presentes en la misma proteína que su cinasa, como se indicó previamente en la sección 'cinasa')^{1285,1306,1307}.

- o **Proteína de la familia fosfatasa de serina/treonina.** Debido al proceso $\text{Ser/P-Thr} + \text{H}_2\text{O} \rightarrow \text{Ser/Thr} + \text{Pi}$ que catalizan las fosfatasa, su denominación formal es fosfoesterasas de Ser/Thr (STPs). En bacterias las proteínas reguladas STPs son del metabolismo central y de mantenimiento o 'housekeeping', aunque hace relativamente poco tiempo se descubrió que las STPs, al igual que su contraparte catalítica (STKs), son ubicuas en los tres dominios de la vida y en algunos virus^{1285,1306,1307,1331}. De la superfamilia de las fosfoesterasas se han descrito sus características generales, de varias se ha obtenido su secuencia y la estructura de algunas, de esas observaciones se desprende que las STPs comparten un módulo común con motivos y topología altamente conservados.
- o Las STPs pertenecen a una superfamilia de fosfohidrolasas asociadas a membrana denominada DDDD por la presencia de 4 Asp estrictamente conservados, en los motivos $[\text{F/I/L/M/A/T/V}]\text{-D-}[\text{I/L/F/R/M/V/Y}]\text{-D-}[\text{G/S/N/D/E}]\text{-}[\text{T/V}]\text{-}[\text{I/L/V/A/M}]\text{-}[\text{A/T/S/V/I/L/M/C}]\text{-x-}[\text{Y/F/W/H/K/R}]\text{-x-}[\text{Y/F/W/H/N/Q}]\text{-x}_{102-191}\text{-}[\text{K/R/H/N/Q-G-D}]\text{-}[\text{F/Y/W/H/I/L/V/M/C}]\text{-}[\text{Q/N/H}]\text{-}[\text{F/W/Y/G/P}]\text{-D-}[\text{P/S/N/Q/Y/W}]$, a lo largo de la secuencia⁴⁷⁹.
- o El dominio fosfoesterasa por su parte está relacionado con algunas polimerasas de DNA y otras enzimas de involucradas en el procesamiento de ácidos nucleicos con las que comparte cuatro motivos; entre los que destaca una díada Hx_4D en el dominio-N e involucrada en la catálisis y en la coordinación de dos cationes metálicos (1° y 2°M^+) que típicamente son Zn^{2+} o Mn^{2+} o Ni^{2+} o Co^{2+} y Fe^{3+} , en la coordinación de dichos cationes metálicos convergen otros residuos His, Asp y Asn, como **D**⁸, E³⁷, **N**³⁸, **N**⁶⁵, H¹⁴⁸, H¹⁷³ e

H¹⁷⁵. De manera particular las fosfoesterasas contienen dos motivos conservados **DxHx₋₂₅GDX₂DX₋₂₅GNH[D/E]** y **GHxc₋₅₀GHx[H/x]**, también el mecanismo de hidrólisis del enlace éster con el fosfato es el mismo en todas ellas e involucra la interacción del substrato con el 1°**M**⁺ seguida por un ataque nucleófilo hacia el fósforo por un ión hidróxido coordinado con el 2°**M**⁺. Debido a los metales involucrados, las fosfatasa se consideran metalofosfoesterasas de núcleo dimetálico, cuya topología consiste de un plegamiento **βαββ** hacia el dominio-C en que los metales se posicionan en los bucles o hélices. Con base en las características descritas se considera que todas las fosfatasa se derivan de un proteína común que utilizaba eficientemente el mecanismo hidrolítico, la cual se modificó y comenzó a divergir en el curso de la evolución hacia enzimas distintas, con especificidad por substratos particulares con base en las necesidades de adaptación de cada grupo de organismos a su ambiente^{1332,1333,1334,1335,1336,1337,1338,1339,1340}.

- o Varios de los aa conservados en las STPs son codificados por tripletes RNY, tanto en la zona que las identifica como parte de la superfamilia DDDD, como en la que indica que son metalofosfoesterasas, dado que los aa conservados en esta última unen cationes metálicos que fácilmente habrían existido desde un ambiente prebiótico, es posible que haya sido la región identificada como ancestral; aunque en realidad no es posible discernir aún, bajo la metodología presentada en esta tesis, cuál de las dos regiones es la que surgió en los albores del Mundo de las RNPs.

T17. Características plausibles de la probable región codificada por tripletes RNY en las proteínas identificadas como componentes del sistema de transducción celular.

Conjunto	Proteína (nombre traducido y adaptado)	Características plausibles de la probable región codificada por tripletes RNY.
Cinasas	Cinasa sensora de histidina	Unión de ATP
	Proteína cinasa de serina/treonina, Stk1	Unión de ATP, o Mg ²⁺ , o ATP-Mg ²⁺
Fosfatasas	Proteína de la familia fosfatasas de Ser/Thr	Quizá unión de cationes metálicos.

6.3.1.8 Resistencia y Supervivencia.

Dentro de esta categoría he considerado aquellas enzimas que actualmente permiten al organismo vivir frente a condiciones adversas del medio extracelular; en este caso (en T18 y T19) una transferasa que posee varias repeticiones internas y cuya reacción catalizada provoca que varios agentes dañinos sean desactivados, una enzima que confiere resistencia frente a una condición particular del medio ambiente y una enzima para la síntesis de un antimicrobiano.

Order	Sc.	E-v	ID	%	PROTEIN / FAMILY / DOMAIN
6	31	3.00E-07	37/39	94%	transferase, hexapeptide repeat family (59.40 %)
3	24	0.004	24/24	100%	tellurite resistance protein TehB, putative (8.00 %)
5	23	0.015	23/23	100%	phenazine biosynthesis protein, PhzF family (57.77 %)

T18. Proteínas involucradas en la resistencia o supervivencia de *SagA909*, que contienen regiones codificadas por tripletes RNY, nomenclatura y código de colores igual que T1 y T2.

o TRANSFERASA DE LA FAMILIA DE REPETICIÓN DE HEXAPÉPTIDOS.

o Las proteínas que contienen patrones repetidos de manera regular son poco usuales, una de esas familias son aciltransferasas (AT) dependientes de coenzima A (CoA), que poseen numerosos pares repetidos de hexapéptidos (Hex) altamente conservados (HexAT), del tipo $x[\mathbf{S/T/A/V}]x[\mathbf{L/I/V}][\mathbf{G/A/E/D}]x$ o bien $[\mathbf{L/I/V}][\mathbf{G/A/E/D}]xx[\mathbf{S/T/A/V}]x$, que se encuentran principalmente en microorganismos y plantas. Las enzimas HexAT son usualmente trímeros en su forma activa y adoptan una conformación estructural de hélices- β paralelas con giro a la izquierda ($L\beta H$), misma que depende intrínsecamente del patrón Hex repetido, por tanto es posible utilizar indistintamente 'HexAT' y ' $L\beta H$ '. Cada monómero contiene sitios de unión al sustrato y a la CoA, de la que se transfiere la parte fosfopanteína sin intermediarios enzimáticos a diversos aceptores susceptibles, como aa, azúcares, metabolitos diversos o agentes exógenos que contengan iones hidroxilo o grupos amino libres, que bien pueden ser antibióticos que son inactivados como resultado del proceso catalítico^{1341,1342,1343,1344}. Cada resaltar que

los modelos usuales de evolución de proteínas no toman en cuenta explícitamente la estructura, pese que diversos estudios sustentan que ésta en muchos casos está más conservada que la secuencia; para suplir tal deficiencia metodológica se desarrolló un método específico y se aplicó a proteínas con patrones de aa repetidos, entre ellas HexAT¹³⁴⁵.

- o Dado que todos los aa conservados en el fragmento repetido de las HexAT son codificados por tripletes RNY, ésta fue probablemente la región identificada como ancestral, dado que la repetición de dicho fragmento conduce directamente a la evolución del plegamiento L β H, tal familia pudiera significar un botón de muestra de que para algunos péptidos la conformación estructural es la primera característica en surgir y evolucionar, no así la capacidad catalítica que fue posterior.

- o **PROTEÍNA DE RESISTENCIA A LA TELURITA.**

- o El telurio (Te) es un elemento traza no esencial, es miembro del grupo 16 de los chalcógenos (formadores de minerales), es rara vez encontrado en su forma elemental no tóxica Te⁰ y aunque los derivados orgánicos e inorgánicos del Te son relativamente raros en el ambiente éstos son muy tóxicos para los organismos, en particular los oxianiones de telurio son relativamente los más estables pero también las formas más tóxicas para la mayoría de las formas de vida^{1346,1347,1348}.

- o Algunos organismos pueden procesar el telurio no esencial; como forma de defensa algunas bacterias y hongos pueden metabolizar las sales de telurio mediante su metabolismo reductivo, originando Telurio inerte (Te⁰), lo que produce depósitos negros de telurio metálico o telurolato (Te²⁻) dentro de la célula, el cual se metila hasta el compuesto volátil del olor a ajo ((H₃C)₂Te)¹³⁴⁶⁻¹³⁴⁸.

- o De los oxianiones de telurio, la sal denominada telurita (TeO₃²⁻) es la forma más hidrosoluble y altamente tóxica para las bacterias incluso a concentraciones de 1 μ g/mL. Actualmente se desconoce la bioquímica precisa de la toxicidad de los

oxianiones del telurio, pero se ha asumido que la toxicidad de la telurita es consecuencia de las fuertes propiedades oxidantes que llevan a la formación de letales especies reactivas de oxígeno (ROS), lo que puede cambiar el potencial RedOx intracelular y desregular la expresión de muchos genes e incluso provocar mutaciones deletéreas; una propuesta alternativa sugiere que la telurita podría reemplazar al azufre en varios procesos celulares¹³⁴⁶⁻¹³⁴⁸.

- o Dado que los compuestos derivados del telurio son relativamente raros en el ambiente no es de esperar que las bacterias tengan contacto con ellos, sin embargo poseen varios determinantes de resistencia que les permiten sobrevivir en presencia de la sal telurita (TeO_3^{2-}); por ejemplo la transforman a una forma menos tóxica mediante distintos mecanismos, ya sea procesada mediante interacciones con tioles celulares como el glutatión, o mediante reacciones catalizadas por metiltransferasas (MTs) dependientes de S-AdoMet¹³⁴⁶⁻¹³⁴⁸.

- o En diversas especies bacterianas se han identificado cinco determinantes de resistencia a la telurita no relacionados entre sí, de los cuales sólo uno es codificado en el cromosoma y los otros cuatro en plásmidos. Esos determinantes median la resistencia a la telurita por un mecanismo aún desconocido; inclusive se ha especulado que la resistencia a la telurita puede no ser la acción principal de ninguno de esos sistemas, sino un efecto secundario de su funcionamiento normal¹³⁴⁶⁻¹³⁴⁸.

- o El determinante del cromosoma es el operón **tehAB**, encontrado en bacterias G+ y G-¹³⁴⁸; el operón **tehAB** se localiza cerca del extremo **ter** del cromosoma y requiere de sobreexpresión para mediar la resistencia a la telurita¹³⁴⁹. TehB es una proteína dimérica soluble de 23kDa que se asocia débilmente con la membrana¹³⁵⁰.

- o Debido a la amplia distribución de homólogos de TehB se ha sugerido que la función de esta proteína podría no ser exclusivamente la resistencia a la telurita, sino que podría ser importante en general para la supervivencia de los organismos

que la poseen^{1348,1351}, por ejemplo se ha mostrado que la proteína TehB es muy importante en la resistencia al daño oxidativo en general¹³⁴⁶⁻¹³⁴⁸ y en algunos organismos es importante para la virulencia¹³⁵².

- o Los homólogos de TehB constan de ~300 aa y comparten del 46% al 58% de identidad y del 52% al 68% de similitud. La proteína TehB contiene tres motivos conservados (I, II y III), característicos de las S-AdoMet MT no dependientes de ácidos nucleicos incluyendo el motivo de unión a S-AdoMet, estudios adicionales de mutagénesis han indicado que la TehB puede también ser una MT^{1348,1351}; dado que TehB puede ser una MT, esta une tanto telurita como S-AdoMet y cualquiera de estas uniones produce cambios en la conformación de la proteína, los cuales parecen involucrar tres Cys conservadas que coordinarían el átomo de Te aunque no estarían involucradas directamente en el mecanismo catalítico de TehB¹³⁵⁰.

- o Los motivos de las S-AdoMet MT se caracterizan por: el motivo I es una región rica en **G** con la secuencia consenso h[**D/E**]h**GxG** (donde h es cualquier residuo hidrófobo); al motivo I le sigue generalmente un residuo ácido como **D** en la parte C-terminal, en particular en TehB hay 16 aa antes del **D**; adicionalmente hay 34 residuos entre el motivo I y el motivo II; el motivo II consiste de 8 aa, la mayoría de ellos aromáticos (Y, W y F) alrededor de un **D** invariante; el motivo III se localiza a 22 residuos del motivo II y consiste de un bloque de 9 aa cuya secuencia consenso es L[R/K]P**GG**[R/**I**/J][L/**I**][L/F/**I**][**I**/L], aunque varios estudios han sugerido que el motivo III no es esencial para la actividad catalítica de las S-AdoMet MT. Dado que es probable que TehB sea una MT, es lógico inferir que existe un mecanismo de desintoxicación que explique la resistencia a la telurita, ya que la modificación del oxianión telurita TeO_3^{2-} con grupos metilo podría reducir o eliminar su toxicidad, además otros estudios parecen confirmar que la telurita es modificada dentro de la célula. Inicialmente se consideraba que la metilación estaba involucrada en la resistencia a la telurita (pues TehB puede unir Te y S-AdoMet), sin

embargo posteriormente se mostró que la S-AdoMet es en realidad un cofactor de TehB, por lo que es necesaria la unión de ambas moléculas para que se lleve a cabo la catálisis¹³⁵³; de hecho recientemente se ha cristalizado exitosamente la proteína TehB de *Escherichia coli* en presencia de S-AdoMet¹³⁵⁴.

o La abundancia del telurio y sus derivados es alta en el universo y aunque baja en la corteza terrestre nada descarta la posibilidad que haya sido más abundante en los inicios de formación de la Tierra; de hecho existen ambientes que actualmente presentan enriquecimiento de este elemento, como en las cercanías de ventas hidrotermales oceánicas y en las minas donde se han identificado diversos tipos de microorganismos resistentes al telurio y sus derivados^{1346-1348,1355,1356,1357,1358}. Las regiones de donde se han aislado organismos resistentes al telurio puede que sean muestra de las condiciones ambientales de la Tierra primitiva, lo que podría explicar que la proteína de resistencia a telurita haya sido identificada como ancestral; no obstante es probable que la región codificada por tripletes RNY en TehB sea la que une el adenosil de la S-AdoMet, pues ésta se ha indicado como una molécula muy antigua que posiblemente ya existía en los albores del Mundo de las RNPs.

o **PhzF, FAMILIA DE LA BIOSÍNTESIS DE LA FENAZINA.**

o Las fenazinas son pigmentos aromáticos nitrogenados triheterocíclicos solubles en agua, cuya estructura central se mantiene pero pueden contener substituciones en varios sitios, lo que permite un amplio rango de derivados estructurales y diversas actividades biológicas^{1359,1360,1361,1362,1363,1364,1365,1366}.

o Las fenazinas son producidas por diversos géneros bacterianos como *Pseudomonas* y *Streptomyces*, algunas especies de *Actinomycetes* marinos, *Erwinia*, *Burkholderia*, *Brevibacterium*, *Pelagibacter* y *Vibrio*^{1359-1363,1366}. Algunas de las fenazinas bacterianas son: endofenazina, 2-hidroxifenazina, ácido fenazina-1-

carboxílico, ácido fenazina-1,6-dicarboxílico y clororafina, aunque posiblemente existen varios tipos de fenazina por descubrir¹³⁵⁹.

o Dichos metabolitos biológicamente activos están involucrados en la competencia ecológica entre los microorganismos que las producen, por lo que conferirían ventaja competitiva durante su fase de crecimiento y contribuir a la supervivencia de éstos en hábitats naturales, ya que las fenazinas poseen propiedades antibióticas de amplio espectro contra bacterias G+ y G-, hongos y tejidos de plantas y animales; además se ha mostrado que poseen cierta actividad antiviral y antitumoral¹³⁵⁹⁻¹³⁶³.

o Se considera que las propiedades antibióticas de las fenazinas son consecuencia de procesos de óxido-reducción que sufren las moléculas, pues actúan como donadores y aceptores de electrones interfiriendo así en la cadena respiratoria y precipitando reacciones que producen especies reactivas de oxígeno (ROS), como iones superóxido y peróxido y radicales hidroxilo¹³⁵⁹⁻¹³⁶³.

o Se precisan siete enzimas para la biosíntesis de la fenazina (PhzA-PhzG), aunque sólo 5, entre ellas PhzF, son esenciales para su anabolismo. Las moléculas fosfoenolpiruvato y D-eritrosa 4-fosfato forman el ácido 7-fosfato 3-desoxi-d-arabino-hetpulosónico (DAHP) que entra a la vía del shikimato y forma ácido fenazina-1 carboxílico (PCA); PhzE cataliza la transformación del ácido corísmico en ácido 2-amino-2-desoxiisocorísmico (ADIC, por sus siglas en inglés) y PhzD cataliza la transformación de ADIC en ácido trans-2,3-dihidro-3-hidroxi-antranílico (DHHA, por sus siglas en inglés). La condensación de dos moléculas derivadas del DHHA va conformando el armazón o estructura central de las fenazinas, dicha condensación es catalizada por PhzA, -B y -G solamente en presencia de PhzF; sin embargo PhzF es suficiente y estrictamente necesaria puede catalizar por sí sola, aunque ineficientemente, la formación de PCA a partir de DHHA utilizando NADH como cofactor, en ese caso PhzF catalizaría el paso inicial en la reacción posiblemente

mediante un re-arreglo alílico que genere un análogo 3-oxo de DHHA que es más reactivo, mientras que los pasos subsecuentes ocurrirían espontáneamente¹³⁵⁹⁻¹³⁶⁶.

o Según la deducción de secuencias microbianas existen varias proteínas semejantes a la fenazina ampliamente distribuidas en los organismos e incluso una misma especie puede presentar múltiples parálogos, de las cuales únicamente PhzF ha sido caracterizada funcional y bioquímicamente. Los residuos cercanos al sitio activo en PhzF son: L⁶⁹, **A**⁷², H¹⁵⁵, **A**¹⁸³, M¹⁹⁸, Y²⁰³, **V**²⁰⁵ y **T**²¹¹ mientras que E⁴⁵, H⁷⁴ y **D**²⁰⁸ son los residuos catalíticos clave, la mayoría de esos residuos tienen contacto hidrófobo con el ligando, no obstante la mayoría de esos aa no están conservados en proteínas con relación distante a PhzF¹³⁶⁴.

o La PhzF es un dímero de ~280 aa con masa de ~32kDa. La identidad en secuencia que comparten sus dominios -N y -C es de ~15%, lo que pudiese indicar un antiguo evento de duplicación interna. La topología central de la PhzF consiste de dos dominios arreglados en una hélice- α central rodeada por un barril de ocho hojas- β , con una hélice- α en el dominio N-terminal y tres hélices- α en C-terminal. Mientras que la secuencia de PhzF carece de homología detectable con otras proteínas, su inusual plegamiento **denominado 'hot dog'** exhibe similitud estructural con enzimas diversas, como la '**sintetasa de 3-desoxi-D-arabinoheptulosoato 7-fosfato**' (DAHP sintetasa), la PhzC, la 'epimerasa de diaminopipemilato' (DapF) y varias proteínas de función desconocida; aunque los residuos clave en sus sitios activos son totalmente distintos, lo que sugiere mayor variedad estructural y funcional dentro de la familia de lo considerado inicialmente^{1362-1364,1367}.

o Mediante el análisis del contenido de GC, los patrones de uso de codones, la frecuencia de nucleótidos, la similitud de secuencias y reconstrucciones filogenéticas, se ha mapeado la historia evolutiva de los genes de fenazina de múltiples especies bacterianas. La distribución filética irregular, evidencia de acoplamiento ('linkage') genético y la gran similitud entre secuencias, sugiere que

las especies de *Pseudomonas*, *Streptomyces cinnamonensis* (*Stcin*), *Pantoea agglomerans* (*Paagg*), *Burkholderia cepacia* (*Bucep*), *Pectobacterium atrosepticum* (*Peatr*), *Brevibacterium linens* (*Brlin*) y *Mycobacterium abscessus* (*Myabs*), contienen el mismo operón *phz*, que posiblemente fue introducido mediante HGT de genes de una de esas especies a las demás. La adquisición de un operón de propiedades antibióticas es significativa, ya que incrementa el grado de adaptación ('fitness') de la especie receptora. Cada uno de los genes del operón *phz* fue caracterizado filogenéticamente; en el caso particular de *phzF* se han propuesto varios escenarios posibles: a) el gen *phzF* se originó de novo en *Pseudomonas* y fue introducido en *Bucep*, *Paagg* y *Peatr* mediante HGT; b) *phzF* se originó en *Bucep*, *Paagg* y *Peatr* y se transfirió de éstas a las demás especies y al ancestro común de las *Pseudomonas*; c) tanto *Pseudomonas* como *Bucep*, *Paagg* y *Peatr* obtuvieron el gen *phzF* de otra fuente aún no representada; lo único claro es que todos los genes de *PhzF* en *Pseudomonas* y *Bucep*, *Paagg* y *Peatr* comparten la misma historia evolutiva y por tanto forman un mismo grupo denominado Clade A; con base en ese análisis filogenético se infiere que además que el gen *phzF* de *Brlin* no fue adquirido de la misma fuente que el del Clade A, de hecho ese gen no es parte del operón *phz* en *Brlin*. En resumen, se puede observar que aún no es posible determinar con certeza la especie donadora original del operón *phz*¹³⁶⁸.

- o Un estudio posterior muestra que, excepto por la arquea *Methanosarcina mazei* que utiliza una fenazina unida a la membrana como acarreador de electrones en la metanogénesis, parece ser que la biosíntesis de fenazina es limitada a géneros bacterianos como *Actinobacteria*, *Betaproteobacteria* y *Gammaproteobacteria*. En el estudio mencionado se utilizaron genes de mantenimiento concatenados para generar filogramas, los cuales coinciden con la filogenia aceptada para las especies bacterianas y sin embargo sólo coinciden parcialmente con el filograma de *phzF*. Los resultados del análisis sugieren que la evolución del operón *phz* pudo haber sido: a)

mediante HGT como se ejemplifica con el evidente agrupamiento de genes *phzF*, evento que es posiblemente el más antiguo que incluso precedió a la especiación de al menos los géneros estudiados o b) mediante la estricta conservación de genes. Debido a las propiedades químicas de las fenazinas éstas pueden influir en la fisiología de los organismos que las producen, más allá de fungir como antimicrobianos, por ejemplo mediando la reoxidación del NADH bajo condiciones de oxígeno limitado como las encontradas en los biofilms, o bien pueden funcionar en la reducción microbiana de minerales como lanzaderas de electrones, facilitando el acceso al hierro y nutrientes como fosfato, minerales traza y compuestos orgánicos asociados con minerales¹³⁶⁹.

o Por lo anteriormente mencionado, es posible que la región identificada como ancestral en la PhzF no sea la que une el PCA ni sus moléculas precursoras sino el sitio que une NADH, pues se ha caracterizado como una molécula de componentes prebióticos; de ello se puede inferir además que el primordio de la PhzF posiblemente ya fungía como lanzadera de electrones en los procesos bióticos primordiales, tal como ocurre actualmente en la reducción microbiana de minerales.

T19. Características plausibles de la probable región codificada por tripletes RNY en las proteínas que actualmente conforman el sistema de defensa o supervivencia.

Conjunto	Proteína (nombre traducido y adaptado)	Características plausibles de la probable región codificada por tripletes RNY.
	Transferasa de la familia de repetición de hexapéptidos	Región de hexapéptidos repetidos, o quizá unión de CoA
	Proteína TehB de resistencia a la telurita, putativa	Unión de S-AdoMet
	Proteína para la biosíntesis de fenazina, familia PhzF	Unión de NADH

6.3.1.9 Proteínas/dominios hipotéticas(os) y/o conservadas(os).

El término hipotético se refiere a proteínas o dominios que se presumen como tal, pese que no se les ha asignado función específica alguna; muchas veces porque la secuencia tiene algunos rasgos que indican que posiblemente sean o un dominio o una proteína y/o porque no se han realizado los correspondientes experimentos *in vitro* con el fin de caracterizarles. Por otro lado el término conservado se refiere a proteínas o dominios que, pese a no estar plenamente caracterizadas(os), se encuentran en organismos de al menos 3 distintas clades^{1370,1371,1372,1373,1374}.

En las investigaciones sistemáticas que se han realizado para tratar de encontrar el genoma mínimo o el genoma medular (core) de un grupo filético (como un pan-genoma) siempre se han encontrado proteínas y dominios hipotéticas(os), por lo que se asume estaban presentes en su último ancestro común. De hecho **poco más del 50% de los ORFs en un genoma corresponden a dichas 'proteínas o dominios hipotéticas(os) y/o conservadas(os)', lo que sin duda** indica que poseen cierta relevancia, máxime cuando existen reacciones que sólo pueden ser explicadas mediante la catálisis o el apoyo estructural de una proteína que sin embargo se desconoce. Pese a la innegable relevancia de tales proteínas/dominios hipotéticas(os) conservadas(os), hasta hace relativamente poco tiempo se comenzaron a delinear metodologías enfocadas a caracterizarlas¹³⁷⁰⁻¹³⁷⁴.

Con base en la metodología descrita en la presente tesis, se encuentra que algunas proteínas/dominios hipotéticas(os) conservadas(os) contiene regiones codificadas por tripletes RNY, lo que implicaría que no sólo son antiguas sino que estuvieron presentes en organismos que existieron antes que el LUCA, posiblemente incluso involucradas en su origen, evolución y creciente complejidad. A continuación se describen las proteínas/dominios de este tipo (en T20 y T21) con su nombre en

http://www.ncbi.nlm.nih.gov/genome/proteins/186?project_id=57935 al ingresar cada 'Gene ID' y la información ahí contenida, así como la literatura concerniente a cada una(o) cuando ésta exista; es necesario señalar que evidentemente para la mayoría no existe una secuencia o motivos caracterizados que puedan inferirse como los ancestrales, bajo la metodología descrita en la presente tesis, es por ello que ni siquiera se mencionan escenarios evolutivos.

Order	Sc.	E-v	ID	%	PROTEIN / FAMILY / DOMAIN
1	31	3.00E-07	42/45	93%	conserved hypothetical protein (98.97 %)
2	28	2.00E-05	28/28	100%	conserved hypothetical protein (77.22 %)
3	27	6.00E-05	27/27	100%	conserved hypothetical protein (82.76 %)
4	25	0.001	25/25	100%	hypothetical protein (25.60 %)
5	24	0.004	24/24	100%	conserved domain protein (32.97 %)
6	24	0.004	24/24	100%	hypothetical protein (84.36 %)
7	24	0.004	24/24	100%	conserved hypothetical protein (1.71 %)
8	24	0.004	23/23	100%	conserved hypothetical protein (17.16 %)
9	23	0.015	22/22	100%	conserved hypothetical protein (80.63 %)
10	22	0.06	22/22	100%	conserved hypothetical protein (64.96 %)
11	22	0.06	22/22	100%	hypothetical protein (68.58 %)
12	22	0.06	22/22	100%	hypothetical protein (31.96 %)

T20. Proteínas o dominios hipotéticos(as) y/o conservados(as), que contienen regiones codificadas por tripletes RNY, nomenclatura y código de colores igual que T1 y T2.

- o **1. PROTEÍNA HIPOTÉTICA CONSERVADA.** Superfamilia de β -lactamasas.
- o Las β -lactamasas pueden clasificarse con base en sus características moleculares o propiedades funcionales, de manera que las clases moleculares A, B, C y D definen una enzima de acuerdo con su secuencia de aa y motivos conservados, mientras que los grupos funcionales 1, 2, y 3 se utilizan para asignar una descripción clínicamente útil para una familia de enzimas, con subgrupos designados de acuerdo con sustrato y el inhibidor de perfiles, otras designaciones se utilizan para definir la funcionalidad de subgrupos específicos; dado que ninguno de esos sistemas proporciona una descripción inequívoca de este conjunto versátil de enzimas, se ha propuesto un sistema de clasificación de las β -lactamasas para incluir sus propiedades microbiológicas, moleculares y

bioquímicas con base en las clases tradicionales A, B, C, y D y los grupos funcionales 1, 2 y 3¹³⁷⁵. Esta proteína conservada pertenece entonces a la superfamilia de las β -lactamasas con base en sus características generales, aunque aún no pude saberse a cuál de ellas se refiere (SAK_0920).

- o **2. PROTEÍNA HIPOTÉTICA CONSERVADA.**

- o Hipotética (SAK_1090).

- o **3. PROTEÍNA HIPOTÉTICA CONSERVADA.** Familia 'nucleotidiltransferasa'.

- o Las proteínas que presentan el plegamiento nucleotidiltransferasa (NT) constituyen una gran y diversa familia involucrada en el procesamiento de ácidos nucleicos, ésta superfamilia se caracteriza por una estructura medular α/β con topología $\alpha\beta\alpha\beta\alpha$ y 3 motivos conservados: **hG[G/S]**, **[D/E]h[D/E]h** y **h[D/E]h** (donde h es un aa hidrófobo), a los que se añaden distintos elementos que permiten distinguir cada uno de los grupos¹³⁷⁶. Se puede observar que los aa en cada uno de los 3 motivos están codificados por tripletes RNY (SAK_0139).

- o **4. PROTEÍNA HIPOTÉTICA.** Plecstrina, superfamilia PH.

- o Los dominios con homología a la plecstrina (PH) son pequeños módulos presentes una o varias veces a lo largo de diversas proteínas involucradas en la señalización celular, siendo los dominios PH la única característica en común de todas esas proteínas con disímiles secuencias. Los dominios PH pueden unir los grupos fosfato de fosfatos de inositol, péptidos con fosfotirosina o hélices de poli-prolina, de esta manera constituyen no sólo intermediarios en las vías de señalización sino también plataformas moleculares para interacciones proteína-proteína. La estructura del dominio PH puede ser vista en dos mitades, cuya parte

terminal-N contiene 4 hojas- β y la parte terminal-C tres hojas- β y una hélice- α . Inicialmente se consideraba que los módulos PH eran exclusivos de eucariontes, no obstante el haberlos identificado en bacterias ha llevado a plantear que éstos son en realidad dominios muy antiguos^{1377,1378}; además es probable que al menos parte de un módulo PH haya surgido en los albores del Mundo de las RNPs, pues es codificada por tripletes RNY (SAK_0937).

- o **5. PROTEÍNA DE DOMINIO CONSERVADO.** MhpC, 'hidrolasa α/β '.
- o La superfamilia de hidrolasas α/β es una de las más grandes familias de proteínas conocidas y se caracteriza por ocho hebras- β interconectadas por horquillas y hélices- α ; el plegamiento hidrolasa α/β es típico de enzimas hidrolíticas de diversos orígenes filogenéticos actuales, que actúan sobre compuestos químicos disímiles, sin otra similitud en secuencia o estructura y que no obstante han divergido de un ancestro común que con certeza poseía ya una tríada catalítica del tipo: aa nucleófilo (**Ser**, la mayoría de las veces) + His + aa ácido (**Asp/Glu**); para la enzima MhpC la tríada se compone de **S**¹¹⁰-(H¹¹⁴ o H²⁶⁴)-**D**²³⁵, ésta proteína actúa sobre enlaces C-C como esterasa y tioesterasa en la vía de degradación del fenilpropionato. Inclusive se ha creado una base de datos específica para proteínas con plegamiento hidrolasa α/β ^{1379,1380,1381,1382,1383,1384}. Como podemos observar, algunos de los aa de la tríada catalítica en esta hidrolasa α/β con codificados por tripletes RNY, aunque no es posible discernir cuál de todos ellos surgió en el Mundo de las RNPs (SAK_0808).

- o **6. PROTEÍNA HIPOTÉTICA.** Superfamilia EssA, proteína del sistema Wss.
- o El sistema de secreción Wss, quizá erróneamente tipificado como tipo VII, es responsable de la secreción de proteínas WXG100; las proteínas de la familia EssA

se consideran parte fundamental de dicho sistema, pues contienen una región central altamente polar cuya secuencia es de baja complejidad que pudiera constituir un dominio transmembranal y las ubicaría directamente embebidas en la membrana, pueden contener adicionalmente motivos LF en repeticiones de ~17aa de longitud y están flanqueadas por dominios-N y -C hidrófobos^{1385,1386,1387,1388,1389,1390} y algunos de ellos pueden ser codificados por tripletes RNY (SAK_1126).

- o **7. PROTEÍNA HIPOTÉTICA CONSERVADA.** Familia 2 de endonucleasa AP, dominio 'barril TIM α/β '.

- o Las endonucleasas de sitios apúricos/apirimidínicos (AP) de la familia 2 (endo IV) participan en la reparación del DNA hidrolizando los enlaces fosfodiéster en esos sitios AP. Dicha endonucleasa posee una estructura denominada dominio 'barril TIM α/β ' que comparte con otras proteínas de varias familias distintas con baja similitud en secuencia pero cuya estructura está altamente conservada, todas ellas provienen de un ancestro común cuya estructura era posiblemente un medio barril α/β que se duplicó y fusionó. Todas las proteínas con 'barril TIM α/β ' comparten un motivo de unión a fosfato y en general el barril TIM está conformado por un 'tejido' de $\alpha_8\beta_8$ en que el fosfato, que puede provenir del substrato o de cofactores, se une entre las hebras 7 y 8; dos sitios similares no idénticos de unión a metales, como Mg^{2+} , Mn^{2+} o Co^{2+} , formados por aminoácidos del tipo **D** y E; mientras el sitio activo se encuentra fuera del barril TIM y es diferente en cada familia de enzimas^{1391,1392,1393,1394,1395,1396,1397}; tanto el fosfato como los cationes metálicos divalentes pudieron existir en el Mundo de las RNPs, de manera que el sitio de unión de cualquiera de ellos es probablemente codificado por tripletes RNY (SAK_1822).

o **8. PROTEÍNA HIPOTÉTICA CONSERVADA.**

- o Hipotética (SAK_0640).

o **9. PROTEÍNA CONSERVADA.**

- o Hipotética. Pertenece a una superfamilia de pequeñas proteínas secretadas, específicas del género *Streptococcus*; es importante resaltar que en todos los genomas secuenciados las proteínas de esta familia aparecen en un contexto genómico conservado junto a una tiorredoxina, una tRNA sintetasa (aaRS) y una proteína de unión a tRNA, empero la implicación funcional de ello no es clara, es notable pues parte de dicha proteína hipotética es codificada por tripletes RNY (SAK_0242).

o **10. PROTEÍNA HIPOTÉTICA CONSERVADA.** Superfamilia 'PASTA'.

- o El dominio PASTA se encuentra en la parte extracelular, el extremo terminal-C, de varias proteínas de unión a la penicilina (PBP) y de varias cinasas de Ser/Thr (STKs) como PknB, en las que está implicado en la unión al anillo beta-lactámico de la penicilina y al substrato D-Ala-D-Ala respectivamente; en cada una de esas proteínas el dominio PASTA se puede encontrar repetido varias veces y pese que comparte sólo ~10.5% de identidad en secuencia, su estructura de 3 hojas- β y una hélice- α está altamente conservada^{1398,1399,1400,1401,1402,1403}; ese dominio PASTA contiene además motivos codificados por tripletes RNY (SAK_1067).

o **11. PROTEÍNA HIPOTÉTICA.** AbiH.

- o Conservada en *Streptococcus agalactiae*, AbiH se predice que consiste de 346aa y ~40kDa, no presenta similitud con ninguna proteína depositada previamente en alguna base de datos pero se ha encontrado que es esencial para la patogenicidad del microorganismo, en particular su capacidad abortiva en el

humano¹⁴⁰⁴; pese a ser una proteína que confiere características patógenas, posee algún motivo que quizá surgió en el Mundo de las RNPs (SAK_1325).

o **12. PROTEÍNA HIPOTÉTICA.** CotH.

o Conservada en *Firmicutes*, CotH se predice que consiste de 1,086 nt y ~40kDa, aparentemente sólo la presentan los organismos que forman esporas pues está involucrada directamente en la biogénesis del recubrimiento proteico de las mismas^{1405,1406}; no obstante alguna porción de dicha proteína contiene una región codificada por tripletes RNY (SAK_1572).

Previamente se mencionó que **se han tratado de caracterizar las 'proteínas o dominios hipotéticas(os) y/o conservadas(os)'**, creando para ello diversas metodologías *in vitro* e *in silico* que pueden resultar muy efectivas¹³⁷⁰⁻¹³⁷⁴; con base en los resultados aquí obtenidos, se observa que algunas proteínas o dominios hipotéticos(as) contienen regiones codificadas por tripletes RNY y tal característica pudiera añadirse a las metodologías ya existentes. Con base en el análisis de los resultados de cada una de las proteínas presentadas a lo largo de esta tesis, es probable que las regiones codificadas por tripletes RNY en las proteínas o dominios hipotéticas(os) constituyan sitios de unión a moléculas de origen prebiótico o bien conformen estructuras para la interacción proteína-proteína. Es posible entonces sugerir que el primordio de algunos polipéptidos que no están caracterizados plenamente tiene su origen en los albores del Mundo de las RNPs y ello vuelve impostergable su caracterización.

T21. Características plausibles de la probable región codificada por tripletes RNY en cada una de las proteínas o dominios consideradas(os) hipotéticas(os).

Conjunto	Proteína (nombre traducido y adaptado)	Características plausibles de la probable región codificada por tripletes RNY.
Proteínas o dominios, conservados e/o hipotéticos	Proteína hipotética conservada	Superfamilia de las β -lactamasas
	Proteína hipotética conservada	Hipotética
	Proteína hipotética conservada	Nucleotidiltransferasa
	Proteína hipotética	Superfamilia tipo PH; capacidad de unir fosfatos de inositol y otras proteínas
	Proteína de dominio conservado	Hidrolasa alpha/beta, MhpC
	Proteína hipotética	EssA
	Proteína hipotética conservada	Barrel TIM, endonucleasa AP, isomerasa del tipo xilosa
	Proteína hipotética conservada	Hipotética
	Proteína hipotética conservada	Contexto conservado junto a una tiorredoxina, una aaRS y proteína de unión a tRNA
	Proteína hipotética conservada	Dominio PASTA
	Proteína hipotética	AbiH
	Proteína hipotética	CotH

6.3.2 GENOMA ALEATORIO, CONTROL.

Es evidente el exiguo número de proteínas (20) que se identifican como antiguas en el genoma aleatorio, de las cuales sólo una, registrada como un **regulador transcripcional putativo**, se ajusta al nivel de corte establecido en $E \leq 0.05$ (en T22 y T23).

Categorías en COG	Subcategorías (COG & KEGG)
DNA, estructura y modificaciones (1).	

T22. Tabla que muestra los resultados del genoma aleatorio que fungirá como control, se indica el tipo de proteína con su código de colores correspondiente; así como, entre paréntesis, el número de resultados en cada conjunto.

La proteína que contiene una región codificada por tripletes RNY, identificada en el genoma (ORFeoma) aleatorio derivado de *SagA909*, es presuntamente el **regulador transcripcional LytR** (que parece ser un atenuador de la transcripción de proteínas involucradas en la síntesis de componentes de la pared celular)¹⁴⁰⁷; sin embargo dado que el sentido biológico de la secuencia se ha eliminado por completo al convertir en aleatorio el genoma, haber identificado una porción de cualquier proteína es posiblemente sólo resultado del azar y por ende es ilógico siquiera considerar posible el identificar su primordio.

Curr.Or.	Sc.	E-v	ID	%	Category	match 1
1	22	0.044	22/22	100%	DNA	transcriptional regulator, putative (58.38 %)

T23. Se indica la única proteína identificada de la secuencia aleatoria, nomenclatura igual que en T1 y T2.

VII. DISCUSIÓN

Algunos investigadores han sugerido que quizá **"nunca podremos conocer la verdadera naturaleza de LUCA y mucho menos acerca de etapas evolutivas previas"**⁴³, no obstante nuestros resultados indican que sí es posible conocer las probables características de los módulos primordiales, al menos de proteínas, e incluso incidir de este modo en la explicación sobre el origen de la vida misma; en particular los resultados de la presente tesis concuerdan con la evolución paralela de los diversos escenarios evolutivos que se han propuesto.

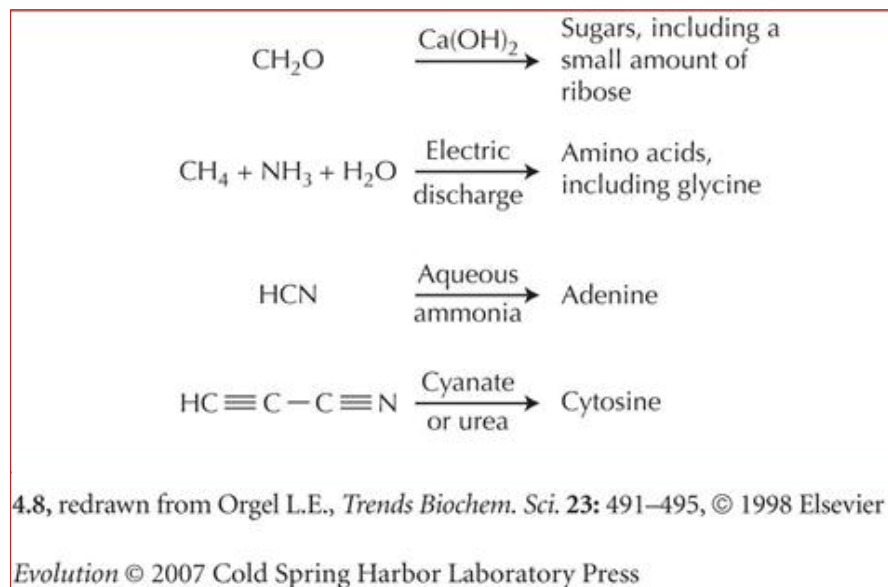
Mediante diversos métodos se ha intentado dilucidar cuáles son los elementos bióticos más antiguos. Con metodologías **"de abajo hacia arriba"** (bottom-up) se trata de esclarecer cómo fue el origen de la vida en la Tierra y la composición genética y/o proteica de los proto-organismos más antiguos, reconstruyendo la cadena de sucesos mediante química prebiótica o estudios de paleogeoquímica; mientras que con metodologías **"de arriba hacia abajo"** (top-down) se realizan reconstrucciones filogenéticas a partir de genómica comparativa de varias especies e incluso con base en aspectos ecológicos, con el fin de averiguar cuál fue el metabolismo ancestral.

DE ABAJO HACIA ARRIBA.

La llamada **'sopa primigenia'** da origen a diversas azúcares (como ribosas y hexosas), aminoácidos (para formar péptidos), hidrocarburos (que originarán cadenas lipídicas) y algunas moléculas fundamentales como adenina o citosina^{Powner, *op.cit.* Pág.14} (éstas incluso a partir de moléculas que son actualmente tóxicas para organismos como el ser humano, tal cual se observa en la siguiente

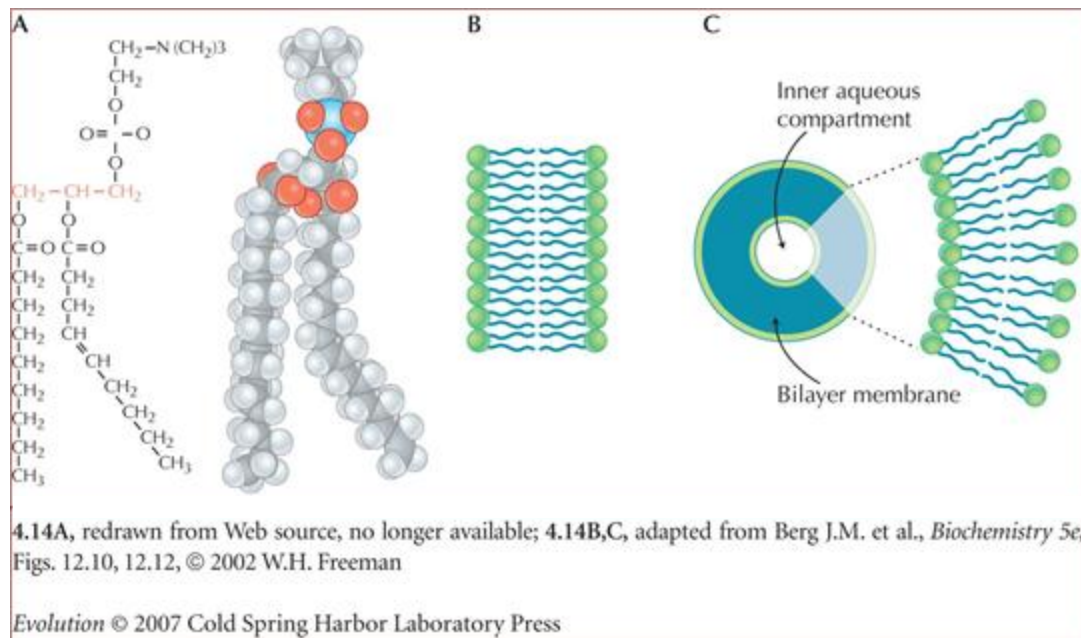
figura); algunas de esas moléculas son parte de cofactores nucleotídicos universales²⁻⁶. Cabe señalar que varios de los aa que pudieron sintetizarse prebióticamente, son también codificados por tripletes RNY y éstos a su vez pudieron polimerizarse formando pequeños péptidos, mismos que corresponderían con los motivos altamente conservados en las proteínas identificadas como ancestrales.

En la siguiente figura (F8) se muestra de hecho la síntesis de varias moléculas bióticas a partir de compuestos prebióticos, así vemos que: a partir de formaldehído, en una reacción catalizada por hidróxido de calcio, pueden sintetizarse algunas azúcares incluyendo pequeñas cantidades de ribosa; a partir de metano, amoníaco y agua, en un ambiente de descargas eléctricas, se sintetizan aminoácidos como la glicina; de manera peculiar es notorio que a partir de compuestos que pueden desprender el anión cianuro (tóxico para los seres vivos que llevan a cabo fosforilación oxidativa) se sinteticen bases nitrogenadas, como adenina a partir de ácido cianhídrico en una solución de amoníaco, o bien citosina a partir de cianoacetileno en una solución de cianato o bien de urea.



F8. Reacciones químicas que muestran la síntesis de moléculas biológicas a partir de compuestos abióticos.

Por otro lado los lípidos y moléculas anfífilas formadas prebióticamente pueden auto-organizarse en supra-estructuras como micelas que al plegarse sobre sí mismas pueden generar liposomas que poseen una bicapa lipídica con un interior acuoso como lo esquematiza la siguiente figura (F9), en la que se observa el esquema de un fosfolípido en A, el esquema de una bicapa lipídica formada por fosfolípidos en B y finalmente una gotícula con bicapa lipídica en C.

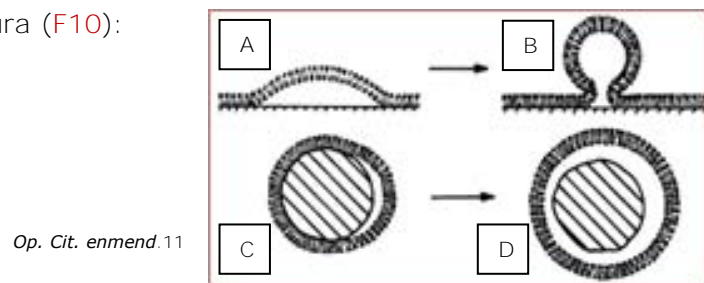


F9. Esquema que muestra la estructura general de un fosfolípido (A), la estructura de una bicapa lipídica de fosfolípidos (B) y la estructura de un liposoma con bicapa lipídica (C).

Si un liposoma contiene moléculas de tipo RNA, éste se autorreplicará, la presión osmótica dentro de la gotícula aumentará y el liposoma se fisionará dando lugar a dos estructuras con cantidades semejantes de material autocatalítico⁷⁻¹⁰.

Si además la química prebiótica tuvo lugar en los resquicios de minerales cercanos a fuentes hidrotermales, pudieron haberse concentrado diversos compuestos polianiónicos, como ATP, NAD/NADP, FMN/FAD, CoA y S-AdoMet que fungen como coenzimas actualmente y el mismo RNA que es un polianión, así como altas concentraciones de iones metálicos, entre ellos posiblemente cationes como Mg²⁺, Mn²⁺, Co²⁺ o Ni²⁺ ¹¹⁻¹⁹. Es importante subrayar que los motivos identificados

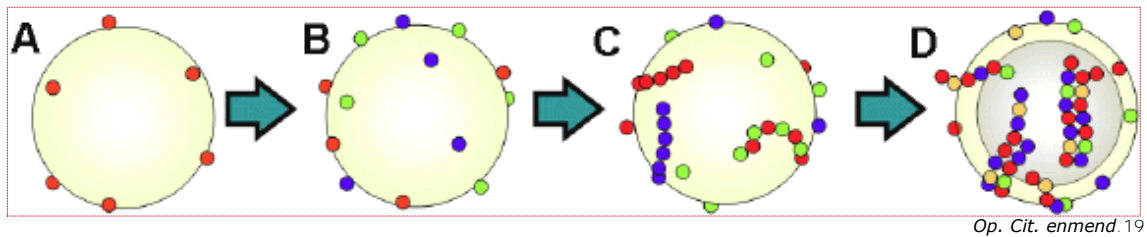
como ancestrales pueden unir esos cofactores universales y/o cationes metálicos. Las moléculas mencionadas, al adherirse a la superficie del mineral comenzaron a interactuar en ciclos autocatalíticos, dando origen a diversos compuestos orgánicos y paulatinamente generando procesos metabólicos elementales autosustentables, con base en la transferencia de e- en contra de un potencial electroquímico; mientras que la membrana celular fue surgiendo como se indicó previamente, con los hidrocarburos organizándose en bicapas lipídicas, mismas que se circularizaron atrapando tanto a los minerales como a las moléculas que se generaron a partir de las coenzimas, proveyendo una barrera semipermeable, proceso que finalmente evolucionó hacia proto-organismos¹¹⁻¹⁹, como se esquematiza en la siguiente figura (F10):



F10: Circularización de una bicapa lipídica, esquemas de la vista sagital del proceso: A) la bicapa comienza a desprenderse de la superficie mineral en que se encuentra adosada, B) se produce lo que se denomina abstricción, que es básicamente una constricción a partir del desprendimiento inicial de la bicapa, C) se desprende la gotícula con moléculas aniónicas que quedaron atrapadas y D) finalmente se forma una célula con membrana de bicapa lipídica.

Las coenzimas son moléculas de bajo peso molecular que funcionan como acarreadores de electrones durante los procesos metabólicos en los diversos organismos actuales; cada coenzima está especializada en aceptar y transportar un sólo tipo de molécula específica, algunas aceptan iones hidrógeno, otras grupos acetilo, amino, etcétera pero una misma coenzima puede unirse a un gran número de enzimas diferentes y, aunque el número de coenzimas es relativamente bajo, dicho sistema funciona eficientemente; es posible además que diversos ciclos metabólicos hayan surgido mediante la sucesión de uniones simbióticas de las coenzimas^{12,13}. Experimentalmente se ha mostrado que los aptámeros

(oligonucleótidos que unen moléculas diana con gran especificidad) de RNA pueden catalizar la formación de complejos RNA-coenzima (-CoA, -NAD y -FAD), utilizando los precursores correspondientes (fosfopanteteína, NMN y FMN), al mismo tiempo que dichas coenzimas se incorporan al RNA mediante transcripción; estos aptámeros son capaces además de captar moléculas que actualmente son metabolitos requeridos universalmente por todos los organismos¹²⁻¹⁸; es decir, las coenzimas pueden describirse como elementos codificantes simples que inicialmente llevaban a cabo tanto funciones hereditarias como metabólicas elementales dentro de microesferas oleicas hechas de hidrocarburos auto-organizados, posteriormente surgirían rutas metabólicas más complejas y el sistema evolucionaría hasta formar proto-organismos definidos, como lo muestra la siguiente figura (F11)^{10,19,28}.



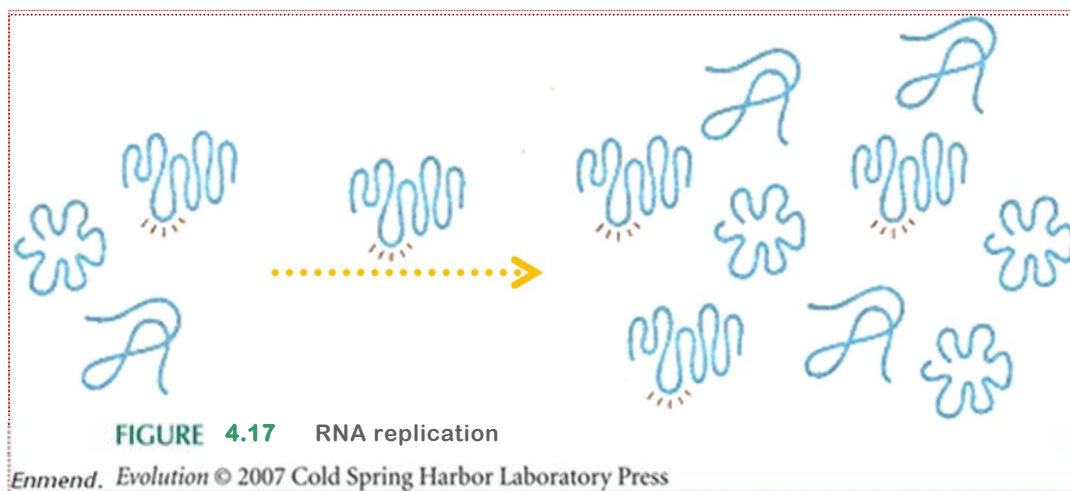
F11. Primeros pasos de la evolución biológica.

Los elementos codificantes (ECs colores rojo, verde y azul) fueron primeramente coenzimas nucleotídicas, que al polimerizarse formaron oligonucleótidos de RNA, que dieron paso a la codificación basada en templado *i.e.* el primer código genético (de RNY de acuerdo con nuestro modelo). Los primeros pasos se llevaron a cabo posiblemente en la superficie de micelas (esfera sobre la que están los ECs), antes de la replicación basada en templado que ya se llevaría a cabo en liposomas (esfera hueca), dentro de los cuales quedaron atrapados los ECs y otras moléculas funcionales (esferitas color amarillo), que migraron desde la superficie al interior del liposoma y quedaron aislados como parte de la primera célula.

A) Inicialmente existió un sólo tipo de EC, B) posteriormente aparecieron múltiples ECs que llevaron a cabo codificación combinatoria, C) dichos ECs se polimerizaron de forma repetitiva o aleatoria y D) finalmente surgió la replicación basada en templado.

Dicha transición del Mundo de las coenzimas a la formación de rutas metabólicas pudiera reflejarse en el hecho que varias enzimas del metabolismo fueron identificadas como ancestrales y de éstas la región codificada por tripletes RNY sea probablemente la que une la coenzima respectiva.

Como se mencionó anteriormente, las coenzimas pueden polimerizarse formando cadenas constituidas por $[1\text{base nitrogenada} + 1\text{ribosa} + 1\text{fosfato}]_n$ que son finalmente cadenas de RNA; este RNA es autocatalítico, debido a la presencia de un grupo hidroxilo (-OH) en el **carbón 2'** y al mismo tiempo sus cualidades son hereditarias; además esas cadenas de RNA podían plegarse en estructuras secundarias y terciarias, lo que le confiere gran estabilidad; finalmente, con base en resultados experimentales de aptámeros, se ha mostrado que el RNA es capaz de unir bases nucleotídicas, coenzimas, aminoácidos simples y compuestos ribofosfatados²³. Por otro lado, la Tierra primitiva carecía de atmósfera y aunque es posible que los rayos UV pudiesen terminar con la incipiente vida, también se ha propuesto que fuesen un factor selectivo a favor de la preservación de cadenas nucleotídicas compuestas por fosfatos+azúcares+bases nitrogenadas¹⁴⁰⁸; el mecanismo general del Mundo del RNA entonces inició probablemente con monómeros que comenzaron a unirse en cadenas, de los cuales surgiría una ligasa que permitiría cadenas más largas, de tal conjunto surgió una ribozima replicasa que catalizaba la replicación de sí misma y de otras moléculas^{15,20-24}, como se esquematiza en la siguiente figura (F12) y explicado en¹⁴⁰⁹.



F12. Esquema de la replicación del RNA. Se muestran distintos tipos de cadenas de RNA (con tres tipos de líneas azules), una de los cuales corresponde a una ribozima replicasa (señalada con pequeñas líneas café en su parte inferior), que se autorreplica y a replica los otros tipos de cadena.

En ese momento evolutivo, los procesos mencionados posiblemente ocurrían en ambientes aislados del medio, en un sistema con compartimentos en los que se generaban además algunas formas de metabolismo basadas en la transferencia de electrones entre los diversos compuestos prebióticos, donde el RNA era la molécula principal que regulaba incluso algunas de las funciones metabólicas tal como los riboswitches actuales, es decir que existían las condiciones óptimas para el surgimiento y desarrollo de proto-organismos basados en RNA como molécula medular *i.e.* riboorganismos^{15,20-24,1410}.

Dado que los péptidos y el RNA sintetizados prebióticamente estaban presentes de manera simultánea en la Tierra primitiva, no es fácil definir una temporalidad exclusiva para el Mundo del RNA e inclusive se plantea que la vida habría comenzado directamente con un Mundo de las RNPs, debido a plausibles interacciones entre coenzimas nucleotídicas, aminoácidos y los derivados de cada uno²⁵⁻²⁸, no obstante ello no explicaría la presencia de ribozimas *i.e.* moléculas de RNA que llevan a cabo procesos autocatalíticos sin necesidad de enzimas peptídicas²¹; de manera que, aunque breve, es posible que existiera el Mundo del RNA puro, no porque no hubiesen en el medio aa u otras moléculas sino porque no existía una interacción directa con ellas. Por un lado las primeras cadenas autocatalíticas de RNA poseían todas la secuencia RNY (R=purina, N=cualquier base, Y=pirimidina)^{60,67-72} y por otro lado algunos péptidos prebióticos, como los formados por GADV (Gly, Ala, Asp, Val), poseían propiedades autocatalíticas⁸² mientras que otros aa pudieran haberse sintetizado mediante reacciones de química prebiótica^{3,5,74,75,292}; dado que ambos ciclos se llevaban a cabo posiblemente en espacios restringidos, ambos procesos comenzaron a acoplarse, emergiendo formalmente el código genético en los albores del Mundo de las RNPs.

De todo lo anterior es posible advertir que las moléculas prebióticas interaccionaron entre sí, algunas formaron polímeros homogéneos o heterogéneos

que a su vez interaccionaron con las moléculas y compuestos existentes, ello generó ciclos metabólicos elementales basados en la transferencia de electrones; el registro evolutivo de dichos procesos se encuentra en los motivos conservados de las enzimas identificadas como ancestrales en el presente trabajo, pues éstos unen moléculas prebióticas y sus derivados, de manera que tal registro corresponde con un Mundo de RNPs y desde ese momento la evolución del código genético coevolucionó con la síntesis metabólica de aminoácidos hasta generar finalmente el repertorio genético y metabólico del LUCA.

DE ARRIBA HACIA ABAJO.

Con el fin de definir el repertorio de proteínas que poseía el último ancestro común, se han comparado diversos genomas y se deduce que las proteínas encontradas en todos los organismos comparados son esenciales, por tanto se considera que constituyen el '**genoma mínimo**' del LUCA; así mismo se pueden realizar **comparaciones del tipo 'genoma mínimo' en el contexto funcional de una célula ancestral**; es posible realizar comparaciones más minuciosas enfocadas a obtener el repertorio mínimo de dominios de proteínas e incluso de motivos de aa.

Quizá el primer trabajo de esta índole se refiera al tamaño mínimo del genoma de un organismo en condiciones óptimas de crecimiento (de 318 a 562 kpb)¹⁴¹¹; aunque los primeros trabajos '**genoma mínimo**' **comenzaron** en realidad con la comparación de los genomas de las primeras bacterias secuenciadas (*Mycoplasma genitalium* y *Hemophilus influenzae*), una de las cuales es un organismo parásito que se considera posee el genoma más pequeño, con estos trabajos se indicaron ~250 genes esenciales^{1412,1413,1414}. A la comparación de secuencias, le siguieron años de trabajos experimentales basados en la manipulación misma del genoma de ese organismo parásito (*M. genitalium*), anulando paulatinamente genes hasta

conservar únicamente el número mínimo con el fin de que el organismo continuase siendo viable en condiciones óptimas de laboratorio, con lo que se definieron ~380 genes indispensables, la mayoría codifican para proteínas y algunos para regiones reguladoras imprescindibles^{1415,1416,1417}. Posteriormente, con base en comparaciones de más de 260 genomas distintos, se propone un genoma mínimo de poco más de 200 genes para un organismo que crezca en condiciones ideales¹⁴¹⁸. Varias de las proteínas identificadas como antiguas bajo ese tipo de metodologías comparativas son coincidentes con la nuestra, por ejemplo se identifican como ancestrales las proteínas ribosomales, proteínas involucradas en la traducción, proteínas de reparación del DNA; diferentes subunidades de la ATP F₀F₁, subunidades esenciales para el transporte específico de péptidos, subunidades del PTS; proteínas del metabolismo de lípidos, de nucleótidos, de la glicólisis y de las pentosas; inclusive están ausentes varias de las enzimas del ciclo TCA que no se identifican con nuestra metodología; no obstante hubo algunas enzimas que no se reportan como antiguas pero si contienen regiones codificadas por tripletes RNY, como las MβL, ello se debe a que no todos los organismos presentan este tipo de enzimas aunque, como se explicará más adelante, no sorprende en realidad el hecho que tal tipo de proteínas contenga una región que si es plausible sea antigua; de esta manera se muestra que varias de las enzimas que contenía LUCA tuvieron su origen en los albores del Mundo de las RNPs. Al tratar de caracterizar el complemento genético del LUCA, comparando algunos genomas de organismos que se considera pertenecen a los primeros taxa, se encuentra que la mayoría de las proteínas más antiguas pertenecen tanto al metabolismo de nucleótidos como al procesamiento de RNA (ATPasas, helicasas tipo DEAD y proteínas ribosomales); de manera que puede inferirse que este tipo de moléculas fueron primordiales en la evolución temprana de la vida³⁸. Dado que varias de las proteínas encontradas contienen regiones codificadas por tripletes RNY, puede indicarse que no solamente son previas a LUCA

sino que su origen se remonta a los albores del Mundo de las RNPs y por ende los módulos que interaccionaban con nt son de los primeros en emerger durante la transición de los procesos prebióticos a los procesos bióticos. Cabe señalar que varios de los genes que previamente se consideraron dispensables no lo son en realidad, pues las condiciones ambientales en que esto se determinaba eran las óptimas para el organismo en cuestión (suficientes recursos administrados de manera constante) y por otro lado hay genes que, dependiendo del ambiente en que se encuentre el organismo y por ende el tipo de metabolismo que se lleve a cabo, pueden ser dispensables o no, todo ello aunado a que hay técnicas cada vez más sensibles y precisas para la detección de genes y comparación de genomas¹⁴¹⁹.

Cuando se analiza el genoma mínimo en términos de los dominios de proteínas, se encuentra que LUCA contenía alrededor de 140, de los cuales aquellos relacionados con la unión de nucleótidos y moléculas nucleotídicas como las coenzimas son de los primeros en surgir, así como también son también antiguos los dominios de unión a metales¹⁴²⁰; todo lo cual concuerda con los resultados de la presente tesis, pues se muestra que varias de las enzimas que contienen regiones codificadas por tripletes RNY, son muy probablemente los que unen ese tipo de moléculas.

Al analizar minuciosamente **la secuencia 'per se' de varios** genomas es posible observar que existen fragmentos totalmente invariantes¹⁴²¹, se puede señalar que varios de los aa conservados en esos segmentos son codificados por tripletes RNY y están arreglados de manera semejante a algunos de los motivos identificados probablemente por la metodología descrita en la presente tesis; adicionalmente se encuentra que los patrones de plegamiento se conservan y ello es posiblemente más importante que la conservación misma de nt o aa¹⁴²²; en ambos casos se observa que las secuencias conservadas miden ~11 aa o de 25 a 30 aa de longitud y en cualquier caso se señala que todas las proteínas han evolucionado por

modularidad. Todo ello refuerza con mayor claridad los resultados de la presente tesis, porque efectivamente la región que se identifica como ancestral en cada proteína no sobrepasa los 50aa y éstos corresponden a módulos o dominios que unen determinadas moléculas de síntesis prebiótica.

Posteriormente, se analizaron proteomas (colección de proteínas) representativos de diversos filos, de los que se detectó que existen 27 motivos, de hexámeros a nonámeros, totalmente invariantes; esos motivos están presentes en proteínas estrictamente conservadas en todos los organismos, sus actividades no son metabólicas y en realidad se restringen a una de tres distintas: suministro de monómeros (transportadores ABC y proteasas), síntesis de proteínas (factores de iniciación y de elongación de la traducción, así como aaRS) y síntesis de RNA (polimerasas), de manera que se consideran las proteínas más antiguas; algunos de los motivos invariantes se encuentran también en proteínas que se catalogan como metabólicas, aunque no están presentes en todos los organismos y el motivo en realidad no incide directamente en el metabolismo, tal es el caso de los dominios de unión a NTPs¹⁴²³. Es notable que no sólo las proteínas identificadas como ancestrales por este método de comparación de proteomas también se consideren antiguas bajo la hipótesis desarrollada en esta tesis, sino que numerosos aa presentes en los dominios identificados como invariantes son codificados por tripletes RNY, incluyendo los motivos de unión a nucleótidos.

METODOLOGÍA 'POR CODONES RNY'

CONTRIBUCIÓN FUNDAMENTAL DE LA TESIS.

Para comenzar, la presente tesis no se basa en el desarrollo de un "genoma mínimo", pues estrictamente un genoma incluye las regiones regulatorias y los genes que no codifican para proteínas y no obstante tienen funciones primordiales en los organismos, de manera que la mayoría de las veces nos estamos refiriendo en realidad al proteoma mínimo de los organismos pues al realizar las comparaciones se toma en cuenta únicamente la parte que codifica para proteínas; razón por la cual he denominado Proteoma Ancestral al conjunto de resultados obtenidos en la presente tesis. Es necesario señalar además que no existe una sinonimia directa entre "genoma mínimo" determinado experimentalmente y el contenido de genes de LUCA o algún organismo o sistema previo; primeramente no existe aún un consenso acerca de las características de aquel organismo ancestral y si fue un organismo o una comunidad; segundo porque el conjunto mínimo de genes identificados, sirve para el desarrollo de un organismo en condiciones ideales de crecimiento en laboratorio, que contrastan directamente con las condiciones sub-óptimas en que se desarrollan los organismos la mayor parte del tiempo y que eran muy probablemente las características ambientales de la Tierra primitiva; tercero porque no se han hecho pruebas generales, más allá de los arreglos filogenéticos basados en comparación de decenas de secuencias, del probable origen temporal de todas las proteínas que se asume son antiguas. Por todo lo anterior, cobra importancia la metodología descrita en la presente tesis, ya que utilizando un genoma ancestral reconstruido se obtienen proteínas que con mayor certeza pueden considerarse ancestrales, puesto que contienen regiones codificadas en los tripletes correspondientes al primer código genético⁶⁰.

La ecléctica colección de proteínas identificadas, por la metodología aquí descrita, se corresponde con diversas rutas metabólicas esenciales para la sobrevivencia de un organismo contemporáneo; no obstante ello no implica que en el riboproteoorganismo existiesen los productos metabólicos finales, pues ninguna ruta metabólica o de procesamiento estaba completa y de dichas proteínas sólo se identificó un fragmento, que dado que el patrón RNY se considera el código genético ancestral, es probable que constituya no solamente la región más conservada en cada péptido sino parte fundamental de cada una de las proteínas que se señalan en la presente tesis.

Al observar los resultados que arrojan diversos métodos, con relación al origen evolutivo de dominios, proteínas o familias, se advierte que varios de los elementos que han sido propuestos como parte del repertorio de LUCA son además codificados parcialmente por tripletes RNY, lo que no sólo sugiere que son ancestrales sino que su origen se remonta a los albores del Mundo de las RNPs; tal es el caso de proteínas completas como los transportadores ABC o la ATP sintasa, o bien dominios e incluso motivos como los de unión a nucleótidos o a ciertos cationes metálicos. Ahora bien, como podemos observar en las tablas que se presentan a manera de resumen al final de cada categoría, la región que probablemente se identificó como ancestral en cada una de las proteínas mencionadas, aún en las putativas y lo considerado hipotético, se encargan precisamente o de la unión de moléculas que tienen componentes prebióticos, como lo son nucleótidos en forma de RNA o cofactores, cationes metálicos, o bien conforman estructuras que sirven de andamiaje a otras proteínas, o forman regiones hidrófobas que interaccionan con lípidos, como los que constituían la proto-membrana celular; por supuesto que es también importante conceder que posiblemente algunas de las proteínas no contienen regiones ancestrales, sino que únicamente contienen aa codificados por tripletes RNY debido a la acumulación de mutaciones y en tal caso puede decirse

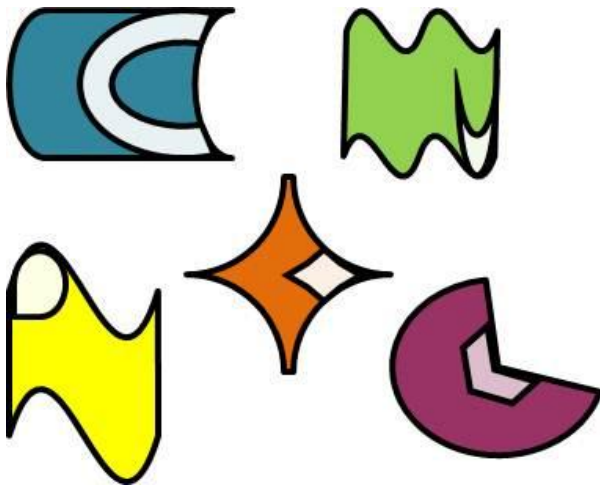
que la metodología descrita en la presente tesis no es lo suficientemente eficiente para discernir ese tipo de eventualidades, lo que no le resta validez como un método importante en la caracterización de secuencias.

En otras palabras, el análisis presentado en esta tesis, con el fin de sustentar las demostraciones teóricas de que el RNY es el código genético primitivo, arroja en lo general que aun cuando más de 100 proteínas contienen alguna región codificada precisamente por tripletes RNY, tales sitios son pequeños y corresponden a uno de sólo 5 tipos distintos de dominios o motivos y, no obstante, su importancia radica en que todos ellos unen moléculas prebióticas y que por ende ya existían en el Mundo de las RNPs.

Recientemente se ha generado una nueva base de datos que retoma las distintas metodologías de comparación masiva de genomas (como las mencionadas **y otras) utilizando las bases de datos existentes 'UniProt', 'ByoCyc' y 'KEGG'; de ese** análisis se obtienen varias conclusiones relevantes: 1) La hipótesis del Mundo del RNA predice que las ribozimas llevaban a cabo la mayoría de las funciones enzimáticas esenciales, antes de la aparición de la traducción; 2) Las funciones enzimáticas que requieren cofactores de nucleótidos puede reflejar la transición de un sistema basado sólo en RNA a un sistema de ribozimas y proteínas; 3) Los aa utilizados inicialmente como cofactores de las ribozimas, pudieron haber desempeñado un papel temprano en la transición hacia un sistema de enzimas proteicas; 4) Los cofactores FeS fueron también importantes desde el inicio de los procesos bióticos, debido a que las superficies minerales de sulfuro de hierro facilitaban la producción de pequeñas moléculas y su polimerización; 5) Los cofactores de Zn pudieron ser esenciales para iniciar prebióticamente los procesos de la química de ácidos nucleicos y de la producción de energía¹⁴²⁴.

Pese a que la base de datos referida (LUCApedia) está enfocada en la reconstrucción de la naturaleza del LUCA, las conclusiones generales nos refieren a

un periodo evolutivo mucho más antiguo; de hecho podemos observar que son reforzados los resultados de la presente tesis, ya que los cofactores mencionados, que iniciaron la transición de un ambiente prebiótico hacia los procesos bióticos, son justamente las moléculas que unen las regiones codificadas por tripletes RNY en cada una de las proteínas identificadas por la metodología descrita en la presente tesis, por lo que podemos llamar a ese primordio el '**sitio de unión y estabilización del cofactor**' (**CSBS**, por las siglas en inglés de '**cofactor stabilizing binding site**') y referirnos como **bindoma** (préstamo del inglés 'binding' o unión) a la colección de todos esos **CSBS**s antiguos (artículo enviado a la revista 'OLEB'), cuya representación se muestra en la siguiente figura (F13).



F13. Representación de los distintos tipos de proteínas, donde el '**sitio de unión y estabilización del cofactor**' (**CSBS**) –el primordio, codificado por tripletes RNY–, se indica por una región más clara en cada figura, *i.e.* **bindoma** ancestral.

Empero, es notable que muy probablemente dicho fragmento, codificado por tripletes RNY, constituye el primordio de cada una de las proteínas identificadas, dichos primordios surgieron en los albores del Mundo de las RNPs, fueron regiones fuertemente seleccionadas positivamente durante el proceso evolutivo y a partir de ellos se generaron las proteínas como actualmente las conocemos; es decir, las proteínas habrían comenzado como péptidos poco específicos que unían una amplia gama de moléculas prebióticas, esos distintos primordios coexistieron simultáneamente e interaccionaron de inmediato (dejando fuera la necesidad de

recurrir a teorías del tipo panspermia que de cualquier manera no explican cómo habría surgido la vida en un remoto mundo); de esta manera la vida rápidamente evolucionó de menor a mayor complejidad, en el tiempo estimado inicialmente y, aunque mediante mecanismos aún por determinar, con una fuerte presión de selección en los módulos primordiales.

VIII. CONCLUSIONES y PROSPECTIVAS

La tesis presentada se basa en la búsqueda de vestigios de proteínas ancestrales a partir del genoma actual de un organismo, bajo la hipótesis de que el primer código genético se basaba en tripletes RNY.

CONCLUSIONES.

Las conclusiones más relevantes, con base en los resultados analizados, son las siguientes:

A) La mayoría de las proteínas que contienen regiones codificadas por tripletes RNY han sido reportadas, mediante otras metodologías, como parte del repertorio proteico de LUCA e inclusive de organismos previos, como los progenotes¹⁴²⁵.

B) Es posible encontrar proteínas ancestrales en genomas (ORFeomas) actuales, ya que parte de ellas está codificada por los codones primordiales RNY.

C) La consecución de tripletes RNY codifica para péptidos, hoy embebidos como dominios en proteínas, que unían cofactores de origen prebiótico y han sido fundamentales durante la evolución de los seres vivos.

D) La porción codificada por tripletes RNY en cada proteína constituye, no sólo un vestigio, sino el primordio a partir del cual evolucionaría la forma actual de cada una; lo que valida el hallazgo teórico de la invariancia de escala de un genoma RNY con respecto al genoma original de cualquier organismo⁶⁷.

E) La metodología descrita en la presente tesis se puede utilizar entonces para rastrear vestigios de varias proteínas, que al mismo tiempo constituyen la porción más antigua de cada una de ellas.

F) No obstante la aparente amplia diversidad de proteínas identificadas, existen pocos módulos primordiales fundamentales, cuyo origen se remonta a los albores del Mundo de las RNPs, a partir de los cuales surgen las proteínas actuales.

G) El origen y evolución de la vida ha sido modular y varios de esos módulos coexistieron simultáneamente e interaccionaron de inmediato aunque es posible reconocer sus primordios desde hace 3.7 mil millones de años de evolución; lo que indica que, al menos antes de la aparición del LUCA, los procesos macro-evolutivos no fueron necesariamente dictados por mecanismos de selección natural aunque sí posiblemente los micro-evolutivos en un contexto en que la transferencia horizontal de genes (HGT) era crucial^{cf. 1426, 1427}.

H) Las diversas interacciones entre aa y nt primigenios habrían generado los **CSBSs**, lo que marca la diferencia entre los procesos abióticos y los bióticos, por lo que el inicio del Mundo de las RNPs supone el inicio formal del código genético.

J) El trabajo mostrado constituye entonces un esfuerzo para dilucidar el proceso evolutivo que habría ocurrido entre los procesos de química prebiótica y la aparición del LUCA, con una temporalidad más cercana al origen de la vida misma.

PROSPECTIVAS.

El siguiente paso es realizar el algoritmo presentado en varios procariontes incluyendo bacterias y arqueas (tanto G+ como G-, parásitos y organismos de vida libre), con el fin de eliminar la mayor cantidad posible de sesgos metodológicos impuestos por las características particulares de cada organismo; del mismo modo que pudieran averiguarse si es posible identificar el mismo tipo de proteínas antiguas con otro tipo de metodologías que también tomen en cuenta la codificación primordial en tripletes RNY.

IX. APÉNDICE

Programa que llama al algoritmo BLAST en su versión instalada, escrito en Perl, para obtener los datos a partir de los cuales se desarrolló la presente tesis:

```
#!/usr/bin/perl

$full_command_name = $0;
if ($full_command_name =~ m/^(^\/)((^\/]*)$/) {
    $my_name = $2;
}
else {
    $my_name = 'MyBlastParser.perl';
}

if ( $ARGV[0] =~ /^-[hH?]/ ) {
    &print_more_help('');
}

( $outname, $ext ) = split /\./, $ARGV[0];
open ( OUT,">$outname.out" );

$/ = "\n>";
open ( BLAST,"$ARGV[0]" ) or die "\n\n\t", 'Could not open ',
$ARGV[0], ': ', $!, "\n\n";
while ( <BLAST> ) {

    if ( $_ =~ /^BLAST\w/ ) {
        @lines = split /\n/, $_;

        foreach $line( @lines ) {
            $line =~ /^Query= .*\|(.*)$/ and $qname = $1;
        }
    }

    $_ =~ /^gb\|.*\|:(.*)\n/ and $id = $1;
    # $_ =~ /^.*\|:(.*)\n/ and $id = $1;

    if ( defined $id ) {
        print "\t\t\t\t\tid es: $id\n";
        @fields = ();
        @fields = split /\n\n\n\n/, $_;

        foreach ( @fields ) {

            /Length = (\d+)\n/ and $tl = $1;
            print "el largo de la secuencia blanco es $tl\n";
            if ( /Score = .*?bits \((\S+)\)\.+Expect.*? =\s+
(\S+)/ ) {
                $bscore = $1;
                $Eval = $2;
            }
            /Identities = (\d+\/\d+\s\((\d+\%\))/ and $idts = $1;

            #####
            ## init and end of alignments
            #####

            /Query:\s+(\d+)/ and $hinit = $1;
        }
    }
}

```

X. REFERENCIAS BIBLIOHEMEROGRÁFICAS

- ¹ http://map.gsfc.nasa.gov/universe/uni_age.html
- ² MILLER SL *et al.* (1997), *Oparin's "Origin of Life": sixty years later*, **J. Mol. Evol.** **94**, 351-353.
- ³ LAZCANO A (2008), *What is life? A brief historical overview*, **Chem. Biodivers.** **5**, 1-15.
- ⁴ BARTON NH *et al.*, **Evolution**, Cold Spring Harbor Laboratory Press, USA, 2007.
- ⁵ LAZCANO (2010), *Historical development of origins research*, **Cold Spring Harb. Perspect. Biol.** **2**, a002089.
- ⁶ MILLER SL *et al.* (1997), *Oparin's "Origin of Life": sixty years later*, **J. Mol. Evol.** **94**, 351-353.
- ⁷ CHEN IA *et al.*, *Protocells: genetic polymers inside membrane vesicles* en **The RNA World**, Gesteland RF, Cech TR & Atkins JF (edit.), 3rd edition, CSHL Press, USA, 2006.
- ⁸ JANAS T *et al.*, *RNA, lipids and membranes* en **The RNA World**, Gesteland RF, Cech TR & Atkins JF (edit.), 3rd edition, CSHL Press, USA, 2006.
- ⁹ SEGRÉ D *et al.* (2002) *The lipid world*, **Orig. Life. Evol. Biosph.** **31**, 119-145.
- ¹⁰ SZATHMÁRY E (2007), *Coevolution of metabolic networks and membranes: the scenario of progressive sequestration*, **Phil. Trans. R. Soc. B** **362**, 1781-1787.
- ¹¹ WÄCHTERSCHAUSER G (1988) *Before Enzymes and Templates: Theory of Surface Metabolism*, **Microb. Rev.** **52**, 452-484.
- ¹² WHITE HB 3rd (1975), *Coenzymes as fossils of an earlier metabolic state*, **J. Mol. Evol.** **7**, 101-104.
- ¹³ KING GAM (1980), *Evolution of the coenzymes*, **Biosystems** **13**, 23-45.
- ¹⁴ WÄCHTERSCHAUSER G (2006) *From volcanic origins of chemoautotrophic life to Bacteria, Archaea and Eukarya*, **Phil. Trans. R. Soc. B** **361**, 1787-1808.
- ¹⁵ <http://www.ncbi.nlm.nih.gov/mesh?term=Aptamers>
- ¹⁶ HUANG F & Yarus M *et al.* (1997), *Versatile 5' phosphoryl coupling of small and large molecules to an RNA*, **PNAS** **94**, 8965-8969.
- ¹⁷ HUANG F *et al.* (2000), *RNA-catalyzed CoA, NAD, and FAD synthesis from phosphopantetheine, NMN, and FMN*, **Biochemistry** **39**, 15548-15555.
- ¹⁸ HUANG F (2003), *Efficient incorporation of CoA, NAD and FAD into RNA by "in vitro" transcription*, **Nuc. Acids Res** **31**, e8.
- ¹⁹ SHAROV A (2009), *Coenzyme autocatalytic network on the surface of oil microspheres as a model for the origin of life*, **Int. J. Mol. Sci** **10**, 1838-1852.
- ²⁰ JOYCE GF & LE Orgel, *Progress toward understanding the origin of the RNA world* en **The RNA World**, Gesteland RF, Cech TR & Atkins JF (edit.), 3rd edition, CSHL Press, USA, 2006.
- ²¹ HOUGLAND JL *et al.*, *How the group I intron works: a case study of RNA structure and function* en **The RNA World**, Gesteland RF, Cech TR & Atkins JF (edit.), 3rd edition, CSHL Press, USA, 2006.
- ²² GILBERT W (1986), *The RNA world*, **Nature** **319**, 618.
- ²³ GESTELAND RF, TR Cech & JF Atkins (edit.), **The RNA World**, 3rd edition, CSHL Press, USA, 2006, Caps. 1-10 y 22-25.
- ²⁴ MOORE PB & TA Steitz, *The roles of RNA in the synthesis of protein* en **The RNA World**, Gesteland RF, Cech TR & Atkins JF (edit.), 3rd edition, CSHL Press, USA, 2006.
- ²⁵ FORTERRE P (2005), *The two ages of the RNA world, and the transition to the DNA world: a story of viruses and cells*, **Biochimie** **87**, 793-803.
- ²⁶ KOONIN EV *et al.* (2006), *The ancient Virus World and evolution of cells*, **BMC Biol. Direct** **1**: 29.
- ²⁷ CECHE *et al.*, *The RNP world* en **The RNA World**, Gesteland RF, Cech TR & Atkins JF (edit.), 3rd edition, CSHL Press, USA, 2006.
- ²⁸ SZATHMÁRY E (1993), *Coding coenzyme handles: A hypothesis for the origin of the genetic code*, **PNAS** **90**, 9916-9920.
- ²⁹ DI GIULIO M (1997), *On the RNA World: Evidence in Favor of an Early Ribonucleopeptide World*, **J. Mol. Evol.** **45**, 571-578.
- ³⁰ JESTIN JL & A Kempf (2009), *Optimization models and the structure of the genetic code*, **J. Mol. Evol.** **69**, 452-457.
- ³¹ LAZCANO A & P Forterre (1999), *The molecular search for the Last Common Ancestor*, **J. Mol. Evol.** **49**, 411-412.
- ³² MAT WK *et al.* (2008), *The genomics of LUCA*, **Front. Biosci.** **13**, 5605-5613.
- ³³ MADIGAN CMT & JM Martinko, **Brock, Biology of microorganisms**, 11th edition, Pearson & Prentice Hall, USA, 2006.
- ³⁴ WOESE C & GE Fox (1977), *Phylogenetic structure of the prokaryotic domain: The primary kingdoms*, **PNAS** **74**, 5088-5090.
- ³⁵ LAZCANO A & P Forterre (1999), *The molecular search for the Last Common Ancestor*, **J. Mol. Evol.** **49**, 411-412.
- ³⁶ KOONIN EV (2003), *Comparative genomics, minimal gene-sets and the last universal common ancestor*, **Nat. Rev. Microbiol.** **1**, 127-136.
- ³⁷ FORTERRE p & S Gribaldo (2007), *The origin of modern terrestrial life*, **HFSP J.** **1**, 156-68.

- ³⁸ DELAYE L *et al.* (2005), *The Last Common Ancestor: What's in a name?*, **Orig. Life Evol. Biosph.** **35**, 537-554.
- ³⁹ GLANDSORF N *et al.* (2009) *The origin of life and last universal common ancestor: do we need a change of perspective?*, **Res. Microbiol.** **160**, 522-528.
- ⁴⁰ FORTERRE P & H Philippe (1999), *The Last Universal Common Ancestor (LUCA), simple or complex?*, **Biol. Bull.** **196**, 373-377.
- ⁴¹ POOLE AM & DT Logan (2005), *Modern mRNA proofreading and repair: clues that the Last Universal Common Ancestor possessed an RNA genome?*, **Mol. Biol. Evol.** **22**, 1444-1455.
- ⁴² PENNY D & A Poole (1999), *The nature of the last universal common ancestor*, **Curr. Opin. Genet. Dev.** **9**, 672-677.
- ⁴³ HOENIGSBERG H (2003), *Evolution without speciation but with selection: LUCA, the Last Universal Common Ancestor in Gilbert's RNA world*, **Genet. Mol. Res.** **4**, 366-375.
- ⁴⁴ WONG Tze-Fei J *et al.* (2007), *Polyphasic evidence delineating the root of life and roots of biological domains*, **Gene** **403**, 39-52.
- ⁴⁵ MORANGE M (2011), *Some considerations on the nature of LUCA, and the nature of life*, **Res. Microbiol.** **162**, 5-9.
- ⁴⁶ WOESE C (1998), *The universal ancestor*, **PNAS** **95**, 6854-6859.
- ⁴⁷ WOESE C (2000), *Interpreting the universal phylogenetic tree*, **PNAS** **97**, 8392-8396
- ⁴⁸ FOURNIER GP *et al.* (2009), *Horizontal gene transfer from extinct and extant lineages: biological innovation and the coral of life*, **Phil. Trans. R. Soc. B** **364**, 2229-2239
- ⁴⁹ ISENBARGER TA *et al.* (2008), *The most conserved genome segments for life detection on Earth and other planets*, **Orig. Life Evol. Biosph.** **38**, 517-533.
- ⁵⁰ DOOLITTLE WF & JR Brown (1994), *Tempo, mode, the progenote, and the universal root*, **PNAS** **91**, 6721-6728.
- ⁵¹ GLANSDORFF N *et al.* (2008), *The Last Universal Common Ancestor: emergence, constitution and genetic legacy of an elusive forerunner*, **BMC Biol. Dir.** **3**, 29.
- ⁵² LAZCANO A (2010), *Which way to life?*, **Orig. Life. Evol. Biosph.** **40**, 161-167.
- ⁵³ CRICK FHC (1968), *The Origin of the Genetic Code*, **J. Mol. Biol.** **38**, 367-379.
- ⁵⁴ <http://www.ncbi.nlm.nih.gov/Taxonomy/taxonomyhome.html/index.cgi?chapter=cgencodes#SG1>
- ⁵⁵ OSAWA S *et al.* (1992), *Recent evidence for evolution of the genetic code*, **Microbiol. Rev.** **56**, 229-64.
- ⁵⁶ JUKES TH & S Osawa (1993), *Evolutionary changes in the genetic code*, **Comp. Biochem. Physiol. B.** **106**, 489-94.
- ⁵⁷ BROWN TA, **Genomes**, 3rd edition, Garland Science, USA, 2006.
- ⁵⁸ CRICK FHC *et al.* (1976), *A speculation on the origin of protein synthesis*, **Orig. Life** **7**, 389-397.
- ⁵⁹ WATANABE F *et al.* (2006), *Sobre el origen del código genético*, **Rev. Del Centro de Inv. (Méx.)** **7**, 93-110.
- ⁶⁰ EIGEN M & P Schuster (1978), *The Hypercycle, A Principle of Natural Self-Organization Part C: The Realistic Hypercycle*, **Die Naturwissenschaften** **65**, 341-369.
- ⁶¹ LEHMAN J (2002), *Amplification of the sequences displaying the pattern RNY in the RNA world: the translation --> translation/replication hypothesis*, **J. Theor. Biol.** **219**, 521-537.
- ⁶² SHEPHERD J (1981), *Periodic correlations in DNA sequences and evidence suggesting their evolutionary origin in a comma-less genetic code*, **J. Mol. Evol.** **17**, 94-102.
- ⁶³ SHEPHERD J (1981), *Method to determine the reading frame of a protein from the purine/pyrimidine genome sequence and its possible evolutionary justification*, **PNAS** **78**, 1596-1600.
- ⁶⁴ EIGEN M *et al.* (1985), *Pattern analysis of 5S rRNA*, **PNAS** **82**, 2437-2442.
- ⁶⁵ WONG Tze-Fei J & R Cedergren (1986), *Natural selection versus primitive gene structure as determinant of codon usage*, **Eur. J. Biochem.** **159**, 175-180.
- ⁶⁶ JUKES TH (1996), *On the prevalence of certain codons ("RNY") in genes for proteins*, **J. Mol. Evol.** **42**, 377-381.
- ⁶⁷ JOSÉ MV, T Govezensky, JA García, JR Bobadilla (2009), *On the evolution of the standard genetic code: vestiges of critical scale invariance from the RNA world in current prokaryote genomes*, **PLoS One** **4**, e4340.
- ⁶⁸ LÓPEZ-VILLASEÑOR I, MV José, J Sánchez (2004), *Three-base periodicity patterns and self-similarity in whole bacterial chromosomes*, **Biochem. Biophys. Res. Comm.** **325**, 467-478.
- ⁶⁹ GARCIA JA & MV José (2005), *Mathematical properties of DNA sequences from coding and noncoding regions*, **Rev. Mex. Fis.** **51**, 122-130.
- ⁷⁰ JOSÉ MV, T Govezensky, JR Bobadilla (2005), *Statistical properties of DNA sequences revisited: the role of inverse bilateral symmetry in bacterial chromosomes*, **Physica A** **351**, 477-498.
- ⁷¹ JOSÉ MV, ER Morgado, T Govezensky (2007), *An extended RNA code and its relationship to the standard genetic code: an algebraic and geometrical approach*, **Bull. Math. Biol.** **69**, 215-243.
- ⁷² JOSÉ MV, ER Morgado, T Govezensky (2011), *Genetic Hotels for the Standard Genetic Code: Evolutionary Analysis Based upon Novel Three-Dimensional Algebraic Models*, **Bull. Math. Biol.** **73**, 1443-1476.
- ⁷³ LAGUNA J & E Piña, **Bioquímica**, 4^a edición, JGH editores, México, 2002.

- ⁷⁴ TRIFONOV EN (2009), *The origin of the genetic code and of the earliest oligopeptides*, **Res. Microbiol.** **160**, 481-486.
- ⁷⁵ TRIFONOV EN (2004), *The triplet code from first principles*, **J. Biomol. Struct. Dyn.** **22**, 1-11.
- ⁷⁶ EIGEN M *et al.* (1989), *How old is the genetic code? Statistical geometry of tRNA provides an answer*, **Science** **244**, 673-679.
- ⁷⁷ MILLER SA (1953), A production of amino acids under possible primitive earth conditions, *Science* **117**, 528-529
- ⁷⁸ WONG Tze-Fei J (1975), *A Co-Evolution Theory of the Genetic Code*, **PNAS** **72**, 1909-1912.
- ⁷⁹ TAYLOR FJ & D Coates (1989), *The code within the codons*, **Biosystems** **22**, 177-87.
- ⁸⁰ DAVIS BK (1999), *Evolution of the genetic code*, **Prog. Biophys. Mol. Biol.** **72**, 157-243.
- ⁸¹ DI Giulio M (2008), *An extension of the coevolution theory of the origin of the genetic code*, **Biol Direct.** **3**:37
- ⁸² IKEHARA K (2009), *Pseudo-Replication of [GADV]-Proteins and Origin of Life*, **Int. J. Mol. Sci.** **10**, 1525-1537.
- ⁸³ GUIMARÃEZ RC *et al.* (2008), *A self-referential model for the formation of the genetic code*, **Theory Biosci.** **127**, 249-270.
- ⁸⁴ DELARUE M (2007), *An asymmetric underlying rule in the assignment of codons: possible clue to a quick early evolution of the genetic code via successive binary choices*, **RNA** **13**, 161-169.
- ⁸⁵ RODIN SN & AS RODIN (2008), *On the origin of the genetic code: signatures of its primordial complementarity in tRNAs and aminoacyl-tRNA synthetases*, **Heredity** **100**, 341-355.
- ⁸⁶ ARQUÉS DG *et al.* (1995), *Identification and simulation of shifted periodicities common to protein coding genes of eukaryotes, prokaryotes and viruses*, **J. Theor. Biol.** **172**, 279-91.
- ⁸⁷ <http://www.ncbi.nlm.nih.gov/genomeprj?Db=genomeprj&cmd=ShowDetailView&TermToSearch=326>
- ⁸⁸ TETTELIN H *et al.* (2005), *Genome analysis of multiple pathogenic isolates of Streptococcus agalactiae: implications for the microbial "pan-genome"*, **PNAS** **102**, 13950-13955.
- ⁸⁹ MEDINI D *et al.* (2005), *The microbial pan-genome*, **Curr. Opin. Genet. Dev.** **15**, 589-594.
- ⁹⁰ TETTELIN H *et al.* (2008), *Comparative genomics: the bacterial pan-genome*, **Curr. Opin. Microbiol.** **11**, 472-477.
- ⁹¹ <http://www.ncbi.nlm.nih.gov/COG/grace/fiew.cgi>
- ⁹² <http://www.genome.jp/kegg/pathway.html>
- ⁹³ ALBERTS B *et al.*, **Molecular Biology of the Cell**, 4th edition, Garland Science, USA, 2002.
- ⁹⁴ NELSON D & M Cox, **Lehninger, Principles of Biochemistry**, WH Freeman, 4th edition, USA, 2004.
- ⁹⁵ LEWIN B, **Genes VIII**, 8th edition, Pearson & Prentice Hall, USA, 2005.
- ⁹⁶ <http://www.ncbi.nlm.nih.gov/proteinclusters>
- ⁹⁷ <http://www.ncbi.nlm.nih.gov/BLAST/tutorial/Altschul-1.html>
- ⁹⁸ ANANTHARAMAN V *et al.* (2002), *Comparative genomics and evolution of protein involved in RNA metabolism*, **Nucleic Acids Res.** **30**, 1427-1464.
- ⁹⁹ HOOGSARTEN CG & M Sumita (2007), *Structure-function relationships in RNA and RNP enzymes: recent advances*, **Biopolymers** **87**, 318-328.
- ¹⁰⁰ KACZANOWSKA M & M Rydén-Aulin (2007), *Ribosome biogenesis and the translation process in Escherichia coli*, **Microb. Mol. Rev.** **71**, 477-494.
- ¹⁰¹ MEARS JA *et al.* (2002), *Modelling a minimal ribosome based on the comparative sequence analysis*, **J. Mol. Biol.** **321**, 325-234.
- ¹⁰² LECOMPTE O *et al.* (2002), *Comparative analysis of ribosomal proteins in complete genomes: an example of reductive evolution at the domain scale*, **Nucleic Acids Res.** **30**, 5382-5390.
- ¹⁰³ MAGUIRE BA & RA Zimmermann (2001), *The ribosome in focus*, **Cell** **102**, 813-816.
- ¹⁰⁴ BRODERSEN DE *et al.* (2002), *Crystal structure of the 30 S ribosomal subunit from Thermus thermophilus: structure of the proteins and their interactions with 16 S RNA*, **J. Mol. Biol.** **316**, 725-768.
- ¹⁰⁵ YAGUCHI M (1975), *Primary structure of ribosomal protein S18 from the small Escherichia coli ribosomal subunit*, **FEBS Lett.** **59**, 217-220.
- ¹⁰⁶ GREGORY RJ *et al.* (1984), Interaction of ribosomal proteins S6, S8, S15 and S18 with the central domain of 16S ribosomal RNA from Escherichia coli, *J. Mol. Biol.* **178**, 287-302.
- ¹⁰⁷ MC DOUGALL J *et al.* (1989), *The complete amino acid sequence of ribosomal protein S18 from the moderate thermophile Bacillus stearothermophilus*, **FEBS Lett.** **245**, 253-260.
- ¹⁰⁸ ZH G *et al.* (2001), *The path of messenger through the ribosome*, **Cell** **106**, 233-241.
- ¹⁰⁹ DZIONARA M *et al.* (1977), *Prediction for secondary structures of ten proteins from the 50S subunit of the Escherichia coli ribosome*, **J. Supramol. Struct.** **7**, 191-204.
- ¹¹⁰ WOWER I *et al.* (1981), *The use of 2-iminothiolane as an RNA-protein cross-linking agent in Escherichia coli ribosomes, and the localisation on 23S RNA of sites cross-linked to proteins L4, L6, L21, L23, L27 and L29*, **Nucleic Acids Res.** **9**, 4285-4302.
- ¹¹¹ GOLDEN BL *et al.* (1993), *Ribosomal protein L6: structural evidence of gene duplication from a primitive RNA binding protein*, **EMBO J.** **12**, 4901-4908.
- ¹¹² DAVIES C *et al.* (1998), *Ribosomal proteins S5 and L6: high-resolution crystal structures and roles in protein synthesis and antibiotic resistance*, **J. Mol. Biol.** **279**, 873-888.

- ¹¹³ UCHIUMI T *et al.* (1999), *Interaction of the sarcin/ricin domain of 23 S ribosomal RNA with proteins L3 and L6*, **J. Biol. Chem.** **274**, 981-686.
- ¹¹⁴ BEN N *et al.* (1999), *Placement of protein and RNA structures into a 5 A-resolution map of the 50S ribosomal subunit*, **Nature** **400**, 841-857.
- ¹¹⁵ CSEMERLY P (1997), *Proteins, RNAs and chaperones in enzyme evolution: a folding perspective*, **TIBS** **22**, 147-149.
- ¹¹⁶ LINDER P (2000), *Quick guide: DEAD-box proteins*, **Curr. Biol.** **10**, R887.
- ¹¹⁷ TANNER NK & P Linder (2001), *DEAD/H box RNA helicases: from generic motors to specific dissociation functions*, **Mol. Cell** **8**, 251-262.
- ¹¹⁸ ROCAK S & P Linder (2004), *DEAD-box proteins: the driving forces behind RNA metabolism*, **Nat. Rev. Mol. Cell Biol.** **5**, 232-241.
- ¹¹⁹ JANKOWSKY E & H Bowers (2006), *Remodeling of ribonucleoprotein complexes with DEH/D RNA helicases*, **Nucleic Acids Res.** **34**, 4181-4188.
- ¹²⁰ OWTTRIM GW (2006), *RNA helicases and abiotic stress*, **34**, **Nucleic Acids Res.** **34**, 3220-3230.
- ¹²¹ JANKOWSKY E & ME Fairman (2007), *RNA helicases- one fold for many functions*, *Curr. Opin. Struct. Biol.* **17**, 316-324.
- ¹²² CHEN Y *et al.* (2008), *DEAD-box proteins can completely separate an RNA duplex using a single ATP*, **PNAS** **105**, 20203-20208.
- ¹²³ LIU F *et al.* (2008), *ATP hydrolysis is required for DEAD-box protein recycling but not for duplex unwinding*, **PNAS** **105**, 20209-20214.
- ¹²⁴ ZHAO X & C Jain (2011), *DEAD-box proteins from Escherichia coli exhibit multiple ATP-independent activities*, **J. Bacteriol.** **193**, 2236-2241.
- ¹²⁵ LÓPEZ-RAMÍREZ V *et al.* (2011), *Phylogenetic distribution and evolutionary history of bacterial DEAD-Box proteins*, **J. Mol. Evol.** **72**, 413-431.
- ¹²⁶ SCHLUCKEBIER G *et al.* (1995), *Universal catalytic domain structure of AdoMet dependent methyltransferases*, **J. Mol. Biol.** **247**, 6-20.
- ¹²⁷ MARTIN J & FM MacMillan (2002), *SAM (dependent) I AM: the S-adenosylmethionine-dependent methyltransferase fold*, **Curr. Opin. Struct. Biol.** **12**, 783-793.
- ¹²⁸ SCHUBERT HL *et al.* (2003), *Many paths to methyltransfer: a chronicle of convergence*, **TRENDS Biochem. Sci.** **28**, 329-335.
- ¹²⁹ MARTÍNEZ-GIMÉNEZ JA *et al.* (1998), *On the function of modified nucleosides in the RNA world*, **J. Theor. Biol.** **194**, 485-490.
- ¹³⁰ CHOW C *et al.* (2007), *Expanding the nucleotide repertoire of the ribosome with post-transcriptional modifications*, **ACS Chem. Biol.** **2**, 610-619.
- ¹³¹ TSCHERNE JS *et al.* (1999), *Purification, cloning, and characterization of the 16S RNA m⁵C967 methyltransferase from Escherichia coli*, **Biochemistry** **38**, 1884-1892.
- ¹³² GU XR *et al.* (1999), *Identification of 16S RNA m⁵C967 methyltransferases from searching genomic and proteomic sequences*, **Nucleic Acids Res.** **27**, 3138-3145.
- ¹³³ REID R *et al.* (1999), *Exposition of a family of RNA m⁵C methyltransferases from Escherichia coli*, **Biochemistry** **38**, 1884-1892.
- ¹³⁴ BUJNICKI JM *et al.* (2004), *Sequence-structure-function studies of tRNA:m⁵C methyltransferase Trm4p and its relationship to DNA:m⁵C and RNA:m⁵U methyltransferases*, **Nucleic Acids Res.** **32**, 2453-2463.
- ¹³⁵ MOTORIN Y *et al.*, (2010), *5-methylcytosine in RNA: detection, enzymatic formation and biological functions*, **Nucleic Acids Res.** **38**, 1415-1430.
- ¹³⁶ ROMANOWSKI MJ *et al.*, (2002), *Crystal structure of the Escherichia coli glucose-inhibited division protein B (GidB) reveals a methyltransferase fold*, **Proteins** **47**, 563-567.
- ¹³⁷ NISHIMURA K *et al.* (2007), *Identification of the RsmG methyltransferase target as 16S rRNA nucleotide G527 and characterization of Bacillus subtilis rsmG mutants*, **J. Bacteriol.** **189**, 6068-6073.
- ¹³⁸ GREGORY ST *et al.* (2010), *Structural and functional studies of the Thermus thermophilus 16S rRNA methyltransferase RsmG*, **RNA** **15**, 1693-1704.
- ¹³⁹ HEURGUE-HAMARD V *et al.* (2002), *The hemK gene in Escherichia coli encodes the N(5)-glutamine methyltransferase that modifies peptide release factors*, **EMBO J.** **21**, 769-778.
- ¹⁴⁰ SCHUBERT HL *et al.* (2003), *Structures along the catalytic pathway of PrmC/HemK, an N5-glutamine AdoMet-dependent methyltransferase*, **Biochemistry** **42**, 5592-5599.
- ¹⁴¹ YANG Z *et al.* (2004), *Structural characterization and comparative Phylogenetic analysis o Structural characterization and comparative phylogenetic analysis of Escherichia coli HemK, a protein (N5)-glutamine methyltransferase*, **J. Mol. Biol.** **340**, 695-706.
- ¹⁴² GRAILLE M *et al.* (2005), *Molecular basis for bacterial class I release factor methylation by PrmC*, *Mol. Cell* **20**, 917-927.
- ¹⁴³ KOZBIAL PZ & AR Mushegian (2005), *Natural history of S-adenosylmethionine-binding proteins*, **BMC Struct. Biol.** **5**:19.
- ¹⁴⁴ LAURSEN R & L Duffy (1978), *The evolution of elongation factors Tu and G by gene duplication*, **FEBS Lett** **92**, 200-202.

- 145 COUSINEAU B *et al.* (1997), *On the origin of protein synthesis factors: a gene duplication/fusion model*, **J. Mol. Evol.** **45**, 661-670.
- 146 CZWORKOWSKI J & PB Moore (1997), *The conformational properties of elongation factor G and the mechanism of translocation*, **Biochemistry** **36**, 10327-10334.
- 147 RODNINA MV *et al.* (1997), *Hydrolysis of GTP by elongation factor G drives tRNA movement on the ribosome*, **Nature** **385**, 37-41.
- 148 PESKE F *et al.* (2004), *Conformational changes of the small ribosomal subunit during elongation factor G-dependent tRNA-mRNA translocation*, **J. Mol. Biol.** **343**, 1183-1194.
- 149 LI W *et al.* (2011), *Molecular dynamics of EF-G during translations*, **Proteins**, **79**, 1478-1486.
- 150 MARGUS T *et al.* (2007), *Phylogenetic distribution of translational GTPases in bacteria*, **BMC Genomics** **8**: 15.
- 151 JANOSI L *et al.* (1996), *Dual functions of ribosome recycling factor in protein biosynthesis: disassembling the termination complex and preventing translational errors*, **Biochemistry** **78**, 959-969.
- 152 FUJIWARA T *et al.* (2001), *Functional mapping of ribosome-contact sites in the ribosome recycling factor: a structural view from a tRNA mimic*, **RNA** **7**, 64-70.
- 153 HIROKAWA G *et al.* (2002), *Post-termination complex disassembly by ribosome recycling factor, a functional tRNA mimic*, **EMBO J.** **21**, 2272-2281.
- 154 AGRAWAL RK *et al.* (2004), *Visualization of ribosome-recycling factor on the Escherichia coli 70S ribosome: functional implications*, **PNAS** **101**, 8900-8905.
- 155 PAI RD *et al.* (2008), *Visualization of ribosome-recycling factor on the Escherichia coli 70S ribosome: functional implications*, **J. Mol. Evol.** **376**, 1334-1337.
- 156 ZAVIALOV AV *et al.* (2003), *Peptidyl-tRNA regulates the GTPase activity of translations factors*, **Cell** **114**, 113-122.
- 157 JOSEPH S (2003), *After the ribosome structure: How does translocation work?* **RNA** **9**, 160-164.
- 158 ITO K *et al.* (2002), *Elongation factor G participates in ribosome disassembly by interacting with ribosome recycling factor at their tRNA-mimicry domains*, **Mol. Cell** **9**, 1263-1272.
- 159 NAKAMURA Y & K Ito (2003), *Making sense of mimic in translation termination*, **TRENDS Biochem. Sci.** **28**, 99-105.
- 160 RUAN B *et al.* (2001), *Genomics and the evolution of aminoacyl-tRNA synthesis*, **Acta Biochim. Pol.** **48**, 313-321.
- 161 DELARUE M & D Moras (1993), *The aminoacyl-tRNA synthetase family: modules at work*, **Bioessays** **15**, 675-687.
- 162 BROWN JR & WF Doolittle (1995), *Root of the universal tree of life based on ancient aminoacyl-tRNA synthetase gene duplications*, **PNAS** **92**, 2441-2445.
- 163 WOESE C *et al.* (2000), *Aminoacyl-tRNA synthetases, the genetic code, and the evolutionary process*, **Microbiol. Mol. Biol. Rev.** **64**, 202-236.
- 164 CUSACK S *et al.* (1997), *Aminoacyl-tRNA synthetases*, **Curr. Opin. Struct. Biol.** **7**, 881-889.
- 165 DELARUE M (1995), *Aminoacyl-tRNA synthetases*, **Curr. Opin. Struct. Biol.** **5**, 48-55.
- 166 BEUNING PJ & K Musier-Forsyth (1999), *Transfer RNA recognition by aminoacyl-tRNA synthetases*, **Biopolymers** **52**, 1-28.
- 167 GEIGÉ R *et al.* (1998), *Universal rules and idiosyncratic features in tRNA identity*, **Nucleic Acids Res.** **26**, 5017-5035.
- 168 WOLF YI *et al.* (1999), *Evolution of aminoacyl-tRNA synthetases--analysis of unique domain architectures and phylogenetic trees reveals a complex history of horizontal gene transfer events*, **Genome Res.** **9**, 689-710.
- 169 IBBA M & D Söll (2000), *Aminoacyl-tRNA synthesis*, **Annu. Rev. Biochem.** **69**, 617-642.
- 170 YAMANE T & JJ Hopfield (1977), *Experimental evidence for kinetic proofreading in the aminoacyl of tRNA by synthetase*, **PNAS** **74**, 2246-2250.
- 171 SZYMAŃSKI M *et al.* (2000), *The new aspects of aminoacyl-tRNA synthetases*, **Acta Biochim. Pol.** **47**, 821-834.
- 172 FRANCKLYN C *et al.* (2002), *Aminoacyl-tRNA synthetases: versatile players in the changing theater of translation*, **RNA** **8**, 1363-1372.
- 173 FRANCKLYN C (2003), *tRNA synthetase paralogs: evolutionary links in the transition from tRNA-dependent amino acid biosynthesis to de novo biosynthesis*, **PNAS** **100**, 9650-9652.
- 174 SZYMAŃSKI M *et al.* (2001), *Aminoacyl-tRNA synthetases database*, **Nucleic Acids Res.** **29**, 288-290.
- 175 JAKÓ É *et al.* (2007), *In silico detection of tRNA sequence features characteristic to aminoacyl-tRNA synthetase class membership*, **Nucleic Acids Res.** **35**, 5593-5609.
- 176 O'DONOGHUE P *et al.* (2003), *On the evolution of structure in aminoacyl-tRNA synthetases*, **Microbiol. Mol. Biol. Rev.** **67**, 550-573.
- 177 SPRINGER M *et al.* (1985), *Autogenous control of Escherichia coli threonyl-tRNA synthetase expression in vivo*, **J. Mol. Biol.** **185**, 93-104.
- 178 ROMBY P *et al.* (1990), *The relation between catalytic activity and gene regulation in the case of E coli threonyl-tRNA synthetase*, **Biochimie** **72**, 485-494.

- 179 BRUNEL C *et al.* (1992), *Domains of the Escherichia coli threonyl-tRNA synthetase translational operator and their relation to threonine tRNA isoacceptors*, **J. Mol. Biol.** **227**, 621-634.
- 180 ROMBY P *et al.* (1992), *Molecular mimicry, in translational control of E. coli threonyl-tRNA synthetase gene. Competitive inhibition in tRNA aminoacylation and operator-repressor recognition switch using tRNA identity rules*, **Nucleic Acids Res.** **20**, 5633-5640.
- 181 SACERDOT C *et al.* (1998), *The Escherichia coli threonyl-tRNA synthetase gene contains a split ribosomal binding site interrupted by a hairpin structure that is essential for autoregulation*, **Mol. Microbiol.** **29**, 1077-1090.
- 182 MOINE H *et al.* (1990), *Escherichia coli threonyl-tRNA synthetase and tRNA(Thr) modulate the binding of the ribosome to the translational initiation site of the thrS mRNA*, **J. Mol. Biol.** **216**, 299-310.
- 183 COMER MM *et al.* (1996), *Growth rate-dependent control, feedback regulation and steady-state mRNA levels of the threonyl-tRNA synthetase gene of Escherichia coli*, **J. Mol. Biol.** **261**, 108-124.
- 184 TORRES-LARIOS A *et al.* (2003), *Growth rate-dependent control, feedback regulation and steady-state mRNA levels of the threonyl-tRNA synthetase gene of Escherichia coli*, **J. Mol. Biol.** **331**, 201-211.
- 185 DOCK-BREGEON AC *et al.* (2000), *Transfer RNA-mediated editing in threonyl-tRNA synthetase. The class II solution to the double discrimination problem*, **Cell** **103**, 877-884.
- 186 SANKARANARAYANAN R *et al.* (1999), *The structure of threonyl-tRNA synthetase-tRNA(Thr) complex enlightens its repressor activity and reveals an essential zinc ion in the active site*, **Cell** **97**, 371-381.
- 187 SANKARANARAYANAN R *et al.* (2000), *Zinc ion mediated amino acid discrimination by threonyl-tRNA synthetase*, **Nat. Struct. Mol. Biol.** **7**, 461-465.
- 188 DOCK-BREGEON AC *et al.* (2004), *Achieving error-free translation; the mechanism of proofreading of threonyl-tRNA synthetase at atomic resolution*, **Mol. Cell** **16**, 375-386.
- 189 MINAJIGI A & C Francklyn (2010), *Achieving error-free translation; the mechanism of proofreading of threonyl-tRNA synthetase at atomic resolution*, **J. Biol. Chem.** **285**, 23810-23817.
- 190 CAILLET J *et al.* (2003), *The modular structure of Escherichia coli threonyl-tRNA synthetase as both an enzyme and a regulator of gene expression*, **Mol. Microbiol.** **47**, 961-974.
- 191 ROMBY P & Springer M *et al.* (2003), *Bacterial translational control at atomic resolution*, **TRENDS Genet.** **19**, 155-161.
- 192 SKOGMAN SG & J Nilsson (1984), *Molecular cloning and characterization of the gene for Escherichia coli valyl-tRNA synthetase*, **Gene** **30**, 219 y 223.
- 193 HÄRTLEIN M *et al.* (1987), *Nucleotide sequence of Escherichia coli valyl-tRNA synthetase gene valS*, **Nucleic Acids Res.** **15**, 9081.
- 194 YANIV M & F Gros (1969), *Studies on valyl-tRNA synthetase and tRNA from Escherichia coli, I Purification and properties of the enzyme from normal Escherichia coli strains*, **J. Mol. Biol.** **44**, 1-15.
- 195 YANIV M & F Gros (1969), *Studies on valyl-tRNA synthetase and tRNA from Escherichia coli, II Interaction between valyl-tRNA synthetase and valine acceptor tRNA*, **J. Mol. Biol.** **44**, 17-30.
- 196 BORJFORD TJ *et al.* (1987), *The valyl-tRNA synthetase from Bacillus stearothermophilus has considerable sequence homology with the isoleucyl-tRNA synthetase from Escherichia coli*, **Biochemistry** **26**, 2480-2486.
- 197 HECK JD & GW Hatfield (1988), *Valyl-tRNA synthetase gene of Escherichia coli K12. Primary structure and homology within a family of aminoacyl-tRNA synthetases*, **J. Biol. Chem.** **263**, 868-877.
- 198 PALLANCK L & LDH Schulman (1991), *Anticodon-dependent aminoacylation of a noncognate tRNA with isoleucine, valine, and phenylalanine in vivo*, **PNAS** **88**, 3872-3876.
- 199 LIN L & P Schimmel (1996), *Mutational analysis suggests the same design for editing activities of two tRNA synthetases*, **Biochemistry** **35**, 5596-5601.
- 200 TAMURA K *et al.* (1994), *Role of the CCA terminal sequence of tRNA(Val) in aminoacylation with valyl-tRNA synthetase*, **J. Biol. Chem.** **269**, 22173-22177.
- 201 LIU M & J Horowitz (1994), *Functional transfer RNAs with modifications in the 3'-CCA end: differential effects on aminoacylation and polypeptide synthesis*, **PNAS** **91**, 10389-10393.
- 202 NOMANBHOY TK & PR Schimmel (2000), *Misactivated amino acids translocate at similar rates across surface of a tRNA synthetase*, **PNAS** **97**, 5119-5122.
- 203 FUKAI S *et al.* (2003), *Mechanism of molecular interactions for tRNA(Val) recognition by valyl-tRNA synthetase*, **RNA** **9**, 100-111.
- 204 HOROWITZ J *et al.* (1999), *Synthetase recognition determinants of E. coli valine transfer RNA*, **Biochemistry**, **38**, 7737-7746.
- 205 TARDIF KD & Horowitz (2002), *Transfer RNA determinants for translational editing by Escherichia coli valyl-tRNA synthetase*, **Nucleic Acids Res.** **30**, 2538.
- 206 FUKAI S *et al.* (2000), *Structural basis for double-sieve discrimination of L-valine from L-isoleucine and L-threonine by the complex of tRNA(Val) and valyl-tRNA synthetase*, **Cell** **103**, 793-803.
- 207 TARDIF KD & J Horowitz (2004), *Functional group recognition at the aminoacylation and editing sites of E. coli valyl-tRNA synthetase*, **RNA** **10**, 493-503.

- 208 FUKUNAGA R & S Yokoyama (2005), *Structural basis for non-cognate amino acid discrimination by the valyl-tRNA synthetase editing domain*, **J. Biol. Chem** **280**, 29937-29945.
- 209 TARDIF KD *et al.* (2001), *Misacylation and editing by Escherichia coli valyl-tRNA synthetase: evidence for two tRNA binding sites*, **Biochemistry** **40**, 8118-8125.
- 210 BHARATHAMAN N *et al.* (2009), *Molecular dynamics simulation study of valyl-tRNA synthetase with its pre- and post-transfer editing substrates*, **Biophys. Chem.** **143**, 34-43.
- 211 SEIGNOVERT L *et al.* (1996), *Asparaginyl-tRNA synthetase from Thermus thermophilus HB8, Sequence of the gene and crystallization of the enzyme expressed in Escherichia coli*, **Eur. J. Biochem.** **239**, 501-508.
- 212 KIM SI *et al.* (1996), *Lactobacillus bulgaricus asparagine synthetase and asparaginyl-tRNA synthetase: coregulation by transcription antitermination?*, **J. Bacteriol.** **178**, 2459-2461.
- 213 KIM SI *et al.* (1996), *Asn-tRNA in Lactobacillus bulgaricus is formed by asparaginylation of tRNA and not by transamidation of Asp-tRNA*, **Nucleic Acid Res.** **24**, 2648-2651.
- 214 SALAZAR JC *et al.* (2001), *A dual-specific Glu-tRNA(Gln) and Asp-tRNA(Asn) amidotransferase is involved in decoding glutamine and asparagine codons in Acidithiobacillus ferrooxidans*, **FEBS Lett** **500**, 129-131.
- 215 RACZNIK G *et al.* (2001), *A single amidotransferase forms asparaginyl-tRNA and glutaminyl-tRNA in Chlamydia trachomatis*, **J. Biol. Chem.** **276**, 45862-45867.
- 216 MIN B *et al.* (2002), *Transfer RNA-dependent amino acid biosynthesis: an essential route to asparagine formation*, **PNAS** **99**, 2678-2683.
- 217 AKOCHY PM *et al.* (2004), *Direct glutaminyl-tRNA biosynthesis and indirect asparaginyl-tRNA biosynthesis in Pseudomonas aeruginosa PAO1*, **J. Bacteriol.** **186**, 767-776.
- 218 BAILLY M *et al.* (2006), *A single tRNA base pair mediates bacterial tRNA-dependent biosynthesis of asparagine*, **Nucleic Acids Res.** **34**, 6083-6094.
- 219 KREBS HA (1950), *Manometric determination of L-aspartic acid and L-asparagine*, **Biochem. J.** **47**, 605 y 613.
- 220 ANSELME J & M Härtlein (1989), *Asparaginyl-tRNA synthetase from Escherichia coli has significant sequence homologies with yeast aspartyl-tRNA synthetase*, **Gene** **84**, 481-485.
- 221 ERNIANI G *et al.* (1990), *Aspartyl-tRNA synthetase from Escherichia coli: cloning and characterisation of the gene, homologies of its translated amino acid sequence with asparaginyl- and lysyl-tRNA synthetases*, **Nucleic Acids Res.** **18**, 7109-7118.
- 222 BERTHET-COLOMINAS C *et al.* (1998), *The crystal structure of asparaginyl-tRNA synthetase from Thermus thermophilus and its complexes with ATP and asparaginyl-adenylate: the mechanism of discrimination between asparagine and aspartic acid*, **EMBO J.** **17**, 2947-2960.
- 223 IWASAKI W *et al.* (2006), *Structural basis of the water-assisted asparagine recognition by asparaginyl-tRNA synthetase*, **J. Mol. Biol.** **360**, 329-342.
- 224 IBBA M *et al.* (2000), *The adaptor hypothesis revisited*, **TRENDS Biochem. Sci.** **25**, 311-316.
- 225 KUMARAN D *et al.* (2005), *Structure and mechanism of ADP-ribose-1''-monophosphatase (Appr-1''-pase), a ubiquitous cellular processing enzyme*, **Prot. Sci.** **14**, 719-726.
- 226 PUTICS A *et al.* (2005), *ADP-ribose-1''-monophosphatase: a conserved coronavirus enzyme that is dispensable for viral replication in tissue culture*, **J. Virol.** **79**, 12721-12731.
- 227 EGLOFF MP *et al.* (2006), *Structural and functional basis for ADP-ribose and poly(ADP-ribose) binding by viral macro domains*, **J. Virol.** **80**, 8493-8502.
- 228 XU Y *et al.* (2009), *Crystal structures of two coronavirus ADP-ribose-1''-monophosphatases and their complexes with ADP-Ribose: a systematic structural analysis of the viral ADRP domain*, **J. Virol.** **83**, 1083-1092.
- 229 KARRAS G *et al.* (2005), *The macro domain is an ADP-ribose binding module*, **EMBO J.** **24**, 1911-1920.
- 230 NEUVONEN M & T Ahola (2009), *Differential activities of cellular and viral macro domain proteins in binding of ADP-ribose metabolites*, **J. Mol. Biol.** **385**, 212-225.
- 231 HUANG N *et al.* (2009), *Structure and function of an ADP-ribose-dependent transcriptional regulator of NAD metabolism*, **Structure** **17**, 939-951.
- 232 ARAVIND L & E Koonin (1999), *DNA polymerase beta-like nucleotidyltransferase superfamily: identification of three new families, classification and evolutionary history*, **Nucleic Acids Res.** **27**, 1609-1618.
- 233 RAYNAL LC & AJ Carposius (1999), *Poly(A) polymerase I of Escherichia coli: characterization of the catalytic domain, an RNA binding site and regions for the interaction with proteins involved in mRNA degradation*, **Mol. Microbiol.** **32**, 765-775.
- 234 JASIECKI J & G Węgrzyn (2005), *Localization of Escherichia coli poly(A) polymerase I in cellular membrane*, **Biochem. Biophys. Res. Commun.** **329**, 598-602.
- 235 YEHUDAI-RESHEFF S & G Schuster (2000), *Characterization of the E.coli poly(A) polymerase: nucleotide specificity, RNA-binding affinities and RNA structure dependence*, **Nucleic Acids Res.** **28**, 1139-1144.
- 236 MOHANTY BK & SR Kushner (1999), *Analysis of the function of Escherichia coli poly(A) polymerase I in RNA metabolism*, **Mol. Microbiol.** **34**, 1094-1108.

- 237 TOH Y *et al.* (2010), *Mechanism for the alteration of the substrate specificities of template-independent RNA polymerases*, **Structure** **19**, 232-243.
- 238 GRIEP MA (1995), *Primase, structure and function*, **Indian J. Biochem. Biophys.** **32**, 171-178.
- 239 RODINA A & N Godson (2006), *Role of conserved amino acids in the catalytic activity of Escherichia coli primase*, **J. Bacteriol.** **188**, 3614-3621.
- 240 KECK JL *et al.* (2000), *Structure of the RNA polymerase domain of the E. coli primase*, **Science** **287**, 2482-2486.
- 241 PODOBNIK M *et al.* (2000), *A TOPRIM domain in the crystal structure of the catalytic core of Escherichia coli primase confirms a structural link to DNA topoisomerases*, **J. Mol. Biol.** **300**, 353-362.
- 242 SOULTANAS P (2005), *The bacterial helicase-primase interaction: a common structural/functional module*, **Structure** **13**, 839-844.
- 243 KUTCHA RD & G Stengel (2010), *Mechanism and evolution of DNA primases*, **Biochim. Biophys. Acta** **1804**, 1180-1189.
- 244 STÜLKE J & W Hillen (1999), *Carbon catabolite repression in bacteria*, **Curr. Opin. Microbiol.** **2**, 195-201.
- 245 DEUTSCHER J (2008), *The mechanisms of carbon catabolite repression in bacteria*, **Curr. Opin. Microbiol.** **11**, 87-93.
- 246 TILBEURGH H & N Declerck (2001), *Structural insights into the regulation of bacterial signaling proteins containing PRDs*, **Curr. Opin. Microbiol.** **11**, 685-693.
- 247 STÜLKE J *et al.* (1998), *PRD-a protein domain involved in PTS-dependent induction and carbon catabolite repression of catabolic operons in bacteria*, **Mol. Microbiol.** **28**, 865-874.
- 248 GREENBERG DB *et al.* (2002), *Domain analysis of transcriptional regulators bearing PTS regulatory domains*, **Res. Microbiol.** **153**, 519-526.
- 249 JONES BE *et al.* (1997), *Binding of the catabolite repressor protein CcpA to its DNA target is regulated by phosphorylation of its corepressor HPr*, **J. Biol. Chem.** **272**, 26530-26535.
- 250 WARNER JB & JS Lolkema (2003), *CcpA-dependent carbon catabolite repression in bacteria*, **Microbiol. Mol. Biol. Rev.** **67**, 475-490.
- 251 CHAMPOUX JJ (2001), *DNA topoisomerases: structure, function, and mechanism*, **Annu. Rev. Biochem.** **70**, 369-413.
- 252 CORBETT KD & JM Berger (2004), *Structure, molecular mechanisms, and evolutionary relationships in DNA topoisomerases*, **Annu. Rev. Biophys. Biomol. Struct.** **33**, 95-118.
- 253 BAKER NM *et al.* (2009), *Structural studies of type I topoisomerases*, **Nucleic Acids Res.** **37**, 693-701.
- 254 FORTERRE P *et al.* (2007), *Origin and evolution of DNA topoisomerases*, **Biochimie** **89**, 427-446.
- 255 FORTERRE P & D Gadelle (2009), *Phylogenomics of DNA topoisomerases: their origin and putative roles in the emergence of modern organisms*, **Nucleic Acids Res.** **37**, 679-692.
- 256 PAO JM & MH Saier Jr. (1995), *Response regulators of bacterial signal transduction systems: selective domain shuffling during evolution*, **J. Mol. Evol.** **40**, 136-154.
- 257 SHERRAT DJ & DB Wigley (1998), *Conserved themes but novel activities in recombinases and topoisomerases*, **Cell** **93**, 149-152.
- 258 YANG W & K Mizuuchi (1997), *Site-specific recombination in plane view*, **Structure** **5**, 1401-1406.
- 259 NÜNES-DÜBY SE *et al.* (1998), *Similarities and differences among 105 members of the Int family of site-specific recombinases*, **Nucleic Acids Res.** **26**, 391-406.
- 260 SWALLA BM *et al.* (2003), *Conservation of structure and function among tyrosine recombinases: homology-based modeling of the lambda integrase core-binding domain*, **Nucleic Acids Res.** **31**, 805-818.
- 261 KWON HJ *et al.* (1997), *Flexibility in DNA recombination: Structure of the lambda integrase catalytic core*, **Science** **276**, 126-131.
- 262 LEU FP *et al.* (2000), *The delta subunit of DNA polymerase III holoenzyme serves as a sliding clamp unloader in Escherichia coli*, **J. Biol. Chem.** **275**, 34609-34618.
- 263 SONG MS *et al.* (2001), *The delta and delta' subunits of the DNA polymerase III holoenzyme are essential for initiation complex formation and processive elongation*, **J. Mol. Chem.** **276**, 35165-35175.
- 264 LEU FP & M O'Donnell (2001), *Interplay of clamp loader subunits in opening the beta sliding clamp of Escherichia coli DNA polymerase III holoenzyme*, **J. Mol. Chem.** **276**, 47185-47194.
- 265 BULLARD JM *et al.* (2002), *A three-domain structure for the delta subunit of the DNA polymerase III holoenzyme delta domain III binds delta' and assembles into the DnaX complex*, **J. Biol. Chem.** **277**, 13246-13256.
- 266 SINGLETON MR *et al.* (1999), *Structure of the adenylation domain of an NAD⁺-dependent DNA ligase*, **Structure** **7**, 35-42.
- 267 TIMSON DJ & DB Wigley (1999), *Functional domains of an NAD⁺-dependent DNA ligase*, **J. Mol. Biol.** **285**, 73-83.
- 268 WILKINSON A *et al.* (2001), *Bacterial DNA ligases*, **Mol. Microbiol.** **40**, 1241-1248.

- 269 SCIOCHETTI SA y PJ Piggot (2000), *A tale of two genomes: resolution of dimeric chromosomes in Escherichia coli and Bacillus subtilis*, **Res. Microbiol.** **151**, 503-511.
- 270 BARRE FX (2007), *FtsK and SpoIIIE: the tale of the conserved tails*, **Mol. Microbiol.** **66**, 1051-1055.
- 271 ERRINGTON J *et al.* (2001), *DNA transport in bacteria*, **Nat. Rev. Mol. Cell Biol.** **2**, 538-544.
- 272 IYER LM *et al.* (2004), *Comparative genomics of the FtsK-HerA superfamily of pumping ATPases: implications for the origins of chromosome segregation, cell division and viral capsid packaging*, **Nucleic Acids Res.** **32**, 5260-5279.
- 273 HENDRIX RW (2002), *Bacteriophages: Evolution of the majority*, **Theor. Popul. Biol.** **61**, 471-480.
- 274 CASJENS SR (2005), *Comparative genomics and evolution of the tailed-bacteriophages*, **Curr. Opin. Microbiol.** **8**, 451-458.
- 275 BRÜSSOW H (2009), *The not so universal tree of life or the place of viruses in the living world*, **Proc. Trans. R. Soc. B** **364**, 2263-2274.
- 276 CHASE JW & JJ Richardson (1974), *Exonuclease VII of Escherichia coli. Mechanism of action*, **J. Biol. Chem.** **249**, 4553-4561.
- 277 VALES LD *et al.* (1982), *Subunit structure of Escherichia coli exonuclease VII*, **J. Biol. Chem.** **257**, 8799-8805.
- 278 CHASE JW *et al.* (1986), *Escherichia coli exonuclease VII. Cloning and sequencing of the gene encoding the large subunit (xseA)*, **J. Biol. Chem.** **261**, 14929-14935.
- 279 MIDON M *et al.* (2011), *Mutational and biochemical analysis of the DNA-entry nuclease EndA from Streptococcus pneumoniae*, **Nucleic Acids Res.** **39**, 623-634.
- 280 MOON AF *et al.* (2011), *Structural insights into catalytic and substrate binding mechanisms of the strategic EndA nuclease from Streptococcus pneumoniae*, **Nucleic Acids Res.** **39**, 2943-2953.
- 281 SCHREIER HJ *et al.* (1989), *Regulation of Bacillus subtilis glutamine synthetase gene expression by the product of the glnR gene*, **J. Mol. Biol.** **210**, 51-63.
- 282 MORTENSEN I *et al.* (1989), *Purification and characterization of the deoR repressor of Escherichia coli*, **EMBO J.** **8**, 325-331.
- 283 LOBELL RB & RF Schleif (1990), *DNA looping and unlooping by AraC protein*, **Science** **250**, 528-532.
- 284 ARAVIND L *et al.* (2005), *The many faces of the helix-turn-helix domain: transcription regulation and beyond*, **FEMS Microbiol. Rev.** **29**, 231-262.
- 285 HSU DS *et al.* (1995), *Structure and function of the UvrB protein*, **J. Biol. Chem.** **270**, 8319-8327.
- 286 THEIS K *et al.* (1999), *Crystal structure of UvrB, a DNA helicase adapted for nucleotide excision repair*, **EMBO J.** **18**, 6899-6907.
- 287 GOOSEN N & GF Moolenaar (2001), *Role of ATP hydrolysis by UvrA and UvrB during nucleotide excision repair*, **Res. Microbiol.** **152**, 401-409.
- 288 TRUGLIO JJ *et al.* (2006), *Structural basis for DNA recognition and processing by UvrB*, **Nat. Struct. Mol. Biol.** **13**, 360-364.
- 289 TRUGLIO JJ *et al.* (2006), *Prokaryotic nucleotide excision repair: the UvrABC system*, **Chem. Rev.** **106**, 233-252.
- 290 SOMERVILLE GA & RA Proctor (2009), *At the crossroads of bacterial metabolism and virulence factor synthesis in Staphylococci*, **Microbiol. Mol. Biol.** **73**, 233-238.
- 291 SRINIVASAN V & HJ Morowitz (2009), *The canonical network of autotrophic intermediary metabolism: minimal metabolome of a reductive chemoautotroph*, **Biol. Bull.** **216**, 126-130.
- 292 LAZCANO A *et al.* (1995), *On the levels of enzymatic substrate specificity: implications for the early evolution of metabolic pathways*, **Adv. Space Res.** **15**, 345-356.
- 293 PEREGRÍN-ALVAREZ JM *et al.* (2009), *The conservation and evolutionary modularity of metabolism*, **Genome Biol.** **10**, R63.
- 294 CAETANO-ANOLLÉS G *et al.* (2007), *The origin of modern metabolic networks inferred from phylogenomic analysis of protein architecture*, **PNAS** **104**, 9358-9363.
- 295 DEAMER DW (1997), *The first living systems: a bioenergetic perspective*, **Microbiol. Mol. Biol. Rev.** **61**, 239-261.
- 296 CASTRESANA J & Saraste M (1995), *Evolution of energetic metabolism: the respiration-early hypothesis*, **TRENDS in Biochem. Sci** **20**, 443-448.
- 297 CASTRESANA J (2001), *Comparative genomics and bioenergetics*, **Biochim. Biophys. Acta** **1506**, 147-162.
- 298 BAYMANN F *et al.* (2003), *The redox protein construction kit: pre-last universal common ancestor evolution of energy-conserving enzymes*, **Phil. Trans. R. Soc. Lond. B** **358**, 267-274.
- 299 http://www.ebi.ac.uk/interpro/potm/2005_12/Page1.htm
- 300 http://www.ebi.ac.uk/interpro/potm/2005_12/Page2.htm
- 301 http://www.ebi.ac.uk/interpro/potm/2005_12/Page3.htm
- 302 LODISH H *et al.*, *Figure 16-28: Model of the structure of ATP synthase (the F0F1 ATPase complex) in the bacterial plasma membrane* in **Molecular Cell Biology**, 4th edition, W H Freeman, USA, 2000.
- 303 <http://www.rcsb.org/pdb/101/motm.do?momID=72>

- 304 SENIOR AE (1988), *ATP synthesis by oxidative phosphorylation*, **Physiol. Rev.** **68**, 177-219.
- 305 PEDERSEN PL *et al.* (2000), *ATP synthases in the year 2000: Defining the different levels of mechanism and getting a grip on each*, **J. Bioener. Biomembr.** **32**, 423-432.
- 306 JONES HM *et al.* (1983), *In vivo 5' terminus and length of the mRNA for the protein-translocating ATPase (unc) operon of Escherichia coli*, **J. Bacteriol.** **155**, 1279-1287.
- 307 McCARN DF *et al.* (1988), *Genes encoding alpha, gamma, delta, and four F0 subunits of ATP synthase constitute an operon in the cyanobacterium Anabaena sp. Strain PCC7120*, **J. Bacteriol.** **170**, 3448-3458.
- 308 BRUSILOV WSA *et al.* (1989), *Organization and sequence of the genes coding for the proton-translocating ATPase of Bacillus megaterium*, **J. Biol. Chem.** **264**, 1528-1533.
- 309 KASIMOGLU E (1996), *Transcriptional regulation of the proton-translocating ATPase (atpIBEFHAGDC) operon in Escherichia coli: control by cell growth rate*, **J. Bacteriol.** **178**, 5563-5567.
- 310 CROSS RL & L Taiz (1990), *Gene duplication as a means for altering H⁺/ATP ratios during the evolution of F₀F₁ ATPases and synthases*, **FEBS** **259**, 227-229.
- 311 CROSS RL & Müller V (2004), *The evolution of A-, F-, and V-type ATP synthases and ATPases: reversals in function and changes in the H⁺/ATP coupling ratio*, **FEBS Lett.** **576**, 1-4.
- 312 MÜLLER V & G GRÜBER (2003), *ATP synthases: structure, function and evolution of unique energy converters*, **Cell Mol. Life Sci.** **60**, 474-494.
- 313 HILARIO E & JP Gogarten (1993), *Horizontal transfer of ATPase genes - the tree of life becomes a net of life*, **BioSystems** **31**, 111-119.
- 314 NAKAMOTO RK *et al.* (2008), *The rotary mechanism of the ATP synthase*, **Arch. Biochem. Biophys.** **476**, 43,44.
- 315 GUNSALUS RP *et al.* (1982), *Gene order and gene-polypeptide relationships of the proton-translocating ATPase operon (unc)*, **PNAS** **79**, 320-324.
- 316 HELLMUT K *et al.* (1992), *Translational coupling varying in efficiency between different pairs of genes in the central region of the atp operon of Escherichia coli*, **Mol. Microbiol.** **5**, p.813.
- 317 GOGOL EP *et al.* (1989), *Molecular architecture of Escherichia coli F1 adenosinetriphosphatase*, **Biochemistry** **28**, 4709-4716.
- 318 GAY NJ & JE Walker (1981), *The atp operon: nucleotide sequence of the region encoding the a-subunit of Escherichia coli ATP-synthase*, **Nucl. Acids Res.** **9**, pc.2187-2192.
- 319 KANAZAVA H *et al.* (1981), *Nucleotide sequence of the genes coding for α , β and γ subunits of the proton-translocating ATPase of Escherichia coli*, **Biochem. Biophys. Res. Commun.** **103**, 604-612.
- 320 FUTAI M *et al.* (2000), *Synthase (H⁺ ATPase): coupling between catalysis, mechanical work, and proton translocation*, **Biochim. Biophys. Acta** **1458**, 276-288.
- 321 NAKAMOTO RK *et al.* (2000), *Molecular mechanisms of rotational catalysis, in the F₀F₁ ATP synthase*, **Biochim. Biophys. Acta** **1458**, 289-299.
- 322 LEYVA JA *et al.* (2003), *Understanding ATP synthesis: structure and mechanism of the F1-ATPase (Review)*, **Mol. Membr. Biol.** **20**, 27-33.
- 323 WALKER JE *et al.* (1982), *Distantly related sequences in the α - and β -subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide-binding fold*, **EMBO J.** **1**, 945-951.
- 324 KOONIN EV (1993), *A superfamily of ATPases with diverse functions containing either classical or deviant ATP-binding motif*, **J. Mol. Biol.** **229**, 1165-1174.
- 325 SHIRAKIHARA Y *et al.* (1997), *The crystal structure of the nucleotide-free $\alpha_3\beta_3$ subcomplex of F₁-ATPase from the thermophilic Bacillus PS3 is a symmetric trimer*, **Structure** **15**, 825-836.
- 326 MUNEYUKI E *et al.* (2000), *F₀F₁-ATP synthase: general structural features of 'ATP' engine and a problem of free energy transduction*, **Biochim. Biophys. Acta** **1458**, 467-481.
- 327 RAO R *et al.* (1988), *Directed mutagenesis of the strongly conserved lysine 175 in the proposed nucleotide-binding domain of α -subunit from Escherichia coli F1-ATPase*, **J. Biol. Chem.** **263**, 15957-15963.
- 328 SOGA S *et al.* (1989), *Mutational replacements of conserved amino acid residues in the α subunit change the catalytic properties of Escherichia coli F1-ATPase*, **Arch. Biochem. Biophys.** **268**, 643-648.
- 329 WEBER J *et al.* (1995), *α -aspartate 261 is a key residue in a noncatalytic sites of Escherichia coli*, **J. Biol. Chem.** **270**, 21045-2149.
- 330 LI W *et al.* (2009), *Role of α -subunit VISIT-DG sequence residues Ser-347 and Gly-351 in the catalytic sites of Escherichia coli ATP synthase*, **J. Biol. Chem.** **284**, 10747-10754.
- 331 MEIER T *et al.* (2005), *Structure of the rotor ring of F-Type Na⁺-ATPase from Ilyobacter tartaricus*, **Science** **308**, 659-662.
- 332 POGORYELOV D *et al.* (2009), *High resolution structure of the rotor ring of a proton-dependent ATP synthase*, **Nature Struct. Mol. Biol.** **16**, 1068-1073.

- 333 NIELSEN J *et al.* (1981), *The nucleotide sequence of the atp genes coding for the F₀ subunits a, b, c and the F₁ subunit d of the membrane bound ATP synthase of Escherichia coli*, **Mol. Gen. Genet.** **184**, 33-39.
- 334 RASTOGI VK & ME Girvin (1999), *Structural changes linked to proton translocation by subunit c of the ATP synthase*, **Nature** **402**, 263-268.
- 335 DECKERS-HEBESTREIT G *et al.* (2000), *The ATP synthase of Escherichia coli: structure and function of F₀ subunits*, **Biochim. Biophys. Acta** **1458**, 364-373.
- 336 IVEY DM & TA Krulwich (1991), *Organization and nucleotide sequence of the atp genes encoding the ATP synthase from alkaliphile Bacillus firmus OF4*, **Mol. Gen. Genet.** **229**, 292-300.
- 337 FILLINGAME RH *et al.* (2000), *Structural interpretations of F₀ rotary function in the Escherichia coli F₁F₀ ATP synthase*, **Biochim. Biophys. Acta** **1458**, 387-403.
- 338 FILLINGAME RH & OY Dmitriev (2002), *Structural model of the transmembrane F₀ rotary sector of H⁺-transporting ATP synthase derived by solution NMR and intersubunit cross-linking in situ*, **Biochim. Biophys. Acta** **1565**, 232-245.
- 339 FILLINGAME RH *et al.* (2003), *Mechanics of coupling proton movements to c-ring rotation in ATP synthase*, **FEBS Lett** **545**, 29-34.
- 340 WEBER J & AE Senior (2003), *ATP synthesis driven by proton transport in F₀F₁-ATP synthase*, **FEBS Lett** **545**, 61-70.
- 341 GROTH G (2000), *Molecular models of the structural arrangement of subunits and the mechanism of proton translocation in the membrane domain of F₁F₀ ATP synthase*, **Biochim. Biophys. Acta** **1458**, 417-427.
- 342 DMITRIEV OY & RH Fillingame (2001), *Structure of Ala²⁰→Pro/Pro⁶⁴→Ala substituted subunit c of Escherichia coli ATP synthase in which the essential proline is switched between transmembrane helices*, **J. Biol. Chem.** **276**, 27449-27454.
- 343 DMITRIEV OY *et al.* (2002), *Structure of Ala²⁴/Asp⁶¹→Asp²⁴/Asn⁶¹ substituted subunit c of Escherichia coli ATP synthase: implications for the mechanism of proton transport and rotary movement in the F₀ complex*, **Biochemistry** **41**, 5537-5547.
- 344 FILLINGAME RH *et al.* (1998), *Subunit organization and structure in the F₀ sector of Escherichia coli F₁F₀ ATP synthase*, **Biochim. Biophys. Acta** **1365**, 135-142.
- 345 FIMMEL AL *et al.* (1983), *The F₁F₀-ATPase of Escherichia coli, Substitution of proline by leucine at position 64 in the c-subunit causes loss of oxidative phosphorylation*, **Biochem. J.** **213**, 451-458.
- 346 RECIPON H *et al.* (1992), *ATP synthase subunit c/III/9 gene sequences as tool for interkingdom and metaphytes molecular phylogenies*, **J. Mol. Evol.** **34**, 292-303.
- 347 KOIKE K *et al.* (1985), *Purification and characterization of NADH oxidase from a strain of Leuconostoc mesenteroides*, **J. Biochem.** **97**, 1279-1288.
- 348 SCHMIDT HL *et al.* (1986), *Isolation and properties of an H₂O-forming NADH oxidase from Streptococcus faecalis*, **Eur. J. Biochem.** **156**, 149-155.
- 349 AHMED SA & A Claiborne (1989), *The streptococcal flavoprotein NADH oxidase, I. Evidence linking NADH oxidase and NADH peroxidase cysteinyl redox centres*, **J. Biol. Chem.** **264**, 19856-19863.
- 350 ROSS RP & A Claiborne (1992), *Molecular cloning and analysis of the gene encoding the NADH oxidase from Streptococcus faecalis 10C1*, **J. Mol. Biol.** **227**, 658-671.
- 351 STANTON TB & NS Jensen (1993), *Purification and characterization of NADH oxidase from Serpulina (Treponema) hyodysenteriae*, **J. Bacteriol.** **175**, 2980-2987.
- 352 LIU XL & RK Scopes (1993), *Cloning, sequencing and expression of the gene encoding NADH oxidase from the extreme anaerobic thermophile Thermoanaerobium brockii*, **Biochim. Biophys. Acta** **1174**, 187-190.
- 353 HIGUCHI M *et al.* (1993), *Identification of two distinct NADH oxidases corresponding to H₂O₂-forming oxidase and H₂O-forming oxidase induced in Streptococcus mutans*, **J. Gen. Microbiol.** **139**, 2343-2351.
- 354 MATSUMOTO J *et al.* (1996), *Molecular cloning and sequence analysis of the gene encoding the H₂O-forming NADH oxidase from Streptococcus mutans*, **Biosci. Biotech. Biochem.** **60**, 39-43.
- 355 SAKAMOTO M & KOMAGATA K (1996), *Aerobic growth of and activities of NADH oxidase and NADH peroxidase in lactic acid bacteria*, **J. Ferment. Bioeng.** **82**, 210-216.
- 356 DE FELIPE FL *et al.* (1998), *Cofactor engineering: a novel approach to metabolic engineering in Lactococcus lactis by controlled expression of NADH oxidase*, **J. Bacteriol.** **180**, 384-388.
- 357 HIGUCHI M *et al.* (1999), *Functions of two types of NADH oxidases in energy metabolism and oxidative stress of Streptococcus mutans*, **J. Bacteriol.** **181**, 5940-5947.
- 358 AUZAT I *et al.* (1999), *The NADH oxidase of Streptococcus pneumoniae: its involvement in competence and virulence*, **Mol. Microbiol.** **34**, 1018-1028.
- 359 GIBSON CM *et al.* (2000), *Contribution of NADH oxidase to aerobic metabolism of Streptococcus pyogenes*, **J. Bacteriol.** **182**, 448-455.

- 360 HIGUCHI M *et al.* (2000), *Molecular biology of oxygen tolerance in lactic acid bacteria: functions of NADH oxidases and Dpr in oxidative stress*, **J. Biosci. Bioeng.** **90**, 484-493.
- 361 ECHENIQUE JR & MC Trombe (2001), *Competence modulation by the NADH oxidase of Streptococcus pneumoniae involves signal transduction*, **J. Bacteriol.** **183**, 768-772.
- 362 CHAPUY-REGAUD S *et al.* (2001), *Competence regulation by oxygen availability and by NOX is not related to specific adjustment of central metabolism in Streptococcus pneumoniae*, **J. Bacteriol.** **183**, 2957-2962.
- 363 YU J *et al.* (2001), *Characterization of the Streptococcus pneumoniae NADH oxidase that is required for infection*, **Microbiology** **147**, 431-438.
- 364 WARD DE *et al.* (2001), *The NADH oxidase from Pyrococcus furiosus, Implications for the protection of anaerobic hyperthermophiles against oxidative stress*, **Eur. J Biochem.** **268**, 5816-5823.
- 365 NEVES AR *et al.* (2002), *Effect of different NADH oxidase levels on glucose metabolism by Lactococcus lactis: kinetics of intracellular metabolite pools determined by in vivo nuclear magnetic resonance*, **Appl. Environ. Microbiol.** **68**, 6332-6342.
- 366 KAWASAKI S *et al.* (2004), *Purification and characterization of an H₂O-forming NADH oxidase from Clostridium aminovalericum: existence of a detoxifying enzyme in an obligate anaerobic bacteria*, **Arch. Microbiol.** **181**, 324-330.
- 367 LOUNTOS GT *et al.* (2004), *Crystallization and preliminary analysis of a water forming NADH oxidase from Lactobacillus sanfranciscensis*, **Acta Crystallogr. D** **60**, 2044-2047.
- 368 KUZU M *et al.* (2005), *Crystallization and preliminary crystallographic analysis of a flavoprotein NADH oxidase from Lactobacillus brevis*, **Acta Crystallogr. F** **61**, 528-530.
- 369 YAMAMOTO Y *et al.* (2006), *The group B Streptococcus NADH oxidase Nox-2 is involved in fatty acid biosynthesis during aerobic growth and contributes to virulence*, **Mol. Microbiol.** **62**, 772-785.
- 370 JIA B *et al.* (2008), *Hexameric ring structure of a thermophilic archaeon NADH oxidase that produces predominantly H₂O*, **FEBS J.** **275**, 5355-5366.
- 371 JIA B *et al.* (2010), *Oxidized NADH oxidase inhibits activity of an ATP/NAD kinase from a thermophilic archaeon*, **Protein J.** **29**, 609-616.
- 372 TACHON S *et al.* (2011), *Identification of a conserved sequence in flavoproteins essential for the correct conformation and activity of the NADH oxidase NoxE of Lactococcus lactis*, **J. Bacteriol.** **193**, 3000-3008.
- 373 WANG L *et al.* (2012), *Comparison of alkyl hydroperoxide reductase and two water-forming NADH oxidases from Bacillus cereus*, **Appl. Microbiol. Biotechnol.** **96**, 1265-1273.
- 374 DERR AM *et al.* (2012), *Mutation of the NADH oxidase gene (nox) reveals an overlap of the oxygen- and acid-mediated stress responses in Streptococcus mutans*, **Appl. Environ. Microbiol.** **98**, 1215-1227.
- 375 ZHANG YW *et al.* (2012), *Cloning and characterization of a thermostable H₂O-forming NADH oxidase from Lactobacillus rhamnosus*, **Enzyme Microbial Technol.** **50**, 255-262.
- 376 GAO H *et al.* (2012), *Characterization of H₂O-forming NADH oxidase from Streptococcus pyogenes and its application in L-rare sugar production*, **Bioorg. Med. Chem. Lett.** **22**, 1931-1935.
- 377 MALLET TC & A Claiborne (1998), *Oxygen reactivity of an NADH oxidase C42S mutant: evidence for a C(4a)-peroxyflavin intermediate and a rate-limiting conformational change*, **Biochemistry** **37**, 8790-8802.
- 378 McKIE JH & KT Douglas (1991), *Evidence for gene duplication forming similar binding folds for NAD(P)H and FAD in pyridine nucleotide-dependent flavoenzymes*, **FEBS** **279**, P 5-8.
- 379 PETSKO GA (1991), *Déjà vu all over again*, **Nature** **352**, 104-105.
- 380 ATYEO RF *et al.* (1999) *Differentiation of Serpulina species by NADH oxidase gene (nox) sequence comparisons and nox-based polymerase chain reaction tests*, **Vet. Microbiol.** **67**, 47-60.
- 381 BRISSON D *et al.* (2001), *Glycerol: a neglected variable in metabolic processes?*, **BioEssays** **23**, 534-542.
- 382 IUCHI S *et al.* (1990), *Multiple regulatory elements for the glpA operon encoding anaerobic glycerol-3-phosphate dehydrogenase and the glpD operon encoding aerobic glycerol-3-phosphate dehydrogenase in Escherichia coli: further characterization of respiratory control*, **J. Bacteriol.** **172**, p.180, fig.1
- 383 KITO M & LI Pizer (1969), *Purification and regulatory properties of the biosynthetic L-glycerol 3-phosphate dehydrogenase from Escherichia coli*, **J. Biol. Chem.** **244**, 3316-3323.
- 384 CRONAN JE & RM Bell (1974), *Mutants of Escherichia coli defective in membrane phospholipid synthesis: mapping of the structural gene for L-glycerol 3-phosphate dehydrogenase*, **J. Bacteriol.** **118**, 598-605.
- 385 TRUNIGER V & W Boos (1994), *Mapping and cloning of gldA, the structural gene of the Escherichia coli glycerol dehydrogenase*, **J. Bacteriol.** **176**, 1796-1800.

- 386 WEINER JH (1974), *The localization of glycerol-3-phosphate dehydrogenase*, **J. Membr. Biol.** **15**, 1-14.
- 387 EDGAR JR & RM Bell (1978), *Biosynthesis in Escherichia coli of sn-glycerol 3-phosphate, a precursor of phospholipid, Purification and physical characterization of wild type and feedback-resistant forms of the biosynthetic sn-glycerol-3-phosphate dehydrogenase*, **J. Biol. Chem.** **253**, 6348-6363.
- 388 CLARK D *et al.* (1980), *Regulation of phospholipid biosynthesis in Escherichia coli, Cloning of the structural gene for the biosynthetic sn-glycerol-3-phosphate dehydrogenase*, **J. Biol. Chem.** **255**, 714-717.
- 389 EDGAR JR & RM Bell (1980), *Biosynthesis in Escherichia coli of sn-glycerol 3-phosphate, a precursor of phospholipid, Further kinetic characterization of wild type and feedback-resistant forms of the biosynthetic sn-glycerol-3-phosphate dehydrogenase*, **J. Biol. Chem.** **255**, 3492-3497.
- 390 MORBIDONI HR *et al.* (1995), *Synthesis of sn-glycerol 3-phosphate, a key precursor of membrane lipids, in Bacillus subtilis*, **J. Bacteriol.** **177**, 5899-5905.
- 391 SAKASEGAWA SI *et al.* (2004), *Structural and functional analysis of the gpsA gene product of Archaeoglobus fulgidus: a glycerol-3-phosphate dehydrogenase with an unusual NADP+ preference*, **Prot. Sci.** **13**, 3161-3171.
- 392 OTTO J *et al.* (1980), *Prediction of secondary structural elements in glycerol elements in glycerol-3-phosphate dehydrogenase by comparison with other dehydrogenases*, **Eur. J. Biochem.** **109**, 325-330.
- 393 WEBER AL (1989), *Model of early self-replication based on covalent complementarity for a copolymer of glycerate-3-phosphate and glycerol-3-phosphate*, **Orig. Life Evol. Biosph.** **19**, 179-186.
- 394 YOUNG TW *et al.* (1998), *Bacillus subtilis ORF yybQ encodes a manganese-dependent inorganic pyrophosphatase with distinctive properties: the first of a new class of soluble pyrophosphatase?*, **Microbiology** **144**, 2563-2571.
- 395 SHINTANI T *et al.* (1998), *Cloning and expression of a unique inorganic pyrophosphatase from Bacillus subtilis: evidence for a new family of enzymes*, **FEBS Lett.** **439**, 263-266.
- 396 MERCKEL MC *et al.* (2001), *Crystal structures of Streptococcus mutans pyrophosphatase: a new fold for an old mechanism*, **Structure** **9**, 289-297.
- 397 SALMINEN A *et al.* (2006), *Two soluble pyrophosphatases in Vibrio cholerae: transient redundancy or enduring cooperation?*, **Biochem. (Mosc.)** **71**, 978-982.
- 398 FABRICHNIY IP *et al.* (2004), *Structural studies of metal ions in family II pyrophosphatases: the requirement for a Janus ion*, **Biochemistry** **43**, 14403-14411.
- 399 HALONEN P *et al.* (2005), *Effects of active site mutations on the metal binding affinity, catalytic competence, and stability of the family II pyrophosphatase from Bacillus subtilis*, **Biochemistry** **44**, 4004-4010.
- 400 FABRICHNIY IP *et al.* (2007), *A trimetal site and substrate distortion in a family II pyrophosphatase*, **J. Biol. Chem.** **282**, 1422-1431.
- 401 ILIAS M & TW Young (2004), *Streptococcus gordonii soluble inorganic pyrophosphatase: an important role for the interdomain region in enzyme activity*, **Biochim. Biophys. Acta** **1764**, 1299-1306.
- 402 FOTHERGILL-GILMORE LA & PAM Michels (1993), *Evolution of glycolysis*, **Prog. Biophys. Mol. Biol.** **59**, 105-235.
- 403 ROMANO AH & T Conway (1996), *Evolution of carbohydrate metabolic pathways*, **Res. Microbiol.** **147**, 448-455.
- 404 HUYNEN MA *et al.* (1999), *Variation and evolution of the citric-acid cycle: a genomic perspective*, **TRENDS Microbiol.** **7**, 281-291.
- 405 SRINIVASAN V *et al.* (2006), *Ancient genes in contemporary persistent microbial pathogens*, **Biol. Bull.** **210**, 1-9.
- 406 HICKMAN J & G Ashwell (1960), *Uronic acid metabolism in bacteria, Purification properties of D-altronic acid and D-mannonic acid dehydrogenases in Escherichia coli*, **J. Biol. Chem.** **235**, 1566-1570.
- 407 PORTALIER R *et al.* (1980), *Regulation of Escherichia coli K-12 hexuronate system genes: exu regulation*, **J. Bacteriol.** **143**, 1095-1107.
- 408 GIJSEGEM FV & A Toussaint (1983), *In vivo cloning of Erwinia carotovora genes involved in the catabolism of hexuronates*, **J. Bacteriol.** **154**, 1227-1235.
- 409 SHULAMI S *et al.* (1999), *The glucuronic acid utilization gene cluster from Bacillus stearothermophilus T-6*, **J. Bacteriol.** **181**, 3695-3704.
- 410 FARMER III JJ & RG Eagon (1969), *Aldohexuronic acid catabolism by a soil Aeromonas*, **J. Bacteriol.** **97**, 97-106.
- 411 LINSTER CL & EV Schaftingen (2004), *A spectrophotometric assay of D-glucuronate based on Escherichia coli uronate isomerase and mannonate dehydrogenase*, **Prot. Expr. Purif.** **37**, 352-360.

- ⁴¹² HUGOUVIEUX-COTTE-PATTAT N & J Robert-Baudouy (1987), *Hexuronate catabolism in Erwinia chrysantemi*, **J. Bacteriol.** **169**, 1223-1231.
- ⁴¹³ ASHWELL G *et al.* (1960), *Uronic acid metabolism in bacteria, Purification properties of uronic acid isomerase in Escherichia coli*, **J. Biol. Chem.** **235**, 1559-1565.
- ⁴¹⁴ RITZENTHALER P *et al.* (1981), *Molecular cloning of Escherichia coli K-12 hexuronate system genes: exu region*, **J. Bacteriol.** **145**, 181-190.
- ⁴¹⁵ MATA-GILSINGER M *et al.* (1983), *Characterization of the operator sites of the exu regulon in Escherichia coli K-12 by operator-constitutive mutations and repressor titration*, **Genetics** **105**, 829-842.
- ⁴¹⁶ SCHWARZENBACHER R *et al.* (2003), *Crystal structure of uronate isomerase (TM0064) from Thermotoga maritima at 2.85 Å resolution*, **Proteins Struct. Funct. Bioinf.** **53**, 142-145.
- ⁴¹⁷ WILLIAMS L *et al.* (2006), *Uronate isomerase: a nonhydrolytic member of the amidohydrolase superfamily with an ambivalent requirement for a divalent metal ion*, **Biochemistry** **45**, 7453-7462.
- ⁴¹⁸ NGUYEN TT *et al.* (2009), *The mechanism of the reaction catalyzed by uronate isomerase illustrates how an isomerase may have evolved from a hydrolase within the amidohydrolase superfamily*, **Biochemistry** **48**, 8879-8890.
- ⁴¹⁹ MANDRAND-BERTHELOT MA & AE Lagarde (1977), *'Hit-and-run' mechanism for D-glucuronate reduction catalyzed by D-mannonate:NAD oxidoreductase of Escherichia coli*, **Biochim. Biophys. Acta** **483**, 6-23.
- ⁴²⁰ PORTALIER R & F Stoeber (1982), *D-mannonate and D-altronate NAD-dehydrogenases from Escherichia coli*, **Methods Enzymol.** **89**, 210-218.
- ⁴²¹ HUGOUVIEUX-COTTE-PATTAT N & Robert-Baudouy J (1983), *Regulation of expression of the uxu operon and of the uxuR regulatory gene in Escherichia coli K12*, **J. General Microbiol.** **129**, 3345-3353.
- ⁴²² BLANCO C *et al.* (1986), *The regulatory region of the uxuAB operon in Escherichia coli K12*, **Mol. Gen. Genet.** **202**, p.112.
- ⁴²³ SUVOROVA IA *et al.* (2011), *Comparative genomic analysis of the hexuronate metabolism genes and their regulation in Gammaproteobacteria*, **J. Bacteriol.** **193**, 3956-3963.
- ⁴²⁴ LATIL-DAMOTTE M & C Lares (1977), *Relative order of glg mutations affecting glycogen biosynthesis in Escherichia coli K12*, **Mol. Gen. Genet.** **150**, 325-329.
- ⁴²⁵ OKITA TW *et al.* (1981), *Biosynthesis of bacterial glycogen, cloning of the glycogen biosynthetic enzyme structural genes of Escherichia coli*, **J. Biol. Chem.** **256**, 6944-6952.
- ⁴²⁶ LEUNG PSC & J Preissi (1987), *Cloning of the ADPglucose pyrophosphorylase (glgC) and glycogen synthase (glgA) structural genes from Salmonella typhimurium LT2*, **J. Bacteriol.** **169**, 4349-4354.
- ⁴²⁷ ROMEO T *et al.* (1988), *Analysis of the Escherichia coli glycogen gene cluster suggests that catabolic enzymes are encoded among the biosynthetic genes*, **Gene** **70**, 363-376.
- ⁴²⁸ KUMAR A (1991), *Biosynthesis of bacterial glycogen and its regulation*, **Curr. Sci.** **60**, p.478.
- ⁴²⁹ ROMEO T *et al.* (1991), *A simple method for cloning genes involved in glucan biosynthesis: isolation of structural and regulatory genes for glycogen synthesis in Escherichia coli*, **Gene** **108**, 23-29.
- ⁴³⁰ KIEL JAKW *et al.* (1994), *Glycogen in Bacillus subtilis: molecular characterization of an operon encoding enzymes involved in glycogen biosynthesis and degradation*, **Mol. Microbiol.** **11**, 203-218.
- ⁴³¹ WILSON WA *et al.* (2010), *Regulation of glycogen metabolism in yeast and bacteria*, **FEMS Microbiol. Rev.** **34**, p.952.
- ⁴³² KAIDOW A *et al.* (1992), *The 55-kilodalton protein in an oriC complex fraction is glycogen synthase*, **J. Bacteriol.** **174**, 5454-5456.
- ⁴³³ KAWAGICHI K *et al.* (1978), *De novo synthesis of Escherichia coli glycogen is due to primer associated with glycogen synthase and activation by branching enzyme*, **Arch. Biochem.** **190**, 385-397.
- ⁴³⁴ KUMAR A *et al.* (1986), *Biosynthesis of bacterial glycogen, primary structure of cloning of Escherichia coli ADP-glucose:1,4-glucan, 4-glucosyltransferase as deduced from the nucleotide sequence of the glgA gene*, **J. Biol. Chem.** **261**, 16256-16259.
- ⁴³⁵ MORÁN-ZORZANO MT *et al.* (2007), *Occurrence of more than one important source of ADPglucose linked to glycogen biosynthesis in Escherichia coli and Salmonella*, **FEBS Lett.** **581**, p. 4423.
- ⁴³⁶ YEP A *et al.* (2004), *Identification and characterization of a critical region in the glycogen synthase from Escherichia coli*, **J. Biol. Chem.** **279**, 8359-8367.
- ⁴³⁷ FURUKAWA K *et al.* (1990), *Identification of lysine 15 at the active site in Escherichia coli glycogen synthase, conservation of a Lys-x-Gly-Gly sequence in the bacterial and mammalian enzymes*, **J. Biol. Chem.** **265**, 2086-2090.

- 438 FURUKAWA K et al. (1994), *Identification of Lys²⁷⁷ at the active site of Escherichia coli glycogen synthase, Application of affinity labelling combined with site-directed mutagenesis*, **J. Biol. Chem.** **269**, 868-871.
- 439 CID E et al. (2002), *Glycogen synthase: towards a minimum catalytic unit*, **FEBS Lett.** **528**, 5-11.
- 440 YEP A et al. (2004), *The active site of the Escherichia coli glycogen synthase is similar to the active site of retaining GT-B glycosyltransferase*, **Biochem. Biophys. Res. Commun.** **316**, 960-966.
- 441 BUSCHIAZZO A et al. (2004), *Crystal structure of glycogen synthase: homologous enzymes catalyze glycogen synthesis and degradation*, **EMBO J.** **23**, 3196-3205.
- 442 HORCAJADA C et al. (2006), *Crystal structure of an archaeal glycogen synthase, Insights into oligomerization and substrate binding of eukaryotic glycogen synthases*, **J. Biol. Chem.** **281**, 2923-2931.
- 443 YEP A et al. (2006), *The ADP-glucose binding site of the Escherichia coli glycogen synthase*, **Arch. Biochem. Biophys.** **453**, 188-196.
- 444 SHENG F et al. (2009), *The crystal structures of the open and catalytically competent closed conformation of Escherichia coli glycogen synthase*, **J. Biol. Chem.** **284**, 17796-17807.
- 445 SHENG F et al. (2009), *Oligosaccharide binding in Escherichia coli glycogen synthase*, **Biochemistry** **48**, 10089-10097.
- 446 PREISS J et al. (1965), *The activation of Escherichia coli ADP-glucose pyrophosphorylase*, **Biochem. Biophys. Res. Commun.** **18**, p.180&185.
- 447 GENTNER N & J Preiss (1967), *Activator-inhibitor interactions in the adenosine diphosphate glucose pyrophosphorylase of Escherichia coli*, **Biochem. Biophys. Res. Commun.** **27**, p.417,422&423.
- 448 HAUGEN TH et al. (1976), *Biosynthesis of bacterial glycogen, Characterization of the subunit structure of Escherichia coli B glucose-1-phosphate adenyltransferase (EC. 2.7.2.27)*, **J. Biol. Chem.** **251**, 7880-7885.
- 449 PARSONS TF & J Preiss (1978), *Biosynthesis of bacterial glycogen, Incorporation of pyridoxal phosphate into the allosteric activator site and an ADP-glucose-protected pyridoxal phosphate binding site of Escherichia coli B ADP-glucose synthase*, **J. Biol. Chem.** **253**, 6197-6202.
- 450 PREISS J et al. (1983), *Regulation of bacterial glycogen synthesis*, **Mol. Cell. Biochem.** **57**, 61-80.
- 451 PREISS J & E Greenberg (1983), *Pyrophosphate may be involved in regulation of bacterial glycogen synthesis*, **Biochem. Biophys. Res. Commun.** **115**, 820-826.
- 452 LEUNG PSC & J Preiss (1987), *Biosynthesis of bacterial glycogen: primary structure of Salmonella typhimurium ADPglucose synthetase as deduced from the nucleotide sequence of the glgC gene*, **J. Bacteriol.** **169**, 4355-4360.
- 453 SMITH-WHITE BJ & J Preiss (1992), *Comparison of proteins of ADP-glucose pyrophosphorylase from diverse plants*, **J. Mol. Evol.** **34**, 449-454.
- 454 WU MX & K Preiss (2001), *Truncated forms of the recombinant Escherichia coli ADP-glucose pyrophosphorylase: The importance of the N-terminal region for allosteric activation and inhibition*, **Arch. Biochem. Biophys.** **389**, 159-165.
- 455 BEJAR CM et al. (2006), *ADP-glucose pyrophosphorylase's N-terminus: Structural role in allosteric regulation*, **Biochem. Biophys. Res. Commun.** **343**, 216-221.
- 456 BEJAR CM et al. (2004), *The ADP-glucose pyrophosphorylase from Escherichia coli comprises two tightly bound distinct domains*, **FEBS Lett.** **573**, 99-104.
- 457 BALLICORA MA et al. (2003), *ADP-glucose pyrophosphorylase, a regulatory enzyme for bacterial glycogen synthesis*, **Microbiol. Mol. Biol. Rev.** **67**, 213-225.
- 458 KO JH et al. (1996), *Purification and characterization of a thermostable ADP-glucose pyrophosphorylase*, **Biochem. J.** **319**, 977-983.
- 459 HAUGEN T et al. (1976), *ADP glucose pyrophosphorylase: evidence for a lysine residue at the activator state of the Escherichia coli B enzyme*, **Biochem. Biophys. Res. Commun.** **69**, p.346,351&352.
- 460 LECKI MP et al. (1984), *Evidence that cyclic AMP stimulates bacterial glycogen synthesis by revealing AMP inhibition of and by increasing the cellular level of ADP-glucose synthetase*, **Arch. Biochem. Biophys.** **235**, 493-503
- 461 MOO-LEE Y et al. (1987), *Amino acid sequence of an Escherichia coli ADPglucose synthetase allosteric mutant as detected from the DNA sequence*, **Nucleic Acids Res.** **15**, p.1063.
- 462 KUMAR A et al. (1989), *Biosynthesis of bacterial glycogen, Determination of the amino acid changes that alter the regulatory properties of a mutant Escherichia coli ADP-glucose synthetase*, **J. Biol. Chem.** **264**, 10464-10471.
- 463 HILL MA et al. (1991), *Biosynthesis of bacterial glycogen, Mutagenesis of a catalytic site residue of ADP-glucose pyrophosphorylase from Escherichia coli*, **J. Biol. Chem.** **266**, 12445-12460.

- 464 GHOSH P *et al.* (1992), *Cloning, expression, and nucleotide sequence of glgC gene from an allosteric mutant of Escherichia coli B*, *Arch. Biochem. Biophys.* **296**, 122-128.
- 465 MEYER CR *et al.* (1993), *Cloning, expression, and sequence of an allosteric mutant ADPglucose pyrophosphorylase from Escherichia coli B*, *Arch. Biochem. Biophys.* **302**, 64-71.
- 466 BEJAR CM *et al.* (2006), *Molecular architecture of the glucose 1-phosphate site in ADP-glucose pyrophosphorylase*, *J. Biol. Chem.* **281**, 40473-40484.
- 467 BALLICORA MA *et al.* (2007), *Identification of regions critically affecting kinetics and allosteric regulation of the Escherichia coli ADP-glucose pyrophosphorylase by modeling and pentapeptide-scanning mutagenesis*, *J. Bacteriol.* **189**, 5325-5333.
- 468 FIGUEROA CM *et al.* (2011), *Understanding the allosteric trigger for the fructose-1,6-bisphosphate regulation of the ADP-glucose pyrophosphorylase from Escherichia coli*, *Biochimie* **93**, 1816-1823.
- 469 HAMAMOTO T *et al.* (1998), *Phosphopentomutase of Bacillus stearothermophilus TH6-2: the enzyme and its gene*, *Biosci. Biotechnol. Biochem.* **62**, 1103-1108.
- 470 PANOSIAN TD *et al.* (2010), *Crystallization and preliminary X-ray analysis of a phosphopentomutase from Bacillus cereus*, *Acta Cryst.* **F66**, 811-814.
- 471 PANOSIAN TD *et al.* (2011), *Bacillus cereus phosphopentomutase is an alkaline phosphatase family member that exhibits an altered entry point into the catalytic cycle*, *J. Biol. Chem.* **286**, 8043-8054.
- 472 IVERSON TM *et al.* (2011), *Molecular differences between a mutase and a phosphatase: investigations of the activation step in Bacillus cereus, phosphomutase*, *Biochemistry* **51**, 1964-1975.
- 473 RASHID N *et al.* (2004), *Presence of a novel phosphopentomutase and a 2-desoxiribose 5-phosphate aldolase reveals a metabolic link between pentoses and a central carbon metabolism in the hyperthermophilic archaeon Thermococcus kodakarensis*, *J. Bacteriol.* **186**, 4185-4191.
- 474 GALPERIN MY *et al.* (1998), *A superfamily of metalloenzymes unifies phosphopentomutase and cofactor-independent phosphoglycerate mutase with phosphatases and sulfatases*, *Prot. Sci.* **7**, 1829-1835.
- 475 KOONIN EV & RL Tatusov (1994), *Computer analysis of bacterial haloacid dehalogenases defines a large superfamily of hydrolases with diverse specificity*, *J. Mol. Biol.* **244**, 125-132.
- 476 BURROUGHS AM *et al.* (2006), *Evolutionary genomics of the HAD superfamily: Understanding the structural adaptations and catalytic diversity in a superfamily of phosphoesterases and allied enzymes*, *J. Mol. Biol.* **361**, 1003-1034.
- 477 HISANO T *et al.* (1996), *Crystal structure of L-2-haloacid dehalogenase from Pseudomonas sp. YL, an α/α hydrolase structure that is different from the α/α hydrolase fold*, *J. Biol. Chem.* **271**, p.20322&20323.
- 478 CALDERONE V *et al.* (2006), *A structure-based proposal for the catalytic mechanism of the bacterial acid phosphatase AphA belonging to the DDDD superfamily of phosphohydrolases*, *J. Mol. Biol.* **355**, 708-721.
- 479 THALLER MC *et al.* (1998), *Conserved sequence motifs among bacterial, eukaryotic, and archaeal phosphatases that define a new phosphohydrolase family*, *Prot. Sci.* **7**, 1647-1652.
- 480 CALDERONE V *et al.* (2004), *The first structure of a bacterial class B acid phosphatase reveals further structural heterogeneity dehalogenase fold*, *J. Mol. Biol.* **335**, p.761&762.
- 481 LAHIRI SD *et al.* (2006), *Diversification of function in the haloacid dehalogenase enzyme superfamily: The role of the cap domain in hydrolytic phosphorus-carbon bond cleavage*, *Bioorg. Chem.* **34**, 394-409.
- 482 ARAVIND L *et al.* (1998), *The catalytic domain of the P-type ATPase has the haloacid dehalogenase fold*, *TRENDS Biochem. Sci.* **23**, 127-129.
- 483 RIDDER IS & BW Dijkstra (1999), *Identification of the Mg²⁺-binding site in the P-type ATPase and phosphatase members of the HAD (haloacid dehalogenase) superfamily by structural of the response regulator protein CheY*, *Biochem. J.* **339**, 223-226.
- 484 KUZNETSOVA E *et al.* (2006), *Genome-wide analysis of substrate specificities of the Escherichia coli haloacid dehalogenase-like phosphatase family*, *J. Biol. Chem.* **281**, 36149-36161.
- 485 http://www.genome.jp/dbget-bin/www_bget?ec:3.1.3.25
- 486 BONE R *et al.* (1992), *Structure of inositol monophosphatase, The putative target of lithium therapy*, *PNAS* **89**, 10031-10035.
- 487 POLLACK SJ *et al.* (1993), *Probing the role of metal ions in the mechanism of inositol monophosphatase by site-direct mutagenesis*, *Eur. J. Biochem.* **217**, 281-287.
- 488 POLLACK ST *et al.* (1994), *Mechanism of inositol monophosphatase, the putative of lithium therapy*, *PNAS* **91**, 5766-5770.

- 489 ATACK JR *et al.* (1995), *Structure and mechanism of inositol monophosphatase*, **FEBS Lett.** **361**, 1-7.
- 490 PARISH T *et al.* (1997), *A Mycobacterium smegmatis mutant with a defective inositol monophosphate phosphatase gene homolog has altered envelope permeability*, **J. Bacteriol.** **179**, 7827-7833.
- 491 NIGOU J *et al.* (2002), *Purification and biochemical characterization of Mycobacterium tuberculosis SuhB, an inositol monophosphatase involved in inositol biosynthesis*, **Biochemistry** **41**, 4392-4398.
- 492 BROWN A *et al.* (2007), *Dimerization of inositol monophosphatase Mycobacterium tuberculosis SuhB is not constitutive, but induced by binding of the activator Mg²⁺*, **BMC Struct. Biol.** **7**(55).
- 493 BATTACHARAYYA S *et al.* (2012), *Crystal structure of the Staphylococcal dual specific inositol monophosphatase/NADP(H) phosphatase (SAS2203) delineates the molecular basis of substrate specificity*, **Biochimie** **94**, 879-890.
- 494 CHEN L & MF Roberts (1999), *Characterization of a tetrameric inositol monophosphatase from the hyperthermophile bacteria Thermotoga maritima*, **Appl. Environ. Microbiol.** **65**, 45559-4567.
- 495 STIEGLITZ KA *et al.* (2007), *Crystal structure of the tetrameric inositol monophosphatase (TM1415) from the hyperthermophile Thermotoga maritima*, **FEBS J.** **274**, 2461-2469.
- 496 CHEN L & MF Roberts (1998), *Cloning and expression of the inositol monophosphatase gene from Methanococcus jannaschii and characterization of the enzyme*, **Appl. Environ. Microbiol.** **64**, 2609-2615.
- 497 STEC B *et al.* (2000), *MJ0109 is an enzyme that is both an inositol monophosphatase and the 'missing' archaeal fructose-1,6-bisphosphatase*, **Nat. Struct. Biol.** **7**, 1046-1050.
- 498 JOHNSON KA *et al.* (2001), *Crystal structure and catalytic mechanism of the MJ0109 gene product: a bifunctional enzyme with inositol monophosphatase and fructose-1,6-bisphosphatase activities*, **Biochemistry** **40**, 618-630.
- 499 LI Z *et al.* (2010), *Mobile loop mutation in an archaeal inositol monophosphatase: modulating three-metal ion assisted catalysis and lithium inhibition*, **Prot. Sci.** **19**, 309-318.
- 500 STIEGLITZ KA *et al.* (2002), *Crystal structure of a dual activity IMPase/FBPase (AF2372) from Archaeoglobus fulgidus, A story of a mobile loop*, **J. Biol. Chem.** **277**, p.22863-22864.
- 501 ZHANG Y *et al.* (1993), *Structural similarities between fructose-1,6-bisphosphatase and inositol monophosphatase*, **Biochem. Biophys. Res. Commun.** **190**, 1080-1083.
- 502 YORK JD *et al.* (1995), *Definition of a metal-dependent/Li²⁺-inhibited phosphomonoesterase protein family based upon a conserved three-dimensional core structure*, **PNAS** **92**, p.5149.
- 503 PATEL S *et al.* (2002), *Crystal structure of an enzyme displaying both inositol-polyphosphate-2-phosphatase and 3-phosphoadenosine-5'-phosphate phosphatase activities: a novel target of lithium therapy*, **J. Mol. Biol.** **315**, p.677-679.
- 504 CUNCHILLOS Ch & G Lecointre (2003), *Evolution of amino acid metabolism inferred through cladistic analysis*, **J. Biol. Chem.** **278**, 47960-47970.
- 505 KIMURA A & Y Inoue (1993), *Glyoxylase I in micro-organisms: molecular characteristics, genetics and biochemical regulation*, **Biochem. Soc. Transact.** **21**, 518-522.
- 506 FEIRBERG I *et al.* (1999), *Energetics of the proposed rate-determining step of the glyoxylase I reaction*, **FEBS Lett.** **453**, p.90.
- 507 CREIGHTON DJ & DS Hamilton (2001), *Brief history of glyoxylase I and what we have learned about metal ion-dependent, enzyme-catalyzed isomerizations*, **Arch. Biochem. Biophys.** **387**, 1-10.
- 508 HIMO F & PEM Siegbahn (2001), *Catalytic mechanism of glyoxylase I: A theoretical study*, **J. Am. Chem. Soc.** **123**, 10280-10289.
- 509 ARMSTRONG RN (2000), *Mechanistic diversity in a metalloenzyme superfamily*, **Biochemistry** **39**, 13625-13632.
- 510 O'YOUNG J *et al.* (2007), *Escherichia coli glyoxylase II is a binuclear zinc-dependent metalloenzyme*, **Arch. Biochem. Biophys.** **459**, 20-26.
- 511 SUTTISANSANEE U & JF Honek (2011), *Bacterial glyoxylase enzymes*, **Semin. Cell Dev. Biol.** **22**, 285-292.
- 512 MULLINGS KY *et al.* (2012), *Ni²⁺-activated glyoxylase I from Escherichia coli: substrate specificity, kinetic isotope effects and evolution within the □□□□ superfamily*, **J. Inorg. Biochem.** **108**, 133-140.
- 513 <http://www.ncbi.nlm.nih.gov/books/NBK22475/>
- 514 HIROTSU K *et al.* (2005), *Dual recognition of aminotransferases*, **Chem. Rec.** **5**, p.160.
- 515 JENSEN RA & DH Calhoun (1981), *Intracellular roles of microbial aminotransferases: overlap enzymes across different biochemical pathways*, **Crit. Rev. Microbiol.** **8**, p.229-252.
- 516 MEHTA PK *et al.* (1993), *Aminotransferases: demonstration of homology and division into evolutionary subgroups*, **Eur. J. Biochem.** **214**, 549-561.

- 517 MEHTA PK *et al.* (1989), *Evolutionary relationships among aminotransferases. Tyrosine aminotransferase, histidinol-phosphate aminotransferase, and aspartate aminotransferase are homologous proteins*, **Eur. J. Biochem.** **186**, 249-253.
- 518 CHRISTEN P *et al.* (1990), *Evolutionary and biosynthetic aspects of aspartate aminotransferase isoenzymes and other aminotransferases*, **Ann. N.Y. Acad. Sci.** **585**, 331-338.
- 519 ALEXANDER FW *et al.* (1994), *Evolutionary relationships among pyridoxal-5-phosphate-dependent enzymes, Regio-specific □, □ and □ families*, **Eur. J. Biochem.** **219**, 953-960.
- 520 JENSEN RA & W Gu (1996), *Evolutionary relationships of biochemically specialized subdivisions of family I within the protein superfamily of aminotransferases*, **J. Bacteriol.** **178**, 2161-2171.
- 521 YOSHIMURA T *et al.* (1996), *Stereospecificity for the hydrogen transfer and molecular evolution of pyridoxal enzymes*, **Biosci. Biotech. Biochem.** **60**, 181-187.
- 522 JANSONIUS JN (1998), *Structure, evolution and actions of vitamin B₆-dependent enzymes*, **Curr. Opin. Struct. Biol.** **8**, 759-769.
- 523 DENESSIOUK KA *et al.* (1999), *Common structural elements in the architecture of the cofactor-binding domains in unrelated families of pyridoxal phosphate-dependent enzymes*, **Proteins** **35**, 250-261.
- 524 SCHNEIDER G *et al.* (2000), *The manifold of vitamin B6 dependent enzymes*, **Structure** **8**, R1-R6.
- 525 SALZMANN D *et al.* (2000), *Rates of evolution of pyridoxal-5'-phosphate-dependent enzymes*, **Biochem. Biophys. Res. Commun.** **270**, 576-580.
- 526 CHRISTEN P & PK Mehta (2001), *From cofactor to enzymes, The molecular evolution of pyridoxal-5'-phosphate-dependent enzymes*, **Chem. Rec.** **1**, 436-447.
- 527 PITTARD J & BJ Wallace (1966), *Distribution and function in genes concerned with aromatic biosynthesis in Escherichia coli*, **J. Bacteriol.** **91**, 1494-1508.
- 528 HERRMANN KM (1995), *The shikimate pathway: early steps in the biosynthesis of aromatic compounds*, **Plant Cell** **7**, 907-919.
- 529 HERRMANN KM & LM Weaver (1999), *The shikimate pathway*, **Annu. Rev. Plant Physiol. Plant Mol. Biol.** **50**, 473-503.
- 530 SKINNER MA *et al.* (1998), *Expression of recombinant aroB-encoded dehydroquinase synthase from pathogenic microorganisms*, **Biochem. Soc. Transact.** **26**, S280.
- 531 Le MARÉCHAL P & R Azerad (1976), *The shikimate pathway, III.- 3-dehydroquinase synthetase of E. coli, Mechanistic studies by kinetic isotope effect*, **Biochimie** **58**, 1123-1128.
- 532 HASAN N & Nester EW (1978), *Dehydroquinase synthase in Bacillus subtilis, An enzyme associated with chorismate synthase and flavin reductase*, **J. Biol. Chem.** **253**, 4999-5004.
- 533 MAITRA US & DB Sprinson (1978), *5-dehydro-3-deoxy-D-arabino-hepulosonic acid 7-phosphohate, an intermediate in the 3-dehydroquinase synthase reaction*, **J. Biol. Chem.** **253**, 5426-5430.
- 534 Le MARÉCHAL P *et al.* (1980), *The interaction of phosphonate and homophosphonate analogues of 5-dehydro-3-deoxy-D-arabino-hepulosonate 7-phosphohate with 3-dehydroquinase synthetase from E. coli*, **Biochem. Biophys. Res. Commun.** **92**, 1104-1109.
- 535 FROST JW *et al.* (1984), *Dehydroquinase synthase from Escherichia coli: purification, cloning, and construction of overproducers of the enzyme*, **Biochemistry** **23**, 4470-4475.
- 536 MEHDI S *et al.* (1987), *3-dehydroquinase synthetase from E. Coli, and its substrate 3-deoxy-D-arabino-hepulosonic acid 7-phosphohate*, **Meth. Enzymol.** **142**, 306-314.
- 537 BENDER SL *et al.* (1989), *Dehydroquinase synthase: the role of divalent metal cations and of nicotinamide adenine dinucleotide in catalysis*, **Biochemistry** **28**, 7555-7560.
- 538 BENDER SL *et al.* (1989), *Dehydroquinase synthase: the use of substrate analogues to probe the early steps of the catalyzed reaction*, **Biochemistry** **28**, 7560-7572.
- 539 WIDLANSKI T *et al.* (1989), *Dehydroquinase synthase: the use of substrate analogues to probe the late steps of the catalyzed reaction*, **Biochemistry** **28**, 7572-7582.
- 540 MILLAR G & JR Coggins (1986), *The complete amino acid sequence of 3-dehydroquinase synthase of Escherichia coli K12*, **FEBS Lett** **200**, 11-17.
- 541 KLEANTHOUS C *et al.* (1990), *Active site labeling of the shikimate pathway enzyme, dehydroquinase*, **J. Biol. Chem.** **265**, 10929-10934.
- 542 CARPENTER EP *et al.* (1998), *Structure of dehydroquinase synthase reveals an active site capable of multistep catalysis*, **Nature** **394**, 299-302.
- 543 BROWN KA *et al.* (2003), *Twists and turns: a tale of two shikimate-pathway enzymes*, **Biochem. Soc. Transact.** **31**, 543-547.
- 544 PARK A *et al.* (2004), *Biophysical and kinetic analysis of wild-type and site-directed mutants of the isolated and native dehydroquinase synthase domain of the AROM protein*, **Protein Sci.** **13**, 2108-2119.
- 545 SUGAHARA M *et al.* (2005), *Crystal structure of dehydroquinase synthase from Thermus thermophilus HB8 showing functional importance of the dimeric state*, **Proteins** **58**, 249-252.

- 546 De MENDONÇA JD *et al.* (2007), *Functional characterization by genetic complementation of aroB-encoded dehydroquinase synthase from Mycobacterium tuberculosis H37Rv and its heterologous expression and purification*, **J. Bacteriol.** **189**, 6246-6252.
- 547 NGO PT *et al.* (2008), *Cloning, expression, crystallization and preliminary X-ray crystallographic analysis of 3-dehydroquinase synthase, Xoo1243, from Xanthomonas oryzae pv. Oryzae*, **Acta Crystallogr. F Struct. Biol. Cryst. Commun.** **64**, p.1128.
- 548 LIU JS *et al.* (2008), *Structure-based inhibitor discovery of Helicobacter pylori dehydroquinase synthase*, **Biochem. Biophys. Res. Commun.** **373**, 1-7.
- 549 NEGRON L *et al.* (2011), *Expression, purification, and characterization of dehydroquinase synthase from Pyrococcus furiosus*, **Enzyme Res.** **ID 134893**.
- 550 GARBE T *et al.* (1991), *The Mycobacterium tuberculosis shikimate pathway genes: evolutionary relationships between biosynthetic and catabolic 3-dehydroquinases*, **Mol. Gen. Genet.** **228**, 385-392.
- 551 HAWKINS AR *et al.* (1993), *The genetics and enzymology of multienzyme systems*, **Biochem. Soc. Transact.** **21**, 181-186.
- 552 NICHOLS CE *et al.* (2004), *Comparison of ligand-induced conformational changes and domain closure mechanisms, between prokaryotic and eukaryotic dehydroquinase synthases*, **J. Mol. Biol.** **343**, 533-546.
- 553 NEGRON L & EJ Parker (2011), *Fluorinated substrates result in variable leakage of a reaction intermediate during catalysis by dehydroquinase synthase*, **Org. Biomol. Chem.** **9**, 2861-2867.
- 554 MORELL H *et al.* (1967), *The enzymic synthesis of chorismic and prephenic acids from 3-enolpyruvylshikimic acid 5-phosphate*, **J. Biol. Chem.** **242**, 82-90.
- 555 FLOSS HG *et al.* (1972), *Stereochemistry of the 3-deoxy-D-arabino-heptulosonate 7-phosphate synthetase reaction and the chorismate synthetase reaction*, **J. Biol. Chem.** **247**, 736-744.
- 556 CHARLES IG *et al.* (1990), *Isolation, characterization and nucleotide sequences of the aroC gene encoding chorismate synthase from Salmonella typhi and Escherichia coli*, **J. Gen. Microbiol.** **136**, 353-358.
- 557 HAWKES TR *et al.* (1990), *Chorismate synthase, Pre-steady-state kinetics of phosphate release from 5-enolpyruvylshikimate 3-phosphate*, **Biochem. J.** **265**, 899-902.
- 558 RAMJEE MK *et al.* (1994), *A continuous, anaerobic spectrophotometric assay for chorismate synthase activity that utilizes photoreduced flavin mononucleotide*, **Analyst. Biochem.** **220**, 137-141.
- 559 BALASUBRAMANIAN S *et al.* (1995), *Observation of a secondary tritium isotope effect in the chorismate synthase reaction*, **Biochemistry** **24**, 341-348.
- 560 BORNEMANN S *et al.* (1995), *Escherichia coli chorismate synthase: a deuterium kinetic-isotope effect under single-turnover and steady-state conditions shows that a flavin intermediate forms before the C-(6proR)-H bond is cleaved*, **Biochem. J.** **305**, 707-710.
- 561 BORNEMANN S *et al.* (1995), *Escherichia coli chorismate synthase: catalyzes the conversion of (6S)-6-fluoro-5-enolpyruvylshikimate-3-phosphate to 6-fluorochorismate*, **J. Biol. Chem.** **270**, 22811-22815.
- 562 MACHERAUX P *et al.* (1996), *Binding of the oxidized, reduced, and radical flavin species to chorismate synthase, An investigation by spectrophotometry, fluorimetry, and electron paramagnetic resonance and electron nuclear double resonance spectroscopy*, **Biochemistry** **35**, 1643-1652.
- 563 BORNEMANN S *et al.* (1996), *The transient kinetics of Escherichia coli chorismate synthase: substrate consumption, product formation, phosphate dissociation, and characterization of a flavin intermediate*, **Biochemistry** **35**, 9907-9916.
- 564 MACHERAUX P *et al.* (1996), *Studies with flavin analogues provide evidence that a protonated reduced FMN is the substrate-induced transient intermediate in the reaction of Escherichia coli chorismate synthase*, **J. Biol. Chem.** **271**, 25850-55858.
- 565 BORNEMANN S *et al.* (2000), *A secondary \square deuterium kinetic isotope effect in the chorismate synthase reaction*, **Bioorg. Chem.** **28**, 191-194.
- 566 DIMITRENKO O *et al.* (2001), *A theoretical study of the chorismate synthase reaction*, **Org. Lett.** **3**, 4137-4140.
- 567 ELY F *et al.* (2008), *The Mycobacterium tuberculosis Rv2540c DNA sequence encodes a bifunctional chorismate synthase*, **BMC Biochem.** **9**:13.
- 568 HORSBURGH MJ *et al.* (1996), *Chorismate synthase from Staphylococcus aureus*, **Microbiology** **142**, 2943-950.
- 569 BORNEMANN S *et al.* (1996), *Escherichia coli chorismate synthase*, **Biochem. Soc. Transact.** **24**, 84-88.
- 570 MACHERAUX P *et al.* (1998), *Evidence for a major structural change in Escherichia coli chorismate synthase induced by flavin and substrate binding*, **Biochem. J.** **335**, 319-327.
- 571 OSBORNE A *et al.* (2000), *Studies with substrate and cofactor analogues provide evidence for a radical mechanism in the chorismate synthase reaction*, **J. Biol. Chem.** **275**, 35825-35830

- 572 AHN HJ *et al.* (2003), *Crystalization and preliminary X-ray crystallographic studies of chorismate synthase from Helicobacter pylori*, **Acta Cryst. D** **59**, p.569.
- 573 BORNEMANN S *et al.* (2003), *A branch point in chorismate synthase research*, **Structure** **11**, 1463-1464.
- 574 MACLEAN J & S Ali (2003), *The structure of chorismate synthase reveals a novel flavin-binding site fundamental to a unique chemical reaction*, **Structure** **11**, 1499-1511.
- 575 DIAS MVB *et al.* (2004), *Crystalization and preliminary X-ray crystallographic studies of chorismate synthase from Mycobacterium tuberculosis*, **Acta Cryst. D** **60**, p.2003.
- 576 DIAS MVB *et al.* (2006), *Structure of chorismate synthase from Mycobacterium tuberculosis*, **J. Struct. Biol.** **154**, 130-143.
- 577 ZHOU H *et al.* (2006), *Homology modelling and molecular dynamics study of chorismate synthase from Shigella flexeneri*, **J. Mol. Graph. Model.** **25**, 434-441.
- 578 FERNANDES CL *et al.* (2007), *A structural model for chorismate synthase from Mycobacterium tuberculosis in complex with coenzyme and substrate*, **Comput. Biol. Med.** **37**, 149-158.
- 579 SCHMIDT J *et al.* (1993), *A novel operon organization involving the genes for chorismate synthase (aromatic biosynthesis pathway) and ribosomal GTPase center proteins (L11, L1, L10, L12: rplKAJL) in cyanobacterium Synechocystis PCC 6803*, **J. Biol. Chem.** **268**, 27-447-27457.
- 580 MACHEROUX P *et al.* (1999), *A unique reaction in a common pathway: mechanism and function of chorismate in the shikimate pathway*, **Planta** **207**, 325-334.
- 581 FITZPATRICK TB *et al.* (2001), *Chorismate synthase from the hyperthermophile Thermotoga maritima combines thermostability and increased rigidity with catalytic and spectral properties similar to mesophilic counterparts*, **J. Biol. Chem.** **276**, 18052-18059.
- 582 AHN HJ *et al.* (2004), *Crystal structure of chorismate synthase: a novel FMN-binding protein fold and functional insights*, **J. Mol. Biol.** **336**, 903-915.
- 583 VIOLA CM *et al.* (2004), *Crystal structure of chorismate synthase from Aquifex aeolicus reveals a novel beta alpha beta sandwich topology*, **Proteins** **54**, 166-169.
- 584 MASSIÈRE F & MA Badet-Denisot (1998), *The mechanism of glutamine-dependent amidotransferases*, **Cell Mol. Life Sci.** **54**, 205-222.
- 585 RAUSHEL FM *et al.* (1999), *The amidotransferase family of enzymes: molecular machines for the production and delivery of ammonia*, **Biochemistry** **38**, 7891-7899.
- 586 JACKSON EN & C Yanofsky (1974), *Localization of two functions of the phosphoribosyl anthranilate transferase of Escherichia coli to distinct regions of the polypeptide chain*, **J. Bacteriol.** **117**, 502-508.
- 587 REINERS JJ & H Zalkin (1975), *Immunological study of anthranilate synthase*, **J. Bacteriol.** **123**, 620-630.
- 588 HOROWITZ H *et al.* (1982), *Nucleotide sequence of the trpD gene, encoding anthranilate synthetase component II of Escherichia coli*, **J. Mol. Biol.** **156**, 245-256.
- 589 ESSAR DW *et al.* (1990), *DNA sequences and characterization of four early genes of the tryptophan pathway in Pseudomonas aeruginosa*, **J. Bacteriol.** **172**, 853-866.
- 590 TANG X *et al.* (1999), *The tryptophan biosynthesis gene cluster trpCDEGFBA from Pyrococcus kodakaraensis KOD1 is regulated at the transcriptional level and expressed as a single mRNA*, **Mol. Gen. Genet.** **262**, 815-821.
- 591 KNÖCHEL T *et al.* (1999), *The crystal structure of anthranilate synthase from Sulfolobus solfataricus: Functional implications*, **PNAS** **96**, 9479-9484.
- 592 SPRAGGON G *et al.* (2001), *The structures of anthranilate synthase of Serratia marcescens crystallized in the presence of (i) its substrates, chorismate and glutamine, and (ii) its end-product inhibitor, L-tryptophan*, **PNAS** **98**, 6021-6026.
- 593 MOROLLO AA *et al.* (2001), *Structure of the cooperative allosteric anthranilate synthase from Salmonella typhimurium*, **Nat. Struct. Biol.** **8**, 243-247.
- 594 <http://www.ncbi.nlm.nih.gov/books/NBK21216/>
- 595 LAMBDEN PL & WT Drabble (1973), *The gua operon of Escherichia coli K-12: evidence for polarity from guaB to guaA*, **J. Bacteriol.** **115**, 992-1002.
- 596 LEE BH & SC Hartman (1974), *Preferential utilization of glutamine for amination of xanthosine 5'-phosphate to guanosine 5'-phosphate by purified enzymes from Escherichia coli*, **Biochem. Biophys. Res. Commun.** **60**, 918-925.
- 597 SPECTOR T *et al.* (1974), *GMP synthetase from Escherichia coli B-96, Interactions with substrate analogs*, **Biochim. Biophys. Acta** **370**, 585-591.
- 598 SPECTOR T & LM Beacham III (1975), *Guanosine monophosphatase synthase from Escherichia coli B-96, Inhibition by nucleosides*, **J. Biol. Chem.** **250**, 3101-3107.
- 599 von der SAAL W *et al.* (1985), *Positional isotope exchange and kinetic experiments with Escherichia coli guanosine-5'-monophosphate synthetase*, **Biochemistry** **24**, 5343-5350.
- 600 SAKAMOTO N (1978), *GMP synthase (Escherichia coli)*, **Meth. Enzymol.** **L1**, 213-218.
- 601 BUCHANAN JM (1973), *The amidotransferases*, *Adv. Enzymol. Areas Mol. Biol.* **39**, p.91.

- 602 GRIMALDI C *et al.* (2000), *Genetic organization and polymorfism of the guaA gene encoding the GMP syhntetase in Lactobacillus rhamnosus*, **Curr. Microbiol.** **40**, 245-249.
- 603 ZALKIN H (1985), *GMP synthetase*, *Meth. Enzymol.* **113**, 273-278.
- 604 TIEDMAN AA *et al.* (1985), *Nucleotide sequence of the guaA gene encoding GMP synthetase of Escherichia coli*, **J. Biol. Chem.** **260**, 8876-8879.
- 605 TESMER JG *et al.* (1994), *Preliminary X-ray analysis of Escherichia coli GMP synthetase, Determination of anomalous scattering factor for a cysteinyl mercury derivative*, **Proteins** **18**, 394-403.
- 606 TESMER JG *et al.* (1996), *The crystal structure of GMP synthetase reveals a novel catalytic triad and is a structural paradigm for two enzyme families*, **Nat. Struct. Biol.** **3**, 74-86.
- 607 FRANCO TMA *et al.* (2012), *Biochemical characterization of of recombinant guaA-encoded guanosine monophosphate synthase (EC 6.3.5.2) from Mycobacterium tuberculosis H37Rv strain*, **Arch. Biochem. Biophys.** **517**, 1-11.
- 608 ZALKIN H & CD Truitt (1977), *Characterization of the glutamine site of Escherichia coli guanosine 5'-monophosphate synthetase*, **J. Biol. Chem.** **252**, 5431-5436.
- 609 ZALKIN H *et al.* (1985), *Identification of a trpG-related glutamine amide transfer domain in Escherichia coli GMP synthetase*, **J. Biol. Chem.** **260**, 3350-3354.
- 610 ABBOTT JL *et al.* (2006), *The effects of removing the GAT domain from E. coli GMP synthetase*, **Protein J.** **25**, 483-491.
- 611 BORK P & EV Koonin (1994), *A P-loop-like motif in a widespreaad ATP pyrophosphatase domain: implications for the evolution of sequence motifs and enzyme activity*, **Proteins** **20**, 347-355.
- 612 HOCHSTADT J (1978), *Hypoxanthine phosphoribosyltransferase and guanine phosphoribosyltransferase from enteric bacteria*, **Meth. Enzymol.** **L1**, 549-558.
- 613 MELTON DW *et al.* (1984), *Structure, expression, and mutation of the hypoxanthine phosphoribosyltransferase gene*, **PNAS** **81**, 2147-2151.
- 614 LEE CC *et al.* (1998), *A single amino acid substitution in the human and a bacterial hypoxanthine phosphoribosyltransferase modulates specificity for the binding of guanine*, **Biochemistry** **37**, 3491-3498.
- 615 FOCIA PJ *et al.* (1998), *Approaching the transition state in the crystal structure of a phosphoribosyltransferase*, **Biochemistry** **37**, 17120-17127.
- 616 CANYUK B *et al.* (2001), *The role for an invariant aspartic acid in hypoxanthine phosphoribosyltransferases is examined using saturation mutagenesis, functional analysis, and X-ray crystallography*, **Biochemistry** **40**, 2754-2765.
- 617 GUDDAT LW *et al.* (2002), *Crystal structures of free, IMP-, and GMP-bound Escherichia coli hypoxanthine phosphoribosyltransferase*, **Prot. Sci.** **11**, 1626-1638.
- 618 KUKIMOTO-NIINO M *et al.* (2005), *Crystal structures of a predicted phosphoribosyltransferase (TT1426) from Thermus thermophilus HB8 at 2.01 Å resolution*, **Prot. Sci.** **14**, 823-827.
- 619 DANTAS S D *et al.* (2008), *Biochemical and structural characterization of the hypoxantine-guanine-xanthine phosphoribosyltransferase from Pyrococcus horikoshii*, **Biochim. Biophys. Acta** **1784**, 953-960.
- 620 MUSHEGIAN AR & EV Koonin (1994), *Unexpected sequence similarity between nucelosidases and phosphoribosyltransferases of different specificity*, **Prot. Sci.** **3**, 1081-1088.
- 621 CRAIG III SP & AE Eakin (2000), *Purine phosphoribosyltransferases*, **J. Biol. Chem.** **275**, 20231-20234.
- 622 SINHA SC & JL Smith (2001), *The PRT family*, **Curr. Opin. Struct. Biol.** **11**, 733-739.
- 623 ESTUPIÑÁN B & VL Schramm (1994), *Guanosine-inosine preferring nucleoside N-glycohydrolase from Crithidia fasciculata*, **J. Biol. Chem.** **269**, 23068-23073.
- 624 GOPAUL DN *et al.* (1996), *Inosine-uridine nucleoside hydrolase from Crithidia fasciculata, Genetic characterization, crystalization, and identification of histidine 241 as a catalytic site residue*, **Biochemistry** **35**, 5963-5070.
- 625 DEGANO M *et al.* (1996), *Three-dimensional structure of the inosine-uridine nucleoside N-ribohydrolase from Crithidia fasciculata*, **Biochemistry** **35**, 5971-5081.
- 626 IOVANE E *et al.* (2008), *Structural basis for substrate specificity in group I nucleoside hydrolase*, **Biochemistry** **47**, 4418-4426.
- 627 VERSÉES W & J Steyaert (2003), *Catalysis by nucleoside hydrolases*, **Curr. Opin. Struct. Biol.** **13**, 731-738.
- 628 KOSZALKA GW *et al.* (1988), *Purification and properties of inosine-guanosine phosphorylase from Escherichia coli K-12*, **J. Bacteriol.** **170**, 3493-3498.
- 629 TKEHARA M *et al.* (1995), *Molecular cloning and nucleotidesequence of purine nucleoside phophorylase and uridine phosphorylase genes from Klebsiella sp.*, **Biosci. Biotech. Biochem.** **59**, 1987-1990.
- 630 HAMAMOTO T *et al.* (1997), *Cloning and expression of purine nuleoside phosphorylase I gene from Bacillus stearothermophilus TH 6-2*, **Biosci. Biotech. Biochem.** **61**, 272-275.

- 631 SCHUCH R *et al.* (1999), *Nucleosides as a carbon source in Bacillus subtilis: characterization of the *drm-pupG* operon*, **Microbiology** **145**, 2957-2966.
- 632 BASSO LA *et al.* (2001), *Purine nucleoside phosphorylase from Mycobacterium tuberculosis, Analysis of inhibition by a transition-state analogue and dissection by parts*, **Biochemistry** **40**, 8196-8203.
- 633 MAO Ch *et al.* (1997), *The crystal structure of Escherichia coli purine nucleoside phosphorylase: a comparison with the human enzyme reveals a conserved topology*, **Structure** **5**, 1373-1383.
- 634 KOELLNER G *et al.* (1998), *Crystal structure of the ternary complex of E. coli purine nucleoside phosphorylase with fosfomycin B, a structural analogue of the substrate inosine, and phosphate (sulphate) at 2.1 Å resolution*, **J. Mol. Biol.** **280**, 153-166.
- 635 TEBBE J *et al.* (1999), *Crystal structure of the purine nucleoside phosphorylase (PNP) from Cellulomonas sp. and its implication for the mechanism of trimeric PNPs*, **J. Mol. Biol.** **294**, 1239-1255.
- 636 KOELLNER G *et al.* (2002), *Open and closed conformation of the E. coli purine nucleoside phosphorylase active center and implications for the catalytic mechanism*, **J. Mol. Biol.** **315**, 351-371.
- 637 NOLASCO DO *et al.* (2004), *Crystallographic structure of PNP from Mycobacterium tuberculosis at 1.9 Å resolution*, **Biochem. Biophys. Res. Commun.** **324**, 789-794.
- 638 GRENHA R *et al.* (2005), *Structure of purine nucleoside phosphorylase (DeoD) from Bacillus anthracis*, **Acta Cryst.** **F61**, p.459&461.
- 639 CACCIAPUOTI G *et al.* (2007), *Biochemical and structural characterization of mammalian-like purine nucleoside phosphorylase from the archaeon Pyrococcus furiosus*, **FEBS J.** **274**, 2482-2495.
- 640 PAULI I *et al.* (2009), *Molecular modeling and dynamics studies of purine nucleoside phosphorylase from Bacteroides fragilis*, **J. Mol. Model.** **15**, 913-922.
- 641 HOU QM *et al.* (2009), *Preliminary crystallographic studies of purine nucleoside phosphorylase from the cariogenic pathogen Streptococcus mutans*, **Acta Cryst.** **F65**, p.1289.
- 642 DUCATI R *et al.* (2010), *Crystallographic and docking studies of purine nucleoside phosphorylase from Mycobacterium tuberculosis*, **Bioorg. Med. Chem.** **18**, 4769-4774.
- 643 PUGMIRE MJ & SE Ealick (2002), *Structural analysis reveal two distinct families of nucleoside phosphorylases*, **Biochem. J.** **361**, 1-25.
- 644 KONRAD A *et al.* (2012), *The evolution of catalytic residues and enzyme mechanism within the bacterial nucleoside phosphorylase superfamily I*, **Gene** **510**, 154-161.
- 645 PERETÓ J *et al.* (2004), *Ancestral lipid biosynthesis and early membrane evolution*, **TRENDS Biochem. Sci.** **29**, 470-477.
- 646 <http://www.ncbi.nlm.nih.gov/books/NBK22581/table/A3058/?report=objectonly>
- 647 YANG SY & XY He, *Molecular mechanisms of fatty acid α -oxidation enzyme catalysis* in **Current views of fatty acid oxidation and ketogenesis: from organelles to point mutations**, Quant & Keaton (edit.), Kluwer Academic / Plenum Publishers, USA, 1999.
- 648 KIM JJ & Battaile KP *et al.* (2002), *Burning fat: the structural basis of fatty acid α -oxidation*, **Curr. Opin. Struct. Biol.** **12**, 721-728.
- 649 AGNIHOTRI G & HW Liu (2003), *Enoyl-CoA hydratase: reaction, mechanism, and inhibition*, **Bioorg. Med. Chem.** **11**, 9-20.
- 650 WATERSON RM & RL Hill (1972), *Enoyl coenzyme A hydratase (crotonase), Catalytic properties of crotonase and its possible regulatory role in fatty acid oxidation*, **J. Biol. Chem.** **247**, 5258-5265.
- 651 MÜLLER-NEWEN G & W Stoffel (1993), *Site-directed mutagenesis of putative active-site amino acid residues of 3,2-enoyl-CoA-isomerase, conserved within the low-homology isomerase/hydratase enzyme family*, **Biochemistry** **32**, 11405-11412.
- 652 D'ORDINE RL *et al.* (1994), *Electronic rearrangement induced by substrate analog substrate binding to the enoyl-CoA hydratase active site: evidence for substrate activation*, **Biochemistry** **33**, 12653-12643.
- 653 D'ORDINE RL *et al.* (1994), *Enoyl-coenzyme A hydratase-catalyzed exchange of the α -proton of coenzyme a thiol esters: a model for an enolized intermediate in the enzyme-catalyzed elimination*, **Biochemistry** **33**, 14733-14742.
- 654 MÜLLER-NEWEN G *et al.* (1995), *Enoyl-CoA hydratase and isomerase form a superfamily with a common active-site glutamate residue*, **Eur. J. Biochem.** **228**, 68-73.
- 655 ENGEL CK *et al.* (1996), *Crystal structure of enoyl-Coenzyme A (CoA) hydratase at 2.5 Å resolution: a spiral fold defines the CoA-binding pocket*, **EMBO J.** **15**, 5135-5146.
- 656 WU WJ *et al.* (1997), *Structure of hexenandiol-CoA bound to enoyl-CoA hydratase determined by transferred nuclear overhauser effect measurements: mechanistic predictions based on the X-ray structure of 4-(chlorobenzoyl)-CoA dehalogenase*, **Biochemistry** **36**, 2211-2220.

- 657 ISHIKAWA M *et al.* (1997), *Reconstitution, morphology and crystalization of a fatty acid \square -oxidation multienzyme complex from Pseudomonas fragi*, **Biochem. J.** **328**, 815-820.
- 658 ENGEL CK *et al.* (1998), *The crystal structure of enoyl-CoA hydratase complexed with octanoyl-CoA reveals the structural adaptations required for binding of a long chain fatty-acid molecule*, **J. Mol. Biol.** **275**, 847-859.
- 659 KIEMA TR *et al.* (1999), *Mutagenic and enzymological studies of the hydratase and isomerase activities of 2-enoyl-CoA hydratase*, **Biochemistry** **38**, 2991-2999.
- 660 HOFSTEIN HA *et al.* (1999), *Roles of glutamate 144 and glutamate 164 in the catalytic mechanism of enoyl-CoA hydratase*, **Biochemistry** **38**, 9508-9516.
- 661 BELL AF *et al.* (2001), *Involvement of glycine 141 in substrate activation by enoyl-CoA hydratase*, **Biochemistry** **40**, 1725-1733.
- 662 BAHNSON BJ *et al.* (2002), *Structural mechanism of enoyl-CoA hydratase: three atoms from a single water are added in either an E1cb stepwise or concerted fashion*, **Biochemistry** **41**, 2621-2629.
- 663 FENG Y *et al.* (2002), *Effect of mutagenesis on the stereochemistry of enoyl-CoA hydratase*, **Biochemistry** **41**, 12883-12890.
- 664 BELL AF *et al.* (2002), *Stereoselectivity of enoyl-CoA hydratase results from preferential activation of one of two bound substrate conformers*, **Chem. Biol.** **9**, 1247-1255.
- 665 ISHIKAWA M *et al.* (2004), *Structural basis for chaneling mechanism of a fatty acid \square -oxidation multienzyme complex*, **EMBO J.** **23**, 2745-2754.
- 666 YU W *et al.* (2006), *Mutation of Lys242 allows Δ^3 - Δ^2 -enoyl-CoA isomerase to acquire enoyl-CoA hydratase activity*, **Biochem. Biophys. Acta** **1760**, 1873-1883.
- 667 XIANG H *et al.* (1999), *Interchange of catalytic activity within the 2-enoyl-coenzyme A hydratase/isomerase superfamily based on a common active site template*, **Biochemistry** **38**, 7638-7652.
- 668 HOLDEN HM *et al.* (2001), *The crotonase superfamily: divergently related enzymes that catalyze different reactions involving acyl coenzyme A thioesters*, **Acc. Chem. Res.** **34**, 145-157.
- 669 HAMED RB *et al.* (2008), *Mechanisms and structures of crotonase superfamily enzymes – how nature controls enolate and oxyanions reactivity*, **Cell. Mol. Life Sci.** **65**, 2507-2527.
- 670 MAGNUSON K *et al.* (1993), *Regulation of fatty acids biosynthesis in Escherichia coli*, **Microbiol. Rev.** **57**, 522-542.
- 671 GAGO G *et al.* (2011), *Fatty acid biosynthesis in actinomycetes*, **FEMS Microbiol. Rev.** **35**, 475-97.
- 672 MOSS J & MD Lane (1971), *The biotin-dependent enzymes*, **Adv. Microbiol. Rel. Areas Mol. Biol.** **35**, p.364-371.
- 673 POLAKIS E *et al.* (1974), *Acetyl coenzyme A carboxylase system of Escherichia coli*, **J. Biol. Chem.** **249**, 6657-6667.
- 674 KONDO H *et al.* (1991), *Acetyl coenzyme A carboxylase from Escherichia coli, Gene organization and nucleotide sequence of the biotin carboxylase subunit*, **PNAS** **88**, 9730-9733.
- 675 GORNICKI P *et al.* (1993), *Genes for two subunits of acetyl coenzyme A carboxylase of Anabaena sp. Strain PCC 7120: biotin carboxylase and biotin carboxyl carrier protein*, **J. Bacteriol.** **175**, 5268-5272.
- 676 BEST E & VC Knauf (1993), *Organization and nucleotide sequences of the genes encoding the biotin carboxylase and biotin carboxyl carrier protein of Pseudomonas aeruginosa acetyl coenzyme A carboxylase*, **J. Bacteriol.** **175**, 6881-6889.
- 677 MARINI P *et al.* (1995), *The genes encoding the biotin carboxyl carrier protein and biotin carboxylase subunit of Bacillus subtilis acetyl coenzyme A carboxylase, the first enzyme in fatty acid synthesis*, **J. Bacteriol.** **177**, 7003-7006.
- 678 JANIYANI K *et al.* (2001), *Function of Escherichia coli biotin carboxylase requires catalytic activity of both subunits of the homodimer*, **J. Biol. Chem.** **276**, 29864-29870.
- 679 ATTWOOD PV & JC Wallace (2002), *Chemical and catalytic mechanisms of carboxyl transfer reaction in biotin-dependent enzymes*, **Acc. Chem. Res.** **35**, 113-120.
- 680 JITRAPAKDEE S & JC Wallace (2003), *The biotin family: conserved structural motifs and domain rearrangements*, **Curr. Prot. Sci.** **4**, 217-239.
- 681 SHEN Y *et al.* (2006), *Is dimerization required for the catalytic activity of bacterial biotin carboxylase*, **Mol. Cell** **22**, 807-818.
- 682 MOCHALKIN I *et al.* (2008), *Structural evidence for substrate-induced synergism and half-sites reactivity in biotin carboxylase*, **Prot. Sci.** **17**, 1706-1718.
- 683 NOVAK BR *et al.* (2009), *Umbrella sampling simulations of biotin carboxylase: is a structure with an open ATP grasp domain stable in solution*, **J. Phys. Chem.** **113**, 10097-10103.
- 684 CHOU CY & L Tong (2011), *Structural and biochemical studies in the regulation of biotin carboxylase by substrate inhibition and dimerization*, **J. Biol. Chem.** **286**, 24417-24425.
- 685 POLYAK SW *et al.* (2012), *Structure, function and selective inhibition of bacterial acetyl-coa carboxylase*, **Appl. Microbiol. Biotechnol.** **93**, 983-992.

- 686 SMITH AC & JE Cronan (2012), *Dimerization of the bacterial biotin carboxylase subunit is required for acetyl coenzyme A carboxylase activity in vivo*, **J. Bacteriol.** **194**, 72-78.
- 687 LI SJ *et al.* (1992), *The gene encoding the biotin carboxylase subunit of Escherichia coli acetyl coenzyme A carboxylase*, **J. Biol. Chem.** **267**, 855-863.
- 688 WALDROP G *et al.* (1994), *Three-dimensional structure of the biotin carboxylase subunit of acetyl-CoA carboxylase*, **Biochemistry** **33**, 10249-10256.
- 689 BLANCHARD CZ *et al.* (1999), *Mutations at four active site residues of biotin carboxylase abolish substrate-induced synergism by biotin*, **Biochemistry** **38**, 3393-3400.
- 690 SLOANE V *et al.* (2001), *Site-directed mutagenesis of ATP binding residues of biotin carboxylase, Insights into the mechanism of catalysis*, **J. Biol. Chem.** **276**, 24991-24996.
- 691 CHOU CY *et al.* (2009), *Crystal structure of biotin carboxylase in complex with substrates and implications for its catalytic mechanism*, **J. Biol. Chem.** **284**, 11690-11697.
- 692 BORDELON T *et al.* (2009), *The utility of molecular dynamics simulations for understanding site-directed mutagenesis of glycine residues in biotin carboxylase*, **Proteins** **74**, 808-819.
- 693 SAMOIS D *et al.* (1988), *Evolutionary conservation among biotin enzymes*, **J. Biol. Chem.** **263**, 6461-6464.
- 694 ARTYMIUK PJ *et al.* (1996), *Biotin carboxylase comes into the fold*, **Nat. Struct. Biol.** **3**, 128-132.
- 695 GALPERIN MY & EV Koonin (1997), *A diverse family of enzymes with ATP-dependent carboxylate-amine/thiol ligase activity*, **Prot. Sci.** **6**, 2639-2643.
- 696 BROWNSEY RW *et al.* (1997), *Isoforms of acetyl-CoA carboxylase: structures, regulatory properties and metabolic functions*, **Biochem. Soc. Transact.** **25**, p.1232.
- 697 JADHAV VR & M Yarus M (2002), *Coenzymes as coribozymes*, **Biochimie** **84**, 877-888.
- 698 HOLLIDAY GL *et al.* (2007), *Evolution of enzymes and pathways for the biosynthesis of cofactors*, **Nat. Prod. Rep.** **24**, 972-987.
- 699 *Coproporphyrinogen dehydrogenase, 1.3.99.22 en Class 1 Oxidoreductases*, **Springer Handbook of Enzymes**, Schomburg D, Schomburg I & Chang A (edit), Vol. S1, 2nd edition, USA, 2009.
- 700 SHEERA JS *et al.* (1983), *Anaerobic and aerobic coproporphyrinogen III oxidases of Rhodospseudomonas spheroides, Mechanism and stereochemistry of vinyl group formation*, **Biochem. J.** **209**, 709-718.
- 701 XU K *et al.* (1992), *The genes required for heme synthesis in Salmonella typhimurium include those encoding alternative functions for aerobic and anaerobic coproporphyrinogen oxidation*, **J. Bacteriol.** **174**, 3953-3963.
- 702 XU K & T Elliot (1994), *Cloning, DNA sequence, and complementation analysis of the Salmonella typhimurium hemN gene encoding a putative oxygen-independent coproporphyrinogen III oxidase*, **J. Bacteriol.** **176**, 3196-3203.
- 703 HOMUTH G *et al.* (1996), *The genes lepA and hemN form a bicistronic operon in Bacillus subtilis*, **Microbiology** **142**, 1641-1669.
- 704 LIEB C *et al.* (1998), *The alcaligenes eutrophus hemN gene encoding the oxygen-independent coproporphyrinogen III oxidase, is required for heme biosynthesis during growth*, **Arch. Microbiol.** **169**, 52-60.
- 705 HOMUTH G *et al.* (1999), *Transcriptional control of Bacillus subtilis hemN and hemZ*, **J. Bacteriol.** **181**, 5922-5929.
- 706 DAILEY HA (2002), *Terminal steps of haem biosynthesis*, **Biochem. Soc. Transact.** **30**, p.590&591.
- 707 LAYER G *et al.* (2002), *Oxygen-independent coproporphyrinogen-III oxidase HemN from Escherichia coli*, **J. Biol. Chem.** **277**, 34136-34142.
- 708 LAYER G *et al.* (2005), *Radical S-adenosylmethionine enzyme coproporphyrinogen III oxidase HemN*, **J. Biol. Chem.** **280**, 29038-29046.
- 709 LAYER G *et al.* (2006), *The substrate radical of Escherichia coli oxygen-independent coproporphyrinogen III oxidase HemN*, **J. Biol. Chem.** **281**, 15727-15734.
- 710 LAYER G *et al.* (2003), *Crystal structure of coproporphyrinogen III oxidase reveals cofactor geometry of Radical SAM enzymes*, **EMBO J.** **22**, 6214-6224.
- 711 MARSH ENG *et al.* (2004), *S-adenosylmethionine radical enzymes*, **Bioorg. Chem.** **32**, 326-340.
- 712 MARQUET A *et al.* (1995-2007), *Iron-sulfur proteins as initiators of radical chemistry*, **Nat. Prod. Rep.** **24**, 1027-1031&1034-1035.
- 713 LAYER G *et al.* (2004), *Structure and function of radical SAM enzymes*, **Curr. Opin. Chem. Biol.** **8**, 468-476.
- 714 WANG SC & PA Frey (2007), *S-adenosylmethionine as oxidant: the radical SAM superfamily*, **TRENDS Biochem. Sci.** **32**, 101-111.
- 715 FREY PA *et al.* (2008), *The radical SAM superfamily*, **Crit. Rev. Biochem. Mol. Biol.** **43**, 63,78&79.

- ⁷¹⁶ YOUNG IG *et al.* (1975), *Biosynthesis of bacterial menaquinone mutant of Escherichia coli*, **Biochemistry** **14**, 399-406.
- ⁷¹⁷ SHINEBERG B & IG Young (1976), *Biosynthesis of bacterial menaquinones: the membrane-associated 1,4-dihydroxy-2-naphthone octaprenyltransferase of Escherichia coli*, **Biochemistry** **15**, 2754-2758.
- ⁷¹⁸ HADDOCK BA & CW Jones (1977), *Bacterial respiration*, **Bacteriol. Rev.** **41**, 74-99.
- ⁷¹⁹ BENTLEY R & R Meganathan (1982), *Biosynthesis of vitamin K (menaquinone) in bacteria*, **Microbiol. Rev.** **46**, 241-243&262-264.
- ⁷²⁰ SIEBERT M *et al.* (1992), *Cloning of the genes coding for chorismate pyruvate-lyase and 4-hydroxybenzoate octaprenyl transferase from Escherichia coli*, **FEBS** **307**, 347-350.
- ⁷²¹ SUSUKI K *et al.* (1994), *Evidence that Escherichia coli ubiA product is a functional homolog of yeast and the regulation of ubiA gene expression*, **Biosci. Biotech. Biochem.** **58**, 1814-1819.
- ⁷²² MELZER M & L Heide (1994), *Characterization of polyprenyldiphosphate: 4-hydroxybenzoate polyprenyltransferase from Escherichia coli*, **Biochim. Biophys. Acta** **1212**, 93-102.
- ⁷²³ SUVARNA K *et al.* (1998), *Menaquinone (vitamin K) biosynthesis: localization and characterization of the menA gene from Escherichia coli*, **J. Bacteriol.** **180**, 2782-2787.
- ⁷²⁴ LIANG PH *et al.* (2002), *Structure, mechanism and function of prenyltransferase*, **Eur. J. Biochem.** **269**, 3339-3354.
- ⁷²⁵ CHANG SY *et al.* (2004), *Substrate binding mode and reaction mechanism of undecaprenyl pyrophosphate synthase deduced from crystallographic studies*, **Prot. Sci.** **13**, 971-978.
- ⁷²⁶ BRÄUER L *et al.* (2004), *Modelling the E. coli 4-hydroxybenzoic acid oligoprenyltransferase (ubiA transferase) and characterization of potential active site*, **J. Mol. Model.** **10**, 317-327.
- ⁷²⁷ HANDA N *et al.* (2005), *Crystal structure of a novel polyisoprenoid-binding protein from Thermus thermophilus HB8*, **Prot. Sci.** **14**, 1004-1010.
- ⁷²⁸ BRÄUER L *et al.* (2008), *A structural model of the membrane-bound aromatic prenyltransferase UbiA from E. coli*, **Chem. Bio. Chem.** **9**, 982-992.
- ⁷²⁹ BONITZ T *et al.* (2011), *Evolutionary relationships of microbial aromatic prenyltransferases*, **PLoS One** **6**(11), e27336.
- ⁷³⁰ WILLIAMS RJ (2002), *The fundamental nature of life as a chemical system: the part played by inorganic elements*, **J. Inorg. Biochem.** **88**, p.241.
- ⁷³¹ GOSH B (2005), *Three-dimensional structures of sulfatases*, **Meth. Enzymol.** **400**, 273-293.
- ⁷³² SARDIELLO M *et al.* (2005), *Sulfatases and sulfatase modifying factors: an exclusive and promiscuous relationship*, **Hum. Mol. Gen.** **14**, 3203,3217.
- ⁷³³ BERTEAU O *et al.* (2006), *A new type of bacterial sulfatase reveals a novel maturation pathway in prokaryotes*, **J. Biol. Chem.** **281**, 22464-22470.
- ⁷³⁴ <http://www.ncbi.nlm.nih.gov/books/NBK2227/>
- ⁷³⁵ <http://www.ncbi.nlm.nih.gov/books/NBK20255/>
- ⁷³⁶ CHEN LH *et al.* (1992), *4-oxalocrotonate tautomerase, an enzyme composed of 62 amino acid residues per monomer*, **J. Biol. Chem.** **267**, 17716-17721.
- ⁷³⁷ ROPER DI *et al.* (1994), *Preliminary crystallographic analysis of 4-oxalocrotonate tautomerase reveals the oligomeric structure of the enzyme*, **J. Mol. Biol.** **243**, 799-801.
- ⁷³⁸ STIVERS JT *et al.* (1996), *4-oxalocrotonate tautomerase homohexamer: backbone and side-chain resonance assignments, solution secondary structure, and location of active site residues by heteronuclear NMR spectroscopy*, **Prot. Sci.** **5**, 729-741.
- ⁷³⁹ STIVERS JT *et al.* (1996), *Catalytic role of the amino-terminal proline in 4-oxalocrotonate tautomerase: affinity labelling and heteronuclear NMR resonance*, **Biochemistry** **35**, 803-813.
- ⁷⁴⁰ STIVERS JT *et al.* (1996), *4-oxalocrotonate tautomerase: pH dependence of catalysis and pK_a values of active site residues*, **Biochemistry** **35**, 814-823.
- ⁷⁴¹ STIVERS JT *et al.* (1996), *NMR relaxation studies of free and inhibitor-bound 4-oxalocrotonate tautomerase: backbone dynamics and entropy changes of an enzyme upon inhibitor design*, **Biochemistry** **35**, 16036-16047.
- ⁷⁴² SOARES TA *et al.* (1999), *Docking of 4-oxalocrotonate tautomerase substrates: implications for the catalytic mechanism*, **Biopolymers** **50**, 319-328.
- ⁷⁴³ HARRIS TK *et al.* (1999), *Kinetic, stereochemical, and structural effects of mutations of the active site arginine residues in 4-oxalocrotonate tautomerase*, **Biochemistry** **38**, 12343-12357.
- ⁷⁴⁴ CZERWINSKI RM *et al.* (1999), *Effects of mutations of the active site arginine residues in 4-oxalocrotonate tautomerase on the pK_a values of active site residues and on the pH dependence of catalysis*, **Biochemistry** **38**, 12358-12366.
- ⁷⁴⁵ CZERWINSKI RM *et al.* (2001), *The structural basis for the perturbed pK_a of the catalytic base in 4-oxalocrotonate tautomerase: kinetic and structural effects of mutations of Phe-50*, **Biochemistry** **40**, 1984-1995.

- ⁷⁴⁶ SILINSKI P *et al.* (2001), *Guanine-induced equilibrium unfolding of a homo-hexameric enzyme 4-oxalocrotonate tautomerase (4-OT)*, **Biochemistry** **40**, 4493-4502.
- ⁷⁴⁷ SILINSKI P & MC Fitzgerald (2002), *A stable dimer in the pH-induced equilibrium unfolding of the homo-hexameric enzyme 4-oxalocrotonate tautomerase (4-OT)*, **Biochemistry** **41**, 4480-4491.
- ⁷⁴⁸ BRIK A *et al.* (2002), *The product of the natural reaction catalyzed by 4-oxalocrotonate tautomerase becomes an affinity label of its mutant*, **Bioorg. Med. Chem.** **10**, 3891-3897.
- ⁷⁴⁹ BRIK A *et al.* (2002), *Mutants of 4-oxalocrotonate tautomerase catalyze the decarboxylation of oxaloacetate through an imine mechanism*, **Chem. BioChem.** **3**, p. 845&850.
- ⁷⁵⁰ CISNEROS GA *et al.* (2004), *The protein backbone makes important contributions to 4-oxalocrotonate tautomerase enzyme catalysis: understanding from theory and experiment*, **Biochemistry** **43**, 6885-6892.
- ⁷⁵¹ AZURMENDI HF *et al.* (2005), *Half-of-the-sites binding of reactive intermediates and their analogues to 4-oxalocrotonate tautomerase and induced structural asymmetry of the enzyme*, **Biochemistry** **44**, 7725-7737.
- ⁷⁵² TUTTLE T *et al.* (2006), *Understanding the enzymatic activity of 4-oxalocrotonate tautomerase and its mutant analogues: a computational study*, **J. Phys. Chem.** **110**, 19685-19695.
- ⁷⁵³ SEVASTIK R & F Himo (2007), *Quantum chemical modeling of enzymatic reactions: the case of 4-oxalocrotonate tautomerase*, **Bioorg. Chem.** **35**, 444-457.
- ⁷⁵⁴ WANG SC *et al.* (2007), *Kinetic and stereochemical analysis of YwhB, a 4-oxalocrotonate tautomerase homologue in Bacillus subtilis: mechanistic implications for the YwhB- and 4-oxalocrotonate tautomerase-catalyzed reactions*, **Biochemistry** **46**, 11919-11929.
- ⁷⁵⁵ RUIZ-PERNÍA JJ *et al.* (2009), *Critical role of substrate conformational change in the proton transfer process catalyzed by 4-oxalocrotonate tautomerase*, **J. Am. Chem. Soc.** **131**, 2687-2698.
- ⁷⁵⁶ ZANDVOORT E *et al.* (2011), *Systematic screening for catalytic promiscuity in 4-oxalocrotonate tautomerase: enamine formation and aldolase activity*, **Chem. BioChem.** **12**, p.602.
- ⁷⁵⁷ WU P *et al.* (2012), *Catalytic mechanism of 4-oxalocrotonate tautomerase: significances of protein-protein interactions on proton-transfer pathways*, **J. Phys. Chem. B** **116**, 6889-6897.
- ⁷⁵⁸ WHITMAN CP (2002), *The 4-oxalocrotonate tautomerase family of enzyme: how nature makes enzymes using α - β - γ structural motif*, **Arch. Biochem. Biophys.** **402**, 1-13.
- ⁷⁵⁹ ALMRUD JJ *et al.* (2002), *The crystal structure of YdcE, a 4-oxalocrotonate tautomerase homologue from Escherichia coli, confirms the structural basis for oligomer diversity*, **Biochemistry** **41**, 12010-12024.
- ⁷⁶⁰ POELARENDIS GJ *et al.* (2006), *Evolution of enzymatic activity in the tautomerase superfamily: mechanistic and structural consequences of the L8R mutation in 4-oxalocrotonate tautomerase*, **Biochemistry** **45**, 7700-7708.
- ⁷⁶¹ POELARENDIS GJ *et al.* (2008), *The chemical versatility of the α - β - γ fold: catalytic promiscuity and divergent evolution in the tautomerase superfamily*, **Cell Mol. Life Sci.** **65**, 3606-3618.
- ⁷⁶² BURKS EA *et al.* (2010), *Kinetic and structural characterization of a heterohexamer 4-oxalocrotonate tautomerase from Chloroflexus aurantiacus J-10-fl: implications for functional and structural diversity in the tautomerase superfamily*, **Biochemistry** **49**, 5016-5027.
- ⁷⁶³ ALMRUD JJ *et al.* (2010), *Kinetic and structural characterization of DmpI from Helicobacter pylori and Archaeoglobus fulgidus, two 4-oxalocrotonate tautomerase family members*, **Bioorg. Chem.** **38**, 252-259.
- ⁷⁶⁴ <http://www.cells.es/Beamlines/XALOC/jbenach/jbenach/Benach-Jordi-PhD-thesis.pdf> p.1-9.
- ⁷⁶⁵ <http://www.ncbi.nlm.nih.gov/books/NBK22593/figure/A2238/?report=objectonly>
- ⁷⁶⁶ PERSSON B *et al.* (1991), *Characteristics of short-chain alcohol dehydrogenases and related enzymes*, **Eur. J. Biochem.** **200**, 537-543.
- ⁷⁶⁷ KROZOWSKI Z (1994), *The short-chain alcohol dehydrogenase superfamily: variations on a common theme*, **J. Steroid Biochem. Molec. Biol.** **51**, 125-130.
- ⁷⁶⁸ JÖRNVALL H *et al.* (1995), *Short-chain dehydrogenases/reductases (SDR)*, **Biochemistry** **34**, 6003-6013.
- ⁷⁶⁹ JÖRNVALL H *et al.* (1999), *SDR and MDR: completed genome sequences show these protein families to be large, of old origin, and of complex nature*, **FEBS Lett.** **445**, 261-264.
- ⁷⁷⁰ FILLING C *et al.* (2002), *Critical residues for structure and catalysis in short-chain dehydrogenases/reductases*, **J. Biol. Chem.** **277**, 25677-25684.
- ⁷⁷¹ KALLBERG Y *et al.* (2002), *Short-chain dehydrogenases/reductases (SDR), Coenzyme-based functional assignments in completed genomes*, **Eur. J. Biochem.** **269**, 4409-4417.
- ⁷⁷² OPPERMAN U *et al.* (2003), *Short-chain dehydrogenases/reductases (SDR): the 2002 update*, **Chem. Biol. Interact.** **143-144**, 247-253.

- ⁷⁷³ PERSSON B *et al.* (2003), *Coenzyme-based functional assignments of short-chain dehydrogenases/reductases (SDR)*, **ChemBiol. Interact.** **143-144**, 271-278.
- ⁷⁷⁴ KHO R *et al.* (2003), *Genome-wide profile of oxidoreductases in viruses, prokaryotes, and eukaryotes*, **J. Proteome. Res.** **2**, 626-632.
- ⁷⁷⁵ KELLER B *et al.* (2006), *Bioinformatic identification and characterization of new members of short-chain dehydrogenase/reductase family*, **Mol. Cell Endocrinol.** **248**, 56-60.
- ⁷⁷⁶ GANI OABSM *et al.* (2008), *Theoretical calculations of the catalytic triad in short-chain alcohol dehydrogenases/reductases (SDR)*, **Biophys. J.** **94**, p.1412.
- ⁷⁷⁷ KAVANAGH KL *et al.* (2008), *The SDR superfamily: functional and structural diversity within a family of metabolic and regulatory enzymes*, **Cell Mol. Life Sci.** **65**, 3895-3906.
- ⁷⁷⁸ PERSSON B *et al.* (2009), *The SDR (short-chain dehydrogenases/reductase and related enzymes) nomenclature initiative*, **Chem. Biol. Interact.** **178**, 94-98.
- ⁷⁷⁹ KALLBERG Y *et al.* (2010), *Classification of the short-chain dehydrogenases/reductase superfamily using hidden Markov models*, **FEBS J.** **277**, 2375-2386.
- ⁷⁸⁰ JÖRNVALL H *et al.* (2010), *Superfamilies SDR and MDR: from early ancestry to present forms, Emergence of three lines: a Zn-metalloenzyme, and distinct variabilities*, **Biochem. Biophys. Res. Commun.** **396**, 125-130.
- ⁷⁸¹ McCORMICK DB (1972), *Fluorescence properties of the falvin moiety within N-(5'-phospho)-pyridoxyl flavin amides: spectroscopic models for pyridoxine (pyridoxamine) 5'-phosphate oxidase*, **Arch. Biochem. Biophys.** **153**, 233-240.
- ⁷⁸² MERRIL AH *et al.* (1979), *Spectroscopic studies of pyridoxamine (pyridoxine) 5'-phosphate oxidase. Equilibrium dissociation constants and spectra for riboflavin 5'-phosphate and analogues*, *Biochemistry* **18**, 3635-3641.
- ⁷⁸³ ZHAO G & ME Winkler (1995), *Kinetic limitation and cellular amount of pyridoxine (pyridoxamine) 5'-phosphate oxidase of Escherichia coli K-12*, **J. Bacteriol.** **177**, 883-891.
- ⁷⁸⁴ Di SALVO M *et al.* (1998), *Expression, purification, and characterization of recombinant Escherichia coli pyridoxine 5'-phosphate oxidase*, **Prot. Expr. Purif.** **13**, 349-356.
- ⁷⁸⁵ MUSAYEV FN *et al.* (1999), *Crystalization and preliminary X-ray crystallographic analysis of pyridoxine 5'-phosphate oxidase complexed with flavin mononucleotide*, **J. Struct. Biol.** **127**, 88-91.
- ⁷⁸⁶ YANG ES & V Schirch (2000), *Tight binding of pyridoxal 5'-phosphate to recombinant Escherichia coli pyridoxine 5'-phosphate oxidase*, **Arch. Biochem. Biophys.** **377**, 109-114.
- ⁷⁸⁷ van den BERG PAW *et al.* (2001), *Fluorescence correlation spectroscopy of flavins and flavoenzymes: photochemical and photophysical aspects*, **Spectr. Acta A** **57**, 2135-2144.
- ⁷⁸⁸ SAFO MK *et al.* (2000), *X-ray structure of Escherichia coli pyridoxine 5'-phosphate oxidase complexed with FMN at 1.8 Å resolution*, **Structure** **8**, 751-762.
- ⁷⁸⁹ SAFO MK *et al.* (2001), *X-ray structure of Escherichia coli pyridoxine 5'-phosphate oxidase complexed with pyridoxal 5'-phosphate at 2 Å resolution*, **J. Mol. Biol.** **310**, 817-826.
- ⁷⁹⁰ Di SALVO ML *et al.* (2002), *Active site structure and stereospecificity of Escherichia coli pyridoxine 5'-phosphate oxidase*, **J. Mol. Biol.** **35**, 385-397.
- ⁷⁹¹ GARRIDO-FRANCO M *et al.* (2002), *Enzyme-ligand complexes of pyridoxine 5'-phosphate oxidase: implications of substrate binding and catalysis*, **J. Mol. Biol.** **321**, 601-612.
- ⁷⁹² Di SALVO ML *et al.* (2003), *Structure and mechanism of Escherichia coli pyridoxine 5'-phosphate oxidase*, **Biochim. Biophys. Acta** **1647**, 76-82.
- ⁷⁹³ PÉDELACQ JD *et al.* (2006), *Crystal structure of a putative pyridoxine 5'-phosphate oxidase (Rv2607) from Mycobacterium tuberculosis*, **Proteins** **62**, 563-569.
- ⁷⁹⁴ CANAAN S *et al.* (2005), *Crystal structure of the conserved hypothetical protein Rv1115 from Mycobacterium tuberculosis*, **FEBS Lett.** **579**, 215-221.
- ⁷⁹⁵ SAFO MK *et al.* (2005), *Structure of Escherichia coli pyridoxine 5'-phosphate oxidase in a tetragonal crystal form: insights into the mechanism pathway of the enzyme*, **Acta Cryst. D** **61**, p.599&600.
- ⁷⁹⁶ BISWAL BK *et al.* (2005), *Structures of Mycobacterium tuberculosis pyridoxine 5'-phosphate oxidase and its complexes with flavin mononucleotide and pyridoxal 5'-phosphate*, **Acta Cryst. D** **61**, p.1492&1493.
- ⁷⁹⁷ BISWAL BK *et al.* (2006), *The molecular structure of RV2074, a probable pyridoxine 5'-phosphate oxidase from Mycobacterium tuberculosis, at 1.6 Å resolution*, **Acta Cryst. F** **62**, p.735&736.
- ⁷⁹⁸ JACKSON CJ *et al.* (2008), *Cloning, expression, purification, crystalization and preliminary X-ray studies of a pyridoxine 5'-phosphate oxidase from Mycobacterium smegmatis*, **Acta Cryst. F** **64**, p.435.
- ⁷⁹⁹ MASHALIDIS EH *et al.* (2011), *Rv2607 from Mycobacterium tuberculosis is a pyridoxine 5'-phosphate oxidase with unusual substrate specificity*, **PLoS ONE** **6**(11), e27643.
- ⁸⁰⁰ Di SALVO ML *et al.* (2011), *Vitamin B₆ salvage enzymes: mechanism, structure and regulation*, **Biochim. Biophys. Acta** **1814**, 1597-1608.

- 801 CHRISTEN P & PK Mehta (2001), *From cofactor to enzymes, The molecular evolution of pyridoxal-5'-phosphate-dependent enzymes*, **Chem. Rec.** **1**, 436-447.
<http://www.ncbi.nlm.nih.gov/books/NBK20266/>.
- 802
- 803 SCOTT JR & TC Barnett (2006), *Surface proteins and how they get there*, **Annu. Rev. Microbiol.** **60**, 397-423.
- 804 JUNGNICKEL B *et al.* (1994), *Protein translocation: common themes from bacteria to man*, **FEBS Lett.** **346**, 73-77.
- 805 ITO K (1996), *The major pathways of protein translocation across membranes*, **Genes Cells** **1**, 337-346.
- 806 DRIESSEN AJM *et al.* (1998), *The Sec system*, **Curr. Opin. Microbiol.** **1**, 216-222.
- 807 ECONOMOU A (2000), *Bacterial protein translocase: a unique molecular machine with an army of substrates*, **FEBS Lett.** **476**, 18-21.
- 808 MÜLLER M *et al.* (2001), *Protein traffic in bacteria: multiples routes from the ribosome to and across the membrane*, **Prog. Nucl. Acids Res. Mol. Biol.** **66**, 107-157.
- 809 de KEYZER J *et al.* (2003), *The bacterial translocase: a dynamic protein channel complex*, **Cell Mol. Life Sci.**, **60**, 2034-2052.
- 810 VEENENDAAL AKJ, *et al.* (2004), *The protein conducting channel SecYEG*, *Biochim. Biophys. Acta* **1694**, 81-95.
- 811 OSBORNE AR *et al.* (2005), *Protein translocation by the Sec61/SecY channel*, **Annu. Rev. Cell Dev. Biol.** **21**, 29-50.
- 812 ROBSON A & I Collinson (2006), *The structure of the Sec complex and the problem of protein translocation*, **EMBO J.** **7**, 1099-1103.
- 813 DRIESSEN AJM & N Nowen (2008), *Protein translocation across the bacterial cytoplasmic membrane*, **Annu. Rev. Biochem.** **77**, 643-667.
- 814 OLIVER D & J Beckwith (1982), *Identification of a new gene (secA) and gene product involved in the secretion of envelope proteins in Escherichia coli*, **J. Bacteriol.** **150**, 686-691.
- 815 SCHMIDT MG *et al.* (1988), *Nucleotide sequence of the secA gene and secA (Ts) mutations preventing protein export in Escherichia coli*, **J. Bacteriol.** **170**, 3404-3414.
- 816 TAKAMATSU H *et al.* (1992), *In vivo and in vitro characterization of the secA gene product of Bacillus subtilis*, **J. Bacteriol.** **174**, 4308-4316.
- 817 LILL R *et al.* (1989), *SecA protein hydrolyzes ATP and is essential component of the protein translocation ATPase of Escherichia coli*, **EMBO J.** **8**, 961-966.
- 818 CHUN SY & LL Randall (1994), *In vivo studies of the role of SecA during protein export in Escherichia coli*, **J. Bacteriol.** **176**, 4197-4203.
- 819 KUNIOKA EI *et al.* (1998), *Cloning and expression of the secA gene of a marine bacterium, Vibrio alginolyticus, and analysis of its function in Escherichia coli*, **Gene** **216**, 303-309.
- 820 SCHMIDT MG & KB Kiser (1999), *SecA: the ubiquitous component of preprotein translocase in prokaryotes*, **Microb. Infect.** **1**, 999-1004.
- 821 VRONTOU E & A Economou (2004), *Structure and function of SecA, the preprotein translocase nanomotor*, **Biochim. Biophys. Acta** **1694**, 67-80.
- 822 MITCHELL C & D Oliver (1993), *Two distinct ATP-binding domains are needed to promote protein export by Escherichia coli SecA ATPase*, **Mol. Microbiol.** **10**, 483-497.
- 823 ECONOMOU A *et al.* (1995), *SecA membrane cycling at SecYEG is driven by distinct ATP binding and hydrolysis events and is regulated by SecD and SecE*, **Cell** **83**, 1171-1181.
- 824 DAPIC V & D Oliver (2000), *Distinct membrane binding properties of N- and C-terminal domains of Escherichia coli SecA ATPase*, **J. Biol. Chem.** **275**, 25000-25007.
- 825 PAPANIKOU E *et al.* (2005), *Identification of the preprotein binding domain of SecA*, **J. Biol. Chem.** **280**, 43209-43217.
- 826 GELIS I *et al.* (2007), *Structural basis for signal-sequence recognition by the translocase motor SecA as determined by NMR*, **Cell** **131**, 756-769.
- 827 KOBAYASHI M *et al.* (1999), *Cloning and nucleotide sequencing of the secA gene from coryneform bacteria*, **Genet. Analysis Biomol. Eng.** **15**, 9-13.
- 828 van der WOLK JPW *et al.* (1997), *The low-affinity ATP binding site of the Escherichia coli SecA dimer is localized at the subunit interface*, **Biochemistry** **36**, 14924-14929.
- 829 NATALE P *et al.* (2004), *Binding of SecA to the SecYEG complex accelerates the rate of nucleotide exchange on SecA*, **J. Biol. Chem.** **279**, 13769-13777.
- 830 COOPER DB *et al.* (2008), *SecA the motor for the secretion machine binds diverse partners on the interactive surface*, **J. Mol. Biol.** **382**, 74-87.
- 831 van der WOLK JPW *et al.* (1997), *The catalytic cycle of the Escherichia coli SecA ATPase comprises two distinct preprotein translocation events*, **EMBO J.** **16**, 7297-7304.
- 832 de KEYZER J *et al.* (2003), *Direct demonstration of ATP dependent release of SecA from a translocating preprotein by surface plasmon resonance*, **J. Biol. Chem.** **278**, 29581-29586.
- 833 TOMKIEWICZ D *et al.* (2006), *SecA supports a constant rate of preprotein translocation*, **J. Biol. Chem.** **281**, 15709-15713.

- 834 KARAMANOV S *et al.* (1999), *A molecular switch in SecA protein couples ATP hydrolysis to protein translocation*, **Mol. Microbiol.** **34**, 1133-1145.
- 835 NAKATOGAWA H (2000), *Two independent mechanisms down-regulate the intrinsic SecA ATPase activity*, **J. Biol. Chem.** **275**, 33209-33212.
- 836 van VOORST F *et al.* (1998), *Translocase-bound SecA is largely shielded from the phospholipid acyl chains*, **Biochemistry** **37**, 12261-12268.
- 837 KIM JS *et al.* (2001), *Effect of divalent cations on the ATPase activity of Escherichia coli SecA*, **FEBS Lett.** **493**, 12-16.
- 838 GOLD VAM *et al.* (2007), *Allosteric regulation of SecA, Magnesium-mediated control of conformation and activity*, **J. Biol. Chem.** **282**, 17424-17437.
- 839 AHN T & CH Yun (2009), *Ca²⁺-induced stimulation of the membrane binding of Escherichia coli SecA and its association with signal peptides of secretory proteins*, **Arch. Biochem. Biophys.** **486**, 125-131.
- 840 OR E *et al.* (2002), *Dissociation of the dimeric SecA ATPase during protein translocation across the bacterial membrane*, **EMBO J.** **21**, 4470-4479.
- 841 PRICE A *et al.* (1996), *Separable ATPase and membrane insertion domain of the SecA subunit of preprotein translocase*, **J. Biol. Chem.** **271**, 31580-31584.
- 842 DING H *et al.* (2003), *Bacillus subtilis SecA ATPase exists as an antiparallel dimer in solution*, **Biochemistry** **42**, 8729-8738.
- 843 VASSYLYEV DG *et al.* (2006), *Crystal structure of the translocation ATPase SecA from Thermus thermophilus reveals a parallel, head-to-head dimer*, **J. Mol. Biol.** **364**, 248-258.
- 844 WANG HW *et al.* (2003), *Ring-like pore structures of SecA: implication for bacterial protein-conducting channels*, **PNAS** **100**, 4221-4226.
- 845 JILAVEANU LB *et al.* (2005), *Dimeric SecA is essential for protein translocation*, **PNAS** **102**, 7511-7516.
- 846 de KEYZER J *et al.* (2005), *Covalently dimerized SecA is functional in protein translocation*, **J. Biol. Chem.** **280**, 35255-35260.
- 847 ZIMMER J *et al.* (2006), *A novel dimer interface and conformational changes revealed by an X-ray structure of B. subtilis SecA*, **J. Mol. Biol.** **364**, 259-265.
- 848 CHEN Y *et al.* (2008), *Full-length Escherichia coli SecA dimerizes in a closed conformation in solution as determined by cryo-electron microscopy*, **J. Biol. Chem.** **283**, 28783-28787.
- 849 OSBORNE AR *et al.* (2004), *A large conformational change of the translocation ATPase SecA*, **PNAS** **101**, 10937-10942.
- 850 OR E *et al.* (2005), *The bacterial ATPase SecA functions as a monomer in protein translocation*, **J. Biol. Chem.** **280**, 9097-9105.
- 851 TZIATZIOS C *et al.* (2003), *The bacterial protein-translocation complex: SecYEG dimers associate with one or two SecA molecules*, **J. Mol. Biol.** **340**, 513-524.
- 852 SHIN JY *et al.* (2006), *Effects of signal peptide and adenylate on the oligomerization and membrane binding of soluble SecA*, **J. Biochem. Mol. Biol.** **39**, 319-328.
- 853 KOONIN EV & AE Gorbalenya (1992), *Autogenous translation regulation by Escherichia coli ATPase SecA may be mediated by an intrinsic RNA helicase activity of this protein*, **FEBS** **298**, 6-8.
- 854 SALAVATI A & D Oliver (1995), *Competition between ribosome and SecA binding promotes Escherichia coli secA translational regulation*, **RNA** **1**, 745-753.
- 855 PARK SK *et al.* (1997), *RNA helicase activity of Escherichia coli SecA protein*, **Biochem. Biophys. Res. Commun.** **235**, 593-597.
- 856 SCHMIDT MO *et al.* (2001), *Escherichia coli SecA helicase activity is not required in vivo or efficient protein translocation or autogenous regulation*, **J. Biol. Chem.** **276**, 37076-37085.
- 857 PAPANIKOU E *et al.* (2004), *Helicase motif III in SecA is essential for coupling preprotein binding to translocation ATPase*, **EMBO** **5**, 807-811.
- 858 NITHIANANTHAM S & BH Shilton (2008), *Analysis of the isolated SecA DEAD motor suggests a mechanism for chemical-mechanical coupling*, **J. Mol. Biol.** **383**, 380-389.
- 859 ESER M & M Ehermann (2003), *SecA-dependent quality control of intracellular protein localization*, **PNAS** **100**, 13231-13234.
- 860 KARAMYSHEV AL & AE Johnson (2005), *Selective SecA association with signal sequences in ribosome-bound nascent chains, A potential role for SecA in ribosome targeting to the bacterial membrane*, **J. Biol. Chem.** **280**, 37930-37940.
- 861 FLOWER AM (2007), *The SecY translocation complex: convergence of genetics and structure*, **TRENDS Microbiol.** **15**, 203-210.
- 862 MANDON EC *et al.* (2009), *Translocation of proteins through the Sec61 and SecYEG channels*, **Curr. Opin. Cell Biol.** **21**, 501-507.
- 863 PÖHLING S *et al.* (1999), *Analysis and regulation of the secY gene from Streptomyces griseus N2-3-11 and interaction of the SecY protein with the SecA protein*, **Biochim. Biophys. Acta** **1447**, 298-302.
- 864 ITO K (1990), *Structure, function and biogenesis of SecY, an integral membrane protein involved in protein export*, **J. Bioenerg. Biomemb.** **22**, 353.

- 865 DUONG F & W Wickner (1997), *Distinct roles of the SecYE, SecG and SecDFyajC subunits of preprotein translocase holoenzyme*, **EMBO J.** **16**, 2757-2768.
- 866 SUH JW *et al.* (1990), *Isolation of a secY homologue from Bacillus subtilis: evidence for a common protein export pathway in eubacteria*, **Mol. Microbiol.** **4**, 305-314.
- 867 NAKAI M *et al.* (1992), *Cloning and characterization of the secY gene from the cyanobacterium Synechococcus PCC7942*, **Biochim. Biophys. Acta** **1171**, 113-116.
- 868 AUER J *et al.* (1991), *Presence of a gene in the archaeobacterium Methanococcus vannielii homologous to secY of eubacteria*, **Biochimie** **73**, 683-688.
- 869 DOUGLAS SE (1992), *A secY homologue is found in the plastid genome of Cryptomonas* □, **FEBS** **298**, 93-96.
- 870 AKIYAMA Y & K Ito (1987), *Topology analysis of the SecY protein, an integral membrane protein involved in export in Escherichia coli*, **EMBO J.** **6**, 3465-3470.
- 871 SHIMOKAWA N *et al.* (2003), *Importance of transmembrane segments in Escherichia coli SecY*, **Mol. Gen. Genom.** **269**, 180-187.
- 872 KOCH HG *et al.* (2000), *Dissecting the translocase and integrase functions of the Escherichia coli SecYEG translocon*, **J. Cell Biol.** **150**, 689-694.
- 873 CHIBA K *et al.* (2002), *Roles of the C-terminal end of SecY protein translocation and viability of Escherichia coli*, **J. Bacteriol.** **184**, 2243-2250.
- 874 van den BERG B *et al.* (2003), *X-ray structure of a protein-conducting channel*, **Nature** **427**, 36-44.
- 875 SAPAROV SM *et al.* (2007), *Determining the conductance of the SecY protein translocation channel for small molecules*, **Mol. Cell** **26**, 501-509.
- 876 LI W *et al.* (2007), *The plug domain of the SecY protein stabilizes the closed state of the translocation channel and maintains a membrane seal*, **Mol. Cell** **26**, 511-521.
- 877 MAILLARD A *et al.* (2007), *Deregulation of the SecYEG translocation channel upon removal of the plug domain*, **J. Biol. Chem.** **282**, 1281-1287.
- 878 ERLANDSON KJ *et al.* (2008), *Analysis of polypeptide movement in the SecY channel during SecA-mediated protein translocation*, **J. Biol. Chem.** **283**, 15709-15715.
- 879 GUMBART J & K Schulten (2008), *The roles of pore ring and plug in the SecY protein-conducting channel*, **J. Gen. Physiol.** **132**, 709-719.
- 880 OSBORNE AR & TA Rapoport (2007), *Protein translocation is mediated by oligomers of the SecY complex with one SecY copy forming the channel*, **Cell** **129**, 97-110.
- 881 SCHNEURING J *et al.* (2005), *The oligomeric distribution of SecYEG is altered by SecA and translocation ligands*, **J. Mol. Biol.** **354**, 258-271.
- 882 RUSH SL & DA Kendall (2007), *Oligomeric states of the SecA and SecYEG core components of the bacterial Sec translocon*, **Biochim. Biophys. Acta** **1768**, 5-12.
- 883 TSUKAZAKI T *et al.* (2008), *Conformational transition of Sec machinery inferred from bacterial SecYE structures*, **Nature** **455**, 988-991.
- 884 CANNON KS (2005), *Disulfide bridge formation between SecY and a translocating polypeptide localizes the translocation pore to the center of SecY*, **J. Cell Biol.** **169**, 219-225.
- 885 CERRETTI DP *et al.* (1983), *The spc ribosomal operon of Escherichia coli: sequence and cotranscription of the ribosomal protein genes export gene*, **Nucl. Acids Res.** **11**, 2599.
- 886 MÉNÉTRÉ JF *et al.* (2007), *Ribosome binding of a single copy of the SecY complex: implications for protein translocation*, **Mol. Cell** **28**, 1083-1092.
- 887 GUMBART J *et al.* (2009), *Regulation of the protein-conducting channel by a bound ribosome*, **Structure** **17**, 1453-1464.
- 888 AKIMARU J *et al.* (1991), *Reconstitution of a protein translocation system containing purified SecY, SecE and SecA from Escherichia coli*, **PNAS** **88**, 6545-6549.
- 889 DOUVILLE K *et al.* (1995), *SecYEG and SecA are stoichiometric components of preprotein translocase*, **J. Biol. Chem.** **270**, 20106-20111.
- 890 MANTING EH *et al.* (1999), *A single amino acid substitution in SecY stabilizes the interaction with SecA*, **J. Biol. Chem.** **274**, 23868-23874.
- 891 BENACH J & JF Hunt (2004), *Shape-shifting protein channel*, **Nature** **427**, 24-26.
- 892 van der SLUIS EO (2006), *Identification of two interaction sites in SecY that are important for the functional interaction with SecA*, **J. Mol. Biol.** **361**, 839-849.
- 893 MORI H & K Ito (2006), *Different modes of SecY-SecA interactions revealed by site-directed in vivo photo-cross-linking*, **PNAS** **103**, 19159-19164.
- 894 ROBSON A *et al.* (2007), *A large conformational change couples the ATP binding site of SecA to the SecY protein channel*, **J. Mol. Biol.** **374**, 965-976.
- 895 KARAMANOU S *et al.* (2008), *Assembly of the translocase motor onto the preprotein-conducting channel*, **Mol. Microbiol.** **70**, 311-322.
- 896 LIANG FCh *et al.* (2009), *Bacterial Sec protein transport is rate-limited by precursor length: a single turnover study*, **Mol. Biol. Cell** **20**, 4256-4266.
- 897 CHEN M *et al.* (2002), *YidC, a newly defined evolutionarily conserved protein, mediates membrane protein assembly in bacteria*, **Biol. Chem.** **383**, 1565-1572.

- ⁸⁹⁸ YI L & RE Dalbey (2005), *Oxa1/Alb3/YidC system for insertion of membrane proteins ins mitochondria, chloroplasts and bacteria (review)*, **Mol. Membr. Biol.** **22**, 101-111.
- ⁸⁹⁹ KEIFER D & A Kun (2007), *YidC as an assembly and multifunctional component in membrane protein assembly*, **Int. Rev. Cyto.** **259**, 113-138.
- ⁹⁰⁰ KOL S *et al.* (2008), *Mehcanisms of YidC-mediated insertion and asebmly of multimeric membrane protein complexes*, **J. Biol. Chem.** **283**, 31269-31273.
- ⁹⁰¹ DONG Y *et al.* (2008), *Functional overlap but lack of complete cross-complementation of Streptococcus mutans and Escherichia coli YidC orthologs*, **J. Bacteriol.** **190**, 2458-2469.
- ⁹⁰² SÄÄF & *et al.* (1998), *Membrane topology of the 60-kDa Oxa1p homologue from Escherichia coli*, **J. Biol. Chem.** **273**, 30415-30418.
- ⁹⁰³ SCOTTI PA *et al.* (2000), *YidC, the Escherichia coli homologue of mitochondrial Oxa1p, is a component of the Sec traslocase*, **EMBO J.** **19**, 542-539.
- ⁹⁰⁴ van der LAAN M *et al.* (2001), *Reconstitution of Sec-dependent membrane protein insertion: nascent FtsQ interacts with YidC in a SecYEG-dependent manner*, **EMBO J.** **20**, 519-523.
- ⁹⁰⁵ NOUWEN N *et al.* (2002), *SecDFyajC forms a heterotetrameric complex with YidC*, **Mol. Microbiol.** **44**, 1397-1405.
- ⁹⁰⁶ de GIER JW & J Luirink (2003), *The ribosome and YidC, New insights into the biogenesis of Escherichia coli inner membrane proteins*, **EMBO Rep** **4**, 939-943.
- ⁹⁰⁷ NAGAMORI S *et al.* (2004), *Role of YidC in folding of polytopic membrane proteins*, **J. Cell. Biol.** **165**, 53-62.
- ⁹⁰⁸ KUHN A *et al.* (2003), *The Alb3/Oxa1/ YidC protein family: membrane-localized chaperones facilitating membrane protein insertion?* **TRENDS Cell. Biol.** **13**, 510-516.
- ⁹⁰⁹ STEPHENSON K (2005), *Sec-dependent protein translocation across biological membranes: evolutionary conservation of an essential protein transport pathway (review)*, **Mol. Membr. Biol.** **22**, 17-28.
- ⁹¹⁰ SAMUELSON JC *et al.* (2000), *YidC mediates membrane protein insertion in bacteria*, **Nature** **406**, 637-641.
- ⁹¹¹ YI L *et al.* (2003), *YidC is strictly requiered for membrane insertion of subunits a and c f the F(1)F(0)ATP synthase and SecE of the SecYEG traslocase*, **Biochemistry** **42**, 10537-10544.
- ⁹¹² ERNST S *et al.* (2011), *YidC driven membrane insertion of single fluorescent Pf3 coat proteins*, **J. Mol. Biol.** **412**, 165-175.
- ⁹¹³ SEREK J *et al.* (2004), *Escherichia coli YidC is a membrane insertase for Sec-independent proteins*, **EMBO J.** **23**, 294-301.
- ⁹¹⁴ van BLOOIS E *et al.* (2005), *The Sec-independent function of Escherichia coli YidC is evolutionarily conserved and essential*, **J. Biol. Chem.** **280**, 12996-13003.
- ⁹¹⁵ RAVAUD S *et al.* (2008), *Purification, crystalization and preliminary structural characterization of the periplasmic domain P1 of the Escherichia coli membrane-protein insertase YidC*, **Acta Cryst.** **F64**, 144-148.
- ⁹¹⁶ LOTZ M *et al.* (2008), *Projection structure of yidC: a conserved mediator os membrane protein assembly*, **J. Mol. Biol.** **375**, 901-907.
- ⁹¹⁷ LUIRINK J. *et al.* (2001), *YidC/Oxa1p/Alb3: evolutionarily conserved mediators of membrane protein assembly*, **FEBS Lett.** **501**, 1-5.
- ⁹¹⁸ FUNES S *et al.* (2011), *Evolution of YidC/Oxa1/Alb3 insertases: three independent gene duplications followed by functional specialization in bacteria, mitochondrial and chloroplasts*, **Biol. Chem.** **392**, 13-19.
- ⁹¹⁹ van der SLUIS EO & AJM Driessen (2006), *Stepwise of the Sec machinery in Proteobacteria*, **TRENDS Microbiol.** **14**, 105-108.
- ⁹²⁰ KUHN A (2009), *From the Sec complex to the membrane insertase YidC*, **Biol. Chem.** **390**, 701-706.
- ⁹²¹ SPECIAL ISSUE (2011), *Protein traslocation across or insertion into membranes*, **Biochim. Biophys. Acta** **1808**(3).
- ⁹²² BIERNE H & P Cossart, *Listeria monocytogenes surface proteins: from genome predictions to function*, **Microbiol. Mol. Biol. Rev.** **71**, 377-397.
- ⁹²³ PALLEN MJ *et al.* (2001), *An embarrassment of sortases – richness of substrates?*, **TRENDS Microbiol.** **9**, 97-101.
- ⁹²⁴ TON-THAT H *et al.* (2003), *An embarrassment of sortases – richness of substrates?*, **TRENDS Microbiol.** **9**, Response.
- ⁹²⁵ MARRAFFINI LA *et al.* (2006), *Sortases and the art of anchoring proteins to the envelopes of Gram-positive bacteria*, **Microbiol. Mol. Biol. Rev.** **70**, 192-221.
- ⁹²⁶ COMFORT D & RT Clubb (2004), *A comparative genome analysis identifies distinct sorting pathways in Gram-positive bacteria*, **Infect. Immun.** **72**, 2710-2722.
- ⁹²⁷ DRAMSI S *et al.* (2005), *Sorting sortases: a nomenclature proposal for the various sortases of gram-positive bacteria*, **Res. Microbiol.** **156**, 289-297.
- ⁹²⁸ TON-THAT H *et al.* (2000), *Anchoring of surface proteins to the cell wall of Staphylococys aureus I, sortase catalyzed in vitro transpeptidation reaciotn using LPXTG peptide and NH2-Gly3 substrates*, **J. Biol. Chem.** **275**, 9876-9881.

- ⁹²⁹ PERRY AM *et al.* (2002), *Anchoring of surface proteins to the cell wall of Staphylococcus aureus III, Lipid II is an in vivo peptidoglycan substrate for sortase-catalyzed surface protein anchoring*, **J. Biol. Chem.** **277**, 16241-16248.
- ⁹³⁰ ROCHE FM *et al.* (2003), *Characterization of novel LPXTG-containing proteins of Staphylococcus aureus identified from genome sequences*, **Microbiology** **149**, 643-654.
- ⁹³¹ TON-THAT H *et al.* (2002), *Anchoring of surface proteins to the cell wall of Staphylococcus aureus II, cysteine 184 and histidine 120 of sortase form a thiolate-imidazolium ion pair for catalysis*, **J. Biol. Chem.** **277**, 7447-7452.
- ⁹³² MARRAFFINI LA *et al.* (2004), *Anchoring of surface proteins to the cell wall of Staphylococcus aureus IV, a conserved arginine residue is required for efficient catalysis of sortase A*, **J. Biol. Chem.** **279**, 37763-37770.
- ⁹³³ ILANGOVA U *et al.* (2001), *Structure of sortase, the transpeptidase that anchors proteins to the cell wall of Staphylococcus aureus*, **PNAS** **98**, 6056-6061.
- ⁹³⁴ NAIK MT *et al.* (2006), *Staphylococcus aureus sortase A transpeptidase, calcium promotes sorting signal binding by altering the mobility and structure of an active site loop*, **J. Biol. Chem.** **281**, 1817-1826.
- ⁹³⁵ ZONG Y *et al.* (2004), *Crystal structure of Staphylococcus aureus sortase A and its substrate complex*, **J. Biol. Chem.** **279**, 31383-31389.
- ⁹³⁶ ZHU J *et al.* (2008), *Single mutation on the surface of Staphylococcus aureus sortase A can disrupts its dimerization*, **Biochemistry** **47**, 1667-1674.
- ⁹³⁷ KOBE B & J Deisenhofer J (1994), *The leucine-rich repeat: a versatile binding motif*, **TIBS** **19**, 415-421.
- ⁹³⁸ KOBE B & J Deisenhofer (1995), *A structural basis for the interactions between leucine-rich repeats and protein ligands*, **Nature** **374**, 183-186.
- ⁹³⁹ KOBE B (1996), *Leucines on a roll*, **Nat. Struct. Biol.** **3**, 977-980.
- ⁹⁴⁰ PETERS JW *et al.* (1996), *A leucine-rich repeat variant with a novel repetitive protein structural motif*, **Nat. Struct. Biol.** **3**, 991-994.
- ⁹⁴¹ KAJAVA AV (1998), *Structural diversity of leucine-rich repeat proteins*, **J. Mol. Biol.** **277**, 519-527.
- ⁹⁴² KOBE B & AV Kajava (2001), *The leucine-rich repeat as a protein recognition motif*, **Curr. Opin. Struct. Biol.** **11**, 725-732.
- ⁹⁴³ KAJAVA AV & B Kobe (2002), *Assesment of the ability to model proteins with leucine-rich repeats in light of the latest structural information*, **Prot. Sci.** **11**, 1082-1090.
- ⁹⁴⁴ ENKHBAYAR P *et al.* (2003), *Structural principles of leucine-rich repeat proteins*, **Proteins Struct. Funct. Bioinf.** **54**, 394-403.
- ⁹⁴⁵ RAO MB *et al.* (1998), *Molecular and biotechnological aspects of microbial proteases*, **Microbiol. Mol. Biol. Rev.** **62**, 597-599.
- ⁹⁴⁶ POTEPA J & RN Pike (2005), *Bacterial peptidases in Enzymes en Concepts in Bacterial Virulence*, Russel H & H Herwald, edit., *Contrib. Microbiol.* **12**, p.132.
- ⁹⁴⁷ RAWLINGS ND *et al.* (2006), *MEROPS: the peptidase database*, **Nucl. Acids. Res.** **34**, D270-D272.
- ⁹⁴⁸ RAWLINGS ND & AJ Barret (1993), *Evolutionary families of peptidases*, **Biochem. J.** **290**, 205-218.
- ⁹⁴⁹ RAWLINGS ND & AJ Barret (1994), *Families of serine peptidases*, **Methods Enzymol.** **244**, 19-21, 37-40.
- ⁹⁵⁰ SIEZEN RJ & JAM Leunissen (1997), *Subtilases: the superfamily of subtilisin-like serine proteases*, **Prot. Sci.** **6**, 501-523.
- ⁹⁵¹ PAGE MJ & E DiCera (2008), *Evolution of peptidase diversity*, **J. Biol. Chem.** **283**, 30010-30014.
- ⁹⁵² DELCOUR J *et al.* (1999), *The biosynthesis and functionality of the cell-wall of lactic acid bacteria, Antoine van Leeuwenhoek* **76**, 159-184.
- ⁹⁵³ SILHAVY TJ *et al.* (2010), *The bacterial cell envelope*, **Cold Spring Harb. Perspect. Biol.**, doi: 10.1101/cshperspect.a000414.
- ⁹⁵⁴ NAVARRE WW & O Schneewind (1999), *Surface proteins of gram-positive bacteria and mechanisms of their targeting to the cell wall envelope*, **Microbiol. Mol. Biol. Rev.** **63**, 174-229.
- ⁹⁵⁵ KOSIC N *et al.* (1993), *Processing of lipid-modified prolipoprotein requires energy and sec gene products in vivo*, **J. Bacteriol.** **175**, 6113-6117.
- ⁹⁵⁶ SUTCLIFFE IC & RRB Russell (1995), *Lipoproteins of gram-positive bacteria*, **J. Bacteriol.** **177**, 1123-1128.
- ⁹⁵⁷ GAUTAM A *et al.* (2011), *Peptidoglycan biosynthesis machinery: A rich source of drug targets*, **Crit. Rev. Biotech.** **31**, 295-336.
- ⁹⁵⁸ van HEIJENOORT J (2001), *Formation of the glycan chains in the synthesis of bacterial peptidoglycan*, **Glycobiology** **11**, 25R-36R.
- ⁹⁵⁹ BARRETEAU H *et al.* (2008), *Cytoplasmic steps of peptidoglycan biosynthesis*, **FEMS Microbiol. Rev.** **32**, 168-207.
- ⁹⁶⁰ EL ZOEIBY A *et al.* (2003), *Structure and function of the Mur enzymes: development of novel inhibitors*, **Mol. Microbiol.** **47**, 1-12.
- ⁹⁶¹ BROWN ED *et al.* (1995), *MurA (MurZ), the enzyme that catalyzes the first committed step in peptidoglycan biosynthesis, is essential in Escherichia coli*, **J. Bacteriol** **177**, 4194-4197.

- ⁹⁶² WANKE C *et al.* (1992), *The UDP-N-acetylglucosamine i-carboxyninyl-transferase of Enterobacter cloacae*, **FEBS** **301**, 271-276.
- ⁹⁶³ MARQUARDT JL *et al.* (1992), *Cloning and sequencing of Escherichia coli murZ and purification of its product, a UDP-N-acetylglucosamine enolpyruvyl transferase*, **J. Bacteriol** **174**, 5748-5752.
- ⁹⁶⁴ EHRT S & W Hillen (1994), *UDP-N-acetylglucosamine 1-carboxyninyl-transferase from Acinetobacter calcoaceticus*, **FEMS Microbiol. Lett.** **117**, 137-142.
- ⁹⁶⁵ BROWN ED *et al.* (1994), *Detection and characterization of a phospholactoyl-enzyme adduct in the reaction catalyzed by UDP-N-acetylglucosamine enolpyruvyl transferase, MurZ*, **Biochemistry** **33**, 10638-10645.
- ⁹⁶⁶ MIZYED S *et al.* (2005), *UDP-N-acetylmuramic acid is a potent inhibitor of MurA (enolpyruvyl-UDP-GlcNAc synthase)*, **Biochemistry** **44**, 4011-4017.
- ⁹⁶⁷ SCHÖNBRUNN E *et al.* (2000), *Role of the loop containing residue 115 in the induced-fit mechanism of the bacterial cell wall biosynthetic enzyme*, **Biochemistry** **39**, 2164-2173.
- ⁹⁶⁸ ESCHENBURG S & E Schönbrunn (2000), *Comparative X-ray of the un-liganded fosfomycin-target MurA*, **Prot. Struct. Funct. Genet.** **40**, 290-298.
- ⁹⁶⁹ YOON HJ *et al.* (2008), *Crystal structure of UDP-N-acetylglucosamine enolpyruvyl transferase from Haemophilus influenzae in complex with of UDP-N-acetylglucosamine and fosfomycin*, **Proteins** **71**, 1032-1037.
- ⁹⁷⁰ SCHÖNBRUNN E *et al.* (1996), *Crystal structure of UDP-N-acetylglucosamine enolpyruvyltransferase, the target of the antibiotic fosfomycin*, **Curr. Biol.** **4**, 1065-1075.
- ⁹⁷¹ SKARZYNSKI T *et al.* (1996), *Structure of UDP-N-acetylglucosamine enolpyruvyl transferase, an enzyme essential for the synthesis of bacterial peptidoglycan, complexed with substrate UDP-N-acetylglucosamine and the drug fosfomycin*, **Curr. Biol.** **4**, 1465-1474.
- ⁹⁷² DU W *et al.* (2000), *Two active forms of UDP-N-acetylglucosamine enolpyruvyl transferase in gram-positive bacteria*, **J. Bacteriol.** **182**, 4146-4152.
- ⁹⁷³ MCCOY A *et al.* (2003), *In vitro and in vivo functional activity of Chlamydia MurA, a UDP-N-acetylglucosamine enolpyruvyl transferase involved in peptidoglycan synthesis and fosfomycin resistance*, **J. Bacteriol.** **182**, 1218-1228.
- ⁹⁷⁴ KIM DH *et al.* (1996), *Analysis of fluoromethyl group chirality establishes a common stereochemical for the enolpyruvyl transfers catalyzed by EPSP synthase and UDP-GlcNAc enolpyruvyl transferase*, **Biochemistry** **35**, 5435-5440.
- ⁹⁷⁵ ESCHENBURG S *et al.* (2003), *A new view of the mechanisms of UDP-N-acetylglucosamine enolpyruvyl transferase (MurA) and 5-enolpyruvylshikimate-3-phosphate synthase (AroA) derived from X-ray structures of their tetrahedral reaction intermediate states*, **J. Biol. Chem.** **278**, 49215-49222.
- ⁹⁷⁶ PALM G *et al.* (2000), *Crystal structure of the 3'-terminal phosphate cyclase, a ubiquitous enzyme with unusual topology*, **Structure** **8**, 13-23.
- ⁹⁷⁷ BOUHSS A *et al.* (1997), *Invariant amino acids in the Mur peptide synthetase of bacterial peptidoglycan synthesis and their modification by site-directed mutagenesis in the UDP-Mur Nac:L-alanine ligase from Escherichia coli*, **Biochemistry** **36**, 11556-11563.
- ⁹⁷⁸ AZZOLINA BA *et al.* (2001), *The cell wall and cell division gene cluster in the mra operon of Pseudomonas aeruginosa: cloning, production, and purification of active enzymes*, **Prot. Exp. Purif.** **21**, 393-400.
- ⁹⁷⁹ SMITH CA (2003), *Structure, function and dynamics in the mur family of bacterial cell wall ligases*, **J. Mol. Biol.** **362**, 640-655.
- ⁹⁸⁰ MILEWSKI S (2002), *Glucosamine-6-phosphate synthase—the multifacets enzyme*, **Biochim. Biophys. Acta** **1597**, 173-192.
- ⁹⁸¹ DURAND P *et al.* (2008), *Highlights of glucosamine-6P synthase catalysis*, **Arch. Biochem. Biophys.** **474**, 302-317.
- ⁹⁸² MOUILLERON S *et al.* (2011), *Dynamics of glucosamine 6-phosphate synthase catalysis*, **Arch. Biochem. Biophys.** **505**, 1-12.
- ⁹⁸³ ZALKIN H (1985), *Glucosamine-6-phosphate synthase*, **Methods Enzymol.** **113**, 278-281.
- ⁹⁸⁴ DENISOT MA *et al.* (1991), *Glucosamine-6-phosphate synthase from Escherichia coli yields two proteins upon limited proteolysis: identification of the glutamine amidohydrolase and 2R ketose/aldose isomerase-bearing domains based on their biochemical properties.*, **Arch. Biochem. Biophys.** **288**, 225-230.
- ⁹⁸⁵ OBMOLOVA G *et al.* (1994), *Crystalization and preliminary X-ray analysis of the two domains of glucosamine-6-Phosphate synthase from Escherichia coli*, **J. Mol. Biol.** **242**, 703-705.
- ⁹⁸⁶ FLOQUET N *et al.* (2009), *Collective motions in glucosamine-6-phosphate synthase: influence of ligand binding and role in ammonia channelling and opening of the fructose-6-phosphate binding site*, **J. Mol. Biol.** **385**, 653-664.
- ⁹⁸⁷ MOUILLERON S *et al.* (2006), *Glutamine binding opens the ammonia channel and activates glucosamine-6P synthase*, **J. Biol. Chem.** **281**, 4404-4412.

- ⁹⁸⁸ MOULLIERON S *et al.* (2007), *Domain motions of glucosamine-6P synthase: comparison of the anisotropic displacements in the crystals and the catalytic hinge-bending rotation*, **Prot. Sci.** **16**, 485-493.
- ⁹⁸⁹ DENG MD *et al.* (2006), *Directed evolution and characterization of Escherichia coli glucosamine synthase*, **Biochimie** **88**, 419-429.
- ⁹⁹⁰ BATEMAN A (1999), *The SIS domain: a phosphosugar-binding domain*, **Trends Biochem. Sci.** **24**, 94-95.
- ⁹⁹¹ CECH T (2004), *RNA finds a simpler way*, **Nature** **428**, 263-264.
- ⁹⁹² MCCARTHY T *et al.* (2005), *Ligand requirements for glmS ribozyme self-cleavage*, **Chem. Biol.** **12**, 1221-1226.
- ⁹⁹³ SOKUP GA (2006), *Core requirements for glmS ribozyme self-cleavage reveal putative pseudoknot structure*, **Nucleic Acids Res.** **34**, 968-975.
- ⁹⁹⁴ MCCOWN PJ *et al.* (2011), *An expanded collection and refined consensus model of glmS ribozymes*, **RNA** **17**, 728-736.
- ⁹⁹⁵ COCHRANE JC *et al.* (2006), *Structural investigation of the GlmS ribozyme bound to its catalytic cofactor*, **Chem. Biol.** **14**, 97-105.
- ⁹⁹⁶ TINSLEY RA *et al.* (2007), *Trans-acting glmS catalytic riboswitch: Locked and loaded*, **RNA** **13**, 468-477.
- ⁹⁹⁷ HAMPEL KJ & MM Tinsley (2006), *Evidence for preorganization of the glmS ribozyme ligand binding pocket*, **Biochemistry** **45**, 7861-7871.
- ⁹⁹⁸ KLEIN DJ & AR Ferré-D'Amaré (2006), *Structural basis of glmS ribozyme activation by glucosamine-6-phosphate*, **Science** **313**, 1752-1756.
- ⁹⁹⁹ JANSEN JA *et al.* (2006), *Backbone and nucleobase contacts to glucosamine-6-phosphate in the glmS ribozyme*, **Nat. Struct. Mol. Biol.** **13**, 517-523.
- ¹⁰⁰⁰ LIM J *et al.* (2006), *Characteristics of ligand recognition by a glmS self-cleaving ribozyme*, **Angew. Chem. Int. Ed. Engl.** **45**, 6689-6693.
- ¹⁰⁰¹ COCHRANE JC *et al.* (2009), *Structural and chemical basis for glucosamine-6-phosphate binding and activation of the glmS ribozyme*, **Biochemistry** **48**, 3239-3246.
- ¹⁰⁰² WATSON PY & MJ Fedor (2011), *The glmS riboswitch integrates signals from activating and inhibitory metabolites in vivo*, **Nat. Struct. Mol. Biol.** **18**, 361-363.
- ¹⁰⁰³ ROTH A *et al.* (2006), *Characteristics of the glmS ribozyme suggest only structural roles for divalent metal cations*, **RNA** **12**, 607-619.
- ¹⁰⁰⁴ BROOKS KM & KJ Hampel (2011), *Rapid steps in the glmS ribozyme catalytic pathway: Cation and ligand requirements*, **Biochemistry** **50**, 2424-2433.
- ¹⁰⁰⁵ BREAKER RR, *Riboswitches and the RNA world* en **The RNA World**, Gesteland RF, Cech TR & Atkins JF (edit.), 3rd edition, CSHL Press, USA, 2006.
- ¹⁰⁰⁶ BEEN MD (2006), *Versatility of self-claving ribozymes*, **Science** **313**, 1745-1747.
- ¹⁰⁰⁷ FERRÉ-D'AMARÉ AR (2011), *Use of a coenzyme by the glmS ribozyme-riboswitch suggests primordial expansion of RNA chemistry by small molecules*, **Phil. Trans. R. Soc. B** **366**, 2492-2448.
- ¹⁰⁰⁸ ZHANG Y *et al.* (2001), *Recombinant PhpA protein, a unique histidine motif-containing protein from Streptococcus pneumoniae, protects mice against intranasal pneumococcal challenge*, **Infect. Immun.** **69**, 3827-3836.
- ¹⁰⁰⁹ RIBOLDI-TUNNICLIFFE A *et al.* (2005), *1.2 Å crystal structure of the S. pneumoniae PhtA histidine triad domain a novel zinc binding fold*, **FEBS Lett.** **579**, 5333-5360.
- ¹⁰¹⁰ MELIN M *et al.* (2010), *Interaction of pneumococcal histidine triad proteins with human complement*, **Infect. Immun.** **78**, 2089-2098.
- ¹⁰¹¹ SYMERSKY J *et al.* (1997), *Structure of the collagen-binding domain from a Staphylococcus aureus adhesin*, **Nat. Struct. Biol.** **4**, 833-838.
- ¹⁰¹² DEIVANAYAGAM CCS *et al.* (1999), *Crystallization and preliminary analysis of B-domain fragments of a Staphylococcus aureus collagen-binding protein*, **Acta Cryst. D** **55**, 525-527.
- ¹⁰¹³ GILLASPY AF *et al.* (1998), *Factors affecting the collagen binding capacity of Staphylococcus aureus*, **Infect. Immun.** **66**, 3170-3178.
- ¹⁰¹⁴ HATFORD O *et al.* (1999), *Matrix-binding proteins of Staphylococcus aureus: functional analysis of mutant and hybrid molecules*, **Microbiology** **145**, 2497-2505.
- ¹⁰¹⁵ SNODGRASS JL *et al.* (1999), *Function analysis of the Staphylococcus aureus collagen adhesin B domain*, **Infect. Immun.** **67**, 3952-3959.
- ¹⁰¹⁶ RICH RL *et al.* (1998), *Domain structure of the Staphylococcus aureus collagen adhesin*, **Biochemistry** **37**, 15423-15433.
- ¹⁰¹⁷ DEIVANAYAGAM CCS *et al.* (1999), *Novel fold and assembly of the repetitive B region of the Staphylococcus aureus collagen-binding surface protein*, **Structure** **8**, 67-78.
- ¹⁰¹⁸ COURTNEY HS *et al.* (1994), *Cloning, sequencing, and expression of fibronectin /fibrinogen-binding protein from Group A Streptococci*, **Infect. Immun.** **62**, 3937-3946.
- ¹⁰¹⁹ MEEHAN M *et al.* (2000), *Neither the A- nor B- repeat regions of the fibrinogen-binding protein of Streptococcus equi supsp. equi are essential for fibrinogen binding*, **FEMS Microbiol.** **190**, 317-321.

- ¹⁰²⁰ JACOBSSON K (2003), *A novel family of fibrinogen-binding proteins in Streptococcus agalactiae*, **Vet. Microbiol.** **96**, 103-113.
- ¹⁰²¹ SCHUBERT A *et al.* (2004), *The fibrinogen receptor FbsA promotes adherence of Streptococcus agalactiae to human epithelial cells*, **Infect. Immun.** **72**, 6197-6205.
- ¹⁰²² JONSSON IM *et al.* (2005), *Role of fibrinogen-binding adhesin expression in septic arthritis and septicemia caused by Streptococcus agalactiae*, **JID** **192**, 1456-1464.
- ¹⁰²³ ROSENAU A *et al.* (2002), *Evaluation of the ability of Streptococcus agalactiae strains isolated from genital and neonatal specimens to bind to human fibrinogen and correlation with characteristics of the fbsA and fbsB genes*, **Infect. Immun.** **75**, 1310-1317.
- ¹⁰²⁴ PIETROCCOLA G *et al.* (2005), *FbsA, a fibrinogen-binding protein from Streptococcus agalactiae mediates platelet aggregation*, **Blood** **105**, 1052-1059.
- ¹⁰²⁵ TENEBBAUM T *et al.* (2005), *Adherence and invasion of human brain microvascular endothelial cells are promoted by fibrinogen-binding protein FbsA of Streptococcus agalactiae*, **Infect. Immun.** **73**, 4404-4409.
- ¹⁰²⁶ PIERNO M *et al.* (2006), *FbsA-driven fibrinogen polymerization: a bacterial "deceiving strategy"*, **Phys. Rev. Lett.** **96**(2).
- ¹⁰²⁷ PIETROCCOLA G *et al.* (2006), *Multiple interactions of FbsA, a surface protein from Streptococcus agalactiae with fibrinogen: affinity, stoichiometry, and structural characterization*, **Biochemistry** **45**, 12840-12852.
- ¹⁰²⁸ SCHUBERT A *et al.* (2002), *A fibrinogen receptor form group B Streptococcus interacts with fibrinogen by repetitive units with novel binding sites*, **Mol. Microbiol.** **2**, 557-569.
- ¹⁰²⁹ RAGUNATHAN P & K Ponnuraj (2011), *Expression, purification and structural analysis of a fibrinogen receptor FbsA from Streptococcus agalactiae*, **Prot. J.** **30**, 159-166.
- ¹⁰³⁰ DEVI AS & K Ponnuraj (2010), *Cloning, expression, purification and ligand binding studies of novel fibrinogen protein FbsB of Streptococcus agalactiae*, **Prot. Exp. Purif.** **64**, 148-155.
- ¹⁰³¹ SZABADOS F *et al.* (2011), *Occurrence of genes of putative fibrinogen binding proteins and hemolysins, as well as of their phenotypic correlates in isolates of S. lugdunensis of different origins*, **BMC Res. Notes** **4**.
- ¹⁰³² GAGNEUX P & A Varki (1999), *Evolutionary considerations in relating oligosaccharide diversity to biological function*, **Glycobiology** **9**, 747-755.
- ¹⁰³³ CAMPBELL JA *et al.* (1997), *A classification of nucleotide-diphospho-sugar glycosyltransferases based on amino acid sequence similarities*, **Biochem. J. Lett.** **326**, 929-942.
- ¹⁰³⁴ BRETON C & A Imberty (1999), *Structure/function studies of glycosyltransferases*, **Curr. Opin. Struct. Biol.** **9**, 563-571.
- ¹⁰³⁵ KAPITONOV D & RK Yu (1999), *Conserved domains of glycosyltransferases*, **Glycobiology** **9**, 961-978.
- ¹⁰³⁶ BOURNE Y & B Henrissat (2001), *Glycoside hydrolases and glycosyltransferases: families and functional modules*, **Curr. Opin. Struct. Biol.** **11**, 593-600.
- ¹⁰³⁷ COUTINHO PM *et al.* (2003), *An evolving hierarchical family classification for glycosyltransferases*, **J. Mol. Biol.** **328**, 307-317.
- ¹⁰³⁸ LAMBERT PA *et al.* (1977), *Occurrence and function of membrane teichoic acids*, **Biochim. Biophys. Acta** **472**, 1-12.
- ¹⁰³⁹ SWOBODA JG *et al.* (2010), *Wall teichoic acid function, biosynthesis, and inhibition*, **ChemBiochem.** **11**, 35-45.
- ¹⁰⁴⁰ WEIDENMAIER C & A Peschel (2008), *Teichoic acids and related cell-wall glycopolymers in Gram-positive physiology and hostinteractions*, **Nature** **6**, 276-287.
- ¹⁰⁴¹ PEREGO M *et al.* (1995), *Incorporation of D-alanine into lipoteichoic acid and wall teichoic acid in Bacillus subtilis*, **J. Biol. Chem.** **270**, 15598-15606.
- ¹⁰⁴² SPATAFORA GA *et al.* (1999), *Regulated expression of the Streptococcus mutans dlt genes correlates with intracellular polysaccharide accumulation*, **J. Bacteriol.** **181**, 2363-2372.
- ¹⁰⁴³ HYYRYLÄINEN HL *et al.* (2000), *D-alanine substitution of teichoic acids as a modulator of protein folding and stability at the cytoplasmic membrane/cell wall interface of Bacillus subtilis*, **J. Biol. Chem.** **275**, 26696-26703.
- ¹⁰⁴⁴ CLEMANS DL *et al.* (1999), *Insertional inactivation of genes responsible for the D-alanylation of lipoteichoic acid in Streptococcus gordonii DL1 (Challlis) affects intragenetic coaggregations*, **Inf. Immun.** **67**, 2464-2474.
- ¹⁰⁴⁵ FISHER N *et al.* (2006), *The dltABCD operon of Bacillus anthracis Sterne is required for virulence and resistance to peptide, enzymatic, and cellular mediators of innate immunity*, **J. Bacteriol.** **188**, 1301-1309.
- ¹⁰⁴⁶ PEREA-VÉLEZ M *et al.* (2007), *Functional analysis of D-alanylation of lipoteichoic acid in the probiotic strain Lactobacillus rhamnosus GG*, **Appl. Environ. Microbiol.** **73**, 3595-3604.
- ¹⁰⁴⁷ MCCORMICK NE *et al.* (2011), *Regulation of D-alanylation of lipoteichoic acid in Streptococcus gordonii*, **Microbiology** **157**, 2248-2256.
- ¹⁰⁴⁸ PESCHEL A *et al.* (1999), *Inactivation of the dlt operon in Staphylococcus aureus confers sensitivity to defensins, protegrins, and other antimicrobials*, **J. Biol. Chem.** **274**, 8405-8410.

- 1049 BOYD DA *et al.* (2000), *Defects in D-alanyl-lipoteichoic acid synthesis in Streptococcus mutans results in acid sensitivity*, **J. Bacteriol.** **182**, 6055-6065
- 1050 DEBABOV DV *et al.* (2000), *Biosynthesis of lipoteichoic acid in Lactobacillus rhamnosus: Role of DltD in D-alanylation*, **J. Bacteriol.** **182**, 2855-2864.
- 1051 HOFMANN K (2000), *A superfamily of membrane-bound O-acyltransferases with implications for wnt signaling*, **Trends Biochem Sci.** **25**, 111-112.
- 1052 HEATON MP & FC Neuhaus, (1994), *Role of the D-alanyl carrier protein in the biosynthesis of D-alanyl-lipoteichoic acid*, **J. Bacteriol.** **176**, 681-690.
- 1053 DEBABOV DV *et al.* (1996), *The D-alanyl carrier protein in Lactobacillus casei: cloning, sequencing, and expression of dltC*, **J. Bacteriol.** **178**, 3869-3876.
- 1054 KIRIUKHIN MY & FC Neuhaus (2001), *D-alanylation of lipoteichoic acid: role of the D-alanyl carrier protein in acylation*, **J. Bacteriol.** **183**, 2051-2058.
- 1055 VOLKMAN BF *et al.* (2001), *Biosynthesis of D-alanyl-lipoteichoic acid: the tertiary structure of apo-D-alanyl carrier protein*, **Biochemistry** **40**, 7964-7972.
- 1056 MAY JJ *et al.* (2005), *Inhibition of the D-alanyl carrier protein ligase from Bacillus subtilis increases the bacterium's susceptibility to antibiotic that target the cell wall*, **FEBS J.** **272**, 2993-3003.
- 1057 STEEN A *et al.* (2005), *Autolysis of Lactococcus lactis is increased upon D-alanine depletion of peptidoglycan and lipoteichoic acids*, **J. Bacteriol.** **187**, 114-124.
- 1058 CAVA F *et al.* (2011), *Emerging knowledge of regulatory roles of D-aminoacids in bacteria*, **Cell. Mol. Life Sci.** **68**, 817-831.
- 1059 MORAR M & GD Wright (2010), *The genomic enzymology of antibiotic resistance*, **Annu. Rev. Genet.** **44**, 25-51.
- 1060 CARFI A *et al.* (1995), *The 3-D structure of a zinc metallo- β -lactamase from Bacillus cereus reveals a new type of protein fold*, **EMBO J.** **14**, 4914-4921.
- 1061 FABIANE SM *et al.* (1998), *Crystal structure of the zinc-dependent β -lactamase from Bacillus cereus at 1.9 Å resolution: binuclear active site with features of a mononuclear enzyme*, **Biochemistry** **37**, 12404-12411.
- 1062 NEUWALD AF *et al.* (1997), *Extracting protein alignment models from sequence database*, **Nucleic Acids Res.** **25**, 1665-1677.
- 1063 ARAVIND L (1998), *An evolutionary classification of the metallo- β -lactamase fold proteins*, **In Silico Biol.** **1**, 69-91.
- 1064 DAIYASU H *et al.* (2001), *Expansion of the zinc metallo-hydrolase family of the β -lactamase fold*, **FEBS Lett.** **503**, 1-6.
- 1065 BEBRONE C (2007), *Metallo- β -lactamases (classification, activity, genetic organization, structure, zinc coordination) and their superfamily*, **Biochem. Pharmacol.** **74**, 1686-1701.
- 1066 WRIGHT GD (2005), *Bacterial resistance to antibiotics: Enzymatic degradation and modification*, **Adv. Drug Deliv. Rev.** **57**, 1461-1470.
- 1067 TOMATIS PE *et al.* (2008), *Adaptive protein evolution grants organismal fitness by improving catalysis and flexibility*, **PNAS** **105**, 20605-20610.
- 1068 AMINIOV RI & RI Mackie (2007), *Evolution and ecology of antibiotic resistance genes*, **FEMS Microbiol. Lett.** **271**, 147-161.
- 1069 AMINOV RI (2009), *The role of antibiotics and antibiotic resistance in nature*, **Environ. Microbiol.** **11**, 2970-2988.
- 1070 WRIGHT GD (2007), *The antibiotic resistome: the nexus of chemical and genetic diversity*, **Nat. Rev.** **5**, 175-186.
- 1071 DAVIES J & D Davies (2010), *Origins and evolution of antibiotic resistance*, **Microbiol. Mol. Biol. Rev.** **74**, 417-433.
- 1072 BROWN MG & DL Balkwill (2009), *Antibiotic resistance in bacteria isolated from deep terrestrial subsurface*, **Microbiol. Ecol.** **57**, 484-493.
- 1073 COLOMER-LLUCH M *et al.* (2011), *Antibiotic resistance genes in the bacteriophage DNA fraction of environmental samples*, **PLoS One** **6**, e17549.
- 1074 RIGDEN DJ *et al.* (2003), *Amidase domains from bacterial and phage autolysins define a family of g-D,L-glutamate-specific amidohydrolases*, **TRENDS Biochem. Sci.** **28**, 230-234.
- 1075 BATEMAN & ND Rawlings (2003), *The CHAP domain: a large family of amidases including GSP amidase and peptidoglycan hydrolases*, **TRENDS Biochem. Sci.** **28**, 234-237.
- 1076 FIRSZUK M & M Bochtler (2007), *Folds and activities of peptidoglycan amidases*, **FEMS Microbiol. Rev.** **31**, 676-691.
- 1077 HUMANN J & LL Lenz (2009), *Bacterial peptidoglycan degrading enzymes and their impact on host muropeptide detection*, **J. Innate Immun.** **1**, 88-97.
- 1078 ZOU Y & C Hou (2010), *Systematic analysis of an amidase domain CHAP in 12 Staphylococcus aureus genomes and 44 staphylococcal phage genomes*, **Comput. Biol. Chem.** **34**, 251-257.
- 1079 LLUL D *et al.* (2006), *Sk1, a novel choline-binding N-acetylmuramoyl-L-alanine amidase from Streptococcus mitis SK137 containing a CHAP domain*, **FEBS Lett.** **580**, 1959-1964.
- 1080 NELSON D *et al.* (2006), *PlyC: A multimeric bacteriophage lysin*, **PNAS** **103**, 10765-10770.

- 1081 DONOVAN DM *et al.* (2006), The cell lysis activity of the Streptococcus agalactiae bacteriophage B30 endolysin relies on the cysteine, histidine-dependent amidohydrolase/peptidase domain, **Appl. Environ. Microbiol.** **72**, 5108-5112.
- 1082 ROSSI P *et al.* (2008), Structural elucidation of the Cys-His-Glu-Asn proteolytic relay in the secreted CHAP domain enzyme from the human pathogen Streptococcus saprophyticus, **Prot. Struct. Funct. Genet.** **#**, 515-519.
- 1083 LAYEC S *et al.* (2009), The CHAP domain of the Cse functions as an endopeptidase that acts at mature septa to promote Streptococcus thermophilus cell separation, **Mol. Microbiol.** **71**, 1205-1217.
- 1084 HORGAN M *et al.* (2009), Phage lysin LysK can be truncated to its CHAP domain and retain lytic activity against live antibiotic-resistant staphylococci, **Appl. Environ. Microbiol.** **75**, 872-874.
- 1085 BECKER SC *et al.* (2009), LysK CHAP endopeptidase domain is required for lysis of live staphylococcal cells, **FEMS Microbiol. Lett.** **294**, 52-60.
- 1086 PAI CH *et al.* (2011), Structure and mechanism of Escherichia coli glutathionylspermidine amidase belonging to hte family of cysteine; histidine-dependent amidohydrolases /peptidases, **Prot. Sci.** **20**, 557-566.
- 1087 FENTON M *et al.* (2011), Characterization of the staphylococcal bacteriophage lysin CHAP_K, **J. Appl. Microbiol.** **111**, 1025-1035.
- 1088 CAVALIER-SMITH T (2006), Cell evolution and Earth history: stasis and revolution, **Phil. Trans. R. Soc. Lond. B** **361**, 969-1006.
- 1089 NAKAIDO H & MH Saier Jr. (1992), Transport proteins in bacteria: common themes in their design, **Science** **268**, 936-942.
- 1090 HEDIGER MA *et al.* (1994), Structure, function and evolution of solute transporters in prokaryotes and eukaryotes, **J. Exp. Biol.** **196**, 15-49.
- 1091 LOKELMA JS *et al.* (1998), Bacterial solute uptake and efflux systems, **Curr. Opin. Microbiol.** **1**, 248-253.
- 1092 SAIER Jr. MH (2000), A functional-phylogenetic classification system for transmembrane solute transporters, **Microbiol. Mol. Biol. Rev.** **64**, 354-411.
- 1093 SAIER Jr. MH (2000), Vectorial metabolism and the evolution of transport systems, **J. Bacteriol.** **182**, 5029-5035.
- 1094 CALAMITA G *et al.* (1995), Molecular cloning and characterization of AqpZ, a water channel from Escherichia coli, **J. Biol. Chem.** **270**, 29063-29066.
- 1095 AGRE P *et al.* (1995), Aquaporin water channels: unanswered questions and unresolved controversies, **Curr. Opin. Cell Biol.** **7**, 472-483.
- 1096 PARK JH & MH Saier Jr. (1996), Phylogenetic comparison of the MIP family of transmembrane channel proteins, **J. Membr. Biol.** **153**, 171-180.
- 1097 HOHMANN S *et al.* (2000), Microbial MIP channels, **TRENDS Microbiol.** **8**, 33-38.
- 1098 CALAMITA G (2000), Understanding microbial MIP channels, **TRENDS Microbiol.** **8**, 104&105.
- 1099 STAHLBERG H *et al.* (2001), C. 2, The aquaporin superfamily: structure and function, **Curr. Topics Membr.** **51**, 39-119(select.).
- 1100 KAYINGO G *et al.* (2001), C. 8, Microbial water channels and glycerol facilitators, **Curr. Topics Membr.** **51**, p.335-364.
- 1101 DUCHESNE L *et al.* (2001), Oligomerization of water and solute channel of the major intrinsic protein (MIP) family, **Kidney Int.** **60**, 422-426.
- 1102 HIGGINS CF *et al.* (1990), Binding protein-dependent transport systems, **J. Bioenerg. Biomembr.** **22**, 571-592.
- 1103 HIGGINS CF *et al.* (1990), Periplasmic binding protein-dependent transport systems: the membrane associated components, **Phil. Trans. R. Soc. Lond. B** **326**, 353-365.
- 1104 TAM R & MH Saier Jr. (1993), Structural, functional, and evolutionary relationships among extracellular-solute-binding receptor of bacteria, **Microbiol. Rev.** **57**, 320-346.
- 1105 FATH MJ & R Kolter (1993), ABC transporters: bacterial exporters, **Microbiol. Rev.** **57**, 995-1017.
- 1106 LINTON KJ & CF Higgins (1998), The Escherichia coli ATP-binding cassette (ABC) proteins, **Mol. Microbiol.** **28**, 5-13.
- 1107 SAURIN W *et al.* (1999), Getting in or out: early segregation between importers and exporters in the evolution of ATP-binding cassette (ABC) transporters, **J. Mol. Evol.** **48**, 22-41.
- 1108 JONES PM & AM George (1999), Subunit interactions in ABC transporters: towards a functional architecture, **FEMS Microbiol. Lett.** **179**, 187-202.
- 1109 YOUNG J & IB Holland (1999), ABC transporters: bacterial exporters-revisited five years on, **Biochim. Biophys. Acta** **1461**, 177-200.
- 1110 DASSA E & E Schneider (2001), The rise of a protein family: ATP-binding cassette (ABC) systems, **Res. Microbiol.** **152**, 203.
- 1111 HIGGINS CF *et al.* (2001), ABC transporters: physiology, structure and mechanism – an overview, **Res. Microbiol.** **152**, 205-2010.

- ¹¹¹² DASSA E & P Bouige (2001), *The ABC of ABCs: a phylogenetic and functional classification of ABC systems in living organisms*, **Res. Microbiol.** **152**, 211-229.
- ¹¹¹³ THOMAS PJ & JE Hunt (2001), *A snapshot of Nature's favorite pump*, **Nat. Struct. Biol.** **8**, 920-923.
- ¹¹¹⁴ BOUIGE P *et al.* (2002), *Phylogenetic and functional classification of ATP-binding cassette (ABC) systems*, **Curr. Prot. Pept. Sci.** **3**, 541-559.
- ¹¹¹⁵ JONES PM & AM George (2004), *The ABC transporter structure and mechanism: perspectives on recent research*, **Cell Mol. Life Sci.** **61**, 682-699.
- ¹¹¹⁶ HOLLAND KA & IB Holland (2005), *Adventures with ABC-proteins: highly conserved ATP-dependent transporters*, **Acta Microbiol. Immunol. Hungarica** **52**, 309-322.
- ¹¹¹⁷ BIEMANS-ODENHINKEL E *et al.* (2006), *ABC transporter architecture and regulatory roles of accessory domains*, **FEBS Lett.** **580**, 1023-1035.
- ¹¹¹⁸ DAWSON RJP *et al.* (2007), *Uptake or extrusion: crystal structures of full ABC transporters suggest a common mechanism*, **Mol. Microbiol.** **65**, 250-257.
- ¹¹¹⁹ HOLLENSTEIN K *et al.* (2007), *Structure and mechanism of ABC transporter proteins*, **Curr. Opin. Struct. Biol.** **17**, 412-418.
- ¹¹²⁰ DAVIDSON AL & PC Maloney (2007), *ABC transporters: how small machines do a big job*, **TRENDS Microbiol.** **15**, 448-455.
- ¹¹²¹ PEDERSEN PL *et al.* (2007), *Transport ATPases into the year 2008: a brief overview related to types, structures, functions and roles in health and disease*, **J. Bioenerg. Biomembr.** **39**, 349-355.
- ¹¹²² DAVIDSON AL *et al.* (2008), *Structure, function, and evolution of bacterial ATP-binding cassette systems*, **Microbiol. Mol. Biol. Rev.** **72**, 317-364.
- ¹¹²³ LOCHER KP (2009), *Structure and mechanism of ATP-binding cassette transporters*, **Phil. Trans. R. Soc. B** **364**, 239-245.
- ¹¹²⁴ REES DC *et al.* (2009), *ABC transporters: the power to change*, **Nat. Rev. Mol. Cell. Biol.** **10**, 218-227.
- ¹¹²⁵ JONES PM *et al.* (2009), *ABC transporters: a riddle wrapped in a mystery inside an enigma*, **TRENDS Biochem. Sci.** **34**, 520-531.
- ¹¹²⁶ SCHNEIDER E *et al.* (2012), *Receptor-transporter interactions of a canonical ATP-binding cassette import systems in prokaryotes*, **Eur. J. Cell Biol.** **91**, 311-317.
- ¹¹²⁷ QUIOCHO FA & PS Ledvina (1998), *Atomic structure and specificity of bacterial periplasmic receptors for active transport and chemotaxis: variation in common themes*, **Mol. Microbiol.** **20**, 17-25.
- ¹¹²⁸ FUKAMI-KBAYASHI K *et al.* (1999), *Domain dislocation: a change of core structure in periplasmic binding proteins in their evolutionary history*, **J Mol. Biol.** **286**, 279-290.
- ¹¹²⁹ SAURIN W & E Dassa (1994), *Sequence relationships between integral inner membrane proteins of binding protein-dependent transport systems: evolution by recurrent gene duplications*, **Prot. Sci.** **3**, 325-344.
- ¹¹³⁰ SAURIN W *et al.* (1994), *Bacterial binding protein-dependent permeases: characterization of distinctive signatures for functionally related integral cytoplasmic membrane proteins*, **Mol. Microbiol.** **12**, 993-1004.
- ¹¹³¹ MIMMACK ML *et al.* (1989), *Energy coupling to periplasmic binding protein-dependent transport systems: Stoichiometry of ATP hydrolysis during transport in vivo*, **PNAS** **86**, 8257-8261.
- ¹¹³² SCHNEIDER E & S Hunke (1998), *ATP-binding-cassette (ABC) ATP-binding-cassette (ABC) transport systems: functional and structural aspects of the ATP-hydrolyzing subunits/domains*, **FEMS Microbiol. Lett.** **22**, 1-20.
- ¹¹³³ KERR ID *et al.* (2002), *Structure and association of ATP-binding cassette transporter nucleotide binding domains*, **Biochim. Biophys. Acta** **1561**, 47-64.
- ¹¹³⁴ JONES PM & AM George (2002), *Mechanism of ABC transporters: a molecular dynamics simulations of well characterized nucleotide-binding subunit*, **PNAS** **99**, 12539-12644.
- ¹¹³⁵ AMBUDKAR SV *et al.* (2006), *The A-loop, a novel conserved aromatic acid subdomain upstream of the Walker A motif in ABC transporters, is critical for ATP binding*, **FEBS Lett.** **580**, 1049-2055.
- ¹¹³⁶ LINTON KJ & CF Higgins (2007), *Structure and function of ABC transporters: the ATP switch provides flexible control*, **Eur. J. Physiol.** **453**, 555-567.
- ¹¹³⁷ GYMESI G *et al.* (2022), *ATP hydrolysis at one of the two sites in ABC transporters initiates transport related conformational transitions*, **Biochim. Biophys. Acta** **1808**, 2954-2964.
- ¹¹³⁸ SWEET GD *et al.* (1984), *Tricarboxylate-binding proteins of Salmonella typhimurium, Purification, crystallization, and physical properties*, **J. Biol. Chem.** **259**, 1586-1592.
- ¹¹³⁹ CHEN CM *et al.* (1990), *Molecular biology of carbon-phosphorus bond cleavage, Cloning and sequencing of the phn (psiD) genes involved in alkylphosphonate uptake and C-P lyase activity in Escherichia coli*, **J. Biol. Chem.** **265**, 4461-4471.

- ¹¹⁴⁰ KURYATOV A *et al.* (1994), *Mutational analysis of the glycine-binding site of the NMDA receptor: structural similarity with bacterial amino acid-binding proteins*, **Neuron** **12**, 1291-1300.
- ¹¹⁴¹ NGUYEN CC & MH Saier Jr. (1995), *Phylogenetic, structural and functional analyses of the LacI-GalR family of bacterial transcription factors*, **FEBS Lett.** **377**, 98-102.
- ¹¹⁴² FELDER CB *et al.* (1999), *The venus flytrap od periplasmic binding proteins: an ancient protein module present in multiple drug receptor*, **AAPS PharmSci.** **1**(2) E2.
- ¹¹⁴³ MERLIN C *et al.* (2002), *The Escherichia coli metD locus encodes an ABC transporter which includes Abc (MetN), YaeE (MetI), and YaeC (MetQ)*, **J. Bacteriol.** **184**, 5513-5517.
- ¹¹⁴⁴ ANTOINE R *et al.* (2003), *Overrepresentation of a gene family encoding extracytoplasmic solute receptors in Bordetella*, **J. Bacteriol.** **185**, 1470-1474.
- ¹¹⁴⁵ FUKAMI-KOBAYASHI K *et al.* (2003), *Parallel evolution of ligand specificity between LacI/GalR family repressors and periplasmic sugar-binding proteins*, **Mol. Biol. Evol.** **20**, 267-77.
- ¹¹⁴⁶ ACHER FC & HO Bertrand (2005), *Amino acid recognition by venus flytrap domains is encoded in an 8-residue motif*, **Biopolymers** **80**, 357-366.
- ¹¹⁴⁷ TRAKHANOV S *et al.* (2005), *Ligand-free and -bound structures of the binding protein (LivJ) of the Escherichia coli ABC leucine/isoleucine/valine transport system: trajectory and dynamics of the interdomain rotation and ligand specificity*, **Biochemistry** **44**, 6597-6608.
- ¹¹⁴⁸ RIZK SS *et al.* (2006), *Identification of cognate ligands for the Escherichia coli pH₂ protein product and engineering of a reagentless fluorescent biosensor for phosphates*, **Prot. Sci.** **15**, 1745-1751.
- ¹¹⁴⁹ WALSHAW DL *et al.* (1997), *Distribution of a sub-class of bacterila ABC polar amino acid transporter and identification of an N-terminal region involved in solute specificity*, **FEBS Lett.** **414**, 397-401.
- ¹¹⁵⁰ HOSSIE AHF & PH Poole (2001), *Bacterial transporters of amino acids*, **Res. Microbiol.** **152**, 259-270.
- ¹¹⁵¹ LOMBARDO MJ *et al.* (1993), *Physical mapping of the Escherichia coli PepT and potABCD genes*, **J. Bacteriol.** **175**, 7745-7746.
- ¹¹⁵² MIERAU I *et al.* (1994), *Tripeptidase gene (pepT) of Lactococcus lactis: molecular cloning and nucleotide sequencing of pepT and construction of a chromosomal deletion mutant*, **J. Bacteriol.** **176**, 2854-2861.
- ¹¹⁵³ RODIONOV DA *et al.* (2006), *Comparative and functional genomic analysis of prokaryotic nickel and cobalt uptake transporters: evidence for a novel group of ATP-binding cassette transporters*, **J. Bacteriol.** **188**, 317-327.
- ¹¹⁵⁴ CHIVERS PT *et al.* (2012), *Identification of Ni-(L-His)₂ as a substrate for NikABCDE-dependent nickel uptake in Escherichia coli*, **Metalomics** **4**, 1043-1050.
- ¹¹⁵⁵ FIEDLER G *et al.* (1998), *The DevBCA exporter is essential for envelope formation in heterocysts of the cyanobacterium Anabaena sp. strain PCC 7120*, **Mol. Microbiol.** **27**, 1193-1202.
- ¹¹⁵⁶ de LEEUW *et al.* (1999), *Molecular characterization of Escherichia coli FtsE and FtsX*, **Mol. Microbiol.** **31**, 983-993.
- ¹¹⁵⁷ SCHMIDT KI *et al.* (2004), *A predicted ABC transporter, FtsX, is needed for a cell division in Escherichia coli*, **J. Bacteriol.** **186**, 785-793.
- ¹¹⁵⁸ XU Y *et al.* (2009), *Crystal structure of the periplasmic region of MacB, a noncanonic ABC transporter*, **Biochemistry** **48**, 5212-5255.
- ¹¹⁵⁹ GOWRISHANKAR J (1989), *Nucleotide sequence of the osmoregulatory proU operon of Escherichia coli*, **J. Bacteriol.** **171**, 1923-1931.
- ¹¹⁶⁰ STIRLING DA *et al.* (1989), *Molecular characterization of the proU loci of Salmonella typhimurium and Escherichia coli encoding osmoregulated glycine betaine transport systems*, **Mol. Microbiol.** **3**, 1025-1038.
- ¹¹⁶¹ LUCHT JM & E Bremer (1994), *Adaptation of Escherichia coli to high osmolarity environments: Osmoregulations of the high-affinity glycine betaine transport system ProU*, **FEMS Microbiol Rev.** **14**, 3-20.
- ¹¹⁶² HAARDT M *et al.* (1995), *The osmoprotectant proline betaine is a major substrate for the binding-protein-dependent transport system ProU of Escherichia coli K-12*, **Mol. Gen. Genet.** **246**, 783-786.
- ¹¹⁶³ KEMPF B & E Bremer (1995), *OpuA, an osmotically regulated binding protein-dependent transport system for the osmoprotectant glycine betaine in Bacillus subtilis*, **J. Biol. Chem.** **270**, 16701-16713.
- ¹¹⁶⁴ HAARDT M & E Bremer (1996), *Use of phoA and lacZ fusions to study the membrane topology of ProW, a component of the osmoregulated ProU transport system of Escherichia coli*, **J. Bacteriol.** **178**, 5370-5381.
- ¹¹⁶⁵ KEMPF B *et al.* (1997), *Lipoprotein from the osmoregulated ABC transport system OpuA of Bacillus subtilis: purification of the glycine betaine binding protein and characterization of a functional lipidless mutant*, **J. Bacteriol.** **179**, 6213-6220.

- ¹¹⁶⁶ KAPPES RM *et al.* (1999), *Two evolutionarily closed related ABC transporters mediate the osmoprotectant glycine betaine in Bacillus subtilis*, **Mol. Microbiol.** **32**, 203-216.
- ¹¹⁶⁷ BOUVIER J *et al.* (2000), *Characterization of OpuA, a glycine-betaine uptake system of Lactococcus lactis*, **J. Mol. Microbiol. Biotechnol.** **2**, 199-205.
- ¹¹⁶⁸ van der HEIDE T & B Poolman (2000), *Osmoregulated ABC-transport system of Lactococcus lactis senses water stress via changes in the physical state of the membrane*, **PNAS** **97**, 7102-7106.
- ¹¹⁶⁹ ALLOING G *et al.* (2006), *Proline betaine uptake in Sinorhizobium meliloti: Characterization of Prb, an opp-like ABC transporter regulated by both proline betaine and salinity stress*, **J. Bacteriol.** **188**, 6308-6317.
- ¹¹⁷⁰ ANNAMALAI T & K Venkitanarayanan (2009), *Role of proP and proU in betaine uptake by Yersinia enterocolitica under cold and osmotic stress conditions*, **Appl. Environ. Microbiol.** **75**, 1471-1477.
- ¹¹⁷¹ SPELLERBERG B *et al.* (1999), *Identification of genetic determinants for the hemolytic activity of Streptococcus agalactiae by ISS1 transposition*, **J. Bacteriol.** **181**, 3212-3219.
- ¹¹⁷² HAAS W & MS Gilmore (1999), *Molecular nature of a novel bacterial toxin: the cytolysin of Enterococcus faecalis*, **Mol. Microbiol. Immunol.** **187**, 183-190.
- ¹¹⁷³ SPELLERBERG B *et al.* (2000), *The cyl genes of Streptococcus agalactiae are involved in the production of pigment*, **FEMS Microbiol. Lett.** **188**, 125-128.
- ¹¹⁷⁴ SEMEDO T *et al.* (2003), *Comparative study using type strains and clinical and food isolates to examine hemolytic activity and occurrence of the cyl operon in enterococci*, **J. Clin. Microbiol.** **41**, 2569-2576.
- ¹¹⁷⁵ GOTTSCHALK B *et al.* (2006), *Transport of multidrug resistance substrates by the Streptococcus agalactiae hemolysin transporter*, **J. Bacteriol.** **188**, 5984-5992.
- ¹¹⁷⁶ NEILANDS JB *et al.* (1995), *Siderophores: structure and function of microbial iron transport compounds*, **J. Biol. Chem.** **270**, 26723-26726.
- ¹¹⁷⁷ GROEGER W & W Köster (1998), *Transmembrane topology of the two FhuB domains representing the hydrophobic components of bacterial ABC transporters involved in the uptake of siderophores, haem and vitamin B12*, **Microbiology** **144**, 2759-2769.
- ¹¹⁷⁸ BRAUN V & H Killmann (1999), *Bacterial solutions to the iron-supply problem*, **TRENDS Biochem. Sci.** **24**, 104-109.
- ¹¹⁷⁹ KÖSTER W (2001), *ABC transporter-mediated uptake of iron, siderophores, heme and vitamin B12*, **Res. Microbiol.** **152**, 291-301.
- ¹¹⁸⁰ CLARKE TE *et al.* (2001), *Structural biology of bacterial iron uptake systems*, **Curr. Top. Med. Chem.** **1**, 7-30(selec.).
- ¹¹⁸¹ ANDREWS SC *et al.* (2003), *Bacterial iron homeostasis*, **FEMS Microbiol. Rev.** **27**, 215-237.
- ¹¹⁸² WANDERSMAN C & P Delepelaire (2004), *Bacterial iron sources: from siderophores to hemophores*, **Annu. Rev. Microbiol.** **58**, 611-647.
- ¹¹⁸³ KREWULAK KD & HJ Vogel (2008), *Structural biology of bacterial iron uptake*, **Biochim. Biophys. Acta** **1778**, 1781-1804.
- ¹¹⁸⁴ SANDY M & A Butler (2009), *Microbial iron acquisition: marine and terrestrial siderophores*, **Chem. Rev.** **109**, 4580-4595.
- ¹¹⁸⁵ DESAI DK *et al.* (2012), *Factors influencing the diversity of iron uptake systems in aquatic microorganisms*, **Front. Microbiol.** **3**(362).
- ¹¹⁸⁶ SCHNEIDER E (2001) *ABC transporters catalyzing carbohydrate uptake*, **Res. Microbiol.** **152**, 303-310.
- ¹¹⁸⁷ WEBB AJ & SHF Hosie (2006), *A member of the second carbohydrate uptake subfamily of ATP-binding cassette transporters is responsible for ribonucleoside uptake in Streptococcus mutans*, **J. Bacteriol.** **188**, 8005-8012.
- ¹¹⁸⁸ WEBB AJ *et al.* (2008), *Two closely related ABC transporters in Streptococcus mutans are involved in disaccharide and/or oligosaccharide uptake*, **J. Bacteriol.** **190**, 168-178.
- ¹¹⁸⁹ GUILLÉN-NAVARRO K *et al.* (2005), *Biotin biosynthesis, transport and utilization in rhizobia*, **FEMS Microbiol. Lett.** **246**, 159-165.
- ¹¹⁹⁰ GUILLÉN-NAVARRO K *et al.* (2005), *The Rhizobium etli bioMNY operon is involved in biotin transport*, **FEMS Microbiol. Lett.** **250**, 209-219.
- ¹¹⁹¹ HEBBLN P *et al.* (2007), *Biotin uptake in prokaryotes by solute transporters with an optional ATP-binding cassette-containing module*, **PNAS** **104**, 2909-2914.
- ¹¹⁹² FINKENWIRTH F *et al.* (2010), *Subunit composition of an energy-coupling-factor-type biotin transporter analysed in bacteria*, **Biochem. J.** **431**, 373-380.
- ¹¹⁹³ NEUBAUER O *et al.* (2011), *Interactions among the A and T units of an ECF-type biotin transporter analyzed by site-specific crosslinking*, **PLoS One** **6**(12), e29087.
- ¹¹⁹⁴ FISHER DJ *et al.* (2012), *Uptake of biotin by Chlamydia Spp. through the use of a bacterial transporter (BioY) and a host-cell transporter (SMVT)*, **PLoS One** **7**(9), e46052.
- ¹¹⁹⁵ BERTSSON RPA *et al.* (2012), *Structural divergence of paralogous S components from ECF-type ABC transporters*, **PNAS** **109**, 13990-13995.

- ¹¹⁹⁶ KIRSCH F *et al.* (2012), *Essential amino acid residues of BioY reveal that dimers are the functional S unit of the Rhodobacter capsulatus biotin transporter*, **J. Bacteriol.** **194**, 4505-4512.
- ¹¹⁹⁷ SCHNEIDER J *et al.* (2012), *Characterization of the biotin uptake system encoded by the biotin-inducible bioYMN operon of Corynebacterium glutamicum*, **BMC Microbiol.** **12**(6).
- ¹¹⁹⁸ FURUCHI T *et al.* (1991), *Characteristics of the gene for a spermidine and putrescine transport system that maps at 15 min on the Escherichia coli chromosome*, **J. Biol. Chem.** **266**, 20928-20932.
- ¹¹⁹⁹ PISTOCCHI R *et al.* (1993), *Characteristics of the gene for a putrescine transport system that maps at 19 minutes on the Escherichia coli chromosome*, **J. Biol. Chem.** **268**, 146-152.
- ¹²⁰⁰ IGARASHI K & K Kashiwagi (1999), *Polyamine transport in bacteria and yeast*, **Biochem. J.** **344**, 633-642.
- ¹²⁰¹ IGARASHI K & K Kashiwagi (2001), *Polyamine uptake systems in Escherichia coli*, **Res. Microbiol.** **152**, 271-278.
- ¹²⁰² HIGASHI K *et al.* (2010), *Identification and functions of amino acid residues in PotB and PotC involved in spermidine uptake activity*, **J. Biol. Chem.** **285**, 39061-39069.
- ¹²⁰³ WU LF *et al.* (1991), *The hydC region contains a multi-cistronic operon (nik) involved in nickel transport in Escherichia coli*, **Gene** **107**, 37-42.
- ¹²⁰⁴ NAVARRO C *et al.* (1993), *The nik operon of Escherichia coli encodes a periplasmic binding-protein-dependent transport system for nickel*, **Mol. Microbiol.** **9**, 1181-1191.
- ¹²⁰⁵ WU LF *et al.* (1994), *Antagonistic effect of nickel on the fermentative growth of Escherichia coli K-12 and comparison of nickel and cobalt toxicity on the aerobic and anaerobic growth*, **Environ. Health. Perspect.** **102**, 297-300.
- ¹²⁰⁶ EITINGER T & MA Mandrand-Berthelot (2000), *Nickel transport systems in microorganisms*, **Arch. Microbiol.** **173**, 1-9.
- ¹²⁰⁷ MULROONEY SB & RP Hausinger (2003), *Nickel uptake and utilization by microorganisms*, **FEMS Microbiol. Rev.** **27**, 239-261.
- ¹²⁰⁸ KAPPES RM *et al.* (1996), *Three transport systems for the osmoprotectant glycine betaine operate in Bacillus subtilis: characterization of OpuD*, **J. Bacteriol.** **178**, 5071-5079.
- ¹²⁰⁹ BONOCOMPAGNI E *et al.* (2000), *Characterization of a Sinorhizobium meliloti ATP-binding cassette histidine transporter also involved in betaine and proline uptake*, **J. Bacteriol.** **182**, 3717-3725.
- ¹²¹⁰ FRASER KR (2000), *Identification and characterization of an ATP binding cassette L-carnitine transporter in Listeria monocytogenes*, **Appl. Environ. Microbiol.** **66**, 4696-4704.
- ¹²¹¹ DUPONT L *et al.* (2004), *The Sinorhizobium meliloti ABC transporter Cho is highly specific for choline and expressed in bacteroids from Medicago sativa nodules*, **J. Bacteriol.** **186**, 5988-5996.
- ¹²¹² HOLLAND BI & Blight MA *et al.* (1999), *ABC-ATPases, ABC-ATPases, adaptable energy generators fuelling transmembrane movement of a variety of molecules in organisms from bacteria to humans*, **J. Mol. Biol.** **293**, 381-399.
- ¹²¹³ OSWALD C *et al.* (2006), *The motor domains of ABC transporters What can structures tell us? Naunyn Schmiedebergs Arch Pharmacol.* **372**, 385-399.
- ¹²¹⁴ LUTSENKO S & JH Kaplan (1996), *P-type ATPases*, **TRENDS Biochem. Sci.** **21**, p.467.
- ¹²¹⁵ MAEDA M *et al.* (1998), *Structures of P-type transportig ATPases and chromosomal locations of their genes*, **Cell Struct. Funct.** **23**, 315-323.
- ¹²¹⁶ ARAVIND L *et al.* (1998), *The catalytic domain of the P-type ATPase has the haloacid dehalogenase fold*, **TRENDS Biochem. Sci.** **23**, 127-129.
- ¹²¹⁷ PALMGREN MG & P Nissen (2011), *P-type ATPases*, **Annu. Rev. Biophys.** **40**, 243-266.
- ¹²¹⁸ BUBLITZ M *et al.* (2012), *P-type ATPases at a glance*, **J. Cell Sci.** **124**, 2515-2519.
- ¹²¹⁹ MARGER MD & MH Saier Jr. (1993), *A major superfamily transmembrane facilitator that catalyse uniport, symport and antiport*, **TREND Biol. Sci.** **18**, 13-20.
- ¹²²⁰ SAIER MH Jr. (1999), *The major facilitator superfamily*, **J. Mol. Microbiol. Biotechnol.** **1**, 257-279.
- ¹²²¹ LAW CJ *et al.* (2008), *Ins and outs of major facilitator superfamily antiporters*, **Annu. Rev. Microbiol.** **62**, 289-305.
- ¹²²² PAO SS *et al.* (1998), *Major facilitator superfamily*, **Microbiol. Mol. Biol. Rev.** **61**, 1-34.
- ¹²²³ HIRAI T *et al.* (2003), *Structural model of the 12-helix transporters belongin to the major facilitator superfamily*, **J. Bacteriol.** **185**, 1712-1718.
- ¹²²⁴ VARDY E *et al.* (2004), *Structural conservation in the major facilitator superfamily as revealed by the comaparative modeling*, **Prot. Sci.** **13**, 1832-1840.
- ¹²²⁵ HOLYOAKE J *et al.* (2006), *Modeling, docking, and simuation of the major facilitator superfamily*, **Biophys. J. Biophys. Lett.** **91**, L84-L86.
- ¹²²⁶ JEON J *et al.* (2009), *Integration of evolutionary features for the identification of functionally important residues in major facilitator superfamily transporters*, **PLoS Comput. Biol.** **5**(10).

- 1227 REDDY VS *et al.* (2011), *The major facilitator superfamily revisited*, **FEBS J.** **279**, 2022-2035.
- 1228 STEWART LMD *et al.* (1985), *Entry coupling to K⁺ uptake via the Trk system in Escherichia coli: the role of ATP*, **J. Gen. Microbiol.** **131**, 77-85.
- 1229 DOSCH DC *et al.* (1991), *Genetic analysis of potassium transport loci in Escherichia coli: evidence for three constitutive systems mediating uptake of potassium*, **J. Bacteriol.** **173**, 687-696.
- 1230 EPSTEIN W *et al.* (1993), *Multiple mechanisms, roles and controls of K⁺ transport in Escherichia coli*, **Biochem. Soc. Transact.** **21**, 1006&1007.
- 1231 SCHLÖSSER A *et al.* (1995), *TrkH and its homolog, TrkG, determine the specificity and kinetics of cation transport by the Trk system of Escherichia coli*, **J. Bacteriol.** **177**, 1908-1910.
- 1232 NAKAMURA T *et al.* (1999), *Cloning of the trkAH gene cluster and characterization of the Trk K⁺-uptake system of Vibrio alginolyticus*, **Microbiology** **144**, 2281-2289.
- 1233 DURELL ST *et al.* (1993), *Evolutionary relationship between K⁺ channels and symporters*, **Biophys. J.** **77**, 775-788.
- 1234 DURELL SR & HR Guy (1999), *Structural models of the KtrB, TrkH, and Trk1,2 symporters based on the structure of the KcsA K⁺ channel*, **Biophys. J.** **77**, 789-807.
- 1235 HARMS C *et al.* (2001), *Identification of the ABC protein SapD as the subunit that confers ATP dependence to the K⁺-uptake systems TrkH and TrkG from Escherichia coli K12*, **Microbiology** **147**, 2991-3003.
- 1236 EPSTEIN W (2003), *The roles and regulation of potassium in bacteria*, **Progr. Nucl. Acids Res.** **75**, 293-314.
- 1237 BUURMAN ET *et al.* (2004), *Multiple paths for nonphysiological transport of K⁺ in Escherichia coli*, **J. Bacteriol.** **186**, 4238-4245.
- 1238 KRAEGLOH A *et al.* (2005), *Potassium transport in a halophilic member of the Bacteria domain: identification and characterization of the K⁺ uptake systems TrkH and TrkI from Halomonas elongata DSM 2581*, **J. Bacteriol.** **187**, 1036-1043.
- 1239 JOHNSON H *et al.* (2009), *The Thermotoga maritima Trk potassium transporter -from frameshift to function*, **J. Bacteriol.** **191**, 2276-2284.
- 1240 SIMONI RD *et al.* (1967), *Genetic evidence for the role of a bacterial phosphotransferase system*, **PNAS** **58**, 1963-1970.
- 1241 ROSE SP & CF Fox (1971), *The \square -glucoside system of Escherichia coli, II Kinetic evidence for a phosphoryl-enzyme II intermediate*, **Biochem. Biophys. Res. Commun.** **45**, 376-380.
- 1242 KUNDIG W & S Roseman (1971), *Sugar transport I, Isolation of a phosphotransferase system from Escherichia coli*, **J. Biol. Chem.** **246**, 1393-1406.
- 1243 KUNDIG W & S Roseman (1971), *Sugar transport II, Characterization of constitutive membrane-bound enzymes II of the Escherichia coli phosphotransferase system*, **J. Biol. Chem.** **246**, 1497-1418.
- 1244 SAIER MH Jr. *et al.* (1977), *Sugar phosphate:sugar transphosphorylation coupled to exchange group translocation catalyzed by the enzyme II complexes of the phosphoenolpyruvate:sugar phosphotransferase system in membrane vesicles of Escherichia coli*, **J. Biol. Chem.** **252**, 8908-8916.
- 1245 SAIER JR Jr. (1980), *Catalytic activities associated with the enzymes II of the bacterial phosphotransferase system*, **J. Supramol. Struct.** **14**, 281-294.
- 1246 SAIER MH Jr. & MR Schmidt (1981), *Vectorial and nonvectorial transphosphorylation catalyzed by enzymes II of the bacterial phosphotransferase system*, **J. Bacteriol.** **145**, 391-397.
- 1247 ROBILLARD GT (1982), *The enzymology of the bacterial phosphoenolpyruvate-dependent sugar transport systems*, **Mol. Cell Biochem.** **46**, p.3.
- 1248 POSTMA PW & JW Lengeler (1985), *Phosphoenolpyruvate:sugar phosphotransferase system of bacteria*, **Microbiol. Rev.** **49**, 232-269.
- 1249 STÜLKE J & W Hillen (1998), *Coupling physiology and gene regulation in bacteria: the phosphotransferase sugar uptake system delivers the signals*, **Naturwissenschaften** **85**, p.583.
- 1250 GUNNEWIJK MGW *et al.* (2001), *Hierarchical control versus autoregulation of carbohydrate utilization in bacteria*, **J. Mol. Microbiol. Biotechnol.** **3**, 401-413.
- 1251 ROBILLARD GT & JS Lokelma (1988), *Enzymes II of the phosphoenolpyruvate-dependent sugar transport systems: a review of their structure and mechanism of sugar transport*, **Biochim. Biophys. Acta.** **947**, 493-519.
- 1252 REIZER J *et al.* (1988), *The phosphoenolpyruvate:sugar phosphotransferase system in gram-positive bacteria: properties, mechanism, and regulation*, **Crit. Rev. Microbiol.** **15**, 297-338.
- 1253 POSTMA PW *et al.* (1993), *Phosphoenolpyruvate:carbohydrate phosphotransferase systems of bacteria*, **Microbiol. Rev.** **57**, 543-594.

- 1254 LENGELER JW *et al.* (1994), *Enzymes II of the phospho enol pyruvate-dependent phosphotransferase systems: their structure and function in carbohydrate transport*, **Biochim. Biophys. Acta.** **1188**, 1-28.
- 1255 ROBILLARD GT & J Broos (1999), *Structure/function studies on the bacterial carbohydrate transporters, enzymes II, of the phosphoenolpyruvate-dependent phosphotransferase system*, **Biochim Biophys Acta.** **1422**, 73-104.
- 1256 KOTRBA P *et al.* (2001), *Bacterial phosphotransferase system (PTS) in carbohydrate uptake and control of carbon metabolism*, **J. Biosci. Bioeng.** **92**, 502-517.
- 1257 ABOULWABA M & MH Saier Jr. (2004), *Characterization of soluble enzyme II complexes of the Escherichia coli phosphotransferase system*, **J. Bacteriol.** **186**, 8453-8462.
- 1258 BARABOTE RD & MH Saier Jr. (2005), *Comparative genomic analysis of the bacterial phosphotransferase system*, **Microbiol. Mol. Biol. Rev.** **69**, 608-634.
- 1259 SAIER MH Jr. *et al.* (1985), *Evidence for the evolutionary relatedness of the proteins of the bacterial phosphoenolpyruvate:sugar phosphotransferase system*, **J. Cell. Biochem.** **27**, 43-56.
- 1260 SAIER MH Jr. *et al.* (1988), *Sugar permeases of the bacterial phosphoenolpyruvate:sugar phosphotransferase system: sequence comparisons*, **FASEB J.** **2**, 199-208.
- 1261 LENGELER JW *et al.* (1990), *Structures and homologies of carbohydrate: phosphotransferase system (PTS) proteins*, **Phil. Transact. R. Soc. Lond. B** **326**, 489-504.
- 1262 REIZER J *et al.* (1992), *Functional interactions between proteins of the phosphoenolpyruvate:sugar phosphotransferase systems of Bacillus subtilis and Escherichia coli*, **J. Biol. Chem.** **267**, 9158-9169.
- 1263 SAIER MH Jr. & J Reizer (1992), *Proposed uniform nomenclature for the proteins and protein domains of the bacterial phosphoenolpyruvate:sugar phosphotransferase system*, **J. Bacteriol.** **174**, 1433-1438.
- 1264 REIZER J *et al.* (1994), *Novel phosphotransferase system genes revealed by bacterial genome analysis: unique, putative fructose- and glucoside-specific systems*, **Protein Sci.** **3**, 440-450.
- 1265 SAIER MH Jr. & J Reizer (1994), *The bacterial phosphotransferase system: new frontiers 30 years later*, **Mol. Microbiol.** **13**, 755-764.
- 1266 SAIER MH Jr. *et al.* (2005), *Evolution of the bacterial phosphotransferase system: from carriers and enzymes to group translocators*, **Biochem. Soc. Trans.** **33**, 220-224.
- 1267 DEUTSCHER J *et al.* (2006), *How phosphotransferase system-related protein phosphorylation regulates carbohydrate metabolism in bacteria*, **Microbiol. Mol. Biol. Rev.** **70**, 939-1031, erratum in: *ibid.* 72, p.555.
- 1268 SAIER MH Jr. (1990), *Evolution of permease diversity and energy-coupling mechanisms with special reference to the bacterial phosphotransferase system*, **Biochim. Biophys. Acta** **1018**, 248-251.
- 1269 CASES I *et al.* (2007), *The ancestral role of the phosphoenolpyruvate-carbohydrate phosphotransferase system (PTS) as exposed by comparative genomics*, **Res. Microbiol.** **158**, 666-670.
- 1270 COMAS I *et al.* (2008), *Unraveling the evolutionary history of the phosphoryl-transfer chain of the phosphoenolpyruvate:phosphotransferase system through phylogenetic analyses and genome context*, **BMC Evol. Biol.** **8**(147).
- 1271 SAIER MH Jr. (1994), *Computer-aided analyses of transport protein sequences: gleaned evidence concerning function, structure, biogenesis, and evolution*, **Microbiol. Rev.** **58**, 71-93.
- 1272 REIZER J & MH Saier Jr. (1997), *Modular multidomain phosphoryl transfer proteins in bacteria*, **Curr.** **7**, 407-415.
- 1273 REN Q & IT Paulsen (2005), *Comparative analyses of fundamental differences in membrane transport capabilities in prokaryotes and eukaryotes*, **PLoS Comput. Biol.** **1**(3), e27.
- 1274 LORCA GL *et al.* ((2007), *Transport capabilities of eleven gram-positive bacteria comparative genomic analyses*, **Biochim. Biophys. Acta** **178**, 1342-1366.
- 1275 REN Q & IT Paulsen (2007), *Large-scale comparative genomic analyses of cytoplasmic membrane transport systems in prokaryotes*, **J. Mol. Microbiol.** **12**, 165-179.
- 1276 MINOR DL *et al.* (2009), *Searching for interesting channels: pairing selection and molecular evolution methods to study ion channel structure and function*, **Mol. Biosyst.** **5**, 802-810.
- 1277 GALPERIN MY (2004), *Bacterial signal transduction network in a genomic perspective*, **Environ. Microbiol.** **6**, 552-567.
- 1278 COOPER GM, *Regulation of protein function* en **The Cell: A molecular approach**, 2nd edition, Sunderland Associates, USA, 2000.
- 1279 BERG JM, JL Tymoczko & L Stryer, *Covalent modification is a means of regulating enzyme activity* en **Biochemistry**, 5th edition, WH Freeman, USA, 2002.
- 1280 KANNAN N *et al.* (2007), *Structural and functional diversity of the microbial kinome*, **PLoS Biol.** **5**(3), e17.

- 1281 CHEEK S *et al.* (2002), *Sequence and structure classification of kinases*, **J. Mol. Biol.** **320**, 855-881.
- 1282 SCHEEF ED & PE Bourne (2005), *Structural evolution of the protein kinase-like superfamily*, **PLoS Comput. Biol.** **1**(5), e49.
- 1283 BOURRET RB & RE Silversmith, *Two-component signal transduction*, **Curr. Opin. Microbiol.** **13**, 113-115.
- 1284 MASCHER T *et al.* (2006), *Stimulus perception in bacterial signal-transducing histidine kinases*, **Microbiol. Mol. Biol. Rev.** **70**, 910-938.
- 1285 CASINO P *et al.* (2009), *Structural insight into partner specificity and phosphoryl transfer in two-component signal transduction*, *Cell* **139**, 325-36.
- 1286 LAUB MT & M Goulian (2007), *Specificity in two-component signal transduction pathways*, **Annu. Rev. Genet.** **41**, 121-145.
- 1287 GREBE TW Y JB Stock (1999), *The histidine protein kinase superfamily*, **Adv. Microb. Physiol.** **41**, p.139.
- 1288 STOCK JB *et al.* (1989), *Protein phosphorylation and regulation of adaptive responses in bacteria*, **Microbiol. Rev.** **53**, p.450&451.
- 1289 ALEX LA & MI Simon (1993), *Protein histidine kinases and signal transduction in prokaryotes and eukaryotes*, **TRENDS Gen.** **10**, p.133&134.
- 1290 CHANG C & RC Stewart (1998), *The two component system*, **Plant Physiol.** **117**, 723.731.
- 1291 STOCK A *et al.* (2000), *Two-component signal transduction*, **Annu. Rev. Biochem.** **69**, 183-195.
- 1292 HOCH JA & Varughese KI (2001), *Keeping signals straight in phosphorelay signal transduction*, **J. Bacteriol.** **183**, 4941-4949.
- 1293 WEST AH & AM Stock (2001), *Histidine kinases and response regulator proteins in two-component signaling systems*, **TRENDS Biochem. Sci.** **26**, 369-376.
- 1294 VARUGHESI KI (2002), *Molecular recognition of bacterial phosphorelay proteins*, **Curr. Opin. Microbiol.** **5**, 142-148.
- 1295 WOLANIN PM *et al.* (2002), *Histidine protein kinases: key signal transducers outside the animal kingdom*, **Genome Biol.** **3**(10).
- 1296 STEWART RC (2010), *Protein histidine kinases: assembly of active sites and their regulation in signaling pathways*, **Curr. Opin. Microbiol.** **13**, 133-141.
- 1297 CASINO P *et al.* (2010), *The mechanisms of signal transduction by two-component systems*, **Curr. Opin. Microbiol.** **20**, 763-771.
- 1298 LANGE R *et al.* (1999), *Domain organization and molecular characterization of 13 two-component systems identified by genome sequencing of Streptococcus pneumoniae*, **Gene** **237**, 223-234.
- 1299 DUTTA R *et al.* (1999), *Histidine kinases: diversity of domain organization*, **Mol. Microbiol.** **34**, 633-640.
- 1300 FOUSSARD M *et al.* (2001), *The molecular puzzle of two-component signaling cascades*, **Microbes Infect.** **3**, 417-424.
- 1301 WUICHET K *et al.* (2010), *Evolution and phyletic distribution of two-component signal transduction systems*, **Curr. Opin. Microbiol.** **13**, 219-225.
- 1302 DUTTA R & M Inouye (2000), *GHKL, an emergent ATPase/kinase superfamily*, **TRENDS Biochem. Sci.** **25**, p.24-28.
- 1303 ARAVIND L & CP Ponting (1999), *The cytoplasmic helical domain of receptor histidine kinase and methyl-accepting proteins is common to many prokaryotic signalling proteins*, **FEMS Microbiol. Lett.** **176**, 111-116.
- 1304 KHOROCHID A & M Ikura (2006), *Bacterial histidine kinase as signal sensor and transducer*, **Int. J. Biochem. Cell Biol.** **38**, 307-312.
- 1305 GALPERIN MY & AN Nikolskaya (2007), *Identification of sensory and signal transducing domains in two-component signaling systems*, **Meth. Enzymol.** **422**, 47-74.
- 1306 GALPERIN MY *et al.* (2001), *Novel domains of the prokaryotic two-component signal transduction systems*, **FEMS Microbiol. Lett.** **203**, 11-21.
- 1307 KIM D & Forst S (2001), *Genomic analysis of the histidine kinase family in bacteria and archaea*, **Microbiology** **147**, 1197-1212.
- 1308 ALM E *et al.* (2006), *The evolution of two-component systems in bacteria reveals different strategies for niche adaptation*, **PLoS Comput. Biol.** **2**(11), e143&e144.
- 1309 KIM S *et al.* (2010), *Identification and classification of a two-component system based on domain structures in bacteria and differences in domain structure between gram-positive and gram-negative bacteria*, **Biosci. Biotechnol. Biochem.** **74**, 716-720.
- 1310 KORETKE KK *et al.* (2000), *Evolution of two-component signal transduction*, **Mol. Biol. Evol.** **17**, 1956-1970.
- 1311 *Protein kinase, 2.7.1.37 en Class 2 Transferases VIII, Springer Handbook of Enzymes*, Schomburg D & Schomburg I (edit), Vol. 35, USA, 2007, p. 423-441.

- 1312 MACEK B *et al.* (2007), *The serine/threonine/tyrosine phosphoproteome of the model bacterium Bacillus subtilis*, **Mol. Cell Proteomics** **6**, 697-707.
- 1313 MACEK B *et al.* (2008), *Phosphoproteome analysis of E. coli reveals evolutionary conservation of bacterial Ser/Thr/Tyr phosphorylation*, **Mol. Cell Proteomics** **7**, 299-307.
- 1314 MOLLE V *et al.* (2010), *An improved method to unravel phosphoacceptors in Ser/Thr protein kinase-phosphorylated substrates*, **Proteomics** **10**, 3910-3915.
- 1315 RAJAGOPAL L *et al.* (2003), *A eukaryotic type serine/threonine kinase and phosphatase in Streptococcus agalactiae reversibly phosphorylate an inorganic pyrophosphatase and affect growth, cell segregation, and virulence*, **J. Biol. Chem.** **278**, 14429-14441.
- 1316 RANTANEN MK *et al.* (2006), *Crystallization and preliminary crystallographic analysis of two Streptococcus agalactiae proteins: the family II inorganic pyrophosphatase and the serine/threonine phosphatase*, **Acta Cryst. F** **62**, p. 891&892.
- 1317 RANTANEN MK *et al.* (2007), *Structure of the Streptococcus agalactiae family II inorganic pyrophosphatase at 2.80 Å resolution*, **Acta Cryst. D** **63**, p.738&730.
- 1318 RAJAGOPAL L *et al.* (2006), *Regulation of cytotoxin expression by converging eukaryotic-type and two-component signalling mechanisms in Streptococcus agalactiae*, **Mol. Microbiol.** **62**, 941-957.
- 1319 LIN WJ *et al.* (2009), *Threonine phosphorylation prevents promoter DNA binding of the Group B Streptococcus response regulator CovR*, **Mol. Microbiol.** **71**, 1477-1495.
- 1320 SILVESTRONI A *et al.* (2009), *Identification of serine/threonine kinase substrates in the human pathogen group B streptococcus*, **J. Proteome Res.** **8**, 2563-2574.
- 1321 VERON M *et al.* (1993), *A conserved helix motif complements the protein kinase core*, **PNAS** **90**, 10618-10622.
- 1322 TAYLOR SS & Radzio-Andzelm E (1994), *Three protein kinase structures define a common motif*, **Structure** **2**, 345-355.
- 1323 OWEN DJ *et al.* (1995), *Two structures of the catalytic domain of phosphorylase kinase: an active protein kinase complexed with substrate analogue and product*, **Structure** **3**, 467-482.
- 1324 HANKS SK & T Hunter (1995), *The eukaryotic protein kinases 6. The eukaryotic protein kinase superfamily: kinase (catalytic) domain structure and classification*, **FASEB J.** **9**, 576-596.
- 1325 LEONARD CJ *et al.* (1998), *Novel families of putative protein kinases in bacteria and archaea: evolution of the "eukaryotic" protein kinase superfamily*, **Genome Res.** **8**, 1038-1047.
- 1326 TAYLOR SS *et al.* (1999), *Catalytic subunit of cyclic AMP-dependent protein kinase structure and dynamics of the active site cleft*, **Pharmacol. Ther.** **82**, 133-141.
- 1327 KRUPA A & N Srinivasan (2005), *Diversity in domain architectures of Ser/Thr kinases and their homologues in prokaryotes*, **BMC Genomics** **6**(129).
- 1328 SHCEEFF ED & PE Bourne (2005), *Structural evolution of the protein kinase-like superfamily*, **PLoS Comput. Biol.** **1**(5), e49.
- 1329 TYAGI N *et al.* (2010), *A framework for classification of prokaryotic protein kinases*, **PLoS One** **5**(5), e10608.
- 1330 MARTIN J *et al.* (2010), *Classification of protein kinases on the basis of both kinase and non-kinase regions*, **PLoS One** **5**(9), e12460.
- 1331 KOONIN EV (1994), *Conserved sequence pattern in a wide variety of phosphoesterases*, **Prot. Sci.** **3**, 356-358.
- 1332 KLABUNDE T *et al.* (1996), *Mechanism of Fe(III)-Zn(II) purple acid phosphatase based on crystal structures*, **J. Mol. Biol.** **259**, 737-748.
- 1333 ARAVIND L & EV Koonin (1998), *Phosphoesterase domain associated with DNA polymerases of diverse origins*, **Nucl. Acids. Res.** **26**, 3746-3752.
- 1334 KNÖFEL T & N Sträter (1999), *X-ray structure of the Escherichia coli periplasmic 5'-nucleotidase containing a dimetal catalytic site*, **Nature Struct. Biol.** **6**, p.448-452.
- 1335 KNÖFEL T & N Sträter (2001), *Mechanism of hydrolysis of phosphate esters by the dimetal center of 5'-nucleotidase based on crystal structures*, **J. Mol. Biol.** **309**, 239-254.
- 1336 SCHULTZ-HEINENBROK (2004), *Trapping a 96° domain rotation in two distinct conformations by engineered disulfide bridges*, **Prot. Sci.** **13**, 1811-1822.
- 1337 SHIN DH (2008), *Structural and enzymatic characterization of DR1281: A calcineurin-like phosphoesterase from Deinococcus radiodurans*, **Proteins** **15**, 1000-1009.
- 1338 SCHENK G *et al.* (2008), *Crystal structures of a purple acid phosphatase, representing different steps of this enzyme's catalytic cycle*, **BMC Struct. Biol.** **8**(6).
- 1339 KEPPEPITOLA N & S Shuman (2008), *A phosphate-binding histidine of binuclear metallophosphodiesterase enzymes is a determinant of 2',3'-cyclic nucleotide phosphodiesterase activity*, **J. Biol. Chem.** **283**, 30942-30949.
- 1340 SHI Y (2009), *Serine/threonine phosphatases: mechanism through structure*, **Cell** **139**, 468-484.
- 1341 <http://www.ncbi.nlm.nih.gov/Structure/cdd/cddsrv.cgi?uid=cl00160&spf=1>

- 1342 VUORIO R *et al.* (1994), *The novel hexapeptide motif found in the acyltransferase LpxA and LpxD of a lipid A biosynthesis is conserved in various bacteria*, **FEBS Lett.** **337**, 289-292.
- 1343 BEAMAN TW *et al.* (1998), *Structure of the hexapeptide acetyltransferase from Pseudomonas aeruginosa*, **Biochemistry** **37**, 6689-6696.
- 1344 PARISI G & J Echave (2004), *The structurally constrained protein evolution model accounts for sequence patterns of the L₁H superfamily*, **BMC Evol.** **4**(41).
- 1345 WENZEL CQ *et al.* (2005), *Evidence that WbpD is a acyltransferase superfamily and an important protein for O-antigen biosynthesis in Pseudomonas aeruginosa PAO1*, **Mol. Microbiol.** **57**, 1288-1303.
- 1346 CUNHA RLOR *et al.* (2009), *A glimpse on biological activities of tellurium compounds*, **An. Acad. Bras. Cienc.** **81**, 393-407.
- 1347 CHASTEEN TG *et al.* (2009), *Tellurite: history, oxidative stress, and molecular mechanisms*, **FEMS Microbiol. Rev.** **33**, 820-832.
- 1348 TAYLOR DE (1999), *Bacterial tellurite resistance*, **TRENDS Microbiol.** **7**, 111-115.
- 1349 TAYLOR DE *et al.* (1994), *Location of a potassium tellurite resistance operon (tehA tehB) within the terminus of Escherichia coli K12*, **J. Bacteriol.** **176**, 2740-2742.
- 1350 DYLLICK-BREZINGER M *et al.* (2000), *The role of cyteine residues in tellurite resistance mediated by the TehAB determinans*, **Biohem. Biophys. Res. Commun.** **277**, 394-400.
- 1351 LIU M & DE Taylor (1999), *Characterization of Gram-positive tellurite resistance encoded by the Streptococcus pneumoniae tehB gene*, **FEMS Microbiol. Lett.** **174**, 385-392.
- 1352 WHITBY PW *et al.* (2010), *Characterization of the Haemophilus influenzae tehB gene and its role in virulence*, **Microbiology** **156**, 1188-1200.
- 1353 LIU M *et al.* (2000), *Escherichia coli TehB requires S-AdoMet cofactor to mediate tellurite resistance*, **J. Bacteriol.** **182**, 6509-6513.
- 1354 CHOUDHURY HG & K Beis (2010), *Crystallization and initial X-ray diffraction analysis of the tellurite-resistance S-adenosyl-L-methionine TehB from Escherichia coli*, **Acta Cryst.** **F66**, 1496-1499.
- 1355 CSOTONYI JT *et al.* (2006), *Anaerobic respirations on tellurate and other metalloids in bacteria from hydrothermal vent-fields in the Eastern Pacific Ocean*, **Appl. Environ. Microbiol.** **72**, p.4950.
- 1356 BAESMAN SM *et al.* (2007), *Formation of tellurium nanocrystals during anaerobic growth of bacteria that use Te oxyanions as respiratory electron acceptors*, **Appl. Environ. Microbiol.** **73**, p.2135.
- 1357 OLLIVIER PRL *et al.* (2008), *Volatilization and precipitation of tellurium by anaerobic, tellurite-resistant marine microbes*, **Appl. Environ. Microbiol.** **74**, p.7163&7164.
- 1358 BA LA *et al.* (2010), *Tellurium: an element with great biological potency and potential*, **Org. BioMol. Chem.** **8**, 4203-4216.
- 1359 MAVRODI DV *et al.* (2006), *Phenazine compounds in fluorescent Pseudomonas spp. biosynthesis and regulation*, **Annu. Rev. Phytopathol.** **44**, 417-445.
- 1360 SCHNEEMANN I *et al.* (2011), *Genetic approach for the fast discovery of phenazine producing bacteria*, **Mar. Drugs** **9**, 772-789.
- 1361 McDONALD M *et al.* (2001), *Phenazine biosynthesis in Pseudomonas fluorescens: branchpoint from the primary shikimate pathway and role of phenazine-1,6-dicarboxylic acid*, **J. Am. Chem. Soc.** **123**, 9459-9460.
- 1362 MAVRODI D *et al.* (2004), *The purification, crystalization and preliminary and preliminary structural characterization of PhzF, a key enzyme in the phenazine-biosynthesis pathway from Pseudomonas fluorescens 2-79*, **Acta Cryst D60**, 184-186.
- 1363 PARSONS JF *et al.* (2004), *Structure and function of the phenazine biosynthesis protein PhzF from Pseudomonas fluorescens 2-79*, **Biochemistry** **43**, 12427-12435.
- 1364 BLANKENFELDT W *et al.* (2004), *Structure and function of the phenazine biosynthetic protein PhzF from Pseudomonas fluorescens*, **PNAS** **101**, 16431-16436.
- 1365 MENTEL M *et al.* (2009), *Of two make one: the biosynthesis of phenazines*, **Chem. BioChem.** **10**, 2295-2304.
- 1366 MAVRODI DV *et al.* (1998), *A seven-gene locus for synthesis of phenazine-1-carboxylic acid by Pseudomonas fluorescens 2-79*, **J. Bacteriol.** **180**, 2541-2548.
- 1367 PIDUGU LS *et al.* (2009), *Analysis of proteins with the 'hot dog' fold: prediction of function and identification of catalytic residues of hypothetical proteins*, **BMC Struct. Biol.** **9**(37).
- 1368 FITZPATRICK DA (2009), *Lines of evidence for horizontal gene transfer of a phenazine producing operon into multiple bacterial species*, **J. Mol. Evol.** **68**, 171-185.
- 1369 MAVRODI DV *et al.* (2010), *Diversity and evolution of the phenazine biosynthesis pathway*, **Appl. Environ. Microbiol.** **76**, 866-879.
- 1370 EISENSTEIN E *et al.* (2000), *Biological function made crystal clear-annotation of hypothetical proteins via structural genomics*, **Curr. Opin. Biotechnol.** **11**, 25-30.
- 1371 ROBERTS RJ *et al.* (2004), *Identifying protein function-a call for community action*, **PLoS Biol.** **2**, 293&294.

- 1372 GALPERIN MY & EV Koonin (2004), 'Conserved hypothetical' proteins: prioritization of targets for experimental study, **Nucleic Acids Res.** **32**, 5452-5463.
- 1373 DOERKS T *et al.* (2004), Functional clues for hypothetical proteins based on a genomic context analysis in prokaryotes, **Nucleic Acids Res.** **32**, 6321-6326.
- 1374 SIVASHANKARI S & P Shanmughavel *et al.* (2006), Functional annotation of hypothetical proteins – A review, **Bioinformatics** **1**, 335-338.
- 1375 BUSH K (2013), The ABCD's of β -lactamase nomenclature, **J. Infect. Chemother.** **19**, 549-559.
- 1376 KUCHTA K *et al.* (2009), Comprehensive classification of nucleotidyltransferase fold proteins: identification of novel families and their representatives in human, **Nucleic Acids Res.** **37**, 7701-7714.
- 1377 COZIER GE *et al.* (2004), Membrane targeting by pleckstrin homology domains, **Curr. Top. Microbiol. Immunol.** **282**, p.49.
- 1378 SCHEFFZEK K & S Welti (2012), Pleckstrin homology (PH) like domains – versatile modules in protein-protein interaction platforms, **FEBS Lett.** **586**, 2662-2673.
- 1379 NARDINI M & BW Dijkstra (1999), \square/\square Hydrolase fold enzymes: the family keeps growing, **Curr. Opin. Struct. Biol.** **9**, 732-737.
- 1380 DUNN G *et al.* (2005), The structure of the C-C bond hydrolase MhpC provides insights into its catalytic mechanism, **J. Mol. Biol.** **346**, 253-265.
- 1381 LI C *et al.* (2008), Catalytic promiscuity in the alpha/beta-hydrolase superfamily: hydroxamic acid formation, C--C bond formation, ester and thioester hydrolysis in the C--C hydrolase family, **ChemBiochem.** **9**, 71-76.
- 1382 HOLMQUIST M (2000), Alpha/beta-hydrolase fold enzymes: structures, functions and mechanisms, **Curr. Prot. Pept. Sci.** **1**, 209-235.
- 1383 HOTELIER T *et al.* (2004), ESTHER, the database \square/\square hydrolase fold superfamily of proteins, **Nucleic Acids Res.** **32**, D145-D147.
- 1384 CARR PD & DL Ollis (2009), Alpha/beta hydrolase fold: an update, **Protein Pept. Lett.** **16**, 1137-1148.
- 1385 BURTS ML *et al.* (2005), EsxA and EsxB are secreted by an ESAT-6-like system that is required for the pathogenesis of Staphylococcus aureus infections, **PNAS** **102**, 1169-1174.
- 1386 DESVAUX M & M Hébraud (2006), The protein secretion systems in Listeria: inside out bacterial virulence, **FEMS Microbiol. Rev.** **30**, 774-805.
- 1387 BURTS ML *et al.* (2008), EsaC substrate for the ESAT-6 secretion pathway and its role in persistent infections of Staphylococcus aureus, **Mol. Microbiol.** **69**, 736-746.
- 1388 DESVAUX M *et al.* (2009), Secretion and subcellular localizations of bacterial proteins: a semantic awareness issue, **TRENDS Microbiol.** **17**, 139-145.
- 1389 DESVAUX M *et al.* (2009), Outer membrane translocation: numerical protein secretion nomenclature in question in mycobacteria, **TRENDS Microbiol.** **17**, 338-340.
- 1390 SUCTLIFE IC *et al.* (2011), New insights into the distribution of WXG100 protein secretion systems, **Antonie Van Leeuwenhoek** **99**, 127-131.
- 1391 HOSFIELD DJ *et al.* (1999), Structure of the DNA repair enzyme endonuclease IV and its DNA complex: double-nucleotide flipping at abasic sites and three-metal-ion catalysis, **Cell** **98**, 397-408.
- 1392 MOL CD *et al.* (2000), Abasic site recognition by two apurinic/apyrimidinic endonuclease families in DNA base excision repair: the 3' ends justify the means, **Mutat. Res.** **460**, p.211.
- 1393 GARCIN ED *et al.* (2008), DNA apurinic-apyrimidinic site binding and excision by endonuclease IV, **Nature Struct. Mol. Biol.** **15**, p.515&516.
- 1394 FARBER GK & GA Petsko (1990), The evolution of an \square/\square barrel enzymes, **TRENDS Biochem. Sci.** **15**, 228-234.
- 1395 WIERENGA RK (2001), The TIM-barrel fold: a versatile framework for efficient enzymes, **FEBS Lett.** **492**, 193-198.
- 1396 NAGANO N *et al.* (2002), One fold with many functions: the evolutionary relationships between TIM barrel families based on their sequences, structures and functions. **J. Mol. Biol.** **321**, 741-765
- 1397 VEGA MC *et al.* (2003), Evolutionary markers in the $(\square/\square)_8$ -barrel fold, **Curr. Opin. Chem. Biol.** **7**, 694-701.
- 1398 YEATS C *et al.* (2002), The PASTA domain: a \square -lactam-binding domain, **TRENDS Biochem. Sci.** **27**, 438-440.
- 1399 HUSSAIN H *et al.* (2006), A eukaryotic-type serine/threonine protein kinase is required for biofilm formation, genetic competence, and acid resistance in Streptococcus mutans, **J. Bacteriol.** **188**, p.1628.
- 1400 JONES G & P Dyson (2006), Evolution of transmembrane protein kinases implicated in coordinating remodeling of gram-positive peptidoglycan: inside versus outside, **J. Bacteriol.** **188**, 7470-7476.

- 1401 FERNANDEZ P *et al.* (2006), *The Ser/Thr protein kinase PknB is essential for sustaining mycobacterial growth*, **J. Bacteriol.** **188**, 7778-7784.
- 1402 NARAYAN A *et al.* (2007), *Serine threonine protein kinase of mycobacterail genus: phylogeny to function*, **Physiol. Genomics** **29**, 66-75.
- 1403 SAUVAGE E *et al.* (2008), *The penicillin-binding proteins: structure and role in peptidoglycan biosynthesis*, **FEMS Microbiol. Rev.** **32**, p.234,240,245.
- 1404 PRÉVOTS F *et al.* (1996), *Cloning and sequencing of the novel abortive infection gene *abiH* of Lactococcus lactis ssp. lactis biovar. diacetylactis S94*, **FEMS Microbiol. Lett.** **142**, 295-299.
- 1405 NACLERIO G *et al.* (1996), *Bacillus subtilis spore coat assembly requires *coth* gene expression*, **J Bacteriol.** **178**,4375-80. Erratum en: J. Bacteriol. 178, 6407.
- 1406 ZILÁHO R *et al.* (1999), *Assembly requirements and role of *CotH* during spore coat formation in Bacillus subtilis*, **J. Bacteriol.** **181**, 2631-2633.
- 1407 LAZAREVIC V *et al.* (1992), *Sequencing and analysis of the Bacillus subtilis *lytRABC* divergon: a regulatory unit encompassing the structural genes of the N-acetylmuramoyl-L-alanine amidase and its modifier*, **J. Gen. Microbiol.** **138**,1949-1961.
- 1408 MULKIDJANIAN AY *et al.* (2003), *Survival of the fittest before the beginning of life: selection of the first oligonucleotide-like polymers by UV light*, **BMC Evol. Biol.** **3**:12.
- 1409 MCGINNESS KE & GF Joyce (2003), *In search of an RNA replicase ribozyme*, **Chem. Biol.** **10**, 5-14.
- 1410 YARUS M (2002), *Primordial genetics: phenotype of the ribocyte*, **Annu. Rev. Genet.** **36**, 125-151.
- 1411 ITAYA M (1995), *An estimation of minimal genome size required for life*, **FEBS Lett.** **362**, 257-260.
- 1412 MUSHEGIAN AR & EV Koonin (1996), *A minimal gene set for cellular life derived by comparison of complete bacterial genomes*, **PNAS** **93**, 10268-10273.
- 1413 KOONIN EV & AR Mushegian (1996), *Complete genome sequences of cellular life: glimpses of theoretical evolutionary genomics*, **Curr. Opin. Genet. Dev.** **6**, 757-762.
- 1414 MUSHEGIAN AR (1999), *The minimal genome concept*, **Curr. Opin. Genet. Dev.** **9**, 709-714.
- 1415 FRASER CM (1995), *The minimal gene complement of Mycoplasma genitalium*, **Science** **270**, 397-403
- 1416 HUTCHISON CA *et al.* (1999), *Global transposon mutagenesis and a minimal Mycoplasma genome*, **Science** **286**, 2165-2169.
- 1417 GLASS JI *et al.* (2006), *Essential genes of a minimal bacterium*, **PNAS** **103**, 425-430.
- 1418 GIL R *et al.* (2004) *Determination of the core of a minimal bacterial gene set*, **Microb. Mol. Biol. Rev** **68(3)**, 518-537.
- 1419 HURST LD & C Pál, *Genomic redundancy and dispensability*, Cap.7 en **Evolutionary Genomics and Proteomics**, Sinauer Associates, USA, 2007, p.141-166.
- 1420 RANEA JAG *et al.* (2006), *Protein superfamily evolution and the Last Universal Common Ancestor (LUCA)*, **J. Mol. Evol.** **63**, 513-525.
- 1421 GOTO N *et al.* (2007), *Analysis of invariant sequences in 266 complete genomes*, **Gene** **401**, 172-180.
- 1422 SOBOLEVSKY Y & EN Trifonov (2006), *Protein modules conserved since LUCA*, **J. Mol. Evol.** **63**, 622-634.
- 1423 SOBOLEVSKY Y *et al.* (2013), *Towards functional repertoire of the earliest proteins*, **J. Biomol. Struct. Dyn.** **31**, 1293-1300.
- 1424 GOLDMAN AD *et al.* (2013) *LUCApedia: a database for the study of ancient life*, **Nucleic Acids Res** **41**, D1079-D1082, doi:10.1093/nar/gks1217
- 1425 WOESE C & GE Fox (1977), *The concept of cellular evolution*, **J. Mol. Evol.** **10**, 1-6.
- 1426 VETSIGIAN K *et al.* (2006), *Collective evolution and the genetic code*, **PNAS** **103**, 10696-10701.
- 1427 GOLDENFELD N & C Woese (2007), *Biology's next revolution*, **Nature** **445**, 369.

**NI LA ENERGÍA NI LA MATERIA SE CREAN NI
SE DESTRUYEN, TAN SÓLO SE TRANSFORMAN**

**SÓLO LA VIDA TIENE SENTIDO SI
ES A LA LUZ DE LA EVOLUCIÓN**

SALUDOS DESOBEDIENTES Y ABRAZOS RESISTENTES

iii MENTE FRÍA, AUNQUE LA SANGRE HIERVA !!!

SALUDOS EVOLUTIVOS

GRACIAS