



**Universidad Nacional Autónoma de México**  
Doctorado en Ciencias Biomédicas  
Facultad de Medicina

**Origen y evolución de las principales rutas  
catabólicas**

Tesis

Que para obtener el grado de:  
DOCTOR EN CIENCIAS

Presenta

Mario Rivas Medrano

**Director de Tesis**

Dr. Antonio Eusebio Lazcano-Araujo Reyes  
Facultad de Medicina

**Comité Tutor**

Dr Diego González Halphen  
Instituto de Fisiología Celular  
Dr. Georges Dreyfus Cortés  
Instituto de Fisiología Celular

Ciudad de México. 04 de Junio, 2018



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.

## Agradecimientos

### Institucionales

- 1) Al Programa de Doctorado en Ciencias Biomédicas por su constante apoyo a través de su personal diligente y siempre dispuesto a colaborar para llevar a buen término todas las actividades académicas que los alumnos desarrollamos.
- 2) A la Facultad de Medicina por su respaldo tanto a mi persona como a mi tutor, sin el cual nos hubiera sido muy difícil llevar a cabo la investigación aquí reportada.
- 3) A la Universidad Nacional Autónoma de México por ser mi *alma máter* y fomentar desde el primer momento en que ingrese como alumno de esta casa de estudios el deseo de superación académica.
- 4) Al CONACYT por el apoyo económico del que fui beneficiario durante mis estudios de doctorado con el Número de Becario 255708.

## **Agradecimientos**

### Personales

- 1) Para María Silvina Rivas Medrano por ser un gran ejemplo de superación y por su apoyo incondicional durante todos mis estudios. Gracias por respetar mis decisiones y brindarme tu apoyo en cada paso.
- 2) Para Antonio Eusebio Lazcano-Araujo Reyes por su generosidad intelectual, su entrega para con sus estudiantes y su apoyo incondicional como tutor y como amigo, hasta en los momentos más difíciles.
- 3) Para Arturo Carlos II Becerra Bracho por su voluntad y paciencia inagotable para ayudar a los estudiantes, así como por el apoyo que demuestra para cada uno de nosotros.
- 4) Para César Antonio Martínez Gutiérrez por su apoyo, confianza y amistad. Seguimos en el camino amigo.
- 5) Para los Doctores Diego González Halphen y Georges Dreyfus Cortés por su ayuda para consolidar un proyecto que gracias a sus observaciones, críticas constructivas y guía durante los exámenes tutorales tomo forma y rumbo.
- 6) Para mis compañeros del laboratorio de Origen de la Vida por fomentar un ambiente de cordialidad y compañerismo, pero sobre todo por escucharme durante los seminarios y hacer observaciones, preguntas y críticas sin las cuales no se puede mejorar.

## Resumen

Comparados con la cantidad de estudios que se han hecho sobre el origen y la evolución del anabolismo, pocos han sido los esfuerzos enfocados al estudio del origen y la evolución temprana de las rutas catabólicas. En el presente trabajo se emplearon técnicas bioinformáticas de genómica comparada así como la información bioquímica disponible en bases de datos públicas, sobre las enzimas que degradan a los monosacáridos glucosa y ribosa, las nucleobases adenina, guanina, citosina, uracilo y timina, así como los 20 aminoácidos canónicos que integran las proteínas contemporáneas. Usando criterios como la dependencia de las enzimas al oxígeno molecular, su distribución y el nivel de conservación a lo largo de la filogenia se propuso un escenario evolutivo sobre el surgimiento y evolución temprana de dichas rutas, así como la antigüedad relativa de cada una de ellas. Nuestros resultados sugieren la presencia de formas ancestrales de algunas rutas catabólicas en el último ancestro común (LCA). Ejemplos de lo anterior son las degradaciones enzimáticas de la ribosa y de las purinas adenina y guanina. Y aunque hoy en día los organismos poseen una gran diversidad de rutas para degradar a los aminoácidos, es de esperarse que no todas sean igualmente antiguas. Ejemplos de rutas catabólicas de aminoácidos que pudieron estar presentes en el LCA son las degradaciones enzimáticas de glutamato, metionina y treonina.

## ÍNDICE

Introducción General.....	6
Metodología General.....	10
Capítulo 1: Sobre la evolución temprana de las rutas catabólicas: un estudio de genómica comparada. I. Los casos de las degradaciones enzimáticas de glucosa, ribosa y las bases nitrogenadas adenina, guanina, citosina, uracilo y timina.....	18
Capítulo 2: Sobre la evolución temprana de las rutas catabólicas: un estudio de genómica comparada. II. Los casos de las degradaciones enzimáticas de los veinte aminoácidos de las proteínas contemporáneas.....	43
Capítulo 3: Publicaciones sobre la evolución temprana de la autotrofia y las implicaciones evolutivas de las características estructurales del RNA.....	88
Referencias.....	105
Anexo 1 – Catálogo del Catabolismo.....	111

## Introducción General

La importancia de las rutas catabólicas reside no sólo en el papel crucial que desempeñan como fuentes de carbono, nitrógeno y otros elementos, sino también en la extracción y transferencia de energía desde las moléculas consideradas de alto contenido energético hacia otras moléculas biológicamente más relevantes como el ATP o el GTP, así como el aprovechamiento y acumulación del poder reductor en moléculas como el NADH o el FADH. Muchos son los casos en los que el producto final de una ruta catabólica es un metabolito de los llamados de encrucijada, que representan el sustrato de rutas biosintéticas de aminoácidos, ácidos nucleicos, ácidos grasos y otras moléculas de importancia biológica. Es así que las rutas degradativas representan en los organismos actuales una forma de emplear la materia orgánica que los rodea como una fuente de los elementos esenciales, necesarios para su supervivencia, por lo tanto es muy probable que también jugarán un papel igualmente importante en la evolución temprana de la vida, cuando se consolidaba la intrincada red de reacciones bioquímicas que hoy en día llamamos metabolismo.

La primera hipótesis sobre el origen del catabolismo de la que se tiene conocimiento fue propuesta por Oparin (1924, 1938). En su hipótesis de un origen heterotrófico de la vida, Oparin propone a la glucólisis como la forma de fermentación más antigua y que podría sostener a un hipotético organismo heterótrofo anaerobio primitivo totalmente dependiente del material orgánico presente en el ambiente gracias a una serie de síntesis prebióticas y posterior acumulación de compuestos orgánicos que tuvieron lugar después de la formación de la tierra (Oparin 1924, 1938; Lazcano 2016). El escenario evolutivo para el surgimiento de la vida propuesto por Oparin se sustenta en premisas como la universalidad de la fermentación, específicamente la glucólisis, que le hace suponer un origen heterotrófico de la vida. Otro aspecto claramente enmarcado en un escenario Darwinista es su propuesta de un proceso de evolución pre-celular

lento y gradual que ocurriría paralelo a la acumulación de los primeros polímeros orgánicos (Lazcano 1989; Lazcano y Miller 1999; Lazcano 2016).

Más de veinte años después Krebs y Kornberg (1957; Krebs 1981) sugirieron que si el escenario de Oparin fuese correcto, entonces la segunda ruta catabólica que debió aparecer tuvo que ser la ruta de las pentosas fosfato, dado el alto grado de interconexión con la glucólisis, como queda claro cuando se toman en cuenta los metabolitos de encrucijada que se comparten entre ambas rutas. Ejemplos de lo anterior son el gliceraldehído 3-fosfato, la fructosa 1,6-bifosfato, la fructosa 6-fosfato o la glucosa 6-fosfato. Posterior a estas propuestas Clarke y Elsdén (1980), usando premisas como la facilidad con la que se sintetizan aminoácidos de manera abiótica y la relativa simplicidad con la que pueden ser fermentados, sugirieron un sistema primitivo basado en reacciones de óxido-reducción de alta eficiencia energética en el que glicina y prolina representan los agentes oxidantes asociados a una fosforilación a nivel sustrato que se llevaría a cabo bajo condiciones anaerobias, y concluyeron que derivadas de este tipo de sistemas puramente químicos resultaba fácil vislumbrar un escenario para la aparición gradual de rutas catabólicas completas.

En este mismo orden de ideas es que Cunchillos y Lecointre (2002) sugirieron que el catabolismo de los aminoácidos alifáticos muy posiblemente representó la primer fuente de energía que utilizarían las “primeras protocélulas”, y con esta hipótesis en mente desarrollaron un estudio cladístico en el que especulan sobre la antigüedad y el surgimiento gradual de los diferentes tipos de desaminasas, aminotransferasas y carboxilasas que participan en la degradación enzimática de dichos aminoácidos. Basados en lo anterior, propusieron una serie de épocas en las que surgieron las rutas degradativas. Así en una primer época aparecieron los catabolismos de aspartato, asparagina, glutamato, glutamina, valina, isoleucina, cisteína y alanina, seguida de una segunda época en la que el catabolismo de prolina y arginina debió aparecer, hasta que finalmente en la última fase surgieron las degradaciones enzimáticas de metionina, treonina, serina,



glicina, leucina y lisina, así como las interacciones enzimáticas que conectan estas rutas catabólicas con el ciclo de Krebs (Cunchillos y Lecointre 2002).

La más reciente propuesta sobre el origen y la antigüedad de las degradaciones enzimáticas de los componentes celulares como hidratos de carbono, aminoácidos y las bases de los ácidos nucleicos fue desarrollada por Schönheit et al (2016). Ellos supusieron un escenario autotrófico para el surgimiento de la vida, y sugirieron que las moléculas orgánicas con mayor probabilidad de ser degradadas por los primeros organismos heterótrofos tuvieron que haber sido aquellas provenientes de los componentes celulares de organismos autótrofos preexistentes, y que las maquinarias celulares de estos primeros autótrofos sufrieron adaptaciones evolutivas que las adecuaron hasta convertirlas en las rutas catabólicas que se asemejan a las actuales rutas degradativas del tipo fermentativo que exhiben organismos contemporáneos como los pertenecientes al género *Clostridium* y que muy probablemente representan las primeras formas de heterotrofia. Al asumir un origen autotrófico de la vida, la posibilidad de degradaciones más tempranas se elimina por completo, ya sean catalizadas por enzimas o por ribozimas que actuarían sobre moléculas orgánicas presentes en la tierra primitiva como consecuencia de síntesis prebióticas o de su acumulación debida al impacto de meteoritos sobre la superficie terrestre.

No obstante los argumentos en contra (Caetano-Anollés, Kim y Mittenthal 2007; Caetano-Anollés y Caetano-Anollés 2013), hasta ahora no hay evidencia contundente que sugiera la existencia de una correlación directa entre los compuestos que pudieron estar presentes en la tierra primitiva y la intrincada red de reacciones bioquímicas que integran el metabolismo de los organismos contemporáneos. Muchos han sido los intentos de extrapolar las rutas anabólicas actuales, cuya presencia en el último ancestro común es altamente probable, hacia el origen de la vida. Ello implicaría que el origen del metabolismo estaría mucho más cercano a la época en la que pudo existir el último ancestro común que a la época en la que se cree aparecieron los primeros seres vivos (Lazcano y

Miller 1999). Por otro lado, no ha habido muchos intentos por entender el papel que jugaron los compuestos de la sopa primitiva en el desarrollo de las primeras rutas catabólicas (Keefe, Lazcano, y Miller 1995). Es razonable pensar que los compuestos resultantes de los experimentos de síntesis prebiótica y los presentes en meteoritos son buenos candidatos para fungir como sustrato de las primeras rutas catabólicas. No obstante, resulta mucho más prudente limitar las extrapolaciones a periodos en los cuales la biosíntesis de proteínas ya había aparecido, sobre todo si se emplean metodologías basadas en la distribución y comparación de proteínas pertenecientes a rutas metabólicas actuales.

También es importante señalar que incluso en el metabolismo contemporáneo un número importante de reacciones no enzimáticas ocurren espontáneamente, y es posible que este tipo de reacciones hayan jugado un papel relevante en la consolidación de las primeras rutas metabólicas (Lazcano y Miller 1999; Keller, Piedrafita, y Ralser 2015). Por lo tanto, es razonable suponer que las rutas metabólicas ancestrales tenían una catálisis semi-enzimática. En otras palabras, el enfoque que se propone en el presente estudio no supone una conexión directa con las rutas metabólicas ancestrales que pudieron haber existido en períodos previos como el mundo del RNA o de pre-RNA, donde las reacciones pudieron haber sido procesos completamente libres de enzimas dependientes de catalizadores inorgánicos como cationes metálicos o incluso facilitados por ribozimas (Becerra et al. 2007).

Tradicionalmente el metabolismo ha sido clasificado en anabolismo y catabolismo para su estudio, aunque en realidad en un gran número de casos la línea que separa a las rutas de degradación de las de síntesis no es perfectamente clara, especialmente cuando se intenta reconstruir la historia evolutiva de las diferentes rutas. Ejemplos de lo anterior son el ciclo reductivo de los ácidos tricarbóxicos, que se lleva a cabo gracias a un gran número de enzimas del ciclo oxidativo que catalizan las mismas reacciones pero en sentido inverso, así como la ruta de degradación de la lisina que inicia usando la mitad de

las enzimas de la ruta biosintética (Michal 1998). En términos evolutivos, es por lo tanto posible que una ruta que se usaba para sintetizar una molécula ahora se utilice para degradarla o viceversa, lo cual vuelve aún más complicada una reconstrucción exacta de los procesos que dieron origen a la compleja red de reacciones bioquímicas que vemos hoy en día.

En el presente trabajo exploramos la distribución filogenética y la evolución molecular de cada una de las enzimas de las rutas catabólicas de los monosacáridos glucosa y ribosa, las nucleobases del material genético adenina, guanina, citosina, uracilo y timina, así como de los veinte aminoácidos que integran las proteínas contemporáneas. También hemos considerado la dependencia al oxígeno molecular de las enzimas catabólicas como un carácter diagnóstico de su antigüedad relativa, ya que sabemos que la acumulación del oxígeno molecular en la atmósfera del Precámbrico medio cambió por completo y para siempre la química del planeta, haciendo posible la aparición de nuevas reacciones y metabolismos más eficientes en términos energéticos (Goldfine 1965; Canfield 2005). Este enfoque nos ha permitido establecer una datación relativa de las enzimas y las rutas catabólicas que muy probablemente se encontraban presentes en el último ancestro común de los seres vivos (LCA, por sus siglas en inglés), así como de las enzimas y rutas catabólicas que se consolidaron de manera posterior a la acumulación del oxígeno molecular en la atmósfera terrestre, en un periodo que probablemente ocurrió hace aproximadamente 2800-2300 millones de años (Canfield 2005; Lyons, Reinhard, y Planavsky 2014).

## **Metodología General**

**Identificación de las enzimas y rutas catabólicas:** Las enzimas, los sustratos y los productos de las rutas catabólicas bajo estudio fueron identificadas contrastando la información del “Biochemical Pathways Atlas” (Michal y

Schomburg 2012), la base de datos de metabolismos “Kyoto Encyclopedia of Genes and Genomes” (KEGG, Kanehisa y Goto 2000; Kanehisa et al. 2014), la base de datos “The Metabolic Pathway Database” (MetaCyc, Caspi et al. 2014) y la base de datos “The Comprehensive Enzyme Information System” (BRENDA, Schomburg et al. 2013). La información derivada de las fuentes anteriores fue en algunos casos complementada con la obtenida de “The Organic Chemistry of Biological Pathways” (McMurry y Begley 2005). Existieron casos en los que la diferencia entre una enzima anabólica y una enzima catabólica no era del todo evidente, dado que era clasificada tanto en la síntesis como en la degradación de un determinado compuesto. Por lo tanto, solo cuando se localizaba una referencia que claramente indicara una actividad degradativa, se anotaba a dicha enzima como exclusiva del catabolismo, mientras que el resto permanecían anotadas como involucradas tanto en síntesis como en degradación de acuerdo a la información de las bases de datos KEGG, MetaCyc y/o BRENDA. En los casos en los que no existe información definitiva sobre la dirección de la reacción que cataliza una enzima, se le incluía dentro del análisis si su actividad catalítica se consideraba indispensable para llevar a cabo la degradación de la molécula bajo estudio, especialmente cuando no existía ninguna otra enzima que tuviera una actividad equivalente reportada en las bases de datos. Todas las reacciones que integran las rutas catabólicas del estudio, los nombres y los números EC de las enzimas que las catalizan, así como el uso de cofactores y/o coenzimas fueron reunidos en un catálogo que se presenta como el Anexo 1, mismo que también se encuentra disponible en la siguiente liga <https://goo.gl/RQR5tC>.

**Reclutamiento de las secuencias de aminoácidos:** Las secuencias correspondientes a la estructura primaria de las enzimas catabólicas fueron colectadas haciendo uso de la clasificación propuesta por la “Enzyme Commission” (EC). El número EC fue empleado para buscar en la base de datos UniProt (The UniProt Consortium 2012), lo que nos permitió reunir tanto secuencias no redundantes anotadas manualmente del sistema UniProtKB/Swiss-Prot (The UniProt Consortium 2014) como secuencias anotadas automáticamente

del sistema UniProtKB/TrEMBL (Magrane y Consortium 2011). En los casos en los que no fue posible identificar alguna secuencia mediante el uso del número EC en la base de datos de UniProt, se realizó una selección aleatoria de una secuencia representativa en Arqueas, Bacterias y Eucariontes (cuando estaban disponibles) que fueron extraídas de la base de datos KEGG. Posteriormente con estas secuencias se realizó una búsqueda de homólogos mediante el algoritmo BLAST (Altschul et al. 1990) usando el sitio web del “National Center for Biotechnology Information” (NCBI, Geer et al. 2010) contra la base de secuencias no redundante. Las secuencias reclutadas de esta manera incluyen a los primeros cien resultados con los mejores valores de “E” y con una cobertura mínima del 70% en la longitud del alineamiento.

**Análisis de las secuencias de aminoácidos:** Una vez que las secuencias de las enzimas fueron colectadas, estas se seleccionaron para eliminar la redundancia de las bases de datos usando el programa CD-HIT (Fu et al. 2012) con los valores  $C=0.7$  y  $N=5$ . Posteriormente se realizó la identificación de los dominios funcionales Pfam (Punta et al. 2012) en las secuencias restantes usando el programa HMMER3 (Mistry et al. 2013) mediante la subrutina hmmer3 en contra de la base de datos Pfam (versión 27.0) de acuerdo a las instrucciones del manual de HMMER. Lo anterior permitió hacer evidente que existe más de una arquitectura de dominios entre las secuencias anotadas bajo un mismo número EC, por lo que las secuencias con arquitecturas similares se agruparon en archivos separados antes de construir los perfiles de búsqueda. Como resultado de lo anterior, más de un perfil de búsqueda fue construido para las secuencias que estaban anotadas bajo un mismo número EC, lo que a su vez permitió llevar a cabo búsquedas independientes con cada uno de los perfiles generados de esta manera y así cubrir la diversidad de arquitecturas de proteínas que se observó.

**Construcción de los perfiles de búsqueda:** El alineamiento múltiple de secuencias es un requisito antes de construir los perfiles de búsqueda y se obtuvo mediante el programa muscle v3.8.31 (Edgar 2004a, 2004b) que fue alimentado

con las secuencias previamente seleccionadas por similitud de arquitecturas de dominios proteicos. La construcción de los perfiles se realizó implementando la rutina hmmerbuild del programa HMMER de acuerdo al manual. Ni en la alineación ni en la construcción de los perfiles de búsqueda se implementaron parámetros que pudieran modificar el comportamiento nativo de los programas.

**Asignando límites para la búsqueda de homólogos:** Con los perfiles obtenidos se llevó a cabo la búsqueda de secuencias homólogas en contra de las secuencias sin alinear, que se emplearon para crear cada perfil, utilizando la rutina hmmsearch de manera local. Se identificó así la secuencia con el peor valor de “E” cuando se enfrentan contra el perfil que deriva del conjunto de secuencias con las que fue construido. Lo anterior permitió reclutar esas secuencias y a su vez incluirlas como “las secuencias límite” de cada uno de los perfiles de búsqueda dentro de la base local de secuencias proteicas en lugar de establecer un valor de corte general arbitrario. Por lo tanto, los resultados de presencia/ausencia de homólogos incluyen exclusivamente las secuencias que obtuvieron valores iguales o mejores a los de las secuencias límite.

**Composición de la base de datos de secuencias de aminoácidos:** La base local de secuencias proteicas se integró exclusivamente con secuencias provenientes de organismos completamente secuenciados. La base local de proteomas se integró con 97 Arqueas, 1201 Bacterias y 146 Eucariontes cuyos archivos de secuencias fueron descargados de la base de datos KEGG durante el año 2011. El grupo de Arqueas estuvo integrado principalmente por miembros pertenecientes a los grupos Euriarqueota y Crenarqueota, aunque los grupos Thaumarqueota, Nanoarqueota y Korarqueota también fueron incluidos en aras de exhaustividad, aunque estos contaban con un número muy limitado de genomas que consideramos no reflejaban la diversidad de estos grupos al momento de integrar la base de datos. La base de datos de Bacterias se integró por miembros de los grupos  $\alpha$ -proteobacteria,  $\beta$ -proteobacteria,  $\delta$ -proteobacteria,  $\epsilon$ -proteobacteria,  $\gamma$ -proteobacteria, Firmicutes, Tenericutes, Actinobacteria,

Chlamidia, Spiroquetas, Acidobacteria, Bacteroidetes, Fusobacteria, Verrucomicrobia, Planctomicetes, Cianobacteria, Bacterias Verdes-Sulfurosas, Bacterias Verdes-no-sulfurosas, Deinococcus-Thermus y Bacterias Hipertermófilas. La base de Eucariontes se integró por especies de los grupos de Vertebrados, Artrópodos, Nematodos, Cnidarios, así como Eudicotiledónias, Monocotiledónias, Algas, Ascomicetes, Basidiomicetes, Coanoflagelados, Amebas, Alveolados, *Euglenas* y Diatomeas.

**Estableciendo la antigüedad relativa de las rutas:** La datación relativa de las rutas catabólicas analizadas en el presente trabajo se realizó utilizando dos grandes criterios. En primer lugar se utilizó la dependencia de las enzimas al oxígeno molecular, es decir, si una enzima cataliza una reacción catabólica y en dicha reacción se requiere del oxígeno molecular como sustrato y a su vez dicha reacción es indispensable para la ruta de forma que se produzca un determinado metabolito y no existe otra enzima (o conjunto de enzimas) que pueda llevar a cabo dicha reacción sin el uso del oxígeno, entonces consideramos que dicha ruta catabólica evolucionó en su forma actual después del enriquecimiento de oxígeno en la atmósfera de la Tierra. El catálogo del catabolismo que aparece en el Anexo 1 del presente trabajo contiene detalles sobre la dependencia al oxígeno molecular de las enzimas aquí analizadas. Las rutas catabólicas con enzimas que no dependen de oxígeno molecular, o que cuentan con una contraparte anoxigénica a las reacciones dependientes de oxígeno fueron clasificadas como universales o no universales dependiendo de la distribución de sus enzimas en los grupos que integran a los dominios de Arqueas y Bacterias (Forterre y Gribaldo 2010; Embley y Williams 2015). La presencia de las enzimas en todos los organismos que integran la base de datos no se definió como un criterio necesario para establecer la universalidad de las enzimas. En cambio, consideramos que la distribución debe ser representativa entre la mayoría de los grupos de Arqueas y Bacterias, lo que implica que las enzimas de una ruta específica deben estar presentes en por lo menos la mitad de los grupos que integran a los dos grandes dominios de la vida, Arqueas y Bacterias. Para que un grupo fuera considerado como positivo a la

presencia de una ruta, las enzimas que la integran debieron estar presentes en por lo menos una tercera parte de los organismos listados dentro del grupo; dicho valor de corte se definió de manera empírica basados en las filogenias que se desarrollaron para detectar eventos de transferencia horizontal de las enzimas entre los grandes grupos de Arqueas y Bacterias. En aras de exhaustividad también se buscaron homólogos dentro de los integrantes de los grandes grupos de Eucariontes y los resultados se reportaron de la misma manera para su evaluación. Este procedimiento nos permitió distinguir entre las enzimas y las rutas que más probablemente se encontraban presentes en el último ancestro común (Last Common Ancestor, LCA), y aquellas que evolucionaron después de la divergencia de los dos grandes dominios procariontes. Las tablas de distribución de las enzimas que integran las rutas catabólicas analizadas en el presente trabajo pueden ser localizadas en la siguiente liga <https://goo.gl/cZtM2h>.

**Relaciones evolutivas entre las enzimas:** Con el fin de rastrear las posibles relaciones evolutivas de las enzimas que integran las rutas catabólicas bajo estudio, a nivel de estructura primaria se utilizó una muestra aleatoria de las secuencias de las colecciones previamente integradas para cada número EC, así como para cada arquitectura de dominios identificada previamente. De esta manera se integró un archivo que contenía por lo menos una secuencia representativa de cada número EC correspondiente a las enzimas que participan en cada una de las degradaciones de glucosa, ribosa, adenina, guanina, uracilo, citosina y timina, así como las rutas degradativas de los veinte aminoácidos canónicos de las proteínas. Por último, se utilizó una copia de la colección de las secuencias representativas como base de datos blanco para la búsqueda de homólogos con el fin de emular los resultados que se obtendrían de la metodología “bidireccional best hit” implementada con el programa BLAST+ v.2.4.0 (Camacho et al. 2009). Los resultados que se desprenden de este análisis fueron filtrados utilizando los siguientes criterios: a) solo las comparaciones cuyo valor de “E” en el alineamiento fuera igual o menor a  $1 \times 10^{-10}$  fueron consideradas; b) de las anteriores, únicamente los alineamientos con un valor de cobertura igual o



superior al 70% fueron considerados; y c) se consideraron solamente las secuencias identificadas como “bidireccional best hits”. Los datos provenientes de este análisis se pueden localizar en la siguiente liga <https://goo.gl/HD25gx>.

**Comparaciones de la estructura cristalográfica de las enzimas:** Cuando la conservación a nivel de estructura primaria no basto para establecer o descartar la posible homología entre las enzimas, y la arquitectura de dominios de las secuencias sugería una historia evolutiva en común basados en la clasificación de los dominios de Pfam en sus respectivos Clanes (Finn et al. 2006), se compararon las estructuras cristalográficas de las enzimas si estas estaban disponibles en bases de datos públicas. Las estructuras cristalográficas fueron reclutadas de la base de datos PDB (Protein Data Bank, Berman et al. 2000) cuyas resoluciones fueron iguales o mejores a 3.0 Å, se prefirieron las estructuras provenientes del mismo organismo o uno relacionado y que a su vez haya sido identificado en nuestra búsqueda de homólogos. Los alineamientos estructurales se hicieron utilizando las herramientas del sitio web del PDB (Prlic et al. 2010), usando el algoritmo jFATCAT-rigid, que es un Java-port del algoritmo original FATCAT (Ye y Godzik 2003). Un resumen sobre las estructuras cristalográficas utilizadas para las comparaciones así como los resultados obtenidos se pueden localizar en la siguiente liga <https://goo.gl/jMKUGJ>. Los tres parámetros que fueron usados para proponer la posible relación de homología entre dos enzimas son:

1. **RMSD:** La desviación del valor cuadrático medio (root-mean-square deviation), que representa la desviación estándar de las diferencias entre los carbonos  $\alpha$  de las estructuras cristalográficas sometidas a comparación. Es ampliamente usado como un valor de referencia al establecer el parecido entre dos estructuras cristalográficas (Chothia y Lesk 1986; Irving, Whisstock, y Lesk 2001). Debido a que en la mayoría de los casos que discutimos aquí las comparaciones se llevaron a cabo con estructuras provenientes de organismos poco relacionados filogenéticamente, e incluso provenientes de dos diferentes linajes celulares, y como ya lo han hecho

notar Chothia y Lesk (1986), la cantidad de cambios estructurales está directamente ligada a la cantidad de cambios en la secuencia y por lo tanto a la cantidad de tiempo que ha transcurrido desde que ambas proteínas divergieron de su ancestro común, en el presente trabajo el valor de corte del RMSD se fijó en 4.0 para no descartar *a priori* enzimas cuya relación evolutiva sea muy distante.

2. SAS: La puntuación del alineamiento estructural (structural alignment score) es la medida de la distancia geométrica que considera el número de residuos alineados en relación con su respectivo RMSD (Subbiah, Laurents, y Levitt 1993). El SAS por lo tanto, pondera los alineamientos estructurales con un buen valor de RMSD contra el número de residuos que se alinearon entre dos estructuras cristalográficas y lo representa en un único valor estandarizado. En el presente trabajo el valor requerido se fijó en 2.0, mismo que pretende contemplar un mínimo de residuos alineados.
3. Porcentaje de residuos alineados: La evolución de proteínas en linajes celulares distintos puede causar que dos proteínas con un ancestro común sean sometidas a diferentes presiones de selección que afecten el tamaño de la secuencia polipeptídica y de la estructura cristalográfica. No obstante, si estas enzimas se mantienen favoreciendo reacciones similares o idénticas es razonable suponer que existan regiones a lo largo de la proteína que se mantengan sin alteraciones mayores. Por lo tanto se estableció que por lo menos el 50% de los residuos que integran a la cadena polipeptídica más corta deberán alinearse con su correspondiente contraparte para que el alineamiento sea considerado significativo y se puedan derivar posibles relaciones de homología.

# Capítulo 1

Sobre la evolución temprana de las rutas catabólicas: un estudio de genómica comparada. I. Los casos de las degradaciones enzimáticas de glucosa, ribosa y las bases nitrogenadas adenina, guanina, citosina, uracilo y timina.

En este capítulo se presenta el artículo de investigación donde se publican los primeros resultados del proyecto doctoral “Origen y evolución de las principales rutas catabólicas” inscrito en el Programa de Doctorado en Ciencias Biomédicas, mismo que se publicó en la revista “Journal of Molecular Evolution” y que se titula “On the early evolution of the catabolic pathways: a comparative genomics approach. I. The cases of glucose, ribose and the nucleobases catabolic routes”.

# *On the Early Evolution of Catabolic Pathways: A Comparative Genomics Approach. I. The Cases of Glucose, Ribose, and the Nucleobases Catabolic Routes*

**Mario Rivas, Arturo Becerra & Antonio Lazcano**

**Journal of Molecular Evolution**

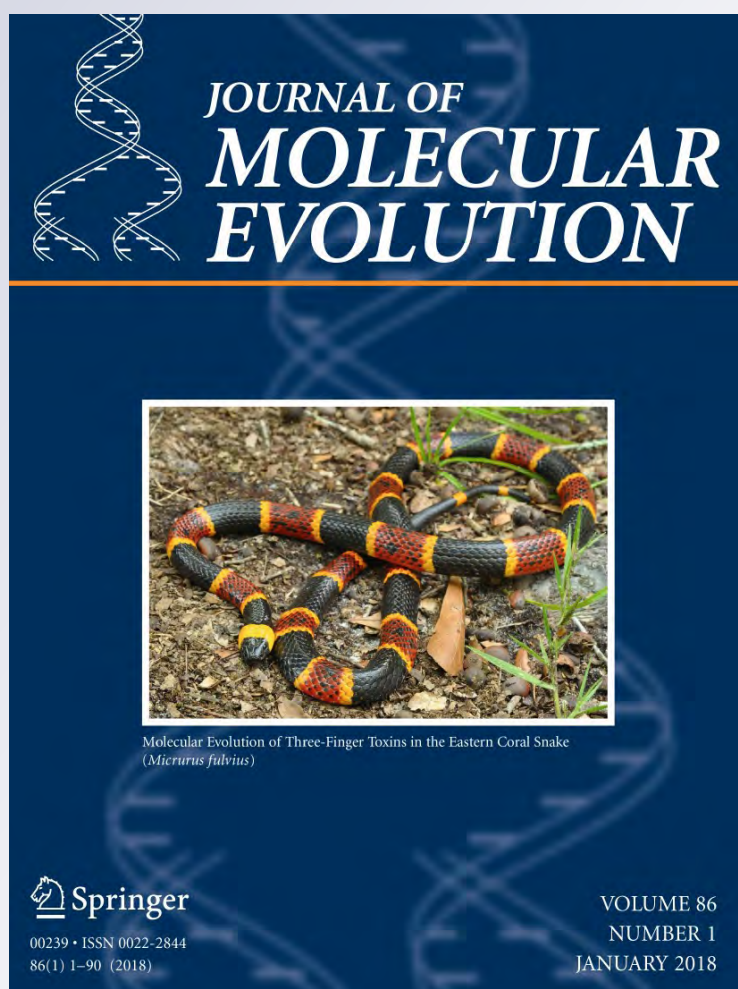
ISSN 0022-2844

Volume 86

Number 1

J Mol Evol (2018) 86:27-46

DOI 10.1007/s00239-017-9822-8



**Your article is protected by copyright and all rights are held exclusively by Springer Science+Business Media, LLC, part of Springer Nature. This e-offprint is for personal use only and shall not be self-archived in electronic repositories. If you wish to self-archive your article, please use the accepted manuscript version for posting on your own website. You may further deposit the accepted manuscript version in any repository, provided it is only made publicly available 12 months after official publication or later and provided acknowledgement is given to the original source of publication and a link is inserted to the published article on Springer's website. The link must be accompanied by the following text: "The final publication is available at [link.springer.com](http://link.springer.com)".**



# On the Early Evolution of Catabolic Pathways: A Comparative Genomics Approach. I. The Cases of Glucose, Ribose, and the Nucleobases Catabolic Routes

Mario Rivas<sup>1</sup> · Arturo Becerra<sup>1</sup> · Antonio Lazcano<sup>1,2</sup>

Received: 18 August 2017 / Accepted: 26 November 2017 / Published online: 30 November 2017  
© Springer Science+Business Media, LLC, part of Springer Nature 2017

## Abstract

Compared with the large corpus of published work devoted to the study of the origin and early development of anabolism, little attention has been given to the discussion of the early evolution of catabolism in spite of its significance. In the present study, we have used comparative genomics to explore the evolution and phylogenetic distribution of the enzymes that catalyze the extant catabolic pathways of the monosaccharides glucose and ribose, as well as those of the nucleobases adenine, guanine, cytosine, uracil, and thymine. Based on the oxygen dependence of the enzymes, their conservation, and evolution, we speculate on the relative antiquity of the pathways. Our results allow us to suggest which catabolic pathways and enzymes may have already been present in the last common ancestor. We conclude that the enzymatic degradations of ribose, as well as those of purines adenine and guanine, are among the most ancient catabolic pathways which can be traced by protein-based methodologies.

**Keywords** Catabolic pathways · Early metabolic evolution · Nucleobase degradation · Carbohydrate degradation · LCA metabolic enzymes

## Introduction

The significance of the extant catabolic pathways resides not only in that they are used by organisms as sources of carbon, nitrogen, and other elements, but also to extract the energy from high-energy molecules and transfer it to other biologically useful molecules like ATP or GTP, as well as in the accumulation of the reducing power in the form of molecules like NADH or FADH. In many cases, the end products of a catabolic pathway are crossroads metabolites, which are taken as substrates of the biosynthetic routes of proteins, nucleic acids, and other biologically relevant

molecules. Through catabolic routes, organisms are able to use the organic material that surrounds them as the source of essential elements to sustain their existence. Therefore, they must have played a key role in the early evolution of life during the establishment of the intricate network of reactions that constitute extant metabolic pathways.

The first discussion on the origin of catabolism was proposed by Oparin (1924, 1938). One of the central tenets of his ideas on the heterotrophic origin of life was the assumption that glycolysis was the oldest metabolic route of a primitive heterotrophic anaerobe that was totally dependent on the organic material present in the surroundings (Oparin 1924, 1938; Lazcano 2016). Based on the apparent ubiquity of glycolytic enzymes and on the relative simplicity of fermentative reactions, Oparin suggested an anaerobic primitive environment in which the prebiotic synthesis and accumulation of organic chemicals occurred (Lazcano and Miller 1999; Lazcano 2016). More than 20 years later, Krebs and Kornberg (1957) and Krebs (1981) argued that if the Oparin's scenario was valid, then the second catabolic process that must have evolved was the pentose phosphate pathway, because of its close relation with glycolysis, as indicated by several shared biochemical products. Later on,

**Electronic supplementary material** The online version of this article (<https://doi.org/10.1007/s00239-017-9822-8>) contains supplementary material, which is available to authorized users.

✉ Antonio Lazcano  
alar@ciencias.unam.mx

<sup>1</sup> Facultad de Ciencias, Universidad Nacional Autónoma de México, Apdo. Postal 70-407, Cd. Universitaria, 04510 Mexico City, Mexico

<sup>2</sup> Miembro de El Colegio Nacional, Mexico City, Mexico

Clarke and Elsdén (1980), based on the easiness of abiotic synthesis of amino acids and the relative simplicity of amino acid fermentation, suggested a “primitive energy-yielding oxido-reduction system” with the usage of glycine and proline as oxidant agents with an associated phosphorylation from which a catabolic pathway could be assembled. The most recent proposal on the origin of enzyme-mediated degradations of cellular components such as carbohydrates, amino acids, and nucleobases was developed by Schönheit et al. (2016). They assumed an autotrophic origin of life, and argued that the most suitable molecules that could be decomposed by early heterotrophs had to be the cellular components of pre-existing autotrophs, and that the enzymatic machineries of those hypothetical autotrophic organisms must have evolved into catabolic pathways resembling for instance the Clostridial-type fermentations which in turn must be the first forms of heterotrophic carbon and energy metabolism. By assuming the autotrophic origin of life, the possibility of earlier degradations, whether enzyme-based or RNA-based, of the organic material present in the primitive earth was dismissed.

In spite of claims on the contrary (Caetano-Anollés et al. 2007; Caetano-Anollés and Caetano-Anollés 2013), it is not easy to correlate the compounds that may have been present in the primitive Earth with the intricate network of reactions that integrate current metabolism. Several attempts have been made based on the extrapolation of extant anabolic routes that may have been part of the metabolic traits of the LCA to the origin of life itself but, in fact, under this view, the origin of metabolism would appear to be closer to the LCA than to the emergence of the first living entities (Lazcano and Miller 1999). On the other side, almost no attention has been given to understand how the constituents of the prebiotic soup lead to the first catabolic pathways (Keefe et al. 1995). It is of course reasonable to assume that the outcome of prebiotic synthesis experiments, combined with the data of organic compounds present in meteorites, provide information about the molecules that may have been available in the primitive environments as substrates of the first catabolic pathways. However, at the time being it may be wiser to limit our extrapolations of extant pathways to a period after protein biosynthesis had already evolved.

It is also important to realize that even in extant biochemical pathways, a certain number of non-enzymatic reactions occur spontaneously, and that it is possible that equivalent reactions played an important role in the establishment of early metabolic pathways (Lazcano and Miller 1999; Keller et al. 2015). It seems likely that a number of reactions of primitive catabolic pathways could have at first taken place semi-enzymatically. In other words, the approach advocated here does not allow any direct inference on the truly primordial catabolic pathways that may have existed during previous stages like the RNA world or a pre-RNA world, where

the reactions could have been entirely enzyme-free processes dependent on inorganic catalysts such as mineral cations or ribozymes (Becerra et al. 2007).

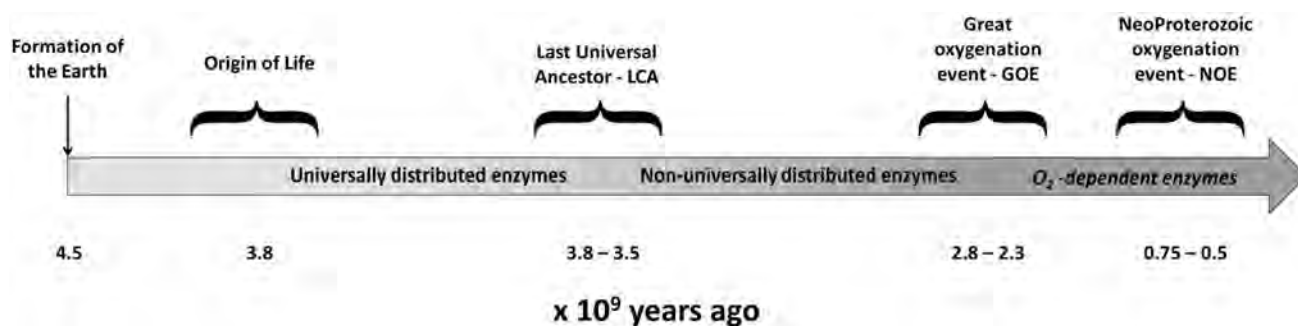
Metabolism has been traditionally divided into anabolism and catabolism, though in a number of cases the line separating pathways involved in the synthesis or degradation of a compound is opaque and narrow. Good examples of the latter include the reductive TCA cycle, which involves almost the same set of oxidative reactions going backwards, as well as the lysine degradation pathway, that begins with half the reactions used in the biosynthetic pathway (Michal 1998). In evolutionary terms, it is possible that a pathway once used to synthesize a molecule has evolved to break down the same molecule or vice versa, making it difficult to reconstruct the processes that gave rise to the complex networks of biochemical reactions we see today.

In the present study, we explore the phylogenetic distribution and molecular evolution of the enzymes that integrate extant catabolic pathways of the monosaccharides glucose and ribose, as well as of the nucleobases adenine, guanine, cytosine, uracil, and thymine. Based on their distribution, we were able to date enzymes and pathways that most likely were present in the LCA and those that evolved within particular phylogenetic groups. We have also considered the oxygen dependence of catabolic enzymes as a diagnostic character of their antiquity, since the first accumulation of O<sub>2</sub> in the Precambrian atmosphere, known as the great oxygenation event (GOE), changed forever the chemistry of the planet leading to the development of more energetically efficient metabolisms (Goldfine 1965; Canfield 2005). Moreover, recent efforts based on paleobiological and biogeochemical evidence now recognize a second increase in the oxygen concentrations in the atmosphere at the end of the Neoproterozoic Era, known as the Neoproterozoic Oxygenation Event (NOE), that may have had an effect on the evolution of multicellular organisms (Knoll and Nowak 2017). This approach allows us to establish a relative date of enzymes and pathways that most likely appeared before and after the oxygen accumulation in the terrestrial atmosphere (Fig. 1), which may have started sometime during the period of 2.8–2.3 billions of years ago (Canfield 2005; Lyons et al. 2014).

## Methodology

### Defining Catabolic Enzymes and Pathways

The catabolic pathways, the enzymes, the substrates, and the products analyzed here have been defined based on the Biochemical Pathways Atlas (Michal and Schomburg 2012), the KEGG database of metabolisms (Kanehisa and Goto 2000; Kanehisa et al. 2014), the MetaCyc database (Caspi



**Fig. 1** Relative dating of the catabolic enzymes and the catabolic pathways. The oxygen dependence of the enzymes and the pathways in which they participate has led us to make a relative dating. They most likely evolved into their present form after the oxygen enrichment of the earth's atmosphere, sometime between 2.8 and 2.3 billions of years ago (Canfield 2005; Lyons et al. 2014; Knoll and Nowak 2017). Distribution of the catabolic enzymes along the phylogenetic groups of the two major domains of life, Bacteria and Archaea (Forterre and Gribaldo 2010; Embley and Williams 2015),

et al. 2014), and the BRENDA database (Schomburg et al. 2013). This information was complemented using the data of McMurry and Begley (2005) when required. In some cases, the distinction between anabolic and catabolic enzymes is not evident, since they may intervene in both synthesis and degradation of a given compound. Accordingly, when a report of a catabolic pathway was found in the literature or in the databases, a distinction was made between enzymes used exclusively for degradation, and those involved in both degradation and synthesis using the biochemical information available in KEGG, MetaCyc, and/or BRENDA databases. There are a few special cases when there are no biochemical data available that allow us to establish the directionality of a reaction. In such cases, the enzyme was annotated as catabolic if it most probably catalyzes a reaction identified as essential within the chain of reactions of the degradation pathway. All reactions of catabolic pathways under examination, including the corresponding enzymes and cofactor/coenzyme requirements, were collected in a catalogue available as supplementary material Catabolism-Catalogue or through the following web address <https://goo.gl/Ts1xpD>.

### Recruitment of Sequences

The recruitment of sequences for homologue searches was performed using the Enzyme Commission (EC) number annotation in the UniProt database (The UniProt Consortium 2012), which allowed us to collect high-quality, manually annotated, and non-redundant protein sequences from UniProtKB/Swiss-Prot (The UniProt Consortium 2014), and high-quality computationally analyzed sequences from UniProtKB/TrEMBL (Magrane and Consortium 2011). In the cases where sequences associated to a particular EC

allowed the classification of the pathways that most likely were present in the LCA as universal, and pathways whose enzyme distribution suggested either an archaeal or bacterial origin as non-universal. Eukaryotes were also included in the database for completeness. The distribution results for the eukaryotic enzymes are also presented in the supplementary material or through the following web addresses <https://goo.gl/dDdLfd>, <https://goo.gl/WDwRQE>, <https://goo.gl/6x7hmo> and <https://goo.gl/JAPI7R>

identifier in the UniProt database were not found, a random seed-sequence was chosen from one representative of Archaea, Bacteria, and Eukarya (when available) out of the KEGG protein web database (Kanehisa and Goto 2000; Kanehisa et al. 2014). A web-search for each seed-sequence was then performed using BLAST software (Altschul et al. 1990) routine *blastp* against the non-redundant database from NCBI (Geer et al. 2010). The sequences gathered in each search included the first one hundred best E-values hits with query coverage above 70%.

### Analyzing the Collection of Sequences

Once the sequences had been collected, they were first filtered to limit the redundancy on the database using the CD-HIT methodology (Fu et al. 2012) with values  $C=0.7$  and  $N=5$ . The identification of the domain architecture of the remaining sequences was done by implementing the recognition of Pfam domains (Punta et al. 2012) through HMMER3 software (Mistry et al. 2013) using the *hmmerscan* routine against the Pfam database (Punta et al. 2012; released 27.0) according to the instructions in the HMMER manual. Identification of domain architecture allowed the recognition of sequences with different domain architectures annotated under a common EC number, which led us to cluster together the sequences with equal domain architectures before further processing. As a result, more than one HMMER profile were built for the sequences that were classified under the same EC number. Every profile represents each one of the previously identified domain architectures, which in turn allowed us to perform independent searches for each profile generated this way.



## Building Profiles for Homologue Search

Since multiple sequence alignment is a prerequisite for profile construction, the aligner muscle v3.8.31 (Edgar 2004a, b) was employed using no modifiers to align each set of the previously curated sequences. A profile was then built for each cluster of sequences using the hmmbuild routine according to the manual instructions (Eddy 2011).

## Assigning a Limit to the Hmmersearch Results

After obtaining the profile for every enzyme under study, a search was performed using the hmmsearch routine against the original unaligned set of sequences that were used to create such profile, in order to detect the sequences that behave in the less efficient way against the model that they gave rise to. This allowed us to isolate and include that sequence as “The Limit” for each particular model inside the protein database, instead of assigning an arbitrary cutoff value. Finally, in the presence/absence results, those sequences that performed equal or better than the limit sequence for each profile were included.

## Protein Database

A database of protein sequences consisting entirely of complete sequenced organisms with representatives from the three domains of life was integrated. It comprises 1201 Bacteria, 97 Archaea, and 146 Eukaryotes, all of them collected from the KEGG database (Kanehisa and Goto 2000; Kanehisa et al. 2014). The Archaea representatives analyzed here include mostly members from the Euryarchaeota and Crenarchaeota major groups; members of the Thaumarchaeota, Nanoarchaeota, and Korarchaeota groups were also analyzed for the sake of completeness. Although we considered that they lack representability over the entire domain due to the few species that they included at the time of the local database integration (2011), results of the searches against their respective organisms are also presented as supplementary material for consideration or through the following web addresses <https://goo.gl/dDdLfd>, <https://goo.gl/WDwRQE>, <https://goo.gl/6x7hmo> and <https://goo.gl/JAPI7R>. The Bacterial members belong to the  $\alpha$ -proteobacteria,  $\beta$ -proteobacteria,  $\delta$ -proteobacteria,  $\epsilon$ -proteobacteria,  $\gamma$ -proteobacteria, Firmicutes, Tenericutes, Actinobacteria, Chlamydiae, Spirochaetes, Acidobacteria, Bacteroidetes, Fusobacteria, Verrucomicrobia, Planctomycetes, Cyanobacteria, Green Sulfur Bacteria, Green non-sulfur bacteria, Deinococcus–Thermus, and Hyperthermophilic-bacteria major groups. The Eukaryote members comprise the Vertebrates, Arthropods, Nematodes, Cnidarians, Eudicots, Monocots, Algae, Ascomycetes, Basidiomycetes, Choanoflagellates, Amoebozoa, Alveolates, Euglenozoa, and

Diatoms major groups according to the classification on the KEGG database.

## Defining the Antiquity of the Pathways

Relative dating of the catabolic pathways was defined using two different criteria. The first obvious one was based on the oxygen dependency of the enzymes that integrate each catabolic pathway. If an enzyme of a particular pathway catalyzes a degradative reaction that requires molecular oxygen as substrate, and that reaction is considered essential for the pathway to produce a compound described in the literature, but has no oxygen-independent counterpart, then we assumed that the entire pathway evolved into its actual form after the oxygen enrichment of the terrestrial atmosphere (see supplementary material Catabolism-Catalogue for details of the oxygen-dependent enzymes in each catabolic pathway). The remaining pathways were then classified as universal and non-universal in accordance to the distribution of their enzymes along the major groups of Bacteria and Archaea domains (Forterre and Gribaldo 2010; Embley and Williams 2015). The presence of enzymes in all organisms of the database was not defined as a requisite. Instead, we considered that their distribution must be representative among the major groups of Archaea and Bacteria, which means that enzymes of a particular pathway must be present in at least half of the known major groups in the Archaea and Bacteria domains. For a particular group to be considered indicative to the presence of a given pathway, the corresponding enzymes must be present in at least one-third of the organisms that integrate each major group; the later value was defined empirically, based on the phylogenies constructed for all catabolic enzymes as a strategy to detect possible cases of horizontal gene transfer between Archaea and Bacteria. For the sake of completeness, enzyme searches in the Eukarya domain were performed and their distribution is reported in the same manner. This allowed us to distinguish between pathways that were most likely present in the LCA (or LUCA, Last Universal Common Ancestor), and those that evolved after the divergence of the two major prokaryotic domains. Detailed group-distribution tables are provided as supplementary material or through the following web addresses <https://goo.gl/dDdLfd>, <https://goo.gl/WDwRQE>, <https://goo.gl/6x7hmo> and <https://goo.gl/JAPI7R>.

## Enzyme Evolution Within and Between the Pathways

We have traced enzyme ancestry within and between the catabolic pathways using sample sequences taken at random from our previously classified collections of enzymes. In this way, sets of sequences that represent all the enzymes

involved in each catabolic pathway were defined. We then merged all sequences corresponding to carbohydrates and nucleobases degradative pathways in a single “*fasta*” file for each category. Finally, copies of the same set of sequences that integrate each category were used as the target database to emulate same type of results as in the “bidirectional best hit” methodology using BLAST + blastp routine v.2.4.0 (Camacho et al. 2009). Results of the blast comparisons were filtered by the following criteria: (a) only those sequences that were identified as “bidirectional hits” were considered; (b) results with *e*-values lesser or equal to  $1 \times 10^{-10}$  were included; and (c) alignments with query coverage above 70% were taken as positives. The data collection that emerged from this comparison is presented as supplementary material or through the following web address <https://goo.gl/ryvTjK>.

### Comparisons of the Crystallographic Structures of Proteins

When protein sequence conservation was not enough to derive or discard homology relationships, and the protein domain architecture suggested a possible common history based on Pfam clan classification (Finn et al. 2006), the crystallographic structures of proteins were compared when available (see supplementary material or the following web addresses <https://goo.gl/fUklBo> and <https://goo.gl/nSoRIZ>). Protein crystallographic structures were recruited from the Protein Data Bank (Berman et al. 2000) with resolutions of at least 3.0 Å, preferably from the same organism (or a phylogenetically related organism) previously identified by our search for homologue sequences when available. Structure alignments were performed with the PDB web tools (Prlic et al. 2010) using the jFATCAT-rigid algorithm, which is a Java port of the original FATCAT algorithm (Ye and Godzik 2003). Three parameters were considered in order to propose homology between two given enzymes:

1. **RMSD** The root-mean-square deviation, which represents the standard deviation of the differences between the equivalent  $\alpha$ -carbons of two crystal structures, and which is widely used as a reference value to establish the likeness between two crystallographic structures (Chothia and Lesk 1986; Irving et al. 2001). In most cases, our comparisons involved proteins of organisms that belong to separate domains of life and, as noted long ago by Chothia and Lesk (1986), the extent of the structural changes is directly related to extent of protein sequence changes and therefore to the amount of time since both proteins diverged from their common ancestor, in this work the RMSD threshold value was set on 4.0, so that distantly related proteins were not discarded without further consideration.

2. **SAS** The structural alignment score is a geometric distance measure that considers the number of aligned residues in relation to their respective RMSD (Subbiah et al. 1993). The SAS value evaluates alignments with good RMSD values against the number of residues that were aligned between the crystallographic structures in one standardized measurement. In this work, the threshold value was set at 2.0.
3. **Percentage of aligned amino acids** The evolution of proteins in separate lineages can cause two homologues to be subject to different selective pressures, leading in many cases to different sizes. However, if they keep on catalyzing similar or identical reactions, then there must be regions along the peptide chain that remain alike. Therefore, we assumed that at least 50% of the residues of the smallest protein under comparison must align with their corresponding protein residues for the comparison to be significant enough to derive homology relationships.

### Results and Discussion

As illustrated in Fig. 1, the methodology used here, based both on enzymes distribution and their dependence (or not) to molecular oxygen, has allowed us to construct a model that lists the pathways that may have been present in the LCA and those that evolved after the oxygen enrichment of the Earth's atmosphere. The possibility of a complete replacement of oxygen-independent enzymes by oxygen-dependent enzymes (Raymond and Blankenship 2004), which could lead to erroneous interpretations in the antiquity of the pathways, should be acknowledged. In our model, enzymes and pathways that have a non-universal distribution but have an oxygen-free biochemistry were placed before this major breaking point in the history of life. This does not necessarily limit their origin to that particular time, since it is possible that those routes are more recent and evolved in anaerobic organisms living in oxygen-free environments. It is important to underline that our methodology and results differ from the reconstructions of the genomic content of the LCA (or LUCA) (Harris et al. 2003; Mirkin et al. 2003; Delaye et al. 2005; Yang et al. 2005; Ranea et al. 2006; Kim and Caetano-Anollés 2011), because we are using a huge number of completely sequenced cellular genomes, and the intersection of the genome content rapidly decreases when an increased number of cellular genomes are used for comparison. Therefore, strict presence of all enzymes in all organisms represented in our database was not a requisite. Instead, as noted above, we have used their distribution along the major phylogenetic groups that integrate the Archaea, Bacteria, and Eukarya domains, since there can always be secondary losses among certain groups,

such as parasites and symbionts, that diminish the effective number of genes that can be traced back to the LCA. While it is true that this approach could introduce a bias in our results, it allows us to classify complete pathways and not only isolated genes. Horizontal gene transfer events can also affect the number of genes that are considered to be part of the genetic content of the LCA. Therefore, we analyzed the phylogenetic history of each enzyme under study; cases detected this way will be discussed further in each section.

## Glucose

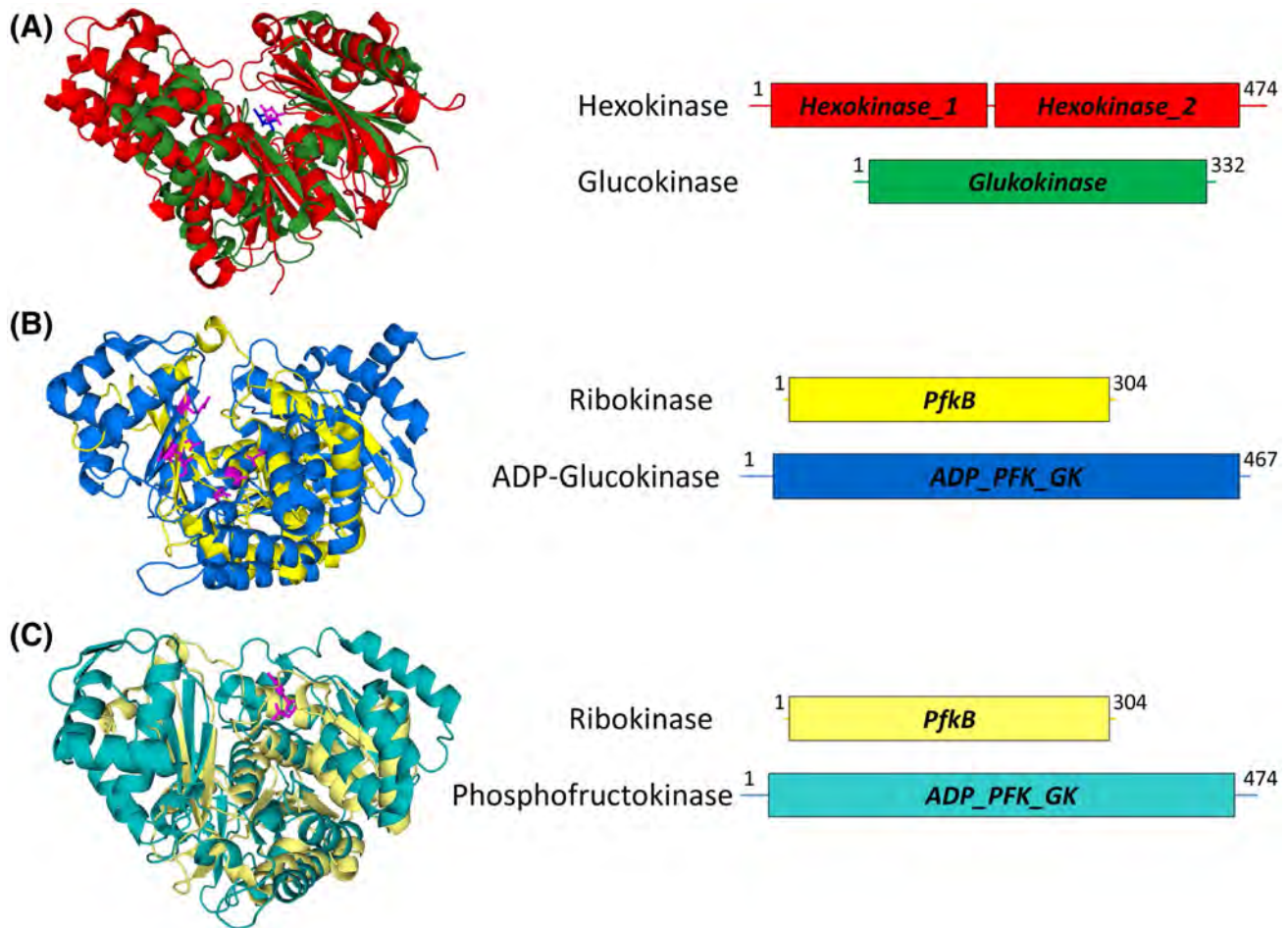
Although it has been long assumed that the fermentation of glucose and ribose were the first catabolic pathways (Oparin 1924, 1938; Krebs and Kornberg 1957; Krebs 1981; Lazcano 2016), the presence of glucose and ribose in the prebiotic environment remains an issue, although the synthesis of smaller sugar compounds such as glycolaldehyde and glyceraldehyde under prebiotically plausible conditions has been reported (Ritson and Sutherland 2012). Even if the first non-enzymatic carbohydrate synthesis was achieved by Butlerov in the well-known “formose reaction” more than 150 years ago (Rauchfuss 2008), it is not easy to envision the synthesis and accumulation of high amounts of carbohydrates like glucose under possible prebiotic conditions. Although ribose synthesis can be achieved through the formose reaction, the overall yield is small (Reid and Orgel 1967; Shapiro 1988), and it is a very unstable carbohydrate with a half-life of  $\approx 73$  min at 100 °C and pH 7, and of 44 years at 0 °C (Larralde et al. 1995). The identification of by-products of the alkaline hydrolysis of ribose in extracts of the Murchison and Murray meteorites has been reported (Cooper et al. 2001), suggesting the ephemeral presence of ribose in the parent body of some carbonaceous meteorites. In the present work, we have analyzed the main catabolic pathways that degrade glucose and ribose because they are considered relevant to understand the early evolution of carbohydrate metabolism, although we recognize that current evidence suggest that it is possible that their prevalence in extant biology resulted from the earliest evolution of anabolic routes, and not necessarily from prebiotic synthesis or delivery from outer space.

The canonical product of glycolysis is pyruvate, a crossroads metabolite that either connects to the Krebs cycle or enters the oxygenic/anoxygenic fermentative pathways in extant organisms. It is well known that the selective advantage of glycolysis is the mining of the reducing power in the form of two molecules of NADH, and the net income of two molecules of ATP per molecule of glucose that is degraded. On the other hand, the main products of extant ribose degradation are fructose 6-phosphate and/or glyceraldehyde-3-phosphate, which are crossroads metabolites connecting directly to the glycolytic pathway (Kanehisa and

Goto 2000; Michal and Schomburg 2012). While the pentose phosphate cycle selective advantage is the production of reducing equivalents, such as NADH, ribose catabolism by its own appears to be related to the production of the intermediaries that feed glycolysis. Although the latter correlates with the Krebs and Kornberg (1957) and Krebs (1981) proposal of the development of the pentose phosphate pathway after the establishment of glycolysis, it does not necessarily imply that those two routes emerged during the very early stages of life.

Glycolysis starts with the phosphorylation of glucose to produce glucose 6-phosphate. There are four different enzymes that can catalyze this reaction, namely, hexokinase (HEX, EC 2.7.1.1), glucokinase (GLUK, EC 2.7.1.2), polyphosphate glucokinase (PGK, EC 2.7.1.63), and ADP-specific glucokinase (AGLU, EC 2.7.1.147). Phylogenetic analysis on the distribution of these enzymes reveals that (a) HEX is restricted to the major groups of Eukarya; (b) GLUK has a universal distribution being present among almost all the major groups of Archaea, Bacteria, and Eukarya; (c) PGK has been found exclusively among some members of the Actinobacteria, Acidobacteria, Bacteroidetes, Verrucomicrobia, Cyanobacteria, and Deinococcus–Thermus bacterial groups; and (d) AGLU appears to be restricted to some members of the Euryarchaeota and some members of the eukaryotic groups of Nematodes and Vertebrates. Such biological distribution suggests that the most ancient form of catalyzing this reaction was through a GLUK-like enzyme that was probably already present in LCA. Sequence comparisons reveal no significant resemblance among these proteins, but domain composition reveals that HEX, GLUK, and PGK are integrated by the *hexokinase\_1*, *hexokinase\_2*, *glucokinase*, and *ROK* domains, all of which belong to the same Pfam clan (CL0108), suggesting that they might share an evolutionary history (Finn et al. 2006). Comparisons of the crystallographic structures of HEX, GLUK, and PGK confirm that these proteins share a common ancestor (see Fig. 2 and supplementary material or visit <https://goo.gl/fUklBo>). On the other hand, AGLU possesses the *ADP\_PFK\_GK* domain, which belongs to a different Pfam clan (CL0118). Its distribution and the lack of resemblance between its crystallographic structure and those of HEX, GLU, and PGK, suggest that it is a later, independent, development.

Isomerization of glucose-6-phosphate into fructose-6-phosphate is catalyzed by glucose 6-phosphate isomerase (EC 5.3.1.9), a universally distributed enzyme that is well represented among all major groups of the Archaea, Bacteria, and Eukarya. Fructose 6-phosphate undergoes a second phosphorylation and produces fructose 1,6-bisphosphate, a reaction that can be catalyzed by two non-homologous enzymes, 6-phosphofructokinase (PPFK, EC 2.7.1.11) and ADP phosphofructokinase (APPFK, EC 2.7.1.146). There is also a PPFK isoenzyme (EC 2.7.1.11/2.7.1.90) that can



**Fig. 2** Comparisons of the crystallographic structures of kinases from the Actin-like ATPase superfamily and the Ribokinase-like superfamily that are involved in the phosphorylation reactions in glycolysis and the ribose catabolic pathway. Alignments are accompanied with a representation of the pfam domain composition of the proteins. **a** Alignment of the crystallographic structures of hexokinase (EC 2.7.1.1, PDB ID 4QS7, red) and glucokinase (EC 2.7.1.2, PDB ID 1S22, forest-green). Both structures are visualized with their respective glucose substrate (blue and magenta, respectively). Although it is difficult to establish homologous relationships based solely on their primary sequence alignments, pfam domain classification suggests a possible common evolutionary history, which is supported by the alignment of both crystallographic structures. **b** Alignment of the crystallographic structures of ribokinase (EC 2.7.1.15, PDB ID 3RY7, yellow) and the ADP-specific glucokinase (EC 2.7.1.147, PDB ID 4B8R, marine-blue). Residues of the glucose binding site of

phosphorylate the fructose-6-phosphate using pyrophosphate (PP<sub>i</sub>) in the absence of ATP (Reeves et al. 1974). This isoenzyme is so similar to the original PPFK that it is not easy to distinguish between them neither by sequence comparisons nor by position-specific weight matrices. Both versions of the PPFK are characterized by the *PFK* domain, which constitutes almost the whole length of their respective sequences, demonstrating their common origin. While PPFK is well distributed among the bacterial and eukaryotic

the ADP-specific glucokinase are highlighted in magenta (N38, D42, E96, Q121, H184, R205, and D451). Although ADP-specific glucokinase and ribokinase phosphorylate different carbohydrates, their overall structure is the same, confirming their common ancestry as suggested by their pfam domain composition classification. **c** Alignment of the crystallographic structures of ribokinase (EC 2.7.1.15, PDB ID 3RY7, pale-yellow) and the archaeal ADP phosphofructokinase (EC 2.7.1.146, PDB ID 3DRW, teal-cyan) with ADP molecule (magenta). The alignment supports the possibility that both enzymes share a common ancestor as suggested by their domain clan assignment. Images presented here were created using The PyMOL Molecular Graphics System, Version 1.8.1.2 Schrödinger, LLC implementing the CE algorithm (Shindyalov and Bourne 1998) and colors correspond to the PyMOL standard color palette. See supplementary material or visit <https://goo.gl/fUklBo> for details on the comparisons of the crystallographic structures

groups, it is almost absent from Archaea. Although there are few Archaea from the Methanomicrobia class in which homologues of PPFK have been identified, their limited distribution and close resemblance to the bacterial enzymes suggest that they are the outcome of two horizontal gene transfer (HGT) events, as shown in Fig. S1. By contrast, APPFK has been identified exclusively in members of the Euryarchaeota group, suggesting an archaeal origin. Sequence comparisons of the PPFK and the APPFK reveal no significant

resemblance that could lead us to establish any evolutionary relationship between them. Analysis of the domain architectures reveals that *PFK* is the only functional domain in the PPFKs, while *ADP\_PFK\_GK* is the only domain identified on the APPFKs. Since these domains are classified under different Pfam clans (CL0240 and CL0118, respectively), common ancestry appears to be unlikely. As shown in Fig. 2, crystallographic comparison reveals no significant structural relationship between them, but in fact suggests that APPFK is related to kinases of the ribokinase-like family. We therefore conclude that there is no universally distributed phosphofructokinase whose ancestor could catalyze the phosphorylation of fructose 6-phosphate in the LCA.

Once fructose 1,6-bisphosphate is synthesized, it undergoes a C–C bond cleavage catalyzed by fructose-bisphosphate aldolase (FBPA, EC 4.1.2.13), releasing glyceraldehyde phosphate and glyceraldehyde-3-phosphate. Although at first the distribution of FBPA among major groups of Bacteria and Eukaryotes suggested its emergence after the two major prokaryotic groups have diverged, detailed analyses of the databases and the sequences of FBPA reveal that they actually belong to two separate classes: Class I is well represented among the major groups of Archaea, and scarcely present on the bacterial lineage, while Class II was found to be well represented among  $\gamma$ -proteobacteria,  $\epsilon$ -proteobacteria, Actinobacteria, Spirochaetes, Bacteroidetes, Verrucomicrobia, and Planctomycetes bacterial groups and nearly absent from archaeal organisms. Such distribution suggests independent evolutionary histories within their respective lineages. Specific sequence comparisons of the two classes of aldolases reveal that the identity and similarity values are in the so-called “twilight zone,” and it is difficult to conclude that they are homologues. However, domain architecture analysis revealed that FBPA Class I possess the *DeoC* domain, while FBPA Class II possess the *F<sub>1,6</sub>P<sub>2</sub> aldolase* domain. Both domains are recognized as members of the same Pfam clan (CL0036), which suggests that they could share a common ancestor (Finn et al. 2006). Comparison between the crystallographic structures of both classes of aldolases does support a common origin (see supplementary material or visit <https://goo.gl/fUklBo>), which in this particular case could date from before the divergence of the two prokaryotic domains, followed by a latter parallel development within their respective prokaryotic groups. This hypothesis is supported by the fact that both crystallographic structures are essentially TIM barrel structures whose functional diversity has been proposed to be the result of an early diversification of the TIM barrel superfamilies prior to the LCA epoch, although the monophyletic origin of all TIM barrel superfamilies remains to be demonstrated (Goldman et al. 2016).

Isomerization of glyceraldehyde phosphate into glyceraldehyde-3-phosphate is an essential step if the route must

reach its optimum energy extraction. Although it seems that this step could have been dispensable in the very early stages of glycolysis evolution, the universal distribution of triosephosphate isomerase (TPI, EC 5.3.1.1), the enzyme responsible for this reaction, suggests that an ancestral form was already present in the LCA. Following the isomerization reaction, two molecules of glyceraldehyde-3-phosphate are further processed, either by triosephosphate dehydrogenase (TPDH, EC 1.2.1.12), or by glyceraldehyde-3-phosphate dehydrogenase (GPDH, EC 1.2.1.59), two universally distributed enzymes that catalyze the reaction where the energy released by the oxidation of the aldehyde group on the glyceraldehyde-3-phosphate is used to synthesize either NADH or NADPH and incorporate a phosphate into the molecule, releasing 1,3-bisphospho-D-glycerate. TPDH and GPDH common ancestry could not be established by sequence comparisons, but domain composition revealed a possible evolutionary relationship since both enzymes possess the *Gp\_dh\_C* domain (CL0139). Although this shared domain represents less than 50% of the total length of each sequence, their common ancestry could be inferred from the comparisons of the corresponding crystallographic structures (see supplementary material or visit <https://goo.gl/fUklBo>). This implies that the common ancestor of TPDH and GPDH dates from a pre-LCA epoch.

In a subsequent step, the 1,3-bisphospho-D-glycerate is used as the phosphate donor to synthesize ATP from ADP, releasing 3-phospho-D-glycerate. This reaction is catalyzed by phosphoglycerate kinase (EC 2.7.2.3) that according to our results is universally distributed. The resulting 3-phospho-D-glycerate is then isomerized into 2-phospho-D-glycerate by two non-homologous enzymes, glycerate phosphomutase (GPM, EC 5.4.2.11) and phosphoglycerate mutase (PGM, EC 5.4.2.12). While PGM has a universal distribution, GPM is found mostly among the major groups of Bacteria and Eukarya. Due to its very restricted distribution in Archaea and the close resemblance of the archaeal sequences with those of bacterial organisms, the few archaeal GPM homologues that were detected by our methodology probably represent a case of HGT from Bacteria to Archaea (see Fig. S2). Neither sequence comparisons nor domain composition revealed any possible evolutionary relationship between GPM and PGM, a conclusion supported by comparisons of their crystallographic structures (see supplementary material or visit <https://goo.gl/fUklBo>). This suggests that GPM evolved independently and after the divergence of the two major prokaryotic domains. The next step in the glycolytic pathway is the dehydration of 2-phospho-D-glycerate into phosphoenolpyruvate by the action of enolase (EC 4.2.1.11), a highly conserved enzyme whose ancestor was probably already present in the LCA, as noted before by Delaye et al. (2005) and supported by our results. Finally, the phosphoenolpyruvate is used as a phosphate

donor to synthesize ATP from ADP, a reaction catalyzed by pyruvate kinase (EC 2.7.1.40), another universally distributed enzyme whose ancestor most likely was also present in the LCA, according to our results.

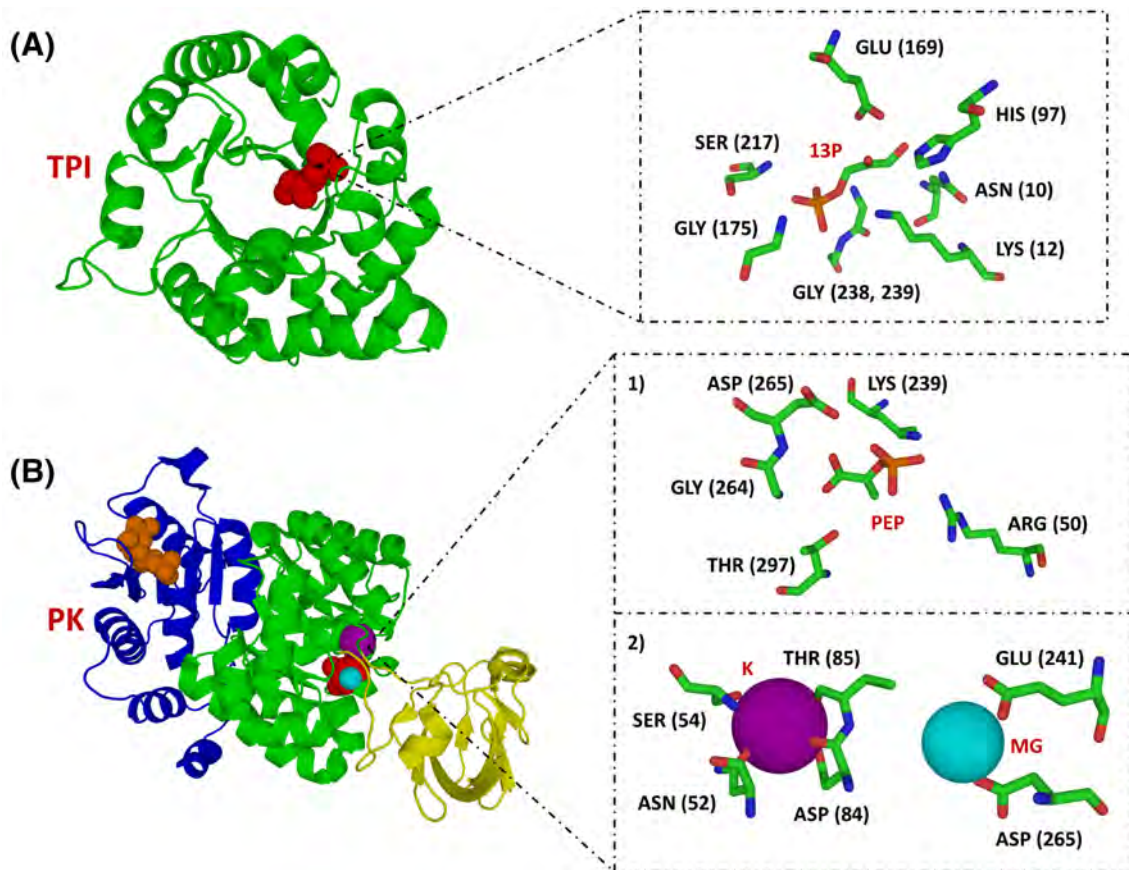
The detailed examination of the phylogenetic distribution of the glycolytic enzymes and their evolutionary relationships discussed here shows how complicated it is to reconstruct the mechanisms that gave rise to the actual pathway. Clearly, the assumption that glycolysis is ubiquitous is no longer valid. As discussed previously by Verhees et al. (2003), several variant enzymes have developed independently within different lineages, like ATP-Glucokinase and ADP-Glucokinase from Bacteria and Archaea, respectively, a hypothesis that appears to be supported by our results. As discussed above, there are indeed non-homologous kinases that catalyze the phosphorylation reaction over glucose, even variants that evolved exclusively among Eukaryotes. Moreover, the phosphorylation of fructose 6-phosphate is a critical glycolytic reaction catalyzed either by PPFK or by APPFK within Bacteria and Archaea, respectively, and whose restricted distribution lead us to conclude that the glycolytic pathway, as we know it today, was not operating in the LCA.

The availability of completely sequenced genomes and the systematic accumulation of biochemical data in public databases allow the recognition of four kinases that catalyze the starting reaction on the glycolytic pathway. The list includes hexokinase (HEX, EC 2.7.1.1), glucokinase (GLUK, EC 2.7.1.2), polyphosphate glucokinase (PGK, EC 2.7.1.63), and ADP-specific glucokinase (AGLU, EC 2.7.1.147). Phylogenetic distribution reveals that, contrary to the canonical version of the pathway, GLUK and not HEX has a universal distribution, and it is quite likely that its ancestor was present in the LCA. Analysis of the domain architecture and crystallographic comparisons show that HEX, GLUK, and PGK most likely share a common ancestor, and that kinases like AGLU and APPFK descend from a different ancestral kinase whose evolutionary history appears to be different from the ancestor of HEX, GLU, and PGK. Phosphoglycerate kinase (EC 2.7.2.3) and pyruvate kinase (PK, 2.7.1.40) represent examples of two other glycolytic kinases that, as shown by sequence comparisons, domain architecture, and crystallographic comparisons, do not exhibit any evolutionary relationship, neither with the previous kinases of the glycolytic pathway nor among them. These results support the notion of a recruitment of different kinases with separate evolutionary origins into the glycolytic pathway, and the available data suggest that this took place at least in five different occasions. This conclusion also supports the hypothesis that the pathway was assembled in a patchwork style within phylogenetic groups, contrary to previous assumptions of a single common origin for the enzymes that catalyze equivalent reactions within a pathway.

Fothergill-Gilmore (1986) suggested that the  $\beta\alpha$ -barrels of triosephosphate isomerase (TPI, EC 5.3.1.1) and pyruvate kinase (PK, EC 2.7.1.40) are in fact the consequence of physicochemical constraints that favor the chemistry of the reactions, and not the direct outcome of evolutionary heritage. The nature of the reaction mechanisms, sequence comparisons, and domain architecture analyses of these enzymes seem to support the latter, but comparisons between the crystallographic structures strongly suggest a common origin between these proteins. The overall connectivity of the  $\beta\alpha$ -elements between the TPI and the TIM-fold of the PK remains very much alike and the TIM-fold, by its own, is known to be associated with additional domains that precede, interrupt, or follow the  $\beta\alpha$ -barrel generating additional variety (Nagano et al. 2002). This seems to be the case of the PK, where the  $\beta\alpha$ -barrel remains as the catalytic part while the following fold functions as the regulatory part of the enzyme (Fig. 3). Even if the reaction mechanisms of the TPI and the PK are different (Fig. 3), the diversity of functions associated with the TIM-fold include five of the six primary classes of enzymes (oxidoreductases, transferases, hydrolases, lyases, and isomerases) (Goldman et al. 2016), which is an example of the great adaptability of the TIM-fold and its functional diversity. Sequences of the proteins that contain the TIM-fold are so diverse that even powerful primary sequence techniques, like the domain composition analyses, have overlooked the evolutionary relationships between TPI and PK. The TIM domain in the TPI and the PK domain of the PK are classified under separate Pfam Clans, although from the crystallographic analysis it seems evident that they are homologues. The issue can be explained by the presence of a 100-residues insertion that interrupts the continuity of the  $\beta\alpha$ -barrel sequence and that constitutes a separate fold within the PK structure (Fig. 3).

## Ribose

Ribose enzymatic degradation starts with a phosphorylation reaction catalyzed by ribokinase (RIBK, EC 2.7.1.15), whose universal distribution suggests that it was already present in the LCA. While sequence comparisons between glycolytic and ribose catabolic enzymes do not suggest any evolutionary relationship, domain identification does suggest common ancestry of RIBK with the glycolytic kinases AGLU (EC 2.7.1.147) and APPFK (EC 2.7.1.146), based on their respective functional domains that belong to the same Pfam clan (CL0118). Comparisons of their corresponding crystallographic structures support this hypothesis (see Fig. 2 and supplementary material or visit <https://goo.gl/fUkIBo>). Once phosphorylated, ribose 5-phosphate reacts with a molecule of xylulose 5-phosphate, releasing glyceraldehyde 3-phosphate, and sedoheptulose 7-phosphate, a reaction that is catalyzed by transketolase (TRK, EC 2.2.1.1),



**Fig. 3** Crystallographic structures of triosephosphate isomerase (TPI, EC 5.3.1.1, PDB ID 4YMZ) and pyruvate kinase (PK, EC 2.7.1.40, PDB ID 4HYV). **a** Crystallographic structure of TPI (green) associated with substrate 1,3-dihydroxyacetonephosphate (red, ligand ID 13P). Close view of the amino-acid residues Asn<sup>10</sup>, Lys<sup>12</sup>, His<sup>97</sup>, Glu<sup>169</sup>, Gly<sup>175</sup>, Ser<sup>217</sup>, Gly<sup>238</sup>, and Gly<sup>239</sup> that according to the crystallographic structure (PDB ID 4YMZ) establish hydrogen bonds to stabilize the 13P in the catalytic core. Isomerization reaction of the 1,3-dihydroxyacetonephosphate into glyceraldehyde 3-phosphate is an example of concerted acid–base catalysis, that in this case is assisted by His<sup>97</sup> acting as the acid and Glu<sup>169</sup> acting as the base. **b** Crystallographic structure of PK associated with substrate phosphoenolpyruvate (red, ligand ID PEP), magnesium ion (cyan, ligand ID MG), and potassium ion (purple, ligand ID K). Two protein folds can be appreciated in association with the  $\beta\alpha$ -barrel (green), the regulatory fold (blue) that in Nature interacts with fructose 1,6-bisphosphate, an allosteric positive regulator, but in the crystallographic

structure is replaced by fructose 2,6-diphosphate (orange, ligand ID FDP) and a fold (yellow) that interrupts the  $\beta\alpha$ -barrel with an unclear function. **b1** Close view of the amino-acid residues Arg<sup>50</sup>, Lys<sup>239</sup>, Gly<sup>264</sup>, Asp<sup>265</sup>, and Thr<sup>297</sup> that according to the crystallographic structure (PDB ID 4HYV) establish hydrogen bonds to stabilize the PEP in the catalytic core. **b2** Close view of the amino-acid residues Asn<sup>52</sup>, Ser<sup>54</sup>, Asp<sup>84</sup>, Thr<sup>85</sup>, Glu<sup>241</sup>, and Asp<sup>265</sup> that according to the crystallographic structure (PDB ID 4HYV) establish hydrogen bonds to stabilize MG and K in the catalytic core, respectively. The PK reaction couples the free energy of PEP cleavage to the generation of ATP during the synthesis of pyruvate and requires the presence of two Mg<sup>+</sup> ions and one K<sup>+</sup> ion to stabilize the negative charges of PEP and ADP within the catalytic core. Images presented here were created using The PyMOL Molecular Graphics System, Version 1.8.1.2 Schrödinger, LLC and colors correspond to the PyMOL standard color palette. See supplementary material or visit <https://goo.gl/fUk-lBo> for details on the comparisons of the crystallographic structures

whose universal distribution suggests that it was already present in the LCA. During this reaction, a molecule of xylulose 5-phosphate is required, that can also be obtained from ribose 5-phosphate, that is first transformed into ribulose 5-phosphate by phosphoriboisomerase (PRI, EC 5.3.1.6), followed by the transformation of ribulose 5-phosphate into xylulose 5-phosphate via phosphoribulose epimerase (PRE, EC 5.1.3.1). While PRI is universally distributed, PRE is restricted almost exclusively to the Bacteria and Eukarya domains. Although some methanogenic Archaea do have a

PRE homologue, its restricted distribution and close resemblance to the corresponding bacterial enzymes suggest that it was acquired by HGT (see Fig. S3). The requirement of xylulose 5-phosphate makes it difficult to think of this enzymatic ribose degradation as an ancient catabolic pathway, especially since the distribution of PRE suggests that it evolved in the Bacteria lineage after the divergence of the two major prokaryotic domains. This poses a problem, since there appears to be no other enzyme capable of synthesizing such an essential product for glyceraldehyde 3-phosphate

formation. However, the chemical nature of the ribulose-5-phosphate and the xylulose-5-phosphate suggests that since they are stereoisomers (epimers), which vary in the relative position of a hydroxyl group on the C3 (see Fig. S4), it is possible that in earlier times the TRK was endowed with a lesser specificity and also used ribulose 5-phosphate as substrate to produce glyceraldehyde 3-phosphate. This suggestion implies that the catabolic pathway that degrades ribose and produces glyceraldehyde 3-phosphate at first required only three universally distributed enzymes (RIBK, TRK, and PRI), opening up the possibility that this pathway was already present in the LCA. In favor of this hypothesis, it should be mentioned that during the complementary phase of the ribose degradation the glyceraldehyde 3-phosphate and the sedoheptulose 7-phosphate produced before, are transformed into erythrose 4-phosphate and fructose 6-phosphate in a reaction catalyzed by transaldolase (TRA, EC 2.2.1.2). Erythrose 4-phosphate then reacts with xylulose 5-phosphate to produce glyceraldehyde 3-phosphate and fructose 6-phosphate in a reaction also catalyzed by TRK (see Fig. S4). This can be seen as an example of the lack of absolute substrate specificity of TRK. During this last stage of ribose degradation, the limiting enzyme is TRA, whose distribution among Bacteria and Eukaryotes indicates that it evolved after the divergence of the two major prokaryotic domains. Although some homologues are present in the methanogenic Archaea, most probably they represent another case of HGT, as suggested by their limited distribution and close resemblance to the bacterial enzymes (see Fig. S5). Accordingly, it can be argued that the second phase of the enzymatic ribose degradation most probably evolved within the bacterial lineage.

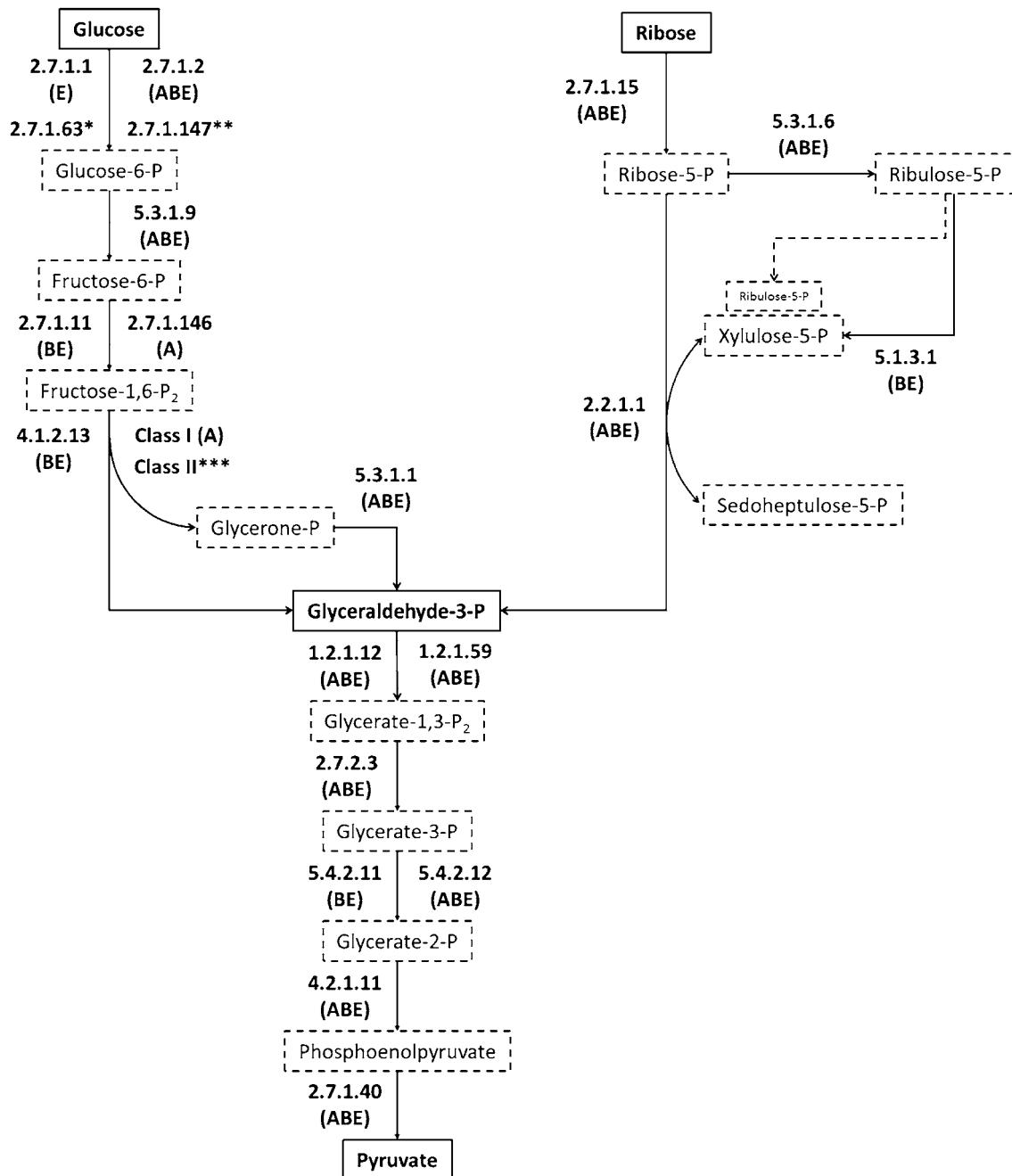
Results on the nature of the kinases that integrate glycolysis and the ribose degradation pathway support the idea that these catabolic routes were assembled by a patchwork mechanism, since they belong to five different families: (a) the actin-like ATPase superfamily that englobes the HEX, GLUK, and PGK of the first phosphorylation reaction on the glycolysis; (b) the PFK-like superfamily, that includes the PPFK, which is one of the enzymes responsible for the phosphorylation of fructose-6-phosphate on the bacterial glycolysis; (c) the ribokinase-like superfamily, that englobes the AGLU, APPFK, and RIBK that are involved in the first and second phosphorylation reaction of the glycolysis, and the first phosphorylation on the ribose catabolic pathway, respectively, (d) the phosphoglycerate kinase (EC 2.7.2.3) that defines its own orphan family (not grouped within a clan) the *PGK* (PF00162), which catalyzes the first synthesis of ATP out of 1,3-bisphospho-D-glycerate, and (e) the pyruvate kinase (EC 2.7.1.40), whose main domain (*PK*) is classified under the CL0151 clan and that synthesizes another molecule of ATP from the phosphoenolpyruvate.

Moreover, if the ribose catabolic pathway integrated by RIBK, TRK, and PRI (proposed here) was present in the LCA, it is plausible to assume that it could be feeding the second part of the glycolysis with glyceraldehyde-3-phosphate, since it is not clear that the first part of the glycolysis was already functioning at that time (see Fig. 4). The universality of the enzymes of the second part of the glycolytic pathway and the ribose catabolic pathway could therefore imply that they are vestiges of an older pathway that used ribose to produce NADH, ATP, and pyruvate. The existence of this hypothetical pathway could explain in part the high connectivity of the pentose phosphate pathway with glycolysis not as a later development (Krebs and Kornberg 1957; Krebs 1981), but as the remnant of an antique pathway. A major drawback of this hypothetical scheme is the overall efficiency of the route, since under this scenario the pyruvate, ATP, and NADH production would be cut in half in comparison to the glycolytic pathway. Even so, it is difficult to extend the antiquity of ribose degradation back in time beyond the LCA, since ribose has to be present in sufficient amounts before it feeds an ancient version of the central metabolism. Therefore, one can assume that either a highly efficient abiotic source of ribose was available or that its anabolic pathway evolved prior to this epoch.

## Nucleobases

To the best of our knowledge, the possibility that the enzymatic degradation of adenine, guanine, cytosine, uracyl, or thymine are among the oldest catabolic routes has not been addressed in the literature. Nevertheless, we have analyzed these catabolic pathways since the five nucleobases play a major role in the metabolism of all living creatures as constitutive elements of RNA, DNA, and as crucial components of bioenergetics processes. Some of them are easily formed in prebiotically plausible conditions, either as free nucleobases (Oró 1960, 1961; Ferris et al. 1968, 1978) or as activated ribonucleotides (Powner et al. 2009) and, as suggested by their presence on meteorites (Callahan et al. 2011; Saladino et al. 2011; Burton et al. 2012) they could be among the earliest components of the primitive environment. Direct uptake of such bases, together with that of many other prebiotically synthesized organic compounds may be considered as the oldest form of heterotrophy (Lazcano 2010). It is therefore possible that they were available as metabolites for the first living entities, which does not necessarily imply that extant catabolic pathways are the direct evolutionary outcome of those early forms of heterotrophy. However, some of the first steps in the enzyme-based degradations of nucleobases may have their enzyme-independent counterpart, and therefore could be envisioned as an ancient enzyme-free form of degradation.





**Fig. 4** Connectivity of the glucose and ribose catabolic pathways. Both produce the essential metabolite glyceraldehyde-3-phosphate. Key crossroads metabolites of the routes are written in bold letters and closed boxes, with other metabolites shown in dotted boxes. Enzymes are represented by their respective EC number. A for Archaea, B for Bacteria, and E for Eukaryotes indicate their biological distribution. (\*) Actinobacteria, Acidobacteria, Bacteroidetes, Verrucomicrobia, Cyanobacteria, and Deinococcus–Thermus; (\*\*) Nematodes and Vertebrates; (\*\*\*)  $\gamma$ -proteobacteria,  $\epsilon$ -proteobacteria, Actinobacteria, Spirochaetes, Bacteroidetes, Verrucomicrobia, and

Planctomycetes. The distribution of some of the enzymes that catalyze the first part of the glycolysis suggests that the pathway, as we know it today, was not operating in the LCA. As discussed in the text, we hypothesized that if the ribose catabolic pathway was sustained by ancestral versions of the ribokinase (EC 2.7.1.15), transketolase (EC 2.2.1.1), and phosphoriboisomerase (EC 5.3.1.6), then it could provide the glyceraldehyde-3-phosphate required for the second part of the glycolysis in an ancient route that could have been present in the LCA

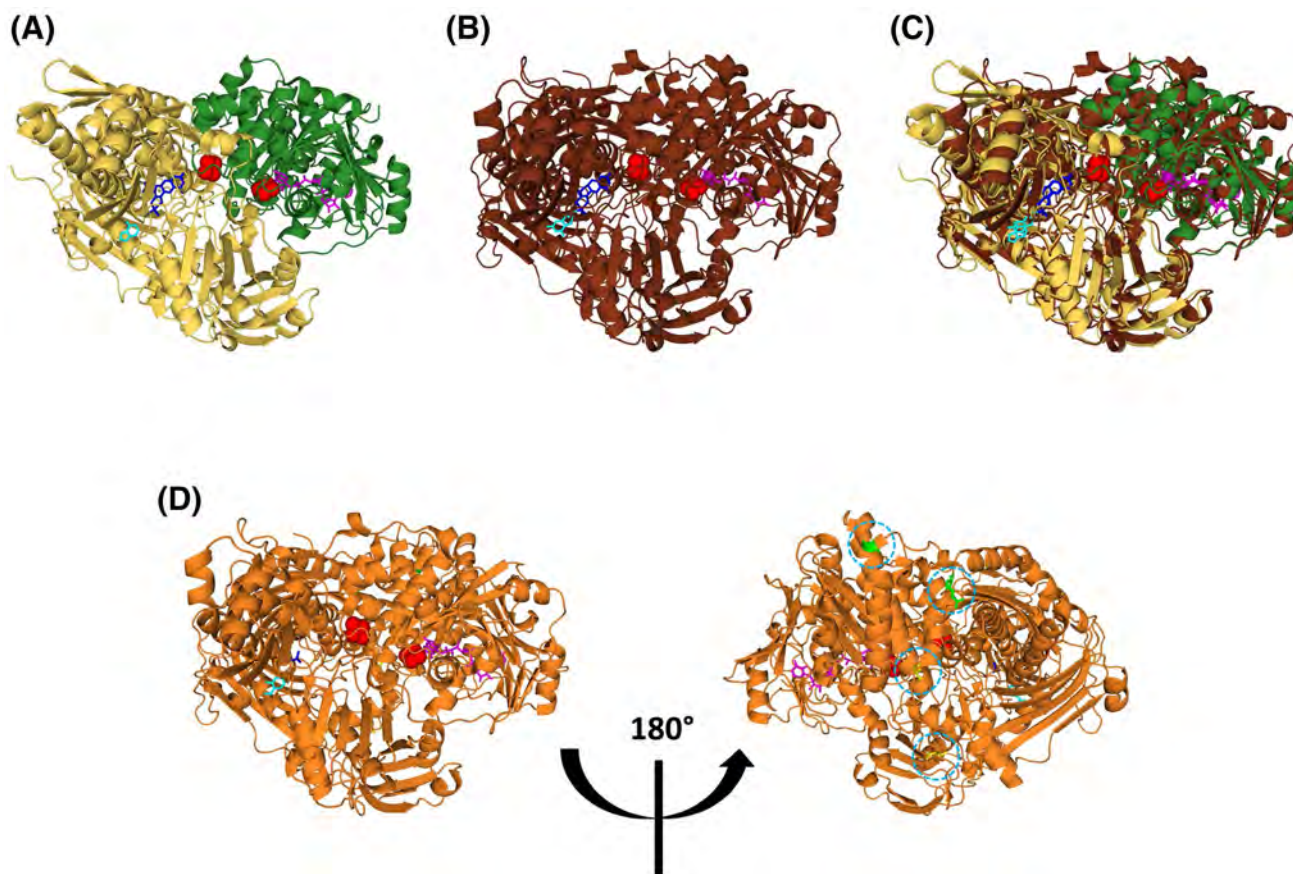
## Catabolism of Adenine and Guanine

In extant organisms, adenine and guanine catabolic pathways start with two hydrolytic reactions that produce first hypoxanthine by adenine deaminase (AD, EC 3.5.4.2), and xanthine by guanine deaminase (GD, EC 3.5.4.3), respectively. As suggested by Becerra and Lazcano (1998) and supported by the results presented here, the universal distribution of these two homologous enzymes suggests that they might share an ancestor that may have been present in the LCA. Domain composition analysis of the AD sequences revealed two distinctive architectures, one with *Amidohydro\_1* and *Adenine\_deam\_C* domains and one with the *A\_deaminase* domain. As suggested by the domain classification of *Amidohydro\_1* and *A\_deaminase* (CL0034) and confirmed by crystallographic comparisons (see supplementary material or visit <https://goo.gl/nSoRIZ>), these enzymes are homologues. Distribution patterns of both versions revealed that the AD whose domain architecture possesses *Amidohydro\_1* and *Adenine\_deam\_C* is the oldest version. By its own, GD is an enzyme that is integrated solely by the *Aminohydro\_1* domain and whose common ancestry with AD is well supported, as confirmed by crystallographic structure comparisons (see supplementary material or visit <https://goo.gl/nSoRIZ>). Moreover, the enzyme-free hydrolysis of adenine and guanine produce hypoxanthine and xanthine, respectively, in significant yields (half-lives of A and G  $\approx$  1 year) (Miller and Orgel 1974; Levy and Miller 1998) opening the possibility that enzyme-free catabolic steps are more ancient than suggested by the distribution of their respective degradative enzymes.

Extant enzyme-mediated degradation of adenine into hypoxanthine is followed by its hydrolysis to xanthine, which in turn undergoes another hydrolysis releasing urate. Both hydrolytic reactions are catalyzed either by xanthine dehydrogenase (XDH, EC 1.17.1.4) or by xanthine oxidase (XO, EC 1.17.3.2), although the latter is an oxygen-dependent enzyme that most likely evolved after the oxygen enrichment of the atmosphere. Analysis of the XDH sequences reveals that it is composed by two subunits (small and large), and that XO is a monomer of over 1000 residues long. Sequence comparison results showed that both subunits of XDH share identity values that range from 24.97 to 45.87% when compared against the XO sequences. Their domain composition reveals that the small subunit of XDH possess the *Fer2*, *Fer2\_2*, *FAD\_binding\_5*, and *Co\_deh\_flav\_C* domain architecture, while the large subunit possesses the *Ald\_Xan\_dh\_C* and *Ald\_Xan\_dh\_C2* domain architecture, and XO seems to be the direct outcome of the fusion between the large and short XDH subunits. The eukaryotic XO appears to be the product of a reversible post-translational modification (RPTM) of the XDH. In this RPTM, two cysteine disulfide bonds are formed between two couples of cysteine residues.

The first bond opens a solvent gate, while the second bond obstructs the NAD<sup>+</sup> binding cavity (Nishino et al. 2005). Analyses of the crystallographic structures of XDH and XO reveal that at least three of the four conserved cysteine residues responsible for the RPTM are the evolutionary outcome of the fusion of the two subunits of XDH, since they are located within the elongated parts of the proteins that arose as result of the fusion (see Fig. 5 and supplementary material or visit <https://goo.gl/nSoRIZ>). These data strongly suggest that XDH was the ancestral form that evolved in the bacteria lineage, and that XO is a later adaptation to the oxygenic conditions in the Earth's atmosphere. The appearance of the XDH also carried the selective advantage of harvesting the reductive power of hypoxanthine and xanthine in the form of a NADH molecule while releasing either xanthine or urate depending on the catalyzed reaction, while the oxidation of hypoxanthine or xanthine by the XO, that also releases xanthine or urea but coupled to the production of H<sub>2</sub>O<sub>2</sub> instead of NADH (see supplementary material Catabolism-Catalogue or go to <https://goo.gl/Ts1xpD>). As suggested by the crystallographic structure of the XDH, the small subunit synthesizes NADH by gathering the energy that is generated in the active site of the large subunit that is indeed the catalytic subunit (see Fig. 5). It is thus possible to envision a simpler form of the enzyme that could catalyze the reactions prior to the appearance of the small subunit. This possibility is supported by the observed distribution results, which show that the large subunit is universal, while the small subunit is restricted to the  $\alpha$ -proteobacteria,  $\beta$ -proteobacteria, and *Deinococcus*–*Thermus* bacterial groups. Thus, one must conclude that the addition of the small subunit is a relatively new event from an evolutionary perspective.

It is noteworthy that in the next step of purine degradation two non-homologous enzymes, urate oxidase (EC 1.7.3.3) and urate hydroxylase (EC 1.14.13.113), catalyze a reaction where urate is transformed into 5-hydroxyisourate in two strictly oxygen-dependent reactions that appeared to lack an oxygen-free equivalent counterpart. This suggests that this reaction was the limiting step in the evolution of purine catabolism prior to the exploitation of O<sub>2</sub> as a more potent oxidative agent by the bacterial organisms. Furthermore, the enzymes that release allantoin, allantoate, urea, CO<sub>2</sub>, and NH<sub>3</sub> appear to be restricted mostly to aerobic members of the proteobacteria phyla, as well as to several eukaryotic groups. These data correlate well with previous reports that identified these metabolites as excretion products of specific eukaryotic organisms (See Fig. S7). The only exception in the distribution patterns is the case of allantoinase (EC 3.5.2.5), that catalyzes the hydrolysis of allantoin producing allantoate. Sequences of this enzyme exhibit the same pfam architecture as the GD (EC 3.5.4.3), where the *Amidohydro\_1* domain constitutes almost the whole length of the protein, and although we created a specific profile search model,



**Fig. 5** Details and alignments of the crystallographic structures of xanthine dehydrogenase (XDH) and xanthine oxidase (XO). **a** Bacterial xanthine dehydrogenase (PDB ID 2W3R, heterodimer), catalytic subunit (yelloworange-colored) with substrate hypoxanthine (cyan sticks), cofactor MTE molecule associated with hydroxy-(dioxo)-molybdenum (blue sticks), and energy harvesting subunit (forest-green) associated with two iron-sulfur centers (red spheres) and prosthetic group FAD (magenta sticks). **b** Eukaryotic xanthine dehydrogenase (PDB ID 3AMZ, chocolate-colored monomer), associated with substrate uric acid (cyan sticks) cofactor MTE molecule covalently linked to a dioxothiomoledium (blue sticks), two iron-sulfur centers (red spheres), and prosthetic group FAD (magenta sticks). **c** Alignment of the bacterial and the eukaryotic xanthine dehydrogenase. Their respective substrates, cofactors, iron-sulfur centers, and FAD prosthetic groups are also shown. Color codes and structures are the same as in (a) and (b). As can be appreciated, the eukaryotic version is the outcome of the fusion of the two different bacterial subunits. **d** Crystal structure of the mutant eukaryotic xanthine dehydrogenase (PDB ID 1WYG, tv\_orange-colored) in which the conversion mechanism into xanthine oxidase has been described (Nishino et al. 2005). Front position shows a molecule of salicylic acid (cyan sticks) occupying the position of the natural substrates (hypoxanthine or xanthine), a molecule of acetic acid (blue sticks) in the place of the naturally occurring MTE cofactor, two iron-sulfur centers (red spheres) and a FAD molecule (magenta sticks). 180 degrees rotation over the Y axis allows the visualization of the residues involved in the conversion mechanism (Light-blue dotted circles). Mutant residues C535A and C992R (green) are involved on the solvent gate formation and mutant residues C1316C and C1324S (yellow) are involved on the obstruction of the NAD binding cavity. Images presented here were created using The PyMOL Molecular Graphics System, Version 1.8.1.2 Schrödinger, LLC implementing the CE algorithm (Shindyalov and Bourne 1998) and colors correspond to the PyMOL standard color palette. See supplementary material or visit <https://goo.gl/nSoRIZ> for details on the comparisons of the crystallographic structures

its resolution proved to be insufficient to distinguish between these homologous enzymes, and in fact we failed to establish in full the exact distribution of this particular enzyme. An alternative possibility is that the GD is a moonlighting enzyme that catalyzes an equivalent hydrolytic reaction over allantoin, although to the best of our knowledge there is no experimental evidence supporting this possibility.

The above strongly suggests that the enzyme evolution of the extant catabolic pathway of purines took place in three different stages. In the first stage, the ancestor of AD (EC 3.5.4.2) and GD (EC 3.5.4.3) evolved prior to the divergence of Archaea and Bacteria and may have thus been present in the LCA. The same probably happened to the ancestral form of the large subunit of the XDH (EC 1.17.1.4), that could catalyze the formation of xanthine out of hypoxanthine and of urate from xanthine. During the second phase, the uric acid oxidase (EC 1.7.3.3) and urate oxidase (EC 1.14.13.113) evolved as a result of the

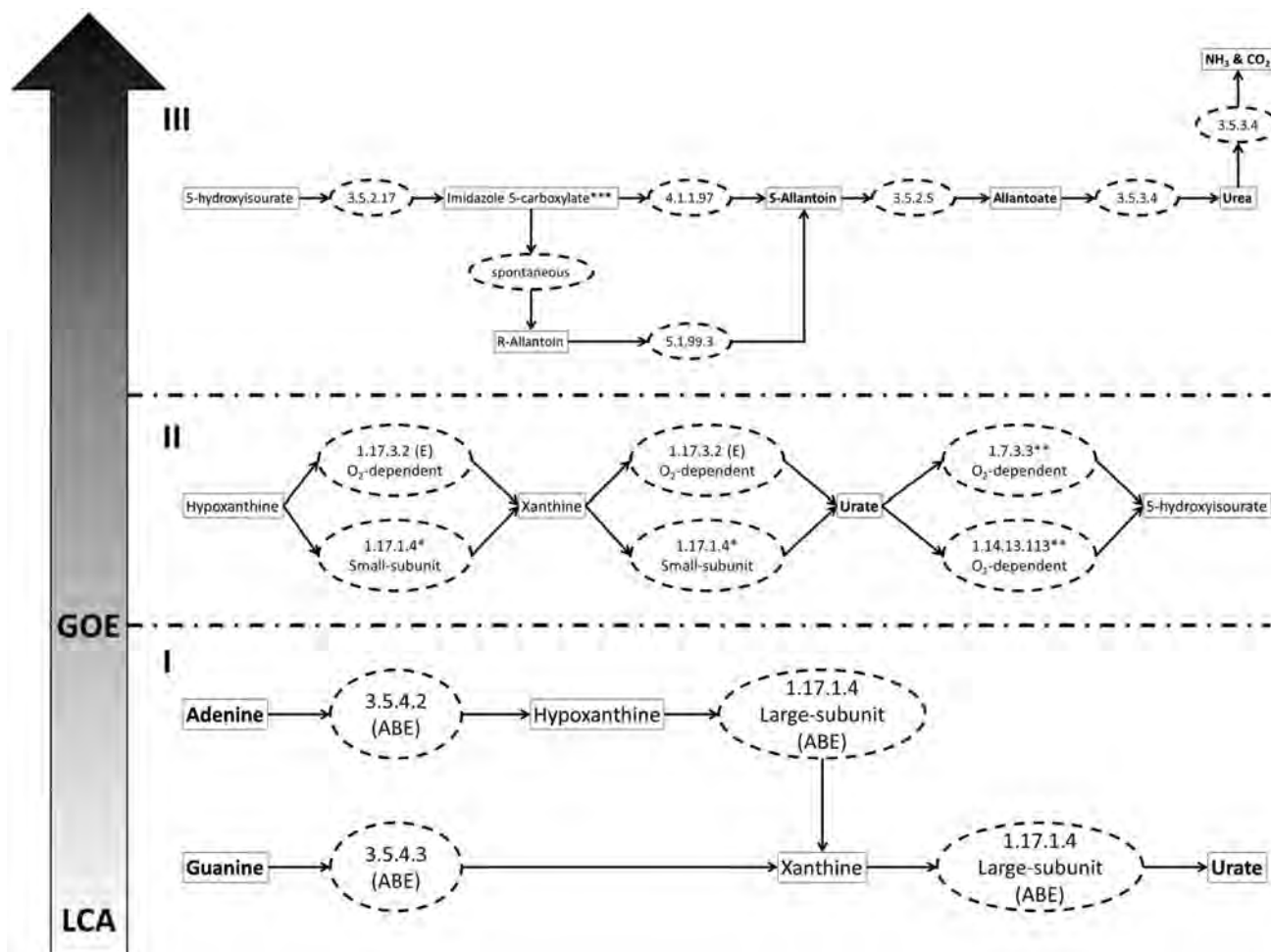
three different stages. In the first stage, the ancestor of AD (EC 3.5.4.2) and GD (EC 3.5.4.3) evolved prior to the divergence of Archaea and Bacteria and may have thus been present in the LCA. The same probably happened to the ancestral form of the large subunit of the XDH (EC 1.17.1.4), that could catalyze the formation of xanthine out of hypoxanthine and of urate from xanthine. During the second phase, the uric acid oxidase (EC 1.7.3.3) and urate oxidase (EC 1.14.13.113) evolved as a result of the

three different stages. In the first stage, the ancestor of AD (EC 3.5.4.2) and GD (EC 3.5.4.3) evolved prior to the divergence of Archaea and Bacteria and may have thus been present in the LCA. The same probably happened to the ancestral form of the large subunit of the XDH (EC 1.17.1.4), that could catalyze the formation of xanthine out of hypoxanthine and of urate from xanthine. During the second phase, the uric acid oxidase (EC 1.7.3.3) and urate oxidase (EC 1.14.13.113) evolved as a result of the

oxygenic conditions of the changing atmosphere, leading to the production of 5-hydroxyisourate. The rest of the enzymes that produce urate, allantoin, allantoate, urea, CO<sub>2</sub>, and NH<sub>3</sub>, from the 5-hydroxyisourate, evolved during the final phase within particular phylogenetic groups. This

creates a more diverse pool of compounds that nowadays are excreted by different organisms (see Figs. 6 and S7).

The possibility that anaerobic bacteria (mostly *Clostridium*) could ferment xanthine, hypoxanthine, and guanine was discussed long ago (Barker and Beck 1941). Later on, Rabinowitz et al. extended their efforts in the study of purine



**Fig. 6** The rising in the concentration of atmospheric oxygen led to the evolution of new enzymes that enable more energetically efficient metabolic routes. This is a case of purine catabolic pathway: after the great oxygenation event (GOE), enzymes including uric acid oxidase (EC 1.7.3.3) and FAD-dependent urate oxidase (EC 1.14.13.113) led to the oxidation of urate into a more diverse set of compounds that represent metabolic waste products. Closed boxes represent the metabolites along the purine catabolic pathway and waste products appear in bold letters. Dotted ellipses represent enzymes that catalyze reactions in the pathway. The oxygen dependence and the phylogenetic distribution of the purine degradative enzymes suggest that the evolution of the extant catabolic pathway took place in three different phases. In the first phase (I), the ancestor of AD (EC 3.5.4.2) and GD (EC 3.5.4.3) evolved prior to the divergence of Archaea and Bacteria. The same probably happened to the ancestral form of the large subunit of the XDH (EC 1.17.1.4) that could catalyze the formation of xanthine out of hypoxanthine and of urate from xanthine. During the second phase (II), xanthine oxidase (EC 1.17.3.2), the uric acid

oxidase (EC 1.7.3.3) and urate oxidase (EC 1.14.13.113) evolved as a result of the oxygenic conditions of the changing atmosphere, leading to the production of 5-hydroxyisourate. The rest of the enzymes that produce allantoin, allantoate, urea, CO<sub>2</sub>, and NH<sub>3</sub>, out of 5-hydroxyisourate evolved during the final phase (III) within particular phylogenetic groups, as described in the text and in Fig. S7. (\*) The small subunit of the XDH (EC 1.17.1.4) is well distributed among *Deinococcus-Thermus*,  $\alpha$ -proteobacteria,  $\beta$ -proteobacteria, and most eukaryotic groups. (\*\*) Urate oxidase (EC 1.7.3.3) is well distributed among  $\alpha$ -proteobacteria,  $\beta$ -proteobacteria, Actinobacteria, Acidobacteria, and *Deinococcus-Thermus* and most eukaryotic groups, while urate hydroxylase (EC 1.14.13.113) is well distributed among  $\alpha$ -proteobacteria,  $\beta$ -proteobacteria,  $\gamma$ -proteobacteria, Actinobacteria, and Cyanobacteria, as well as in Diatoms, Algae, Eudicots, Monocots, Cnidarians, and Vertebrates. (\*\*\*) Systematic name is *imidazole 5-carboxylate 5-Hydroxy-2-oxo-4-ureido-2,5-dihydro-1H-imidazole-5-carboxylate*

fermentations so successfully (Rabinowitz 1956; Rabinowitz and Barker 1956a, b; Rabinowitz and Pricer 1956a, b), that their work remains as a canonical reference and has served as a guide to establish the possible fermentative pathways in some metabolic databases like KEGG (Kanehisa and Goto 2000). Unfortunately, those pioneering efforts were not continued, resulting in a poorly characterized biochemistry and the lack of knowledge about the enzymes that may catalyze those fermentative reactions. As of today, there is no sequence available in public databases that allow phylogenomic analysis of those enzymes.

### Catabolism of Cytosine, Uracil, and Thymine

In extant organisms, the cytosine catabolic pathway starts with the deamination reaction catalyzed by cytosine deaminase (EC 3.5.4.1) that produces uracil and ammonia. Spontaneous deamination of cytosine is an efficient reaction that clearly preceded the evolution of cytosine deaminase and could have been the initial step in cytosine degradation. The half-life for cytosine deamination to uracil is  $\approx 340$  years at 25 °C (Shapiro 1999) or 19 days at 100 °C (Levy and Miller 1998), but these rates do not compare with the reaction yield by cytosine deaminase (CYDA), whose activity varies from 0.1 to 94  $\mu\text{m}/\text{min}/\text{mg}$  (Sakai et al. 1975; Katsuragi et al. 1986; Porter and Austin 1993). Analysis of domain composition revealed that CYDA is a single domain protein, but their sequences are divided in those that possess the *Amidohydro\_1* domain, and those that exhibit the *dCMP\_cyt\_deam\_1* domain. Sequence comparisons and domain classification (CL0034 and CL0109, respectively) suggest that both versions of the enzyme evolved independently. Results on their phylogenetic distribution suggest a universal distribution and since neither of them depend on molecular oxygen, it is quite possible that both versions were already present in the LCA.

The final product of the uracil catabolic pathway is  $\beta$ -alanine, which is a well-known crossroads metabolite that is connected to other metabolic pathways such as the pantothenate-, propanoate-, Coenzyme A-, acetyl-CoA-, and fatty acid biosynthetic pathways. The production of  $\beta$ -alanine from uracil is a three-step process catalyzed by pyrimidine reductase (PYR, EC 1.3.1.1) [or by dihydropyrimidine dehydrogenase (DPYD, EC 1.3.1.2)], dihydropyrimidinase (DHPY, EC 3.5.2.2), and  $\beta$ -ureidopropionase (UP, EC 3.5.1.6), respectively. While PYR (EC 1.3.1.1) requires the reducing power of NADH, DPYD (EC 1.3.1.2) requires NADPH to transform uracil into 5,6-dihydrouracil. Domain architecture analysis on the nature of the DPYD sequences reveals that these represent two different types of sequences annotated as EC 1.3.1.2. The first is a protein of  $\approx 681$  residues in which the *Fer4\_20* and *Pyr\_redox\_2* domains have been identified. The second is a longer sequence of  $\approx 1059$  residues with four well-identified domains, *Fer4\_20*,

*Pyr\_redox\_2*, *DHO\_dh*, and *Fer4\_21*. The last two domains in this long protein have the same domain architecture as PYR sequences (*DHO\_dh* and *Fer4\_21*), but no significant result is obtained when the two smaller versions of the proteins PYR ( $\approx 500$  residues) and DPYD ( $\approx 681$  residues) are aligned. This is consistent with the *Pyr\_redox\_2* and *DHO\_dh* domains classification in separate clans (CL0063 and CL0036, respectively). Unfortunately, there is no crystallographic structure for PYR (EC 1.3.1.1) available in the public databases that could allow crystallographic comparisons between PYR and DPYD. Based on the sequence comparisons and domain architecture analysis, it can be concluded that PYR ( $\approx 500$  residues) and DPYD ( $\approx 681$  residues) are not homologous, and that the larger version of the DPYD is the evolutionary outcome of the fusion of the shorter versions of the PYR and DPYD. The phylogenetic distribution of the PYR suggests a recent development within the bacteria lineage. It is present mostly in organisms that belong to  $\alpha$ -proteobacteria,  $\beta$ -proteobacteria,  $\gamma$ -proteobacteria, and Firmicutes clades, together with few members of the Acidobacteria, Verrucomicrobia, and Deinococcus–Thermus, and it is well represented among the major eukaryotic groups. The short version of the DPYD has been found among organisms that belong to  $\alpha$ -proteobacteria,  $\beta$ -proteobacteria,  $\gamma$ -proteobacteria,  $\delta$ -proteobacteria, Actinobacteria, Acidobacteria, and Deinococcus–Thermus bacterial groups, as well as the major eukaryotic groups, suggesting a bacterial origin. The long version of the DPYD is restricted to few organisms that belong to  $\gamma$ -proteobacteria group and very well represented within eukaryotic groups such as Vertebrates, Arthropods, Nematodes, Cnidarians, and Amoebozoa, suggesting that it evolved recently in evolutionary terms.

In the subsequent reaction, dihydropyrimidinase (DHPY, EC 3.5.2.2) hydrolyzes 5,6-dihydrouracil and releases 3-ureidopropionate. Domain architecture analysis revealed that DHPY exhibit the *Amidohydro\_1* domain, which comprises over half the residues that integrate the sequence length. Alignment of the crystallographic structures of the GD and DHPY do support the common ancestry of these proteins (see supplementary material or visit <https://goo.gl/nSoRIZ>). Phylogenetic distribution results reveal a universal distribution but, as in the case of allantoinase (EC 3.5.2.5), where its sequences exhibit the exact same pfam architecture as the GD (EC 3.5.4.3), where the *Amidohydro\_1* domain constitutes almost the whole length of the protein; the profile search model created for DHPY proved to be insufficient to distinguish between DHPY and GD and therefore we could not establish in full the exact distribution of DHPY. Finally,  $\beta$ -ureidopropionase (UP, EC 3.5.1.6) catalyzes the hydrolysis of 3-ureidopropionate, releasing  $\beta$ -alanine,  $\text{CO}_2$ , and  $\text{NH}_3$ . UP sequences were identified as possessing two different domain architectures, one that includes the *Peptidase\_M20*

as the only domain along the sequence and one with the *CN\_hydrolase* domain which comprises over half the residues that integrate the sequence length. Phylogenetic distribution results suggest that the version of the UP that possesses the *Peptidase\_M20* domain is restricted to a limited number of groups within Bacteria (mostly Proteobacteria), while the UP that possesses the *CN\_hydrolase* domain and that has a distribution that suggests that it is possible to assume that it could be present in the LCA.

Biochemical data indicate that the enzymes that degrade uracil are quite unspecific and, not surprisingly, can also produce 3-aminoisobutyrate as the final product of the thymine catabolic pathway. The 3-aminoisobutyrate can be incorporated into the last part of the valine catabolic pathway, which in turn produces succinyl-CoA. Although both PYR (EC 1.3.1.1) and DHPY (EC 1.3.1.2) catalyze oxygen-free reactions, they have a very restricted distribution mostly among proteobacterial and eukaryotic groups. Such distribution suggests a recent evolution of the extant cytosine-, uracil-, and thymine catabolic pathways. Although in the archaeal domain two members, *Sulfolobus tokodaii* and *Bathyarchaeota archaeon*, have the enzymes necessary to complete the pyrimidine degradation process, our phylogenetic analysis indicate that, like pyrimidine reductase (EC 1.3.1.1), they were most probably acquired by horizontal gene transfer from Bacteria (see Fig. S6).

Uracil and thymine degradation can also be achieved through other routes like the oxidative uracil/thymine degradation pathway (Hayaishi and Kornberg 1952; Lara 1952) that involves the action of uracil/thymine oxidase (EC 1.17.99.4), barbiturase (EC 3.5.2.1), and ureidomalonnase (EC 3.5.1.95), in this oxidative pathway uracil is reduced to malonate and urea via barbiturate. Unfortunately, this route is poorly characterized, and there are no sequences available in the public databases, associated neither with uracil–thymine oxidase (EC 1.17.99.4) nor with ureidomalonnase (EC 3.5.1.95) that could be used to study the phylogenetic distribution of the pathway. A third pathway, known as the *Rut* pathway, was discovered much later in *Escherichia coli* (Loh et al. 2006). Although the enzymes of this pathway need further characterization before being subjected to phylogenetic analyses, the fact that the pathway depends on the enzyme pyrimidine oxygenase (EC 1.14.99.46), which is an strictly oxygen-dependent catalyst, suggests that it must have evolved into its actual form only after the oxygen enrichment of the Earth's atmosphere.

## Conclusions

Glucose catabolic pathway used to be considered ubiquitous, but, as noted before, not all the essential enzymes have a universal distribution, suggesting an independent evolutionary

origin in the Archaea and the Bacteria lineages (Verhees et al. 2003; Schönheit et al. 2016). However, new insights have been obtained from our analysis of recent genomic data. As discussed here, on a first approach, the distribution of the archaeal and bacterial versions of the fructose phosphate aldolase (class I and class II, respectively, EC 4.1.2.13) could lead to the hasty conclusion that they have independent evolutionary origins. However, their respective protein domain architecture and the crystallographic structure comparisons of these proteins strongly suggest that they may share a common ancestor from before the divergence of the two major prokaryotic domains.

The analyses presented here suggest that the glycolytic pathway was still evolving at the time of the LCA, and that it underwent further adaptations within specific phylogenetic groups. These adaptations complicate our attempts to establish the relative antiquity of the pathway. For instance, the distribution of the non-homologous kinases PPFK and APPFK (EC 2.7.1.11 and EC 2.7.1.146) suggests an independent evolution within the archaeal and the bacterial lineages. Therefore, the enzymatic phosphorylation of fructose-6-phosphate still limits our understanding of the actual antiquity of the pathway and, as suggested by its phylogenetic distribution, the recruitment of the ADP-dependent glucokinase (EC 2.7.1.147), a member of the ribokinase-like superfamily of kinases, into the glycolytic pathway took place recently in the Euryarchaeota group. The presence of kinases from the ribokinase-like superfamily in archaeal metabolisms is sometimes misinterpreted as a signal of a separate evolution within different lineages and as a signal of a late development of the carbohydrate metabolism (Schönheit et al. 2016). While the distribution of the APPFK (EC 2.7.1.146), another member of the ribokinase-like superfamily of kinases, support the latter possibility, our results show that the universality of GLUK (EC 2.7.1.2) makes the presence or absence of the ADP-dependent glucokinase (EC 2.7.1.147), in the archaeal organisms, irrelevant for the discussion on the antiquity of the glycolytic pathway.

Contrary to the assumption that Archaea lack the enzymes involved in the direct phosphorylation of free pentoses, which has been interpreted as evidence that free ribose degradation is a recent evolutionary invention (Schönheit et al. 2016). As shown here, core enzymes of the ribose catabolic pathway do exhibit a universal distribution, which suggests that they were already present in the LCA. The pentose phosphate pathway principal selective advantage appears to be NADH synthesis, while ribose enzymatic degradation represents no real selective advantage on its own other than feeding the second part of the glycolysis with glyceraldehyde-3-phosphate. It is therefore possible that its ancestral function during the LCA epoch was to feed the second part of the glycolysis leading to the generation of ATP, NADH, and pyruvate through a hybrid ancestral route

(see Fig. 4). Both the first part of the glycolysis and the complementary enzymes of the ribose catabolic pathway appear to have evolved after the division of the two major phylogenetic groups, not necessarily at the same time. For instance, the emergence of transaldolase (EC 2.2.1.2), whose distribution among the Bacteria and Eukaryotes suggests that it was incorporated into the ribose catabolic pathway *after* the divergence of the two major prokaryotic domains. The selective advantage of a complementary phase of the ribose catabolic pathway may have been the condensation of the first molecule of glyceraldehyde-3-phosphate and the secondary product sedoheptulose-7-phosphate into a second molecule of glyceraldehyde-3-phosphate and fructose-6-phosphate, another product that can be incorporated into the first part of the glycolytic pathway.

On the other hand, the evidence discussed here suggested that extant adenine and guanine catabolic pathways are the outcome of a patchwork addition of enzymes at different epochs. As exemplified by the oxygen-free highly conserved enzymes adenine deaminase (EC 3.5.4.2), guanine deaminase (EC 3.5.4.3) and the large subunit of the xanthine dehydrogenase (EC 1.17.1.4) there are purine catabolic enzymes whose origin seemed to date at least from the LCA epoch. During a second epoch, the accumulation of oxygen in the atmosphere allowed the evolution of enzymes like uric acid oxidase (EC 1.7.3.3) or urate oxidase (EC 1.14.13.113), that further degrade the urate released by xanthine dehydrogenase. Finally, several enzymes like hydroxymethylglutaryl-CoA lyase (EC 3.5.2.17), allantoinase (EC 3.5.2.5), or urease (EC 3.5.1.5) appear to have evolved recently within some classes of proteobacteria or even in some eukaryotic groups (see Fig. 6 and S7).

On the other hand, not all enzymes of the cytosine, uracil, and thymine catabolic pathways are universally distributed. One exception is cytosine deaminase (EC 3.5.4.1) whose biological distribution suggests that its ancestor was already present in the LCA. Deamination of cytosine can take place in the absence of any enzymatic catalyst (Levy and Miller 1998; Shapiro 1999) suggesting that this could have been the very first way to convert cytosine to uracil. According to their limited distribution either in the proteobacteria or the acidobacteria phyla, enzymes like pyrimidine reductase (EC 1.3.1.1) and dihydropyrimidine dehydrogenase (EC 1.3.1.2) are recent developments. Therefore, contrary to recent proposals (Schönheit et al. 2016), the conversion of uracil into  $\beta$ -alanine does not represent a widespread catabolic pathway, and it is very unlikely that this pathway was in operation during the early stages of the evolution of heterotrophy.

Detailed examination of the enzyme distribution and evolution, together with the biochemical data available, has led us to estimate the relative antiquity of the glycolytic and the ribose catabolic routes, as well as the guanine, adenine, cytosine, uracil, and thymine catabolic pathways. The data

presented here strongly suggest that enzymes that are currently involved in these routes come from different epochs, and that the extant pathways are the result of patchwork recruitment throughout their evolutionary history. Accordingly, the metabolic abilities of the LCA that can be derived from our data include (a) ribose degradation coupled with the second part of the glycolytic pathway that together served as a source of ATP, NADH, and pyruvate (see Fig. 4); (b) the production of hypoxanthine, xanthine, and urate from guanine and adenine (see Fig. 6) that could serve as substrates in the synthesis of free mononucleotides, as well as parts of an ancient excretion mechanism that releases the excess of nitrogen; and (c) the production of uracil from cytosine as an efficient source of uracil to be incorporated into RNA molecules and ribonucleotide derivatives.

**Acknowledgements** Mario Rivas Medrano is a doctoral student from the Programa de Doctorado en Ciencias Biomédicas, Universidad Nacional Autónoma de México (UNAM) and received fellowship 255708 from CONACYT. Financial support to AB and AL from the PAPIIT-UNAM (IN223916) is gratefully acknowledged. We are indebted to Ricardo Hernandez-Morales for his technical support in the preparation of the manuscript. We appreciate the advice of Claudia Alvarez-Carreño while analyzing the crystallographic structures of proteins, and the programming skills of César Antonio Martínez Gutiérrez while developing the scripts and routines used in the present work.

## References

- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* 215(3):403–410
- Barker HA, Beck JV (1941) The fermentative decomposition of purines by *Clostridium acidi-urici* and *Clostridium cylindrosporium*. *J Biol Chem* 141:3–27
- Becerra A, Lazcano A (1998) The role of gene duplication in the evolution of purine nucleotide salvage pathways. *Orig Life Evol Biosph* 28(4–6):539–553
- Becerra A, Delaye L, Islas S, Lazcano A (2007) The very early stages of biological evolution and the nature of the last common ancestor of the three major cell domains. *Annu Rev Ecol Evol Syst* 38(1):361–379
- Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H et al (2000) The Protein Data Bank. *Nucleic Acids Res* 28(1):235–242
- Burton AS, Stern JC, Elsila JE, Glavin DP, Dworkin JP (2012) Understanding prebiotic chemistry through the analysis of extraterrestrial amino acids and nucleobases in meteorites. *Chem Soc Rev* 41(16):5459–5472
- Caetano-Anollés K, Caetano-Anollés G (2013) Structural phylogenomics reveals gradual evolutionary replacement of abiotic chemistries by protein enzymes in purine metabolism. *PLoS ONE* 8(3):e59300
- Caetano-Anollés G, Kim HS, Mittenenthal JE (2007) The origin of modern metabolic networks inferred from phylogenomic analysis of protein architecture. *Proc Natl Acad Sci USA* 104(22):9358–9363
- Callahan MP, Smith KE, Cleaves HJ, Ruzicka J, Stern JC, Glavin DP et al (2011) Carbonaceous meteorites contain a wide range of extraterrestrial nucleobases. *Proc Natl Acad Sci USA* 108(34):13995–13998

- Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al (2009) BLAST+: architecture and applications. *BMC Bioinformatics* 10(1):421
- Canfield DE (2005) The early history of atmospheric oxygen: homage to Robert M. Garrels. *Annu Rev Earth Planet Sci* 33(1):1–36
- Caspi R, Altman T, Billington R, Dreher K, Foerster H, Fulcher CA et al (2014) The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. *Nucleic Acids Res* 42(Database issue):D459–D471
- Chothia C, Lesk AM (1986) The relation between the divergence of sequence and structure in proteins. *EMBO J* 5(4):823–826
- Clarke PH, Elsdon SR (1980) The earliest catabolic pathways. *J Mol Evol* 15(4):333–338
- Cooper G, Kimmich N, Belisle W, Sarinana J, Brabham K, Garrel L (2001) Carbonaceous meteorites as a source of sugar-related organic compounds for the early Earth. *Nature* 414(6866):879–883
- Delaye L, Becerra A, Lazcano A (2005) The last common ancestor: what's in a name? *Orig Life Evol Biosph* 35(6):537–554
- Eddy SR (2011) Accelerated profile HMM searches. *PLoS Comput Biol* 7(10):e1002195
- Edgar RC (2004a) MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* 5:113
- Edgar RC (2004b) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32(5):1792–1797
- Embley TM, Williams TA (2015) Steps on the road to eukaryotes. *Nature* 521(7551):169–170
- Ferris JP, Sanchez RA, Orgel LE (1968) Studies in prebiotic synthesis: synthesis of pyrimidines from cyanoacetylene and cyanate. *J Mol Biol* 33(3):693–704
- Ferris JP, Joshi PC, Edelson EH, Lawless JG (1978) HCN: A plausible source of purines, pyrimidines and amino acids on the primitive earth. *J Mol Evol* 11(4):293–311
- Finn RD, Mistry J, Schuster-Böckler B, Griffiths-Jones S, Hollich V, Lassmann T et al (2006) Pfam: clans, web tools and services. *Nucleic Acids Res* 34(Database issue):D247–D251
- Forterre P, Gribaldo S (2010) Bacteria with a eukaryotic touch: a glimpse of ancient evolution? *Proc Natl Acad Sci USA* 107(29):12739–12740
- Fothergill-Gilmore LA (1986) The evolution of the glycolytic pathway. *Trends Biochem Sci* 11(1):47–51
- Fu L, Niu B, Zhu Z, Wu S, Li W (2012) CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* 28(23):3150–3152
- Geer LY, Marchler-Bauer A, Geer RC, Han L, He J, He S et al (2010) The NCBI BioSystems database. *Nucleic Acids Res* 38(Database issue):D492–D496
- Goldfine H (1965) The evolution of oxygen as a biosynthetic reagent. *J Gen Physiol* 49(1):Suppl:253–274
- Goldman AD, Beatty JT, Landweber LF (2016) The TIM barrel architecture facilitated the early evolution of protein-mediated metabolism. *J Mol Evol* 82(1):17–26
- Harris JK, Kelley ST, Spiegelman GB, Pace NR (2003) The genetic core of the universal ancestor. *Genome Res* 13(3):407–412
- Hayaishi O, Kornberg A (1952) Metabolism of cytosine, thymine, uracil, and barbituric acid by bacterial enzymes. *J Biol Chem* 197(2):717–732
- Irving JA, Whisstock JC, Lesk AM (2001) Protein structural alignments and functional genomics. *Proteins* 42(3):378–382
- Kanehisa M, Goto S (2000) KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* 28(1):27–30
- Kanehisa M, Goto S, Sato Y, Kawashima M, Furumichi M, Tanabe M (2014) Data, information, knowledge and principle: back to metabolism in KEGG. *Nucleic Acids Res* 42(Database issue):D199–D205
- Katsuragi T, Sakai T, Tonomura K (1986) Affinity chromatography of cytosine deaminase from *Escherichia coli* with immobilized pyrimidine compounds. *Agric Biol Chem* 50(7):1713–1719
- Keefe AD, Lazcano A, Miller SL (1995) Evolution of the biosynthesis of the branched-chain amino acids. *Orig Life Evol Biosph* 25(1–3):99–110
- Keller MA, Piedrafita G, Ralser M (2015) The widespread role of non-enzymatic reactions in cellular metabolism. *Curr Opin Biotechnol* 34C:153–161
- Kim KM, Caetano-Anollés G (2011) The proteomic complexity and rise of the primordial ancestor of diversified life. *BMC Evol Biol* 11(1):140
- Knoll AH, Nowak MA (2017) The timetable of evolution. *Sci Adv* 3(5):e1603076
- Krebs HA (1981) The evolution of metabolic pathways. In: Carlile JF (ed) *Molecular and cellular aspects of microbial evolution*. Cambridge University Press, Cambridge, pp 215–228
- Krebs HA, Kornberg HL (1957) *Energy transformations in living matter*. Springer, Berlin
- Lara FJS (1952) On the decomposition of pyrimidines by bacteria. II. Studies with cell-free enzyme preparations. *J Bacteriol* 64(2):279–285
- Larralde R, Robertson MP, Miller SL (1995) Rates of decomposition of ribose and other sugars: implications for chemical evolution. *Proc Natl Acad Sci USA* 92(18):8158–8160
- Lazcano A (2010) Which way to life? *Orig Life Evol Biosph* 40(2):161–167
- Lazcano A (2016) Alexandr I: oparin and the origin of life—a historical reassessment of the heterotrophic theory. *J Mol Evol* 83(5–6):214–222
- Lazcano A, Miller SL (1999) On the origin of metabolic pathways. *J Mol Evol* 49(4):424–431
- Levy M, Miller SL (1998) The stability of the RNA bases: implications for the origin of life. *Proc Natl Acad Sci USA* 95(14):7933–7938
- Loh KD, Gyaneshwar P, Markenscoff Papadimitriou E, Fong R, Kim K-S, Parales R et al (2006) A previously undescribed pathway for pyrimidine catabolism. *Proc Natl Acad Sci USA* 103(13):5114–5119
- Lyons TW, Reinhard CT, Planavsky NJ (2014) The rise of oxygen in Earth's early ocean and atmosphere. *Nature* 506(7488):307–315
- Magrane M, Consortium U (2011) UniProt knowledgebase: a hub of integrated protein data. *Database* 2011(0):bar009
- McMurry J, Begley TP (2005) *The organic chemistry of biological pathways*. Roberts and Company Publishers, Englewood
- Michal G (ed) (1998) *Biochemical pathways: an atlas of biochemistry and molecular biology (First)*. Wiley, New York
- Michal G, Schomburg D (eds) (2012) *Biochemical pathways: an atlas of biochemistry and molecular biology (Second)*. Wiley, Hoboken
- Miller SL, Orgel LE (1974) *The origins of life on the Earth*. Prentice-Hall, Englewood Cliffs
- Mirkin B, Fenner T, Galperin M, Koonin E (2003) Algorithms for computing parsimonious evolutionary scenarios for genome evolution, the last universal common ancestor and dominance of horizontal gene transfer in the evolution of prokaryotes. *BMC Evol Biol* 3(1):2
- Mistry J, Finn RD, Eddy SR, Bateman A, Punta M (2013) Challenges in homology search: HMMER3 and convergent evolution of coiled-coil regions. *Nucleic Acids Res* 41:e121
- Nagano N, Orengo CA, Thornton JM (2002) One fold with many functions: the evolutionary relationships between TIM barrel families based on their sequences, structures and functions. *J Mol Biol* 321(5):741–765
- Nishino T, Okamoto K, Kawaguchi Y, Hori H, Matsumura T, Eger BT et al (2005) Mechanism of the conversion of xanthine dehydrogenase to xanthine oxidase: identification of the two cysteine disulfide bonds and crystal structure of a



- non-convertible rat liver xanthine dehydrogenase mutant. *J Biol Chem* 280(26):24888–24894
- Oparin AI (1924) The origin of life (Russian). Moscow Worker, Moscow
- Oparin AI (1938) The origin of life (English) (First Engl). The Macmillan Company, New York
- Oró J (1960) Synthesis of adenine from ammonium cyanide. *Biochem Biophys Res Commun* 2(6):407–412
- Oró J (1961) Mechanism of synthesis of adenine from hydrogen cyanide under possible primitive Earth conditions. *Nature* 191(4794):1193–1194
- Porter DJ, Austin EA (1993) Cytosine deaminase: the roles of divalent metal ions in catalysis. *J Biol Chem* 268(32):24005–24011
- Powner MW, Gerland B, Sutherland JD (2009) Synthesis of activated pyrimidine ribonucleotides in prebiotically plausible conditions. *Nature* 459(7244):239–242
- Prlic A, Bliven S, Rose PW, Bluhm WF, Bizon C, Godzik A, Bourne PE (2010) Pre-calculated protein structure alignments at the RCSB PDB website. *Bioinformatics* 26(23):2983–2985
- Punta M, Cogill PC, Eberhardt RY, Mistry J, Tate J, Boursnell C et al (2012) The Pfam protein families database. *Nucleic Acids Res* 40(Database issue):D290–D301
- Rabinowitz JC (1956) Purine fermentation by *Clostridium cylindrosporum*: III—4-Amino-5-imidazolecarboxylic acid and 4-aminoimidazole. *J Biol Chem* 218(1):175–187
- Rabinowitz JC, Barker HA (1956a) Purine fermentation by *Clostridium cylindrosporum*: I—tracer experiments on the fermentation of guanine. *J Biol Chem* 218(1):147–160
- Rabinowitz JC, Barker HA (1956b) Purine fermentation by *Clostridium cylindrosporum*: II—purine transformations. *J Biol Chem* 218(1):161–173
- Rabinowitz JC, Pricer WE (1956a) Purine fermentation by *Clostridium cylindrosporum*: IV—4-ureido-5-imidazolecarboxylic acid. *J Biol Chem* 218(1):189–199
- Rabinowitz JC, Pricer WE (1956b) Purine fermentation by *Clostridium cylindrosporum*: V—formiminoglycine. *J Biol Chem* 222(2):537–554
- Ranea JAG, Sillero A, Thornton JM, Orengo CA (2006) Protein superfamily evolution and the last universal common ancestor (LUCA). *J Mol Evol* 63(4):513–525
- Rauchfuss H (2008) Chemical evolution and the origin of life (First). Springer, Heidelberg
- Raymond J, Blankenship RE (2004) Biosynthetic pathways, gene replacement and the antiquity of life. *Geobiology* 2(4):199–203
- Reeves RE, South DJ, Blytt HJ, Warren LG (1974) Pyrophosphate: D-fructose 6-phosphate 1-phosphotransferase: a new enzyme with the glycolytic function of 6-phosphofructokinase. *J Biol Chem* 249(24):7737–7741
- Reid C, Orgel LE (1967) Model for origin of monosaccharides: Synthesis of sugars in potentially prebiotic conditions. *Nature* 216(5114):455–455
- Ritson D, Sutherland JD (2012) Prebiotic synthesis of simple sugars by photoredox systems chemistry. *Nat Chem* 4(11):895–899
- Sakai T, Yu T-S, Taniguchi K, Omata S (1975) Purification of cytosine deaminase from *Pseudomonas aureofaciens*. *Agric Biol Chem* 39(10):2015–2020
- Saladino R, Crestini C, Cossetti C, Di Mauro E, Deamer D (2011) Catalytic effects of murchison material: prebiotic synthesis and degradation of RNA precursors. *Orig Life Evol Biosph* 41(5):437–451
- Schomburg I, Chang A, Placzek S, Söhngen C, Rother M, Lang M, ... Schomburg D (2013) BRENDA in 2013: integrated reactions, kinetic data, enzyme function data, improved disease classification: new options and contents in BRENDA. *Nucleic Acids Res* 41(Database issue):D764–D772
- Schönheit P, Buckel W, Martin WF (2016) On the origin of heterotrophy. *Trends Microbiol* 24(1):12–25
- Shapiro R (1988) Prebiotic ribose synthesis: a critical analysis. *Orig Life Evol Biosph* 18(1–2):71–85
- Shapiro R (1999) Prebiotic cytosine synthesis: a critical analysis and implications for the origin of life. *Proc Natl Acad Sci USA* 96(8):4396–4401
- Shindyalov IN, Bourne PE (1998) Protein structure alignment by incremental combinatorial extension (CE) of the optimal path. *Protein Eng* 11(9):739–747
- Subbiah S, Laurents DV, Levitt M (1993) Structural similarity of DNA-binding domains of bacteriophage repressors and the globin core. *Curr Biol* 3(3):141–148
- The UniProt Consortium (2012) Reorganizing the protein space at the universal protein resource (UniProt). *Nucleic Acids Res*, 40(Database issue), D71–D75
- The UniProt Consortium (2014) Activities at the universal protein resource (UniProt). *Nucleic Acids Res*, 42(Database issue), D191–D198
- Verhees CH, Kengen SWM, Tuininga JE, Schut GJ, Adams MWW, De Vos WM, Van Der Oost J (2003) The unique features of glycolytic pathways in Archaea. *Biochem J* 375(Pt 2):231–246
- Yang S, Doolittle RF, Bourne PE (2005) Phylogeny determined by protein domain content. *Proceedings of the National Academy of Sciences of the USA*, 102(2):373–378
- Ye Y, Godzik A (2003) Flexible structure alignment by chaining aligned fragment pairs allowing twists. *Bioinformatics* 19(Suppl 2):ii246–ii255

## Conclusiones del Capítulo 1

La glucólisis tradicionalmente había sido vista como distribuida universalmente entre los organismos contemporáneos, pero como lo sugieren los resultados de nuestro trabajo no todas las enzimas esenciales de la ruta cuentan con una distribución así. Los patrones de distribución de algunas de estas enzimas sugieren fuertemente un origen independiente al interior de los grandes grupos que integran a las Arqueas o las Bacterias. Por lo tanto, hemos sugerido que la ruta glicolítica se encontraba en un proceso de consolidación durante la época del último ancestro común (LCA).

No obstante algunas propuestas previas a nuestro trabajo, que sugerían que la ausencia de enzimas que catalizan la fosforilación de la ribosa en las Arqueas era evidencia de un origen reciente para este mecanismo degradativo, nuestros resultados muestran que de hecho estas enzimas se encuentran universalmente distribuidas y que por lo tanto es posible se encontraran entre las habilidades metabólicas del LCA. Más aún, dada la distribución errática de las enzimas de la primera fase de la glucólisis y la buena distribución de las enzimas esenciales que degradan la ribosa y que producen al gliceraldehído 3 fosfato, que es utilizado como el sustrato de la segunda fase de la glucólisis, hemos propuesto una ruta híbrida hipotética que podría haber estado operando en el LCA (Figura 4 del manuscrito).

Los resultados presentados en el presente trabajo sugieren que las rutas catabólicas contemporáneas de la adenina y la guanina son el resultado de un reclutamiento de enzimas a lo largo de la historia evolutiva de los organismos. Ello incluye aquellas que se encontraban presentes en el LCA, las que surgieron a consecuencia de la adaptación de los organismos por la acumulación del oxígeno molecular y las que se reclutaron en algunos linajes de bacterias y eucariontes, gracias a la mayor diversidad de compuestos creada por la catálisis de las enzimas oxígeno-dependientes (Figura 6 del manuscrito).

Por otro lado las enzimas que integran la ruta que degrada a la citosina, al uracilo y a la timina resultaron tener una distribución que sugiere un origen reciente entre las proteobacterias o las acidobacterias, excepto por la citosina desaminasa (EC 3.5.4.1) que estaría presente en el LCA como una forma eficiente e obtener uracilo, aunque la desaminación de la citosina es conocida por llevarse a cabo de manera espontánea en ausencia de enzimas.

El análisis detallado de la distribución y de la bioquímica de las enzimas de las rutas catabólicas contemporáneas nos ha permitido establecer la edad relativa de sus enzimas, entendiendo que todas ellas provienen de épocas diferentes de la historia de la tierra. De esta manera hemos podido establecer las posibles habilidades catabólicas del último ancestro común: a) la degradación enzimática de la ribosa, que alimenta a la segunda parte de la glucólisis con la ventaja evolutiva de una producción de ATP, NADH y piruvato, b) la producción de xantina, hipoxantina y ácido úrico a partir de adenina y /o guanina, que son utilizados como sustratos o precursores de la síntesis de mononucleótidos y c) la producción facilitada de uracilo a partir de citosina que es incorporado en el RNA y los derivados de ribonucleótidos.

## Capítulo 2

Sobre la evolución temprana de las rutas catabólicas: un estudio de genómica comparada. II. Los casos de las degradaciones enzimáticas de los veinte aminoácidos de las proteínas contemporáneas.

En el presente capítulo se presentan los resultados más recientes del proyecto doctoral “Origen y evolución de las principales rutas catabólicas” inscrito en el Programa de Doctorado en Ciencias Biomédicas. Dichos resultados están siendo utilizados en la preparación de un segundo manuscrito que será sometido a la brevedad para su revisión y posterior publicación con el título tentativo “On the early evolution of the catabolic pathways: a comparative genomics approach. II. The cases of the catabolic routes of the twenty amino acids of extant proteins”.

## Preámbulo del Capítulo 2

Como lo sugieren los resultados de distintas síntesis prebióticas y los análisis químicos de meteoritos carbonáceos como el Murchison, la presencia en la tierra prebiótica de una gran variedad de aminoácidos proteicos y no proteicos es indiscutible. En el experimento Miller la síntesis prebiótica de glicina, alanina y aspartato fue reportada por primera vez (Miller 1953). Recientemente en una variante del experimento Miller llamada “el experimento volcánico” la aparición de serina, valina y fenilalanina fue reportada de manera adicional a los aminoácidos previamente identificados en el experimento original (Johnson et al. 2008). En otra variante del experimento de Miller, que incluía al ácido sulfhídrico ( $H_2S$ ) como reactante, se reportó la síntesis de aminoácidos como treonina, metionina, isoleucina y leucina (Parker et al. 2011). Como resultado de estas variantes de síntesis prebióticas se ha incrementado el número de aminoácidos presentes en la tierra primitiva. Adicionalmente, los análisis químicos del meteorito Murchison han llevado al reconocimiento de dichos cuerpos como una fuente externa de aminoácidos como glicina, alanina, aspartato, glutamato, prolina y leucina (Pizzarello et al. 2004). Por su parte, el estudio de meteoritos como el Murray y el Orgueil, entre otros, ha revelado una gran variedad de aminoácidos proteicos y no proteicos que son estimados con una diversidad que ya supera los 90 diferentes tipos de aminoácidos que se pueden identificar en este tipo de cuerpos. El peso molecular de dichos aminoácidos proteicos es muy variado y va desde los más ligeros como la glicina hasta los más pesados como la fenilalanina (Burton et al. 2012). Finalmente, es importante hacer notar que la polimerización del ácido cianhídrico (HCN), que constituye un proceso abiótico de síntesis, permite la síntesis de aminoácidos como glicina, alanina, aspartato, valina, isoleucina y leucina (Ferris et al. 1978). Todo ello constituye una fuerte evidencia de la presencia de una gran gama de aminoácidos en la tierra primitiva, entre los cuales figuran aquellos que constituyen a las proteínas contemporáneas entre los que se encuentran glicina, alanina, aspartato, serina, valina, fenilalanina, treonina, metionina, isoleucina, leucina, glutamato y prolina.

Aunque los aminoácidos tienen diferentes estabilidades químicas, y factores como la temperatura y el pH tienen un gran impacto en su tasa natural de descomposición, los más estables suelen ser los aminoácidos alifáticos. Por ejemplo, la vida media de la alanina es de aproximadamente 3 mil millones de años a 25°C y en el caso de la glicina su vida media es de aproximadamente mil millones de años a 25°C (Alexandrova y Jorgensen 2011; Miller y Orgel 1974). Por lo general, se considera que los procesos más comunes en la descomposición química de los aminoácidos son las desaminaciones y las descarboxilaciones. En muchas de las rutas catabólicas contemporáneas que degradan a los aminoácidos proteínicos las primeras reacciones también suelen ser las desaminaciones seguidas en frecuencia por las descarboxilaciones, lo que podría indicar restricciones fisicoquímicas con las que la maquinaria enzimática de los primeros organismos tuvo que lidiar y que el metabolismo contemporáneo no ha podido reemplazar del todo. No obstante, las rutas contemporáneas son muy variadas en su bioquímica, ya que han podido explotar otros mecanismos enzimáticos que permiten una mayor diversidad de reacciones que intervienen en una ruta catabólica. Por lo general se considera que las rutas catabólicas contemporáneas que degradan aminoácidos proteicos producen metabolitos de encrucijada como el oxaloacetato o el piruvato, aunque en algunos otros casos pueden ser utilizados para dar origen a otros aminoácidos como el producto principal de su degradación (Caspi et al. 2014; McMurry y Begley 2005; Michal y Schomburg 2012).

## **Resultados**

Los resultados del presente trabajo sugieren fuertemente que las enzimas que integran a las rutas catabólicas contemporáneas y que puede considerarse se encontraban presentes en el último ancestro común (LCA) son las que participan en las degradaciones de alanina, arginina, asparagina, aspartato, cisteína, glutamato, glutamina, glicina, isoleucina, metionina, prolina, serina y treonina,

como se detalla en la Tabla 1. Así mismo, los aminoácidos cuya degradación enzimática se consolidó de manera posterior a la divergencia entre Arqueas y Bacterias son histidina, leucina, lisina, triptófano y valina, dado que la distribución, de las enzimas que integran a sus respectivas rutas, se encuentra claramente restringida al interior de ciertos grupos de bacterias. Del mismo modo las rutas catabólicas de fenilalanina y tirosina parecen haberse consolidado después del gran evento oxidativo (GOE), ya que la dependencia al oxígeno molecular de enzimas como la 3-carboxietilcatecol 2,3-dioxigenasa (EC 1.13.11.16) o la homogentisato 1,2-dioxigenasa (EC 1.13.11.5), ambas indispensables para las rutas catabólicas que degradan fenilalanina y tirosina respectivamente, nos hace suponer un origen reciente en un ambiente rico en O<sub>2</sub> (ver Tabla 1).

Universales	No-universales	Dependientes de O <sub>2</sub>
Alanina (A1 y A2) Arginina (R1) Asparagina (N1, N2, N3 y N4) Aspartato (D1, D5, D6 y D7) Cisteína (C1, C2 y C6) Glutamato (E1, E2, E3, E4 y E5) Glutamina (Q1, Q3, Q5 y Q6) Glicina (G3* y G4*) Isoleucina (I2 e I4) Metionina (M1 y M2) Prolina (P1) Serina (S1) Treonina (T1, T2, T3 y T4)	Alanina (A3 y A4) Arginina (R2) Asparagina (N5) Cisteína (C3, C4, C7 y C8) Glutamina (Q2 y Q4) Histidina (H1 y H2) Isoleucina (I1 e I3) Leucina (L1 y L2) Lisina (K1 y K3) Serina (S2 y S3) Triptófano (W1) Valina (V1, V2, V3 y V4)	Aspartato (D2, D3** y D4**) Cisteína (C5) Glicina (G1 y G2**) Lisina (K2) Fenilalanina (F1 y F2) Prolina (P2) Triptófano (W2 y W3) Tirosina (Y1)

Tabla 1. Clasificación por antigüedad relativa de las rutas catabólicas que degradan a los veinte aminoácidos canónicos de las proteínas contemporáneas. La distribución de los homólogos en los diversos grupos filogenéticos de Arqueas y Bacterias, así como la dependencia al oxígeno molecular de las enzimas fueron los criterios que nos llevaron a establecer esta clasificación. Las claves entre paréntesis corresponden a las rutas catabólicas como se muestran en el Catálogo del Catabolismo (Anexo 1). (\*) Rutas en las que la glicina se utiliza como sustrato para la síntesis de L-serina. (\*\*) Rutas en las que la distribución de las enzimas oxígeno dependientes parece ser universal. Ambos casos se describen a detalle en el texto.

A continuación se presenta una descripción detallada sobre las enzimas, las reacciones, los hallazgos sobre la distribución de sus homólogos, la dependencia con el oxígeno molecular y su contraste con los patrones de distribución, así como también las posibles relaciones evolutivas que guardan entre sí las enzimas catabólicas (derivadas de las comparaciones a nivel de estructura primaria) de cada una de las rutas catabólicas de los veinte aminoácidos canónicos de las proteínas contemporáneas.

*Sobre los aminoácidos que presentan por lo menos una ruta catabólica que se especula se encontraba presente en el último ancestro común (LCA).*

### **Alanina**

Como resultado de su desaminación el producto principal de la degradación enzimática de la L-alanina es el piruvato (ver Anexo 1, rutas A1 a A4). Existen cuatro enzimas capaces de facilitar tal desaminación: I) la L-alanina: glioxilato aminotransferasa (EC 2.6.1.44), cuya actividad de transaminasa produce piruvato y glicina, II) la L-alanina: 2-oxoglutarato aminotransferasa (EC 2.6.1.2), cuya actividad de transaminasa produce piruvato y L-glutamato, III) la L-alanina: NAD<sup>+</sup> oxidoreductasa (EC 1.4.1.1), cuya actividad de oxidoreductasa produce piruvato, NH<sub>3</sub> y NADH; y IV) la D-alanina aminotransferasa (EC 2.6.1.21), cuya actividad de transferasa produce piruvato y D-glutamato, aunque para ello se requiere que su actividad sea precedida de una isomerasa llamada alanina racemasa (EC 5.1.1.1) que produce D-alanina a partir de la L-alanina. Las comparaciones a nivel de estructura primaria entre las enzimas que son capaces de degradar a la L-alanina mientras producen piruvato no permitieron especular sobre las posibles relaciones de ancestría-descendencia entre ellas y han sido clasificadas como no homólogas, aunque la distribución universal de las aminotransferasas (EC 2.6.1.44, EC 2.6.1.2 y EC 2.6.1.21) parece indicar que probablemente se encontraban presentes en el último ancestro común (LCA) y que la oxidoreductasa (EC 1.4.1.1) se originó de



manera posterior a la divergencia de los dos grandes grupos de procariontes, dado que su distribución la ubica únicamente entre los grandes grupos de Bacterias y Eucariontes, lo que hace suponer un origen posterior a la divergencia de los dos grandes linajes celulares, Arqueas y Bacterias.

### ***Arginina***

Existen por lo menos dos rutas que degradan a la L-arginina y que respectivamente producen dos aminoácidos proteínogénicos el L-glutamato y la L-prolina (ver Anexo 1, rutas R1 y R2). La primera degradación enzimática produce L-glutamato y poder reductor en forma de una molécula de NADH, mientras que en la segunda degradación enzimática se produce L-prolina y NH<sub>3</sub>. En ambas rutas la hidrólisis de la cadena lateral de la L-arginina es la primera reacción. Tal hidrólisis constituye una descarboxilación que es facilitada por la acción de la arginasa (EC 3.5.3.1), misma que produce L-ornitina y urea. Posteriormente la L-ornitina puede ceder el grupo amino de la cadena lateral a una molécula de 2-oxoglutarato por la actividad de la ornitina aminotransferasa (EC 2.6.1.13) misma que produce L-glutamato 5-semialdehído y L-glutamato. En cambio si la L-ornitina es tomada como sustrato de la ornitina ciclasa (EC 4.3.1.12) se produce L-prolina mientras se libera el grupo amino de la cadena lateral, este paso constituye el fin de la segunda ruta degradativa (ver Anexo 1, ruta R2).

Cuando el L-glutamato 5-semialdehído se deshidrata de manera espontánea se produce una molécula de H<sub>2</sub>O y (S)-1-pirrolina-5-carboxilato, que finalmente es tomado por la L-pirrolina-5-carboxilato oxidoreductasa dependiente de NAD<sup>+</sup> (EC 1.2.1.88) y cuya actividad produce L-glutamato y una molécula de NADH, productos que constituyen el fin de la primer ruta catabólica (ver Anexo 1, R1). La comparación a nivel de estructura primaria de las enzimas que degradan a la L-arginina no permite especular sobre las relaciones de ancestría-descendencia que pudieran existir entre ellas, sin embargo, los resultados de la distribución de las enzimas que participan en ambas rutas degradativas sugieren que la ruta que produce L-glutamato y NADH, como productos finales, era parte de las habilidades metabólicas del último ancestro común (LCA), ya que las enzimas que participan

en dicha ruta se distribuyen en los tres grandes linajes celulares Arqueas, Bacterias y Eucariontes, mientras que la distribución de las enzimas de la ruta que produce L-prolina sugieren que se originó de manera posterior a la divergencia entre Arqueas y Bacterias debido a que la ornitina ciclasa (EC 4.3.1.12), enzima indispensable para esta ruta, se encuentra ampliamente distribuida entre los grandes grupos de Arqueas y Eucariontes.

### ***Asparagina***

Existen cinco degradaciones enzimáticas de la L-asparagina (ver Anexo 1, rutas N1 a N5), cuyos productos principales son el oxaloacetato y el L-aspartato. En la primera ruta la actividad catalítica de la asparagina-oxoácido transaminasa (EC 2.6.1.14) produce 2-oxosuccinamato y un aminoácido inespecífico, dado que esta enzima utiliza como sustrato para la transaminación cualquier oxoácido disponible. Posteriormente, el 2-oxosuccinamato es hidrolizado por la enzima omega-amidasa (EC 3.5.1.3), cuya actividad produce oxaloacetato y NH<sub>3</sub>. Desafortunadamente la enzima asparagina-oxoácido transaminasa (EC 2.6.1.14) ha sido poco caracterizada, al punto que no existe una secuencia de su estructura primaria que permita la búsqueda de sus homólogos en la base de datos de proteomas. Por lo tanto, la distribución de esta ruta no se pudo establecer con certeza y se asumió una distribución hipotética basada únicamente en los datos de la distribución de la enzima omega-amidasa (EC 3.5.1.3), que se encuentra presente en los tres grandes dominios Arqueas, Bacterias y Eucariontes, lo que hace suponer se encontraba presente en el último ancestro común (LCA).

En las dos degradaciones siguientes se produce L-aspartato por la desaminación de la L-asparagina, producto de la catálisis de dos enzimas no homólogas, la asparaginasa (EC 3.5.1.1) y la glutaminasa-asparaginasa (EC 3.5.1.38), cuya distribución en los tres grandes dominios hace suponer que ambas se encontraban presentes en el LCA. Finalmente, dos enzimas no homólogas cuya catálisis es considerada de síntesis pero que hipotéticamente pueden llevar a cabo la degradación de la L-asparagina, al invertir la dirección de la reacción que tradicionalmente se les asigna, son la asparagina sintetasa (EC 6.3.5.4) y la

aspartato-amonio ligasa (EC 6.3.1.1) que utilizarían la energía de la hidrólisis de la L-asparagina al liberar el grupo  $\text{NH}_3$  de la cadena lateral y así sintetizar ATP y L-aspartato. La distribución de estas dos últimas enzimas revela que la asparagina sintetasa (EC 6.3.5.4) se encuentra distribuida en los tres grandes dominios, y por lo tanto se puede asumir era parte de las habilidades metabólicas del LCA, mientras que la aspartato-amonio ligasa (EC 6.3.1.1) tiene una distribución muy limitada entre algunos grupos de bacterias como las  $\gamma$ -proteobacteria y los firmicutes, así como en algunos grupos de eucariontes como euglenas y diatomeas, por lo que especulamos tiene un origen reciente.

### **Aspartato**

La degradación enzimática del L-aspartato ocurre en los organismos contemporáneos mediante siete diferentes mecanismos enzimáticos (ver Anexo 1, rutas D1 a D7) que producen oxaloacetato o fumarato, respectivamente. Una primera degradación consiste en la transaminación del L-aspartato hacia una molécula de 2-oxoglutarato por la acción de la enzima aspartato transaminasa (EC 2.6.1.1) resultando en la liberación de oxaloacetato y L-glutamato como productos de la reacción. La distribución de esta enzima entre los grupos de Arqueas, Bacterias y Eucariontes hace suponer que esta transaminasa estaba presente en el último ancestro común (LCA). La L-aminoácido oxidasa (EC 1.4.3.2) y la L-aspartato oxidasa (EC 1.4.3.16) son un par de oxidoreductasas no homólogas dependientes de oxígeno molecular que representan las dos siguientes degradaciones del L-aspartato, siendo el oxaloacetato, el  $\text{NH}_3$  y el  $\text{H}_2\text{O}_2$  los productos de su catálisis, y aunque son enzimas cuyo origen parece estar ligado al aumento en la concentración de oxígeno molecular en la atmósfera de la Tierra, solo la L-aminoácido oxidasa (EC 1.4.3.2) cuenta con una distribución restringida al interior de algunos grupos de bacterias como actinobacterias, acidobacterias y cianobacterias, así como en algunos grupos eucariontes como vertebrados, cnidarios, algas, ascomicetos y basidiomicetos, lo que respalda un origen reciente posterior al gran evento oxidativo (GOE), mientras que la distribución de la L-aspartato oxidasa (EC 1.4.3.16) indican un origen más antiguo que se remonta

hasta el LCA. Lo anterior parece contradecir a la hipótesis sobre el origen de las enzimas dependientes de oxígeno molecular, pero no debemos olvidar que es posible que la actividad catalítica de esta enzima no haya cambiado completamente con el incremento del oxígeno molecular en la atmósfera terrestre, pero que se adaptó a un nuevo agente oxidante, más eficiente, como lo es el oxígeno molecular en lugar de otras moléculas orgánicas como el fumarato, mismo que ha sido reportado *in vitro* como potencial agente oxidante en la misma L-aspartato oxidasa (EC 1.4.3.16) de *Escherichia coli* (Tedeschi et al 1996), un organismo aerobio facultativo.

Por otro lado, la isomerización del L-aspartato en D-aspartato, que es una reacción facilitada por la aspartato racemasa (EC 5.1.1.13) al producir D-aspartato junto con su desaminación oxidativa catalizada por la D-aspartato oxidasa (EC 1.4.3.1), representa una ruta degradativa adicional, misma que produce oxaloacetato,  $\text{NH}_3$  y  $\text{H}_2\text{O}_2$ . Aunque la distribución de la oxidasa parecer ser universal, en alguna medida puede tratarse de un caso parecido al de la L-aspartato oxidasa (EC 1.4.3.16). La distribución de la ruta se encuentra limitada por la distribución de la aspartato racemasa (EC 5.1.1.13), que se encuentra presente principalmente entre Bacterias y Eucariontes, lo que hace suponer evolucionó de manera posterior a la divergencia entre Arqueas y Bacterias. Aunque la distribución universal de la D-aspartato oxidasa (EC 1.4.3.1) nos hace cuestionarnos su relación evolutiva con la L-aspartato oxidasa (EC 1.4.3.16), los resultados derivados de la comparación a nivel de estructura primaria no nos permiten especular sobre su posible relación evolutiva.

Tanto el fumarato como el  $\text{NH}_3$  son el producto de otra degradación enzimática del L-aspartato a cargo de la enzima aspartato amoníaco-liasa (EC 4.3.1.1), cuya distribución entre los grandes grupos de Arqueas, Bacterias y Eucariontes sugiere que también se encontraba presente en el LCA. Otra forma de producir fumarato a partir de la degradación del L-aspartato es mediante la catálisis de la adenilosuccinato sintetasa (EC 6.3.4.4), que transforma una molécula de L-aspartato, junto con GTP e IMP en GDP, ortofosfato ( $\text{H}_3\text{PO}_4$ ) y N<sup>6</sup>-(1,2-dicarboxietil)-AMP, que posteriormente es tomado como sustrato por la

adenilosuccinato liasa (EC 4.3.2.2) que lo transforma en fumarato y AMP. La distribución universal de ambas enzimas sugiere que debieron haber estado presentes en el último ancestro común. Finalmente, la actividad catalítica de la arginosuccinato sintetasa (EC 6.3.4.5), que utiliza ATP, L-citrulina y L-aspartato para producir AMP, difosfato ( $H_4P_2O_7$ ) y 2-(Nomega-L-arginino)-succinato, que junto con la actividad de la arginosuccinato liasa (EC 4.3.2.1), que transforma al 2-(Nomega-L-arginino)-succinato en fumarato y L-arginina, representan otra forma de obtener fumarato a partir del L-aspartato. Dada su distribución universal ambas enzimas de esta última ruta parecen haber estado presentes en el LCA.

### ***Cisteína***

Existen por lo menos ocho formas descritas en la literatura en las que la L-cisteína se puede degradar enzimáticamente teniendo como producto principal al piruvato (ver Anexo 1, rutas C1 a C8). En primer lugar la acción de la enzima cistationina gamma liasa (EC 4.4.1.1) produce piruvato,  $NH_3$  y  $SH_2$ . La actividad catalítica de la cistationina beta liasa (EC 4.4.1.8) representa otra forma de obtener los mismos productos a partir de la degradación de la L-cisteína. La distribución de ambas enzimas sugiere que pudieron haber estado presentes en el último ancestro común (LCA). De hecho, las comparaciones de su estructura primaria muestran que existe una relación evolutiva entre ambas enzimas con valores que consideramos significativos (ver metodología) y con porcentajes de identidad que rondan el 50%. Por otro lado, la enzima cistina reductasa (EC 1.8.1.6) facilita la síntesis de L-cistina a la par de la acumulación de poder reductor en forma de NADH utilizando dos moléculas de L-cisteína y junto a la acción de la cistationina gamma liasa (EC 4.4.1.1) o la cistationina beta liasa (EC 4.4.1.8) sobre la L-cistina, producen piruvato,  $NH_3$  y tiocisteína. La antigüedad de esta última ruta parece estar acotada por la distribución de la enzima cistina reductasa (EC 1.8.1.6), presente en unos pocos grupos de bacterias como las  $\alpha$ -proteobacterias,  $\beta$ -proteobacterias y  $\gamma$ -proteobacterias, así como en grupos eucariontes como eudicotiledóneas y monocotiledóneas, lo que hace suponer un origen reciente de la enzima y por consiguiente de la ruta.

Transformar a la L-cisteína isomerizándola en D-cisteína mediante la catálisis de la enzima L-aminoácido racemasa (EC 5.1.1.10) y posteriormente producir piruvato,  $\text{NH}_3$  y  $\text{SH}_2$  mediante la catálisis de la enzima D-cisteína desulfhidrasa (EC 4.4.1.15) representa otra forma de degradación enzimática. No obstante, mientras la racemasa (EC 5.1.1.10) se encuentra principalmente distribuida en Arqueas y algunos grupos de bacterias como las  $\delta$ -proteobacterias,  $\epsilon$ -proteobacterias y actinobacterias así como en algunos grupos eucariontes como artrópodos, nematodos, algas y amebas, la distribución de la D-cisteína desulfhidrasa (EC 4.4.1.15) se encuentra acotada al interior de algunos grupos de bacterias como las  $\beta$ -proteobacterias,  $\gamma$ -proteobacterias,  $\delta$ -proteobacterias y los bacteroidetes así como en la mayoría de los grupos de eucariontes, lo que hace suponer un origen reciente para ambas enzimas y para la ruta.

Las enzimas cisteína oxidasa (EC 1.13.11.20) y transaminasa A (EC 2.6.1.1) son parte de otro mecanismo de degradación de la L-cisteína. La cisteína oxidasa (EC 1.13.11.20) utiliza oxígeno molecular para transformar a la L-cisteína en L-cisteína-sulfinato, mismo que por la acción catalítica de la transaminasa A (EC 2.6.1.1) produce L-glutamato y 3-sulfinilpiruvato, que finalmente se hidroliza de manera espontánea liberando ácido sulfuroso ( $\text{H}_2\text{SO}_3$ ) y piruvato. A pesar de la distribución universal de la transaminasa A (EC 2.6.1.1), tanto la clara dependencia al  $\text{O}_2$  de la enzima cisteína oxidasa (EC 1.13.11.20) como su distribución principalmente entre Eucariontes y algunas bacterias de divergencia reciente como las  $\beta$ -proteobacterias, nos hace suponer un origen reciente de la ruta, probablemente posterior a la acumulación del oxígeno molecular en la atmósfera del planeta. Por otro lado las enzimas transaminasa A (EC 2.6.1.1) y cisteína transaminasa (EC 2.6.1.3), dos enzimas no homólogas que utilizan a la L-cisteína como donador del grupo amino para producir L-glutamato y mercaptopiruvato, que es tomado por la 3-mercaptopiruvato sulfuro-transferasa (EC 2.8.1.2) para producir piruvato y tiocianato, representan otra ruta catabólica. Tanto la transaminasa A (EC 2.6.1.1) como la 3-mercaptopiruvato sulfuro-transferasa (EC 2.8.1.2) parecen tener una distribución universal, y por lo tanto es posible se encontraran presentes en el último ancestro común como parte de sus

habilidades metabólicas. Este no parece ser el caso de la cisteína transaminasa (EC 2.6.1.3), cuya distribución se encuentra acotada entre los principales grupos de Eucariontes y algunas bacterias como las  $\beta$ -proteobacterias y las  $\gamma$ -proteobacterias.

Otra forma de degradar a la L-cisteína es mediante la acción de la cisteína liasa (EC 4.4.1.10), que produce  $\text{SH}_2$  y L-cisteato, así como la actividad catalítica de la L-cisteato sulfo-liasa (EC 4.4.1.25) cuya acción transforma al L-cisteato en  $\text{NH}_3$ , un ion bisulfito ( $\text{HSO}_3^-$ ) y piruvato. Mientras que la distribución de la cisteína liasa (EC 4.4.1.10) parece estar acotada al grupo de las  $\alpha$ -proteobacterias, la L-cisteato sulfo-liasa (EC 4.4.1.25) se encuentra presente en algunos grupos de bacterias como las  $\alpha$ -proteobacterias,  $\beta$ -proteobacterias y  $\gamma$ -proteobacterias así como en algunas algas, lo que hace suponer un origen reciente de la ruta. La última estrategia enzimática para degradar a la L-cisteína emplea la actividad catalítica de la cisteína liasa (EC 4.4.1.10) quién, como ya se describió previamente, es capaz de generar L-cisteato mismo que es tomado como sustrato de la transaminasa A (EC 2.6.1.1), que transfiere el grupo amino a una molécula de 2-oxoglutarato para sintetizar L-glutamato liberando una molécula de 3-sulfopiruvato. El 3-sulfopiruvato es luego transformado en (2R)-3-sulfolactato, ya sea por la actividad catalítica de la D-2-hidroxiácido deshidrogenasa (EC 1.1.1.272) o por la actividad de la (2R)-3-sulfolactato deshidrogenasa (EC 1.1.1.338), que utilizan el poder reductor de una molécula de NADH o NADPH respectivamente, para llevar a cabo la catálisis. Por último, el (2R)-3-sulfolactato es transformado en piruvato y un ion bisulfito ( $\text{HSO}_3^-$ ) por la actividad de la enzima (2R)-sulfolactato sulfo-liasa (EC 4.4.1.24). En este caso la posible antigüedad de la ruta se ve limitada por la distribución de las enzimas cisteína liasa (EC 4.4.1.10) cuyos homólogos fueron localizados exclusivamente en algunas  $\alpha$ -proteobacterias, así como los homólogos de la (2R)-sulfolactato sulfo-liasa (EC 4.4.1.24) que fueron localizados en grupos de bacterias como las  $\alpha$ -proteobacterias,  $\beta$ -proteobacterias y los planctomicetos, lo que sugiere fuertemente que la ruta evolucionó de manera reciente, posiblemente entre las  $\alpha$ -proteobacterias. Los resultados de las comparaciones de estructura primaria sugieren una relación de

ancestría entre las enzimas D-2-hidroxiácido deshidrogenasa (EC 1.1.1.272) y (2R)-3-sulfolactato deshidrogenasa (EC 1.1.1.338) con valores que consideramos significativos (ver metodología) y con porcentajes de identidad que rondan el 35% siendo probablemente la D-2-hidroxiácido deshidrogenasa (EC 1.1.1.272) la forma ancestral dada su distribución universal que sugiere es más antigua que su contraparte. De igual manera, los resultados sugieren que las enzimas L-cisteato sulfo-liasa (EC 4.4.1.25) y D-cisteína desulfhidrasa (EC 4.4.1.15), participantes en dos de las rutas previamente descritas (ver Anexo 1, rutas C4 y C7), que degradan a la L-cisteína de manera independiente, guardan relaciones evolutivas que sugieren un ancestro común con valores que consideramos significativos (ver metodología) y con porcentajes de identidad que superan el 67%. Ello resulta consistente con los resultados de la distribución, mismos que revelan una distribución acotada para ambas enzimas al interior de algunos pocos grupos de Bacterias y Eucariontes, lo que a su vez sugiere un origen reciente.

### ***Glutamato***

Existen cinco formas en las que el L-glutamato es degradado enzimáticamente a la vez que se libera 2-oxoglutarato como producto principal de todas ellas (ver Anexo 1, rutas E1 a E5). La degradación enzimática del L-glutamato esencialmente consiste en su desaminación, misma que se encuentra acoplada a la acumulación de poder reductor en forma de NADH o NADPH y que puede ser catalizada en primer instancia por tres enzimas no homólogas, la glutamato deshidrogenasa dependiente de  $\text{NAD}^+$  (EC 1.4.1.2), la glutamato deshidrogenasa dependiente de  $\text{NADP}^+$  (EC 1.4.1.4) y la glutamato deshidrogenasa (EC 1.4.1.3) que puede utilizar al  $\text{NAD}^+$  o al  $\text{NADP}^+$  indistintamente como agente oxidante. Además de este mecanismo de degradación enzimática del L-glutamato se puede añadir la transaminación que libera 2-oxoglutarato cuando dos moléculas de glutamato son utilizadas para sintetizar glutamina. En dicho mecanismo una molécula de glutamato funge como aceptor del grupo amino mientras que la otra funge como donador, sintetizando una molécula de glutamina mientras se degrada una de glutamato dando lugar al



2-oxoglutarato como producto de la reacción. Esta transaminación se encuentra acoplada a la acumulación del poder reductor en una molécula de NADH o de NADPH. Dicha reacción puede ser catalizada por dos enzimas no homólogas, la glutamato sintasa dependiente de NADP<sup>+</sup> (EC 1.4.1.13) o la glutamato sintasa dependiente de NAD<sup>+</sup> (EC 1.4.1.14). La síntesis de glutamato a partir de una molécula de 2-oxoglutarato y una de glutamina mediante la inversión del poder reductor de NADH o de NADPH es posiblemente la función más comúnmente asociada a estas dos últimas enzimas. No obstante, este es un perfecto ejemplo de como una reacción puede ser vista como de síntesis o de degradación dependiendo de la molécula a la que se haga referencia, así como del sentido de la reacción mismo que depende de las condiciones de crecimiento y que hipotéticamente puede ir en sentido contrario. La comparación a nivel de estructura primaria de estas cinco enzimas capaces de producir 2-oxoglutarato y poder reductor a partir de la desaminación/transaminación del glutamato no nos permiten especular sobre las posibles relaciones evolutivas que podrían existir entre ellas. Sin embargo debido a que todas ellas se encuentran bien distribuidas en los principales grupos de Arqueas, Bacterias y Eucariontes, es altamente probable que se encontraran presentes en el último ancestro común.

### ***Glutamina***

Existen seis formas en las cuales la L-glutamina se degrada enzimáticamente, una en la que el producto final de la ruta es el 2-oxoglutarato y cinco en la que el producto es el L-glutamato (ver Anexo 1, rutas Q1 a Q6). La primer ruta está compuesta por dos enzimas la L-glutamina transaminasa (EC 2.6.1.15), que produce 2-oxoglutaramato y L-alanina en una reacción de transaminación, y la 2-oxoglutaramato amidohidrolasa (EC 3.5.1.3), que produce 2-oxoglutarato y NH<sub>3</sub> mediante la hidrólisis del grupo amino de la cadena lateral. La distribución universal de ambas enzimas hace suponer que eran parte de las habilidades metabólicas del último ancestro común (LCA). Por su parte las enzimas glutaminasa (EC 3.5.1.2) y glutaminasa-asparaginasa (EC 3.5.1.38) son dos hidrolasas, sin aparente relación evolutiva, capaces de hidrolizar a la L-

glutamina liberando L-glutamato y  $\text{NH}_3$ . De las dos, la enzima glutaminasa-asparaginasa (EC 3.5.1.38) es la que cuenta con una amplia distribución en los tres grandes dominios celulares Arqueas, Bacterias y Eucariontes, lo que permite suponer que estaba presente en el LCA. Por su parte, la glutaminasa (EC 3.5.1.2) se distribuye en grupos de bacterias tales como  $\alpha$ -proteobacterias,  $\gamma$ -proteobacterias, fusobacterias y verrucomicrobia así como en grupos de eucariontes como vertebrados, artrópodos y nematodos lo cual hace suponer un origen más bien reciente.

Dos oxidoreductasas no homólogas, la glutamato sintasa dependiente de NADPH (EC 1.4.1.13) y la glutamato sintasa dependiente de NADH (EC 1.4.1.14) son capaces de transferir al grupo amino de la cadena lateral a una molécula de 2-oxoglutarato utilizando el poder reductor de su coenzima respectiva para producir dos moléculas de L-glutamato. Como ya se ha descrito anteriormente (ver Anexo 1, rutas E4 y E5), la distribución de estas enzimas permite sugerir que ambas estaban presentes en el LCA. Finalmente, si consideramos un escenario en el que la actividad de la enzima glutamina sintetasa (EC 6.3.1.2) pudiera llevarse a cabo en sentido inverso al que tradicionalmente le ha sido asignada, es posible que la degradación de la L-glutamina produjera L-glutamato y  $\text{NH}_3$  a la vez que sintetiza una molécula de ATP. La distribución universal de la enzima glutamina sintetasa (EC 6.3.1.2) parece indicar que su actividad catalítica se encontraba entre las habilidades metabólicas del LCA.

### ***Glicina***

Existen tres formas en las que los organismos contemporáneos degradan a la glicina y en las que el producto final de la ruta es el glioxilato (ver Anexo 1, rutas G1 a G3). Las enzimas L-aminoácido oxidasa (EC 1.4.3.2) y glicina oxidasa (EC 1.4.3.19) son dos oxidoreductasas no homólogas capaces de oxidar a la glicina dando lugar a la generación de glioxilato,  $\text{NH}_3$  y  $\text{H}_2\text{O}_2$  aprovechando el potencial oxidativo de una molécula de  $\text{O}_2$ . Tanto la dependencia al oxígeno molecular como la limitada distribución de la enzima L-aminoácido oxidasa (EC 1.4.3.2), descrita previamente entre las rutas que degradan al L-aspartato, sugieren un origen más

bien reciente y que contrasta con la buena distribución de los homólogos de la enzima glicina oxidasa (EC 1.4.3.19), mismos que se encuentran presentes en los tres grandes dominios, Arqueas, Bacterias y Eucariontes, lo que sugiere su presencia en el último ancestro común (LCA). Aunque se puede especular que tal distribución se debe a una adaptación de la glicina oxidasa (EC 1.4.3.19) a las condiciones oxidantes de la atmósfera como en el caso de la L-aspartato oxidasa (EC 1.4.3.16), ya antes detallado, es importante señalar que en este caso no hemos sido capaces de encontrar evidencia experimental *in vivo* ni *in vitro* de tal actividad asociada con otro agente oxidante distinto al oxígeno molecular. En las bases de datos se tiene reporte de la enzima glicina oxidasa NAD dependiente (EC 1.4.1.10), que presenta una actividad similar a la de la glicina oxidasa (EC 1.4.3.19) asociada a la acumulación de poder reductor en forma de una molécula de NADH, desafortunadamente hasta el momento no existe asociada con este número enzimático (EC) a una sola secuencia que nos permita la comparación con las oxidoreductasas antes mencionadas o la búsqueda de la distribución de sus posibles homólogos. Finalmente, una forma distinta de producir glioxilato a partir de la glicina es mediante su transaminación que es facilitada por la enzima serina-glioxilato transaminasa (EC 2.6.1.45) quien es capaz de transferir el grupo amino de la L-glicina a una molécula de 3-hidroxipiruvato produciendo L-serina y glioxilato. La distribución universal de esta enzima sugiere que formaba parte de las habilidades metabólicas del LCA.

### ***Isoleucina***

Existen cuatro rutas catabólicas, o variantes de una sola (dependiendo las enzimas que se tomen en cuenta), que degradan a la L-isoleucina, y que la oxidan hasta formar acetil-CoA o succinil-CoA (ver Anexo 1, rutas I1 a I4). En una primer fase se elimina enzimáticamente al grupo amino de la L-isoleucina, ya sea por una desaminación oxidativa dependiente de oxígeno molecular, que es catalizada por la L-aminoácido oxidasa (EC 1.4.3.2) y que libera (S)-3-metil-2-oxopentanoato, NH<sub>3</sub> y H<sub>2</sub>O<sub>2</sub>, o mediante una transaminación catalizada por la transaminasa de los aminoácidos de cadena ramificada (EC 2.6.1.42) y que produce al (S)-3-metil-2-

oxopentanoato junto con una molécula de L-glutamato. Como es de esperarse por su dependencia al O<sub>2</sub>, la oxidasa (EC 1.4.3.2) presenta una distribución limitada, descrita previamente entre las rutas que degradan al L-aspartato, lo que contrasta con la distribución universal de la transaminasa (EC 2.6.1.42), por lo que especulamos esta representa la forma ancestral de iniciar la ruta catabólica.

Posteriormente, la enzima 3-metil-2-oxobutanoato deshidrogenasa (EC 1.2.7.7) genera en una sola reacción una molécula de CO<sub>2</sub> y el (S)-2-metilbutanoil-CoA. Otra forma de obtener al (S)-2-metilbutanoil-CoA a partir del (S)-3-metil-2-oxopentanoato es la que presentan un conjunto de tres enzimas que acoplan tres reacciones descritas a detalle en el Anexo 1 (rutas I2 e I4), y que son la deshidrogenasa de cetoácido de cadena ramificada (EC 1.2.4.4), la dihidrolipoil transacilasa (EC 2.3.1.168) y la dihidrolipoil deshidrogenasa (EC 1.8.1.4). De acuerdo con los resultados de la distribución la forma ancestral de producir al (S)-2-metilbutanoil-CoA parece ser mediante el conjunto de estas tres enzimas ya que se encuentran universalmente distribuidas, mientras que la 3-metil-2-oxobutanoato deshidrogenasa (EC 1.2.7.7) se encuentra ampliamente distribuida entre los principales grupos de Arqueas, así como en pocos grupos de bacterias, tales como las δ-proteobacterias, las acidobacterias y las bacterias hipertermófilas, lo que sugiere un origen reciente.

Posteriormente, el (S)-2-metilbutanoil-CoA es transformado en 2-metilbut-2-enoil-CoA por tres enzimas homólogas, la butiril-CoA deshidrogenasa (EC 1.3.8.1), la 2-metilacil-CoA deshidrogenasa (EC 1.3.99.12) y la acil-CoA deshidrogenasa (EC 1.3.8.7). Mientras que la butiril-CoA deshidrogenasa (EC 1.3.8.1) y la acil-CoA deshidrogenasa (EC 1.3.8.7) tienen una distribución universal y por lo tanto suponemos se encontraban presentes en el LCA, la 2-metilacil-CoA deshidrogenasa (EC 1.3.99.12) parece tener un origen más reciente dado que sus homólogos se distribuyen principalmente en las α-proteobacterias y algunos grupos eucariontes como los vertebrados, las eudicotiledóneas y las monocotiledóneas. Más tarde, el 2-metilbut-2-enoil-CoA es transformado en (S)-3-hidroxi-2-metilbutiril-CoA gracias a la actividad catalítica de la enoil-CoA hidratasa (EC 4.2.1.17), cuya distribución universal también hace suponer se encontraba

entre el repertorio de enzimas del LCA. Por su parte, el (S)-3-hidroxi-2-metilbutiril-CoA es transformado en una molécula de 2-metilacetoacetilo-CoA al mismo tiempo que se sintetiza una molécula de NADH por la actividad catalítica de dos enzimas no homólogas, la  $\beta$ -hidroxiacil deshidrogenasa (EC 1.1.1.35) y la 3-hidroxi-2-metilbutiril-CoA deshidrogenasa (EC 1.1.1.178). Mientras que la primera cuenta con una distribución universal que hace suponer se encontraba presente en el LCA, la segunda posee una distribución que se limita a unos pocos grupos de Bacterias, como las  $\alpha$ -proteobacterias,  $\beta$ -proteobacterias y actinobacterias, así como en la mayoría de los principales grupos de Eucariontes, por lo que especulamos tiene un origen reciente. Después, el 2-metilacetoacetilo-CoA es transformado en una molécula de acetil-CoA así como una de propanoil-CoA por la acción catalítica de la acetil-CoA C-acil transferasa (EC 2.3.1.16) cuya distribución universal indica que es muy probable que se encontrara en el LCA. De esta manera se obtiene el acetil-CoA que es considerado el metabolito de encrucijada de la primera parte de la ruta degradativa de la L-isoleucina.

Posteriormente, el propanoil-CoA es transformado en (S)-metilmalonil-CoA por la actividad catalítica de la propanoil-CoA carboxilasa (EC 6.4.1.3) que utiliza la energía proveniente de una molécula de ATP para llevar a cabo una carboxilación a partir de una molécula de bicarbonato ( $\text{HCO}_3^-$ ). El (S)-metilmalonil-CoA que resulta de esta última reacción es isomerizado por dos enzimas no homólogas que catalizan dos reacciones consecutivas. La metilmalonil-CoA epimerasa (EC 5.1.99.1) produce (R)-metilmalonil-CoA y la metilmalonil-CoA mutasa (EC 5.4.99.2) produce succinil-CoA. Estas tres últimas enzimas necesarias para la producción del succinil-CoA, como segundo metabolito de encrucijada de la ruta catabólica de la L-isoleucina, presentan una distribución universal por lo que especulamos el último ancestro común era capaz de aprovechar al propanoil-CoA transformándolo en succinil-CoA.

### ***Metionina***

La degradación enzimática de la L-metionina ocurre en los organismos contemporáneos en dos fases de una “única” ruta catabólica que libera L-cisteína

y succinil-CoA como principales productos (ver Anexo 1, rutas M1 y M2). En la primera fase es necesario producir L-homocisteína en una transcarboxilación catalizada por cuatro enzimas aparentemente no homólogas que ocupan diferentes sustratos con el fin de transformar a la L-metionina en L-homocisteína. Las enzimas betaína-homocisteína S-metiltransferasa (EC 2.1.1.5) y homocisteína S-metiltransferasa (EC 2.1.1.10) transfieren reversiblemente el grupo metilo unido al azufre de la cadena lateral de la L-metionina a una molécula de dimetilglicina o a otra molécula de L-metionina respectivamente. Por su parte las enzimas metionina sintasa (EC 2.1.1.13) y metionina sintasa independiente de cobalamina (EC 2.1.1.14) usualmente son vistas, como su nombre lo indica como enzimas biosintéticas, aunque en principio pueden transferir de manera reversible al mismo grupo metilo de la cadena lateral en la L-metionina, con lo cual sintetizan 5-metiltetrahidrofolato a partir de tetrahidrofolato o 5-metiltetrahidropteroiltri-L-glutamato a partir de tetrahidropteroiltri-L-glutamato respectivamente, mientras liberan L-homocisteína. Estas últimas enzimas podrían representar la forma ancestral en la que tentativamente el último ancestro común (LCA) llevaba a cabo el inicio de la ruta catabólica dada su distribución universal. Por su parte es probable que las enzimas betaína-homocisteína S-metiltransferasa (EC 2.1.1.5) y homocisteína S-metiltransferasa (EC 2.1.1.10) evolucionaran de manera posterior a la divergencia entre Arqueas y Bacterias, ya que su distribución es exclusiva de los grandes grupos de Bacterias y Eucariontes.

Después, la L-homocisteína es transformada en L-cistationina, al incorporar a una molécula de L-serina, reacción que es catalizada por la enzima cistationina  $\beta$ -sintasa (EC 4.2.1.22) y que cuenta con una amplia distribución en los grandes grupos de Arqueas, Bacterias y Eucariontes por lo que es posible que su actividad catalítica estuviera presente en el LCA. Por su parte, la L-cistationina es transformada en L-cisteína, 2-oxobutanoato y  $\text{NH}_3$  mediante la actividad catalítica de la enzima cistationina  $\gamma$ -liasa (EC 4.4.1.1), que posee una distribución universal, lo que sugiere su presencia en el LCA. La L-cisteína que se libera en este paso representa el producto de la primera fase de la ruta degradativa y puede

incorporarse a la síntesis de proteínas o ser degradada por cualquiera de las rutas catabólicas que ya se han descrito para este aminoácido.

Cuando se continúa con la oxidación del 2-oxobutanoato, que proviene de la cadena de carbonos de la L-metionina, las enzimas no homólogas formato C-acetiltransferasa (EC 2.3.1.54) y piruvato sintasa (EC 1.2.7.1) son capaces de sintetizar propanoil-CoA a partir de una molécula de 2-oxobutanoato y CoA mientras se libera ácido fórmico ( $\text{CH}_2\text{O}_2$ ). La forma ancestral de catalizar esta reacción es probablemente mediante la enzima piruvato sintasa (EC 1.2.7.1) que cuenta con una amplia distribución en los principales grupos de los dos grandes linajes celulares, Arqueas y Bacterias. En cambio, la limitada distribución, al interior de las  $\gamma$ -proteobacterias,  $\delta$ -proteobacterias y los firmicutes así como de los vertebrados y las algas, de la enzima formato C-acetiltransferasa (EC 2.3.1.54) sugiere un origen reciente. El propanoil-CoA es transformado en (S)-metilmalonil-CoA por la acción catalítica de tres enzimas homólogas, la transcarboxilasa (EC 2.1.3.1), la metilmalonil-CoA descarboxilasa (EC 4.1.1.41) y la propionil-CoA carboxilasa (EC 6.4.1.3). Como lo indica la distribución de sus homólogos, las tres enzimas parecen haber estado presentes en el último ancestro común. Finalmente, el (S)-metilmalonil-CoA es transformado en succinil-CoA por la acción catalítica, previamente descrita entre las enzimas que degradan a la L-isoleucina (ver Anexo 1, rutas I3, I4 y M2), de las isomerasas no homólogas metilmalonil-CoA epimerasa (EC 5.1.99.1) y metilmalonil-CoA mutasa (EC 5.4.99.2) cuya distribución universal sugiere su presencia en el LCA.

### ***Prolina***

Existen dos degradaciones enzimáticas de la L-prolina, una que produce como metabolito principal al L-glutamato, y otra que produce 2-oxoglutarato (ver Anexo 1, rutas P1 y P2). La primer ruta catabólica puede ser iniciada por dos enzimas no homólogas que catalizan la transformación de la L-prolina en (S)-1-pirrolina-5-carboxilato, mientras que la prolina oxidasa (EC 1.5.1.2) acumula el potencial reductor en forma de NADH, la prolina deshidrogenasa (EC 1.5.5.2) sintetiza una molécula de hidroquinona. La distribución de estas enzimas sugiere

que la forma ancestral de llevar a cabo esta reacción es mediante la actividad catalítica de la prolina oxidasa (EC 1.5.1.2), que se encuentra ampliamente distribuida en los grandes grupos de Arqueas, Bacterias y Eucariontes, lo que contrasta con la distribución de la prolina deshidrogenasa (EC 1.5.5.2) cuyos homólogos se encuentran presentes entre los grandes grupos de Bacterias y Eucariontes y que a su vez sugiere que evolucionó de manera posterior a la divergencia entre los dos grandes linajes de organismos procariontes. El (S)-1-pirrolina-5-carboxilato, por su parte, es usado como el sustrato de la L-pirrolina-5-carboxilato oxidoreductasa (EC 1.2.1.88), que acumula poder reductor en forma de otra molécula de NADH durante la formación del L-glutamato. La distribución universal de esta enzima hace suponer que también se encontraba presente en el último ancestro común (LCA). Las comparaciones de estructura primaria sugieren una relación evolutiva entre las enzimas prolina deshidrogenasa (EC 1.5.5.2) y L-pirrolina-5-carboxilato oxidoreductasa (EC 1.2.1.88) con buenos valores de comparación (ver Metodología) y porcentajes de identidad que van desde un 27% y hasta un 48%, siendo la oxidoreductasa (EC 1.2.1.88) la más antigua de las dos dado su patrón de distribución.

La segunda ruta catabólica inicia por a la actividad de la procolágeno-prolina dioxigenasa (EC 1.14.11.2) que utiliza a la L-prolina, junto a una molécula de 2-oxoglutarato y oxígeno molecular para producir hidroxiprolina (trans-hidroxi-L-prolina), succinato y CO<sub>2</sub>. Posteriormente la enzima 4-hidroxiprolina epimerasa (EC 5.1.1.8) transforma a la trans-hidroxi-L-prolina en cis-hidroxi-D-prolina. Tanto la dependencia al O<sub>2</sub> de la dioxigenasa (EC 1.14.11.2) como la particular distribución de ambas enzimas (EC 1.14.11.2 y EC 5.1.1.8), que inician la ruta, sugieren un origen reciente luego de la acumulación de oxígeno en la atmósfera de la tierra, ya que los homólogos de estas enzimas parecen estar presentes principalmente dentro de las β-proteobacterias, así como dentro de grupos eucariontes como vertebrados, artrópodos, nematodos y cnidarios. Posteriormente, la cis-hidroxi-D-prolina es tomada como sustrato de la D-aminoácido oxidasa (EC 1.4.3.3) para producir 1-pirrolina-4-hidroxi-2-carboxilato y H<sub>2</sub>O<sub>2</sub> utilizando una molécula de O<sub>2</sub>. La distribución universal de esta enzima



sugiere una historia evolutiva en la que la actividad de la enzima se ha adaptado a las condiciones oxidantes de la atmósfera terrestre, como en el caso de la L-aspartato oxidasa (EC 1.4.3.16) que ya se ha detallado con anterioridad, aunque tampoco se cuenta con evidencia experimental sobre la actividad con un agente oxidante diferente al O<sub>2</sub>. De hecho, las comparaciones a nivel de estructura primaria no han sido capaces de encontrar relaciones de ancestría/descendencia con alguna otra oxidoreductasa dependiente o independiente de O<sub>2</sub> dentro de las rutas catabólicas analizadas en el presente trabajo. Cuando ya se ha obtenido al 1-pirrolina-4-hidroxi-2-carboxilato este mismo es utilizado como sustrato de la 1-pirrolina-4-hidroxi-2-carboxilato desaminasa (EC 3.5.4.22) cuya actividad produce al 2,5-dioxopentanoato y NH<sub>3</sub>. Esta enzima también presenta una distribución limitada entre algunos pocos organismos de las β-proteobacterias así como de los planctomicetos, por lo que su origen parece ser reciente. Finalmente, el 2,5-dioxopentanoato es transformado en 2-oxoglutarato por la actividad de dos enzimas no homólogas, la NAD-aldehído deshidrogenasa (EC 1.2.1.3) y la NADP-aldehído deshidrogenasa (EC 1.2.1.26), que acumulan el poder reductor en forma de NADH o NADPH respectivamente. Mientras que la distribución de la enzima NAD-aldehído deshidrogenasa (EC 1.2.1.3) es universal y por lo tanto especulamos se encontraba entre las enzimas del LCA, la enzima NADP-aldehído deshidrogenasa (EC 1.2.1.26) cuenta con una buena distribución al interior de los principales grupos de las Arqueas, por lo que suponemos su origen es posterior a la divergencia de los dos grandes grupos de organismos procariontes. Las comparaciones de la estructura primaria de las enzimas que participan en ambas degradaciones de la L-prolina sugieren que la NAD-aldehído deshidrogenasa (EC 1.2.1.3) posee relaciones de ancestría/descendencia con la enzima L-pirrolina-5-carboxilato oxidoreductasa (EC 1.2.1.88) y por ende con la prolina deshidrogenasa (EC 1.5.5.2), que son utilizadas en la primer ruta catabólica de la L-prolina. De las tres anteriores enzimas las más antiguas por su patrón de distribución son la NAD-aldehído deshidrogenasa (EC 1.2.1.3) y la L-pirrolina-5-carboxilato oxidoreductasa (EC 1.2.1.88).

## ***Serina***

Al igual que en la degradación de la L-cisteína, las rutas catabólicas de la L-serina producen como principal metabolito al piruvato y existen tres maneras en las cuales la L-serina es degradada enzimáticamente (ver Anexo 1, rutas S1 a S3). En la primera de ellas la serina desaminasa (EC 4.3.1.19) cataliza la desaminación de la L-serina produciendo piruvato y  $\text{NH}_3$ . La distribución universal de esta enzima hace suponer que esta es la forma en la que el último ancestro común descomponía a la L-serina. Una segunda enzima la L-serina amonía liasa (EC 4.3.1.17) produce igualmente piruvato y  $\text{NH}_3$ . La amplia distribución dentro de los principales grupos de Bacterias y algunos grupos de eucariontes como cnidarios y basidiomicetes hace suponer que esta enzima evolucionó de manera posterior a la divergencia de los dos grandes grupos de organismos procariontes. Una tercer forma de degradar a la L-serina, depende de dos isomerasas no homólogas transforman a la L-serina en D-serina, la aminoácido racemasa (EC 5.1.1.10) y la serina racemasa (EC 5.1.1.18). La serina racemasa (EC 5.1.1.18) parece ser la forma ancestral de obtener al isómero dada su distribución universal, lo que no es el caso de la aminoácido racemasa (EC 5.1.1.10), cuya distribución se limita casi exclusivamente al grupo de las fusobacterias. Finalmente, la D-serina amonía liasa (EC 4.3.1.18) cataliza la liberación del grupo amino y del piruvato a partir de la D-serina. Esta última enzima parece haber evolucionado al interior de un grupo de bacterias de divergencia reciente, las  $\gamma$ -proteobacterias. Por su parte, las comparaciones de la estructura primaria de estas enzimas sugieren que la serina desaminasa (EC 4.3.1.19) y la aminoácido racemasa (EC 5.1.1.10) guardan relaciones de ancestría/descendencia, siendo la serina desaminasa (EC 4.3.1.19) la más antigua de las dos dada su distribución universal.

## ***Treonina***

La degradación enzimática de la L-treonina genera diversos intermediarios metabólicos, entre los que figuran la glicina, el acetil-CoA y el succinil-CoA, que se generan en cuatro diferentes rutas catabólicas (ver Anexo 1, rutas T1 a T4). La primer ruta catabólica es un proceso de dos reacciones que está integrado por dos

enzimas no homólogas, la L-treonina deshidrogenasa (EC 1.1.1.103), que cataliza la transformación de la L-treonina en L-2-amino-acetoacetato mientras se reduce una molécula de  $\text{NAD}^+$ , y la L-2-amino-acetoacetato CoA ligasa (EC 2.3.1.29), que cataliza la formación de acetyl-CoA y glicina. La distribución de ambas enzimas en los grandes grupos de Arqueas, Bacterias y Eucariontes permite suponer que la ruta era parte de las habilidades metabólicas del último ancestro común (LCA).

La enzima L-treonina aldolasa (EC 4.1.2.5) representa otra forma en la que la L-treonina es degradada, por una descarboxilación, produciendo glicina y acetaldehído. La distribución universal de esta enzima sugiere que también se encontraba en el repertorio de las enzimas del LCA. Para oxidar al acetaldehído aún más es necesario transformarlo primero en acetato, una reacción que puede ser catalizada por cinco enzimas, la aldehído deshidrogenasa dependiente de  $\text{NAD}^+$  (EC 1.2.1.3), la aldehído deshidrogenasa dependiente de  $\text{NADP}^+$  (EC 1.2.1.4), la aldehído deshidrogenasa que utiliza indistintamente  $\text{NAD}^+$  o  $\text{NADP}^+$  (EC 1.2.1.5), la aldehído deshidrogenasa asociada a la pirroloquinolina quinona (PQQ) (EC 1.2.99.3) y la carboxilato reductasa (EC 1.2.99.6). Los patrones de distribución sugieren que las enzimas cuya actividad catalítica se encontraba en el LCA son la aldehído deshidrogenasa dependiente de  $\text{NAD}^+$  (EC 1.2.1.3) y la aldehído deshidrogenasa que utiliza indistintamente al  $\text{NAD}^+$  o al  $\text{NADP}^+$  (EC 1.2.1.5). En cambio, la aldehído deshidrogenasa dependiente de  $\text{NADP}^+$  (EC 1.2.1.4) se encuentra distribuida principalmente entre  $\alpha$ -proteobacterias,  $\beta$ -proteobacterias,  $\gamma$ -proteobacterias y actinobacterias, lo que sugiere un origen reciente; por su parte, la aldehído deshidrogenasa asociada a la PQQ (EC 1.2.99.3) se encuentra bien distribuida dentro de los grandes grupos de las Bacterias y los Eucariontes, por lo que su origen parece ser posterior a la divergencia de los dos grandes linajes de organismos procariontes. La carboxilato reductasa (EC 1.2.99.6) presenta una distribución acotada a sólo algunos miembros de las actinobacterias, lo que sugiere un origen muy reciente al interior de este particular grupo de bacterias. Las comparaciones a nivel de estructura primaria entre las anteriores enzimas sugieren una relación evolutiva entre la aldehído deshidrogenasa dependiente de  $\text{NAD}^+$  (EC 1.2.1.3) y la aldehído

deshidrogenasa asociada a la PQQ (EC 1.2.99.3). Mientras que su distribución sugiere que la aldehído deshidrogenasa dependiente de  $\text{NAD}^+$  (EC 1.2.1.3) es probablemente la más antigua de estas dos.

El acetato que se desprende como producto de la oxidación del acetaldehído es más tarde ligado a la coenzima A (CoA), una reacción que puede llevarse a cabo por distintos mecanismos mediante tres transferasas, dos ligasas y una hidrolasa. En primer lugar la propionato CoA-transferasa (EC 2.8.3.1), la acetato CoA-transferasa (EC 2.8.3.8) y la citrato CoA-transferasa (EC 2.8.3.10) utilizan respectivamente al propanoil-CoA, al acetoacetil-CoA o al (3S)-citril-CoA como sustratos para transferir la CoA al acetato y producir acetil-CoA a la vez que se libera propanoato, acetoacetato o citrato. De estas tres transferasas la acetato CoA-transferasa (EC 2.8.3.8) representa la que probablemente se encontrará presente en el LCA dada su distribución universal, mientras que la propionato CoA-transferasa (EC 2.8.3.1) se distribuye principalmente entre las  $\alpha$ -proteobacterias,  $\beta$ -proteobacterias y  $\gamma$ -proteobacterias; y la citrato CoA-transferasa (EC 2.8.3.10) se distribuye entre las  $\gamma$ -proteobacterias, lo que sugiere que ambas tienen un origen reciente al interior de estos grupos. Las comparaciones a nivel de la estructura primaria sugieren que la propionato CoA-transferasa (EC 2.8.3.1) y la acetato CoA-transferasa (EC 2.8.3.8) cuentan con un ancestro común, pero la distribución de ambas sugiere que la más antigua entre ellas es la acetato CoA-transferasa (EC 2.8.3.8). La actividad de la acetil-CoA sintetasa (EC 6.2.1.1) y de la acetil-CoA sintetasa formadora de ADP (EC 6.2.1.13) consiste en la formación del enlace entre el acetato y la CoA a expensas de una molécula de ATP. Ambas ligasas no homólogas cuentan con una distribución universal, que sugiere que se encontraban presentes en el LCA. La formación del acetil-CoA por la actividad catalítica de la acetil-CoA hidrolasa (EC 3.1.2.1) requiere que la hidrólisis de la acetil-CoA sea llevada a cabo en sentido contrario al que normalmente se asume para esta enzima, en un hipotético escenario. No obstante, el patrón de distribución sugiere que su origen es posterior a la divergencia de Arqueas y Bacterias, por lo que no es una candidata a ser una enzima del repertorio del LCA.

La última ruta catabólica de la L-treonina libera succinil-CoA como principal metabolito de encrucijada. La ruta comienza con la desaminación de la L-treonina por parte de la treonina desaminasa (EC 4.3.1.19), que libera 2-oxobutanoato y  $\text{NH}_3$ , y cuya distribución universal hace suponer se encontraba entre las enzimas del LCA. El 2-oxobutanoato es luego transformado en propanoil-CoA por la actividad catalítica de la formato C-acetiltransferasa (EC 2.3.1.54), que libera propanoil-CoA y ácido fórmico ( $\text{CH}_2\text{O}_2$ ), o por la actividad de la piruvato sintasa (EC 1.2.7.1), que libera propanoil-CoA,  $\text{CO}_2$  y  $2\text{H}^+$  utilizando dos ferredoxinas como agentes oxidantes. De este par de enzimas no homólogas, la piruvato sintasa (EC 1.2.7.1) es quien se encuentra ampliamente distribuida entre los grandes grupos de Arqueas y Bacterias por lo que es probable la que se encontrase presente en el LCA, como ya se discutió previamente en la sección sobre la degradación de la L-metionina (ver Anexo, rutas M2 y T4). Por su parte, el propanoil-CoA puede ser transformado en (S)-metilmalonil-CoA por la acción catalítica de tres enzimas homólogas, la transcarboxilasa (EC 2.1.3.1), la metilmalonil-CoA descarboxilasa (EC 4.1.1.41) y la propionil-CoA carboxilasa (EC 6.4.1.3) que han sido descritas previamente entre las enzimas de la degradación de la L-metionina (ver Anexo, rutas M2 y T4). Finalmente el (S)-metilmalonil-CoA es transformado en succinil-CoA por la acción catalítica, previamente descrita entre las enzimas que degradan a la L-isoleucina (ver Anexo 1, rutas I3, I4 y M2), de las isomerasas no homólogas metilmalonil-CoA epimerasa (EC 5.1.99.1) y metilmalonil-CoA mutasa (EC 5.4.99.2), cuya distribución universal sugiere formaron parte de las habilidades metabólicas del LCA.

*Sobre los aminoácidos cuyas rutas catabólicas evolucionaron de manera posterior a la divergencia de los dos grandes linajes procariontes y que por lo tanto no formaban parte de las habilidades metabólicas del último ancestro común (LCA).*

### **Fenilalanina**

Las rutas catabólicas de la L-fenilalanina producen como principal metabolito al fumarato mediante dos grupos de enzimas que constituyen variantes de una estrategia común (ver Anexo 1, rutas F1 y F2). Las enzimas no homólogas fenilalanina deaminasa (EC 4.3.1.24) y fenilalanina/tirosina amonía-liasa (EC 4.3.1.25) catalizan la desaminación de la L-fenilalanina mientras se producen trans-cinamato y NH<sub>3</sub>. Los homólogos de la fenilalanina deaminasa (EC 4.3.1.24) se distribuyen principalmente entre los grupos de bacterias como las  $\delta$ -proteobacterias y los firmicutes, así como los grupos eucariontes de eudicotiledóneas, monocotiledóneas, ascomicetos, basidiomicetos y amebas. Los homólogos de la fenilalanina/tirosina amonía-liasa (EC 4.3.1.25) se encuentran en algunas  $\gamma$ -proteobacterias, así como entre las eudicotiledóneas, monocotiledóneas, ascomicetos y basidiomicetos. La distribución entre grupos bacterianos que se consideran de divergencia tardía así como entre grupos eucariontes sugiere que ambas formas de iniciar la ruta catabólica tienen un origen reciente.

A partir de este punto se siguen dos estrategias para obtener una molécula de trans-2,3-dihidroxicinamato, que es un intermediario previo a la producción de fumarato. La primera de ellas aprovecha la reacción aparentemente espontánea que transforma al trans-cinamato en trans-3-hidroxicinamato y la coordina con las capacidades catalíticas de la enzima 3-(3-hidroxifenil)propanoato hidroxilasa (EC 1.14.13.127), que utiliza el poder reductor de una molécula de NADH y el poder oxidante de una molécula de O<sub>2</sub> para producir una molécula de H<sub>2</sub>O y al trans-2,3-dihidroxicinamato. La distribución de ésta hidroxilasa (EC 1.14.13.127) sugiere su presencia en el último ancestro común (LCA), por lo que es posible que su actividad catalítica se adaptara a las condiciones oxigénicas de la atmósfera terrestre. La segunda estrategia emplea las capacidades catalíticas de la 3-

fenilpropanoato dioxigenasa (EC 1.14.12.19), que utiliza una molécula de NADH y una de O<sub>2</sub> para producir al cis-3-(3-carboxietenil)-3,5-ciclohexadieno-1,2-diol, que a su vez es transformado por la 3-(cis-5,6-dihidroxiciclohexa-1,3-dien-1-il)propanoato deshidrogenasa (EC 1.3.1.87) produciendo al trans-2,3-dihidroxicinamato mientras que restaura el poder reductor en forma de una molécula de NADH. La dependencia al oxígeno molecular y la distribución de la dioxigenasa (EC 1.14.12.19), entre las  $\alpha$ -proteobacterias,  $\beta$ -proteobacterias,  $\gamma$ -proteobacterias, actinobacterias, acidobacterias, planctomycetes y algunas bacterias verdes no-sulfurosas así como entre eudicotiledóneas, monocotiledóneas y ascomicetos, sugiere un origen posterior a la gran oxidación de la atmósfera terrestre (GOE). Ello contrasta con una distribución universal de la deshidrogenasa (EC 1.3.1.87) y que sugiere su presencia en el LCA.

Una vez que se ha obtenido al trans-2,3-dihidroxicinamato, la actividad catalítica de la 3-carboxietilcatecol 2,3-dioxigenasa (EC 1.13.11.16), que utilizando al O<sub>2</sub> como agente oxidante produce 2-hidroxi-6-cetononatrienediotato, que a su vez es hidrolizado por la actividad de la 2-hidroxi-6-oxonona-2,4-diendiotato hidrolasa (EC 3.7.1.14) y que libera una molécula de 2-hidroxi-2,4-pentadienoato y otra de fumarato. También en éste caso tanto la dependencia al oxígeno molecular como la distribución de la dioxigenasa (EC 1.13.11.16), que se limita a unos pocos organismos bacterianos dentro de las  $\alpha$ -proteobacterias, las  $\beta$ -proteobacterias, las  $\gamma$ -proteobacterias y las actinobacterias, sugiere un origen reciente posterior al GOE. No así la hidrolasa (EC 3.7.1.14) cuyos homólogos se encuentran distribuidos en los grandes grupos de Arqueas, Bacterias y Eucariontes por lo que su presencia en el LCA resulta altamente probable.

### ***Histidina***

Existen dos maneras de degradar a la L-histidina, una en la que se produce 2-oxoglutarato y otra que genera L-glutamato (ver Anexo 1, rutas H1 y H2). Ambas rutas comienzan con un par de enzimas no homólogas, la histidil amonía-liasa (EC 4.3.1.3), que produce urocanato y NH<sub>3</sub> a partir de L-histidina, y la urocanasa (EC 4.2.1.49) que hidrata al urocanato para formar 4-imidazolona-5-propanoato. La

amplia distribución de ambas enzimas entre los grandes grupos de Bacterias y Eucariontes sugiere un origen posterior a la divergencia de los dos grandes linajes celulares, Bacterias y Arqueas. A partir de este punto el 4-imidazolona-5-propanoato es degradado por dos vías distintas.

La primera es una vía se considera libre de enzimas y es donde se hidroliza espontáneamente a la molécula en dos ocasiones consecutivas, en una primera hidrólisis se libera 4-oxoglutamato, ácido fórmico ( $\text{CH}_2\text{O}_2$ ) y  $\text{NH}_3$ , mientras que en la segunda hidrólisis se liberan 2-oxoglutarato y  $\text{NH}_3$ .

En la segunda vía el 4-imidazolona-5-propanoato es tomado como sustrato de la imidazonapropionasa (EC 3.5.2.7), que hidrata a la molécula dejando como productos a un catión de hidrógeno ( $\text{H}^+$ ) y al N-formimino-L-glutamato, que a su vez es tomado como sustrato de un par de enzimas no homólogas que mediante dos diferentes mecanismos liberan al L-glutamato. La enzima formimidoilglutamasa (EC 3.5.3.8) hidroliza al N-formimino-L-glutamato produciendo formamida ( $\text{CH}_3\text{NO}$ ) y L-glutamato, mientras que la enzima glutamato formimidoil-transferasa (EC 2.1.2.5) utiliza al tetrahidrofolato como aceptor de la formamida produciendo 5-formiminotetrahidrofolato y L-glutamato. La enzima imidazonapropionasa (EC 3.5.2.7) cuenta con una amplia distribución entre los grandes grupos de Arqueas y Bacterias, por lo que es probable formará parte del catálogo de enzimas del último ancestro común. Por su parte la distribución de las enzimas que finalizan la ruta parece indicar un origen dentro de las Arqueas para la formimidoilglutamasa (EC 3.5.3.8) y un origen dentro algunos grupos bacterianos, como las acidobacterias y las bacterias hipertermófilas, para la imidazonapropionasa (EC 3.5.2.7).

### ***Leucina***

La ruta degradativa de la L-leucina produce como metabolito de encrucijada al acetil-CoA mediante dos variantes de la misma ruta (ver Anexo 1, rutas L1 y L2). Son varias las enzimas que parecen ser poco específicas y que participan también en la degradación de la L-isoleucina, así que en esta sección no se profundizará ni en sus relaciones evolutivas ni en su distribución por considerar



que ya han sido descritas en el apartado correspondiente a la L-isoleucina (mismas que serán identificadas con un asterisco junto a su respectivo número \*EC). La ruta inicia con la desaminación/transaminación de la L-leucina que es catalizada por la leucina deshidrogenasa (EC 1.4.1.9), que produce 4-metil-2-oxopentanoato y  $\text{NH}_3$  mientras acumula poder reductor en forma de una molécula de NADH, o por la transaminasa de los aminoácidos de cadena ramificada (\*EC 2.6.1.42), que ocupa una molécula de 2-oxoglutarato como aceptor del grupo amino liberando L-glutamato y 4-metil-2-oxopentanoato. La distribución universal de la deshidrogenasa (EC 1.4.1.9) sugiere que se encontraba entre el catálogo de enzimas del último ancestro común (LCA). Luego el 4-metil-2-oxopentanoato es transformado en 3-metilbutanoil-CoA ya sea por la actividad catalítica de la 3-metil-2-oxobutanoato deshidrogenasa (\*EC 1.2.7.7) en una sola reacción o por la interacción de tres enzimas, la deshidrogenasa de cetoácido de cadena ramificada (\*EC 1.2.4.4), la dihidrolipoil transacilasa (\*EC 2.3.1.168) y la dihidrolipoil deshidrogenasa (\*EC 1.8.1.4), cuya actividad se detalla en el Anexo 1 (rutas I2, I4 y L2).

En esta ruta se han descrito dos oxidoreductasas homólogas la isovaleril-CoA deshidrogenasa (EC 1.3.8.4) y la acil-CoA deshidrogenasa (\*EC 1.3.8.7), que son capaces de transformar al 3-metilbutanoil-CoA en 3-metilbut-2-enoil-CoA mientras acumulan el poder reductor en forma de FADH. La distribución de ambas enzimas sugiere que se encontraban entre el catálogo de enzimas del LCA. El 3-metilbut-2-enoil-CoA es transformado por la catálisis de la metilcrotonil-CoA carboxilasa (EC 6.4.1.4) en 3-metilglutaconil-CoA a expensas de la energía de una molécula de ATP y el carbono de un ion de bicarbonato ( $\text{HCO}_3^-$ ). La distribución universal de esta carboxilasa (EC 6.4.1.4) la coloca como una enzima que muy probablemente se encontraba presente en el LCA. El 3-metilglutaconil-CoA es hidratado por la enzima metilglutaconil-CoA hidratasa (EC 4.2.1.18) produciendo (S)-3-hidroxi-3-metilglutaril-CoA mismo que es tomado como sustrato de la enzima hidroximetilglutaril-CoA liasa (EC 4.1.3.4) que finalmente libera como productos al acetil-CoA y al acetoacetato. Los homólogos de la hidratasa (EC 4.2.1.18) están distribuidos universalmente, por lo que suponemos también se encontraba

presente en el LCA. Lo mismo no puede ser cierto para la liasa (EC 4.1.3.4) cuya distribución parece estar limitada dentro de algunos grupos de bacterias como  $\alpha$ -proteobacterias,  $\beta$ -proteobacterias,  $\gamma$ -proteobacterias,  $\delta$ -proteobacterias, actinobacterias, acidobacterias, bacteroidetes, bacterias verdes no-sulfurosas y deinococcus-thermus, así como entre los grandes grupos de Eucariontes, por lo que su origen parece ser reciente, tal vez al interior de alguno de los grupos bacterianos.

### ***Lisina***

Las rutas catabólicas que degradan a la L-lisina producen acetil-CoA como producto principal (ver Anexo 1, rutas K1 a K3). Tanto la lisina-2-oxoglutarato reductasa (EC 1.5.1.7) como la lisina-cetoglutarato reductasa (EC 1.5.1.8), dos enzimas homólogas, son capaces de producir sacaropina a partir de 2-oxoglutarato y de L-lisina, con la inversión del poder reductor de una molécula de NADH o NADPH. La primera reductasa (EC 1.5.1.7) se encuentra ampliamente distribuida entre los principales grupos de Bacterias y Eucariontes, lo que sugiere que evolucionó de manera posterior a la divergencia de los dos grandes linajes de organismos procariontes. Por su parte, la segunda reductasa (EC 1.5.1.8) cuenta con una distribución que la ubica entre los grandes grupos de Eucariontes y unos pocos organismos bacterianos que forman parte de los bacteroidetes o las cianobacterias, lo que sugiere que tiene un origen reciente. Luego, la sacaropina es tomada como sustrato de dos enzimas no homólogas que son capaces de transformarla independientemente en L-glutamato y L-2-aminoadipato 6-semialdehído mientras se regenera una molécula de NADH o NADPH, respectivamente. La sacaropina deshidrogenasa (EC 1.5.1.9) se distribuye entre grupos de bacterias como  $\alpha$ -proteobacterias,  $\delta$ -proteobacterias,  $\epsilon$ -proteobacterias, actinobacterias, bacteroidetes, verrucomicrobia y deinococcus-thermus así como entre los grandes grupos de Eucariontes lo que sugiere un origen reciente. Por otro lado, la sacaropina reductasa (EC 1.5.1.10) cuenta con una amplia distribución entre los grandes grupos de Eucariontes y entre algunos pocos organismos del grupo de los bacteroidetes, lo que sugiere un origen aún más

reciente. El L-2-aminoadipato 6-semialdehído es tomado por la aminoadipato semialdehído deshidrogenasa (EC 1.2.1.31) que libera L-2-aminoadipato mientras acumula poder reductor en una molécula de NADH o NADPH. La distribución de sus homólogos principalmente entre las  $\beta$ -proteobacterias y los bacteroidetes, así como entre los vertebrados sugiere un origen reciente para esta deshidrogenasa (EC 1.2.1.31). La 2-aminoadipato transaminasa (EC 2.6.1.39) luego transfiere al grupo amino a una molécula de 2-oxoglutarato con lo que libera L-glutamato y 2-oxoadipato. Esta transaminasa (EC 2.6.1.39) parece estar universalmente distribuida, por lo que es posible se encontrara dentro del repertorio metabólico del último ancestro común (LCA). Por su parte, dos enzimas homólogas, la 2-oxoadipato deshidrogenasa (EC 1.2.4.2) y la dihidrolipoil transuccinilasa (EC 2.3.1.61) son capaces de catalizar la síntesis de glutaril-CoA, CO<sub>2</sub> y NADH a partir del 2-oxoadipato, una molécula de la CoA y una de NAD<sup>+</sup>. La distribución universal de la transferasa (EC 2.3.1.61) parece indicar que se encontraba presente en el LCA, mientras que la deshidrogenasa (EC 1.2.4.2) se distribuye entre los grandes grupos de Bacterias y Eucariontes, lo que sugiere un origen posterior a la divergencia entre Arqueas y Bacterias. Solo entonces, la glutaril-CoA deshidrogenasa (EC 1.3.8.6), cataliza la descarboxilación del glutaril-CoA produciendo crotonoil-CoA y CO<sub>2</sub>, misma que presenta una distribución acotada a los grupos bacterianos como  $\alpha$ -proteobacterias,  $\beta$ -proteobacterias, actinobacterias, acidobacterias, bacteroidetes y deinococcus-thermus, así como entre los grandes grupos de Eucariontes, lo que sugiere un origen reciente entre los grupos bacterianos. Dos enzimas homólogas catalizan el siguiente par de reacciones de la ruta, la enoil-CoA hidratasa (EC 4.2.1.17), como su nombre lo indica, hidrata al crotonoil-CoA en (S)-3-hidroxi-butanoil-CoA, mismo que es tomado como sustrato de la 3-hidroxiacil-CoA deshidrogenasa (EC 1.1.1.35) que lo transforma en acetoacetyl-CoA mientras acumula el poder reductor en forma de una molécula de NADH. Ambas enzimas cuentan con una distribución universal por lo que es altamente probable que fueran parte del catálogo de enzimas del LCA. Finalmente, la enzima acetyl-CoA C-acetyltransferasa (EC 2.3.1.9) utiliza una molécula extra de la coenzima-A (CoA) junto con el acetoacetyl-CoA para producir

dos moléculas de acetil-CoA. La distribución universal de esta transferasa (EC 2.3.1.9) sugiere que también formaba parte de las habilidades metabólicas del LCA.

La segunda ruta catabólica de la L-lisina comienza con la enzima lisina monooxigenasa (EC 1.13.12.2) que cataliza la descarboxilación del grupo carboxilo produciendo 5-aminopentanamida,  $\text{CO}_2$  y  $\text{H}_2\text{O}$ . Como es de esperarse dada su dependencia con el oxígeno molecular, esta monooxigenasa (EC 1.13.12.2) se encuentra distribuida en solo algunos pocos integrantes de grupos bacterianos como las  $\gamma$ -proteobacterias y las actinobacterias, lo que inmediatamente apunta a un origen posterior al gran evento oxidativo (GOE) de la atmósfera de la Tierra. La 5-aminopentanamida es transformada en 5-aminopentanoato al hidrolizar el grupo amino gracias a la actividad catalítica de la 5-aminopentanamidasa (EC 3.5.1.30), cuya distribución se limita a los grupos de bacterias como  $\alpha$ -proteobacterias,  $\gamma$ -proteobacterias y actinobacterias, lo que sugiere un origen reciente al interior de alguno de estos grupos. El 5-aminopentanoato es utilizado por la 5-aminovalerato transaminasa (EC 2.6.1.48) para transferir el grupo amino de la cadena lateral a una molécula de 2-oxoglutarato liberando L-glutamato y glutarato semialdehído. Esta transaminasa presenta una distribución limitada en algunos miembros de las  $\gamma$ -proteobacterias y los firmicutes, lo que sugiere una evolución reciente. La glutarato-semialdehído deshidrogenasa (EC 1.2.1.20) al transformar al glutarato semialdehído produce glutarato, dos cationes de hidrógeno ( $\text{H}^+$ ) y acumula poder reductor en forma de NAD(P)H. Los patrones de distribución de esta deshidrogenasa (EC 1.2.1.20) sugieren un origen reciente dado que sus homólogos se identificaron exclusivamente dentro de algunos pocos organismos de las  $\gamma$ -proteobacterias y vertebrados. Al continuar la ruta el glutarato es ligado covalentemente a la coenzima A (CoA) por la actividad catalítica de la glutaril-CoA sintetasa (EC 6.2.1.6) que gracias a la energía de una molécula de ATP produce glutaril-CoA y cuya distribución entre algunos pocos organismos que pertenecen a las  $\alpha$ -proteobacterias,  $\beta$ -proteobacterias,  $\gamma$ -proteobacterias, actinobacterias y bacterias verdes no-sulfurosas, así como entre cnidarios, eudicotiledóneas y basidiomicetes,

sugiere un origen reciente. A partir de este punto la transformación del glutaril-CoA es llevada por el mismo conjunto de enzimas que se describió para la primer ruta catabólica de la L-lisina hasta producir dos moléculas de acetil-CoA, es decir las enzimas glutaril-CoA deshidrogenasa (EC 1.3.8.6), enoil-CoA hidratasa (EC 4.2.1.17), 3-hidroxiacil-CoA deshidrogenasa (EC 1.1.1.35) y acetil-CoA C-acetiltransferasa (EC 2.3.1.9).

La tercera ruta catabólica de la L-lisina es iniciada por la actividad de dos isomerasas no homólogas que catalizan dos reacciones contiguas. Primero la enzima lisina 2,3-aminomutasa (EC 5.4.3.2) transforma a la L-lisina en L- $\beta$ -lisina, que es después isomerizada en (3S, 5S)-3,5-diaminohexanoato por la actividad de la  $\beta$ -lisina mutasa (EC 5.4.3.3). La distribución de los homólogos de la 2,3-aminomutasa (EC 5.4.3.2) sugiere un origen posterior a la divergencia entre Arqueas y Bacterias, mientras que la distribución de la  $\beta$ -lisina mutasa (EC 5.4.3.3) se limita a algunos pocos organismos pertenecientes a las  $\delta$ -proteobacterias, los firmicutes, las actinobacterias, los bacteroidetes y las bacterias hipertermófilas. La actividad de la L-3,5-diaminohexanoato deshidrogenasa (EC 1.4.1.11) transforma al (3S, 5S)-3,5-diaminohexanoato en (S)-5-amino-3-oxohexanoato y NH<sub>3</sub>, mientras acumula poder reductor en forma de NADH. La distribución de esta deshidrogenasa (EC 1.4.1.11) sugiere un origen reciente entre las  $\beta$ -proteobacteria, los firmicutes, las actinobacteria, los bacteroidetes, las fusobacterias y las bacterias hipertermófilas, así como en los grandes grupos Eucariontes. Por su parte el (S)-5-amino-3-oxohexanoato es transformado en L-3-aminobutiril-CoA y acetoacetato a expensas de una molécula de acetil-CoA mediante la actividad de la enzima de escisión del 3-ceto-5-aminohexanoato (EC 2.3.1.247), cuya distribución entre algunos pocos miembros de los grupos de  $\alpha$ -proteobacterias,  $\beta$ -proteobacterias,  $\gamma$ -proteobacterias, firmicutes, bacteroidetes, fusobacterias y bacterias hipertermófilas sugiere un origen reciente. La transformación en crotonoil-CoA mediante la pérdida del grupo amino del L-3-aminobutiril-CoA es catalizada por la enzima 3-aminobutiril-CoA amonía-liasa (EC 4.3.1.14) cuya distribución entre algunos pocos integrantes de las  $\beta$ -proteobacterias,  $\delta$ -proteobacterias, firmicutes, actinobacterias, bacteroidetes,

fusobacterias y bacterias hipertermófilas también sugiere un origen reciente para esta enzima. A partir de este punto se utilizan las capacidades catalíticas de las enzimas enoil-CoA hidratasa (EC 4.2.1.17), 3-hidroxiacil-CoA deshidrogenasa (EC 1.1.1.35) y acetil-CoA C-acetiltransferasa (EC 2.3.1.9) para transformar al crotonoil-CoA hasta producir las dos moléculas de acetil-CoA como se ha descrito para las rutas anteriores que degradan a la L-lisina.

### ***Triptófano***

Las rutas catabólicas del L-triptófano producen como principales metabolitos al piruvato y al acetil-CoA (ver Anexo 1, rutas W1 a W3). La enzima triptofanasa (EC 4.1.99.1) es conocida por catalizar la formación de indol, piruvato y  $\text{NH}_3$  a partir de una molécula de L-triptófano. En realidad, la enzima cataliza la ruptura entre la cadena ramificada y el compuesto heterocíclico indol, lo que libera como productos al indol y al 2-aminoacrilato, que sufre una isomerización espontánea que lo transforma en 2-iminopropanoato, que luego sufre una hidratación espontánea que termina por liberar piruvato y  $\text{NH}_3$ . La distribución de los homólogos de la triptofanasa (EC 4.1.99.1) principalmente entre algunos miembros de las actinobacterias, las verrucomicrobia y las bacterias verdes no-sulfurosas sugiere una aparición reciente de la enzima.

La segunda ruta degradativa es iniciada por dos enzimas no homólogas, la triptófano oxigenasa (EC 1.13.11.11) y la triptófano pirrolasa (EC 1.13.11.52), que catalizan de manera independiente la oxidación del L-triptófano utilizando  $\text{O}_2$  para producir N-formil-L-kinurenina. La dependencia al oxígeno molecular se correlaciona con la acotada distribución de ambas enzimas, mientras que la distribución de la oxigenasa (EC 1.13.11.11) al interior de pocos integrantes de los grupos de bacterias como  $\alpha$ -proteobacteria,  $\beta$ -proteobacteria,  $\delta$ -proteobacteria,  $\gamma$ -proteobacteria, firmicutes, actinobacteria, acidobacteria, bacteroidetes, planctomicetos, bacterias verdes no-sulfurosas y deinococcus-thermus, así como en vertebrados, artrópodos, nematodos, cnidarios, algas, coanoflagelados y amebas sugiere un origen reciente al interior de los grupos bacterianos. Por su parte, la distribución de los homólogos de la pirrolasa (EC 1.13.11.52) entre los

vertebrados, los cnidarios, los ascomicetos y los basidiomicetos, sugiere un origen al interior de los grupos eucariontes. Después, la kinureina formamidasa (EC 3.5.1.9) hidroliza la N-formil-L-kinurenina formando ácido fórmico ( $\text{CH}_2\text{O}_2$ ) y L-kinureina, y aunque en general su distribución apunta a un origen entre los principales grupos de las Arqueas, también algunas bacterias miembros de las  $\beta$ -proteobacteria,  $\delta$ -proteobacteria, firmicutes, actinobacteria, acidobacteria, bacterias verdes sulfurosas y bacterias verdes no-sulfurosas, así como cnidarios, basidiomicetos y euglenas, cuentan con algunos homólogos de la enzima. Luego, la kinureina monooxigenasa (EC 1.14.13.9) transforma a la L-kinureina en 3-hidroxi-L-kinureina al utilizar el poder reductor de una molécula de NADH y el poder oxidativo del  $\text{O}_2$ . La distribución de esta monooxigenasa (EC 1.14.13.9) entre algunos miembros de  $\beta$ -proteobacteria,  $\delta$ -proteobacteria,  $\gamma$ -proteobacteria, actinobacteria, acidobacteria, bacteroidetes y cianobacterias, así como entre los principales grupos de Eucariontes sugiere junto con su dependencia al oxígeno molecular un origen reciente posterior al gran evento oxidativo (GOE). A continuación la 3-hidroxi-L-kinureina es transformada en L-alanina y 3-hidroxi-antranilato por la kinureinasa (EC 3.7.1.3), cuya distribución entre las  $\beta$ -proteobacterias, las actinobacterias, las bacterias verdes no-sulfurosas y los *deinococcus-thermus*, así como entre los principales grupos de Eucariontes sugiere un origen reciente. La 3-hidroxi-antranilato deshidrogenasa (EC 1.13.11.6) utiliza una molécula de  $\text{O}_2$  para catalizar la síntesis del 2-amino-3-carboximuconato semialdehído a partir del 3-hidroxi-antranilato. La distribución de esta enzima entre las  $\alpha$ -proteobacteria,  $\beta$ -proteobacteria,  $\delta$ -proteobacteria,  $\gamma$ -proteobacteria, firmicutes, actinobacteria, acidobacteria y bacteroidetes, así como entre los vertebrados, nematodos, cnidarios, ascomicetos, basidiomicetos y las amebas, sugiere, junto con su dependencia al oxígeno molecular, un origen reciente posterior al GOE. La actividad catalítica de la enzima aminocarboximuconato-semialdehído descarboxilasa (EC 4.1.1.45) transforma al 2-amino-3-carboximuconato semialdehído en 2-aminomuconato semialdehído y  $\text{CO}_2$ . Su distribución entre  $\beta$ -proteobacteria,  $\delta$ -proteobacteria, actinobacteria, acidobacteria, verrucomicrobia y las bacterias verdes no-sulfurosas, así como

entre los grandes grupos de Eucariontes, sugiere que evolucionó recientemente entre los grupos bacterianos. Una vez sintetizado el 2-aminomuconato semialdehído, es tomado como sustrato de la aminomuconato-semialdehído deshidrogenasa (EC 1.2.1.32) que lo transforma en 2-aminomuconato mientras se acumula poder reductor en forma de NADH. La distribución universal de esta deshidrogenasa (EC 1.2.1.32) la coloca hasta el momento como la única enzima de esta ruta catabólica que probablemente se encontraba en el último ancestro común (LCA). La transformación del 2-aminomuconato en 2-oxodipato y  $\text{NH}_3$  a expensas del poder reductor de una molécula de NAD(P)H es catalizada por una enzima que ha sido poco caracterizada al punto que su clasificación y posible nomenclatura no han sido completados (EC 1.5.1.-) y que hasta el momento no ha sido relacionada con ninguna secuencia que nos permitiera buscar la distribución de sus posibles homólogos. A partir de este punto el 2-oxodipato es procesado hasta generar dos moléculas de acetil-CoA por las enzimas 2-oxoadipato deshidrogenasa (EC 1.2.4.2), glutaril-CoA deshidrogenasa (EC 1.3.8.6), enoil-CoA hidratasa (EC 4.2.1.17),  $\beta$ -hidroxiacil deshidrogenasa (EC 1.1.1.35) y acetil-CoA C-acetiltransferasa (EC 2.3.1.9), que ya han sido descritas en la primer ruta degradativa de la L-lisina (ver Anexo 1, ruta K1).

La tercera ruta catabólica del L-triptófano inicia con el mismo grupo de enzimas que en la segunda ruta y lo transforman mediante varias reacciones secuenciales descritas previamente en 2-aminomuconato (ver Anexo 1, ruta W2 y W3). Las enzimas triptófano oxigenasa (EC 1.13.11.11), triptófano pirrolasa (EC 1.13.11.52), kinureina formamidasa (EC 3.5.1.9), kinureina monooxigenasa (EC 1.14.13.9), kinureinasa (EC 3.7.1.3), 3-hidroxiantranilato deshidrogenasa (EC 1.13.11.6), aminocarboximuconato semialdehído descarboxilasa (EC 4.1.1.45) y aminomuconato-semialdehído deshidrogenasa (EC 1.2.1.32) son las responsables de la síntesis del 2-aminomuconato. Después, el 2-aminomuconato es hidrolizado por la 2-aminomuconato desaminasa (EC 3.5.99.5) con lo que se produce (3E)-2-oxohex-3-enedioato y  $\text{NH}_3$ . La distribución universal de la desaminasa (EC 3.5.99.5) sugiere que ya se encontraba en el LCA. Por su parte, el (3E)-2-oxohex-3-enedioato sufre una descarboxilación con la que se liberan 2-oxopen-4-enoato y



CO<sub>2</sub>, misma que es catalizada por la 2-oxo-3-hexenedioato descarboxilasa (EC 4.1.1.77) y cuya distribución entre las β-proteobacterias, actinobacterias, bacterias verdes no-sulfurosas y deinococcus-thermus sugiere un origen reciente. El 2-oxopen-4-enoato se hidrata y se genera como producto al 4-hidroxi-2-oxopentanoato mediante la actividad de la 2-oxopent-4-enoato hidratasa (EC 4.2.1.80), cuya distribución universal sugiere que también se encontraba presente en el LCA. La actividad catalítica de la enzima 4-hidroxi-2-oxovalerato aldolasa (EC 4.1.3.39) transforma al 4-hidroxi-2-oxopentanoato en acetaldehído y piruvato. La distribución de los homólogos de la aldolasa (EC 4.1.3.39) dentro de α-proteobacteria, β-proteobacteria, γ-proteobacteria, actinobacteria, planctomicetos, bacterias verdes no-sulfurosas y deinococcus-thermus, así como entre las eudicotiledóneas, monocotiledóneas, ascomicetos y basidiomicetos sugiere un origen reciente entre los grupos bacterianos. Finalmente, el acetaldehído se une a una molécula de CoA produciendo acetil-CoA mientras se acumula poder reductor en forma de NADH gracias a la actividad catalítica de la acetaldehído deshidrogenasa (EC 1.2.1.10), cuya amplia distribución entre los principales grupos de Arqueas y Bacterias sugieren se encontraba presente en el LCA.

### ***Tirosina***

La ruta catabólica que degrada a la L-tirosina genera acetoacetato y piruvato como principales productos (ver Anexo 1, ruta Y1). La ruta inicia con la desaminación de la L-tirosina mediante la actividad de dos enzimas no homólogas, la fenilalanina deshidrogenasa (EC 1.4.1.20), que produce 3-(4-hidroxifenil)piruvato, NH<sub>3</sub> y NADH, o la L-aminoácido oxidasa (EC 1.4.3.2) que utilizando una molécula de O<sub>2</sub> produce 3-(4-hidroxifenil)piruvato, NH<sub>3</sub> y H<sub>2</sub>O<sub>2</sub>. Mientras que la buena distribución en los grandes grupos de Bacterias y Eucariontes de los homólogos de la deshidrogenasa (EC 1.4.1.20) sugiere un origen posterior a la divergencia entre Arqueas y Bacterias, la oxidasa (EC 1.4.3.2), cuya distribución ya ha sido descrita con anterioridad entre las rutas que degradan al L-aspartato, junto con su dependencia al oxígeno molecular parecen indicar un origen reciente, posterior al gran evento oxidativo (GOE). El 3-(4-

hidroxifenil)piruvato sufre una descarboxilación oxidativa que lo transforma en homogentisato y CO<sub>2</sub> gracias a las capacidades catalíticas de la 4-hidroxifenilpiruvato dioxigenasa (EC 1.13.11.27), cuya distribución universal sugiere que se trata de una enzima que probablemente se ha adaptado a las grandes concentraciones de O<sub>2</sub> en la atmósfera y que muy probablemente se encontraba entre las enzimas del último ancestro común (LCA). Después, el homogentisato es tomado como producto de la homogentisato 1,2-dioxigenasa (EC 1.13.11.5) que utilizando una molécula de O<sub>2</sub> produce 4-maleilacetoacetato. Tanto la dependencia de oxígeno molecular como la distribución de la dioxigenasa (EC 1.13.11.5), entre  $\alpha$ -proteobacteria,  $\beta$ -proteobacteria, actinobacteria, acidobacteria y bacteroidetes, así como entre los grandes grupos de Eucariontes sugiere un origen reciente posterior al GOE. La maleilacetoacetato isomerasa (EC 5.2.1.2) transforma al 4-maleilacetoacetato en 4-fumarilacetoacetato y su distribución ubica a sus homólogos entre las  $\alpha$ -proteobacterias,  $\beta$ -proteobacterias,  $\gamma$ -proteobacterias,  $\delta$ -proteobacterias, firmicutes, actinobacterias, cianobacterias, bacterias verdes no-sulfurosas y deinococcus-thermus, así como entre los grandes grupos de Eucariontes apuntan a un origen más bien reciente. Finalmente el 4-fumarilacetoacetato es hidrolizado por la actividad de la fumarilacetoacetasa (EC 3.7.1.2) que produce acetoacetato y fumarato como productos finales de la ruta y cuya distribución universal sugiere que formaba parte del catálogo de enzimas presentes en el LCA.

### **Valina**

La ruta degradativa de la L-valina produce como metabolito de encrucijada al succinil-CoA mediante cuatro variantes de la misma ruta (ver Anexo 1, rutas V1 a V4). Son varias las enzimas que parecen ser poco específicas y que participan también en la degradación de la L-isoleucina o la L-leucina, lo que no resulta extraño dado el parecido estructural de estas moléculas, así que en esta sección no se profundizará ni en sus relaciones evolutivas ni en su distribución por considerar que ya han sido descritas en el apartado correspondiente (mismas que serán identificadas con un asterisco junto a su respectivo número \*EC). La ruta

inicia con la desaminación/transaminación de la L-valina que es catalizada por la valina deshidrogenasa (EC 1.4.1.23), que produce 4-metil-2-oxobutanoato y  $\text{NH}_3$  mientras acumula poder reductor en forma de una molécula de NADH, o por la transaminasa de los aminoácidos de cadena ramificada (\*EC 2.6.1.42), que ocupa una molécula de 2-oxoglutarato como aceptor del grupo amino liberando L-glutamato y 4-metil-2-oxobutanoato. La distribución universal de la deshidrogenasa (EC 1.4.1.23) sugiere que se encontraba en el último ancestro común (LCA). El 4-metil-2-oxobutanoato es luego transformado en isobutiril-CoA ya sea por la actividad catalítica de la 3-metil-2-oxobutanoato deshidrogenasa (\*EC 1.2.7.7) en una sola reacción o por la interacción de tres enzimas, la deshidrogenasa de cetoácido de cadena ramificada (\*EC 1.2.4.4), la dihidrolipoil transacilasa (\*EC 2.3.1.168) y la dihidrolipoil deshidrogenasa (\*EC 1.8.1.4), cuya actividad se detalla en el Anexo 1 (rutas I2, I4, L2, V2 y V4).

En esta ruta se han descrito cuatro enzimas homólogas, la acil-CoA deshidrogenasa (\*EC 1.3.8.7), la butiril-CoA deshidrogenasa (\*EC 1.3.8.1), la 2-metilacil-CoA deshidrogenasa (\*EC 1.3.99.12) y la 2-metilpropanoil-CoA 2,3-oxidoreductasa (EC 1.3.99.-), que son capaces de transformar el isobutiril-CoA en metacrilil-CoA. La muy acotada distribución de la última oxidoreductasa (EC 1.3.99.-) sugiere que de entre este conjunto de enzimas homólogas que catalizan la formación del metacrilil-CoA es la de más reciente aparición ya que se distribuye entre algunos pocos firmicutes y algunos pocos vertebrados y nematodos. Posteriormente, el metacrilil-CoA es hidratado por la actividad de la enzima enoil-CoA hidratasa (\*EC 4.2.1.17) para producir (S)-3-hidroxiisobutiril-CoA que a su vez es hidrolizado en (S)-3-hidroxiisobutirato y Coenzima A (CoA) por la actividad de la 3-hidroxiisobutiril-CoA hidrolasa (EC 3.1.2.4), cuya distribución entre los grandes grupos de Arqueas y Eucariontes sugiere un origen posterior a la divergencia entre Arqueas y Bacterias, no obstante también ciertos organismos pertenecientes a las  $\alpha$ -proteobacterias,  $\beta$ -proteobacterias,  $\gamma$ -proteobacterias, actinobacterias, acidobacterias y bacterias verdes no-sulfurosas han sido identificados como poseedores de homólogos de esta enzima. El (S)-3-hidroxiisobutirato es transformado en (S)-metilmalonato semialdehído por dos

enzimas no homólogas que de manera independiente catalizan esta conversión junto con la acumulación de poder reductor en forma de NADH. Tanto la 3-hidroxiisobutirato deshidrogenasa (EC 1.1.1.31) como la  $\beta$ -hidroxiacil deshidrogenasa (EC 1.1.1.35) cuentan con una buena distribución entre Arqueas, Bacterias y Eucariontes, lo que sugiere que ambas se encontraban entre el catálogo de enzimas del LCA. Posteriormente dos enzimas no homólogas transforman al (S)-metilmalonato semialdehído en metilmalonato, la aldehído deshidrogenasa dependiente de NAD<sup>+</sup> (EC 1.2.1.3) aprovecha la energía de la reacción para acumular poder reductor en forma de NADH, mientras que la aldehído oxidasa (EC 1.2.3.1) utiliza al O<sub>2</sub> para llevar a cabo la misma reacción liberando peróxido de hidrógeno (H<sub>2</sub>O<sub>2</sub>). La distribución universal de ambas enzimas sugiere que se encontraban presentes en el LCA, aunque en el caso de la oxidasa (EC 1.2.3.1) podemos suponer que se ha adaptado a las condiciones oxidantes de la atmósfera de la tierra. Por su parte, la metilmalonato:CoA ligasa (EC 6.2.1.-) cataliza el enlace covalente entre el metilmalonato y la CoA utilizando la energía de la hidrólisis de una molécula de ATP para producir (R)-metilmalonil-CoA. Los homólogos de esta ligasa (EC 6.2.1.-) se encuentran distribuidos entre las  $\alpha$ -proteobacterias,  $\beta$ -proteobacterias y las bacterias verdes no-sulfurosas, así como entre los principales grupos de los Eucariontes, lo que sugiere un origen reciente. Finalmente, la metilmalonil-CoA mutasa (\*EC 5.4.99.2) transforma al (R)-metilmalonil-CoA en succinil-CoA que es el producto final de la ruta.

Una alternativa para completar la ruta anterior una vez que se ha obtenido el (S)-metilmalonato semialdehído, es mediante la enzima metilmalonato-semialdehído deshidrogenasa (EC 1.2.1.27) que produce propanoil-CoA, CO<sub>2</sub> y NADH con lo que la ruta degradativa de la L-valina se conecta con la última parte de la ruta degradativa de la L-isoleucina. Tanto la metilmalonato-semialdehído deshidrogenasa (EC 1.2.1.27), como las enzimas propanoil-CoA carboxilasa (\*EC 6.4.1.3), metilmalonil-CoA epimerasa (\*EC 5.1.99.1) y metilmalonil-CoA mutasa (\*EC 5.4.99.2) constituyen un conjunto alternativo de enzimas por las que se obtiene el succinil-CoA a partir del (S)-metilmalonato semialdehído. Este último conjunto de enzimas se encuentra universalmente distribuido, lo que implica que

sus capacidades catalíticas muy probablemente se encontraban presentes en el LCA.

## **Discusión del Capítulo 2**

A pesar de las propuestas más articuladas y coherentes donde sistemas puramente químicos precedieron la aparición de enzimas, como los basados en reacciones de óxido-reducción donde la participación de aminoácidos como la glicina y la prolina representan agentes oxidantes asociados a una fosforilación a nivel sustrato (Clarke y Elsdén, 1980), y donde se concluye que “resulta fácil vislumbrar un escenario para la aparición gradual de rutas catabólicas completas a partir de este tipo de sistemas”, hasta hoy no existe evidencia contundente que permita especular sobre la hipotética relación de los sistemas enzimáticos contemporáneos con los compuestos que pudieron estar presentes en la tierra primitiva (Lazcano y Miller, 1999). Y aunque los resultados del presente trabajo sugieren que en efecto existen conjuntos de enzimas, capaces de degradar tanto a la glicina como a la prolina, cuya antigüedad se ha especulado podría ubicarlas entre las habilidades metabólicas del último ancestro común (LCA), resultaría excesivo tratar de extrapolar estos resultados a etapas mucho más tempranas de la evolución de la vida o incluso al origen de la vida mismo.

Utilizando un estudio cladístico donde las desaminasas, las transaminasas y las descarboxilasas son utilizadas como caracteres diagnósticos, Cunchillos y Lecointre (2002) han querido extrapolar las habilidades metabólicas que vemos hoy en día en los organismos contemporáneos hasta etapas muy tempranas en las que las “primeras protocélulas” utilizarían a los aminoácidos alifáticos como fuente de energía. No obstante que en el presente trabajo coincidimos en que las rutas catabólicas de L-alanina, L-aspartato, L-asparagina, L-glutamato, L-glutamina, L-cisteína y L-isoleucina se pueden considerar junto con otras como

rutas antiguas, reconocemos que su antigüedad las ubica no en el origen de la vida si no entre las habilidades metabólicas del LCA, quien dista muchísimo de ser una “protocélula” y hoy en día es reconocido más como un organismo parecido a un procarionte contemporáneo.

Más recientemente, la propuesta de Schönheit et al (2016) sobre la antigüedad de las degradaciones enzimáticas de componentes celulares como los aminoácidos favorece un escenario en el que los primeros organismos son considerados como autótrofos primordiales que adaptaron su maquinaria enzimática para dar lugar a la aparición de los primeros heterótrofos quienes aprovecharían los componentes celulares de otros organismos mediante reacciones fermentativas como lo hacen hoy algunos organismos del género *Clostridium*. No obstante que, bajo un escenario que supone el surgimiento autótrofo de la vida resulta tentador tratar de relacionar metabolismos fermentativos, por su relativa simplicidad, con mecanismos puramente químicos o con hipotéticas rutas fermentativas presentes en los primeros organismos; es importante hacer notar que estos escenarios carecen de un marco evolutivo ya que muchas de las fermentaciones que ocurren en organismos del género *Clostridium* suelen ser exclusivas de este grupo y tanto ellos como sus parientes más cercanos esencialmente constituyen a grupos que consideramos de divergencia tardía por lo que no existe una conexión clara en términos de ancestría/descendencia entre este tipo de fermentaciones y aquellas que se han propuesto como los primeros metabolismos fermentativos. Más aún, Schönheit et al (2016) sugieren que la evolución de las primeras fermentaciones de los aminoácidos tuvieron lugar en ambientes anaerobios, no obstante, jamás puntualizan sobre la antigüedad relativa de estas fermentaciones dado que el ambiente anaerobio fue predominante en la Tierra desde su formación hasta que fue transformado por la actividad bacteriana hace aproximadamente 2800-2300 millones de años (Canfield 2005; Lyons, Reinhard y Planavsky 2014).

Aunque la fermentación de la L-valina y de la L-lisina son mencionadas por Schönheit et al (2016) como parte de estas primeras rutas fermentativas, nuestros resultados sugieren que las rutas catabólicas de este par de aminoácidos se consolidaron de manera posterior a la divergencia de las Arqueas y las Bacterias, por lo que resulta poco probable que estas sean rutas antiguas.

## **Conclusiones del Capítulo 2**

La mayor diversidad de estrategias enzimáticas, que han sido estudiadas en el presente trabajo, se encuentra en las rutas catabólicas de los aminoácidos. La mayoría de estas estrategias han reclutado enzimas de diferentes épocas de la historia de la tierra. Las condiciones geológicas cambiantes de nuestro planeta han propiciado la aparición de nuevas formas en que los organismos lidian con esos cambios para sobrevivir, en algunos casos provocados por los mismos organismos que habitan la tierra. Hoy reconocemos la existencia de nuevas enzimas o incluso enzimas antiguas que se han adaptado a las condiciones oxidantes de la atmósfera aprovechando a un potente agente oxidante como lo es el oxígeno molecular que se han integrado a la maquinaria enzimática de ciertos grupos de organismos, como las *proteobacterias*, quienes han innovado estrategias metabólicas que ha conducido a una gran diversificación, misma que a su vez ha permitido la colonización de diversos nichos ecológicos.

Nuestros resultados nos han permitido concluir que las habilidades metabólicas del último ancestro común (LCA) muy posiblemente incluían a las rutas catabólicas que degradan trece de los veinte aminoácidos canónicos de las proteínas, L-alanina, L-arginina, L-asparagina, L-aspartato, L-cisteína, L-glutamato, L-glutamina, glicina, L-iso-leucina, L-metionina, L-prolina, L-serina y L-treonina. Las rutas catabólicas de los siete aminoácidos restantes cuentan con enzimas cuya antigüedad es variable y van desde aquellas que formaban parte de

las enzimas del LCA y hasta las que hoy reconocemos como una innovación de grupos de divergencia tardía dentro de las Bacterias o los Eucariontes. Tal vez algunas de esas enzimas representan las formas ancestrales que el LCA podía utilizar para degradar L-histidina, L-Leucina, L-lisina, L-treonina, L-triptófano, L-tirosina o L-valina. Más aún, la existencia de conjuntos de enzimas degradativas poco específicas nos hace suponer que en el LCA estaba evolucionando rutas que si bien no las encontramos completamente distribuidas, están estrechamente relacionadas con otras que sí lo están, como el caso de las rutas catabólicas de L- isoleucina, L-leucina y L-valina.

*Puntos por completar para el siguiente manuscrito:*

- Para la preparación del manuscrito estamos incluyendo estrategias para confirmar los posibles casos de transferencia horizontal de algunas enzimas cuya peculiar distribución sugiere un flujo de información entre grupos distantes de organismos.
- Se están implementando también para el manuscrito, las comparaciones de estructuras cristalográficas en los casos en los que las comparaciones a nivel de estructura primaria de las proteínas han apuntado a relaciones de homología pero que no han sido concluyentes dados nuestros criterios de inclusión o casos en los que la bioquímica de las enzimas sugiere mecanismos enzimáticos similares o idénticos.



## Capítulo 3

Publicaciones sobre la evolución temprana de la autotrofia y las implicaciones evolutivas de las características estructurales del RNA.

En este capítulo se presentan dos trabajos publicados en revistas internacionales arbitradas, que resultaron de la colaboración con el Dr. Arturo Becerra del laboratorio de Origen de la Vida en la Facultad de Ciencias de la Universidad Nacional Autónoma de México, y otro más resultado de una estancia de investigación en el laboratorio del Dr. George E. Fox de la Universidad de Houston en Estados Unidos, misma que fue posible gracias al apoyo de mi tutor el Dr Antonio Lazcano, y que fue financiada por el Programa de Biología Planetaria de la NASA (PBI).

- A phylogenetic approach to the early evolution of autotrophy: the case of the reverse TCA and the reductive acetyl-CoA pathways.
- Nanometer scale pores similar in size to the entrance of the ribosomal exit cavity are a common feature of large RNAs.

# A phylogenetic approach to the early evolution of autotrophy: the case of the reverse TCA and the reductive acetyl-CoA pathways

Arturo Becerra,<sup>1\*</sup> Mario Rivas,<sup>1</sup> Carlos García-Ferris,<sup>2</sup>  
Antonio Lazcano,<sup>1</sup> Juli Peretó<sup>2</sup>

<sup>1</sup>School of Sciences, National Autonomous University of Mexico, Mexico DF, Mexico.

<sup>2</sup>Department of Biochemistry and Molecular Biology and Cavanilles Institute of Biodiversity and Evolutionary Biology, University of Valencia, Valencia, Spain

Received 20 March 2014 · Accepted 4 June 2014

**Summary.** In recent decades, a number of hypotheses on the autotrophic origin of life have been presented. These proposals invoke the emergence of reaction networks leading from CO or CO<sub>2</sub> to the organic molecules required for life. It has also been suggested that the last (universal) common ancestor (LCA or LUCA) of all extant cell lineages was a chemolitho-autotrophic thermophilic anaerobe. The antiquity of some carbon fixation pathways, the phylogenetic basal distribution of some autotrophic organisms, and the catalytic properties of iron-sulfur minerals have been advanced in support of these ideas. Here we critically examine the phylogenetic distribution and evolution of enzymes that are essential for two of the most ancient autotrophic means of metabolism: the reductive tricarboxylic acid (rTCA) cycle and the reductive acetyl-CoA pathway. Phylogenetic analysis of citryl-CoA synthetase and of citryl-CoA lyase, key enzymatic components of the rTCA cycle, and of CO dehydrogenase/acetyl-CoA synthase, a key enzyme in the reductive acetyl-CoA pathway, revealed that all three enzymes have undergone major lateral transfer events and therefore cannot be used as proof of the LCA's metabolic abilities nor as evidence of an autotrophic origin of life. [Int Microbiol 2014; 17(2):91-97]

**Keywords:** autotrophic pathways · reverse Krebs cycle · Wood–Ljungdahl pathway · origin of life · last common ancestor (LCA, LUCA)

## Introduction

Although a heterotrophic origin of life based on the prebiotic synthesis and accumulation of organic compounds is supported by several major lines of evidence [2], other, competing

alternatives that advocate an autotrophic emergence of living systems have been suggested as well [20,28,30,34]. An autotrophic origin of life, i.e., the hypothesis that the first organisms fed on CO<sub>2</sub> as sole carbon source, was proposed in the 19th century. However, in recent decades this proposal has been reassessed, based on biochemical analyses and new geochemical data. Both sources invoke the emergence of reaction networks leading from CO or CO<sub>2</sub> to the organic molecules required for life [23]. One of the most well-articulated proposals was that of Wächtershäuser [34], who argued that life began without genetic information and with the appearance of

\*Corresponding author: A. Becerra  
Facultad de Ciencias  
Universidad Nacional Autónoma de México  
Ciudad Universitaria  
04510 México DF, México  
email: abb@ciencias.unam.mx

a prebiotic autocatalytic reductive tricarboxylic acid (rTCA) cycle (also called the reductive citric acid cycle, reverse Krebs cycle, or Arnon cycle), which is assumed to be originally based on the formation of the highly insoluble mineral pyrite in sulfur-rich hydrothermal environments. The fact that this mode of carbon fixation is found in the most deeply divergent bacteria, i.e., Aquificales, has been used as evidence of its primitive character [35]. The geochemical emergence of a primitive version of the reductive acetyl-CoA pathway (or Wood–Ljungdahl pathway) associated with the FeS-rich mineral boundaries of alkaline hydrothermal vents was proposed by Russell and Hall [20,28]. Accordingly, it has also been suggested that the last common ancestor (LCA) of all extant cell lineages was a chemolitho-autotrophic thermophilic anaerobe [19,31,36].

Nonetheless, the hypothesis that the first organisms on Earth could fix carbon is far from proven, since the ultimate nature of the first life form is unknown. Phylogenetic analysis using comparative genomics offers clues to the nature of the LCA and could provide evidence of its ability to fix carbon. Even so, the attributes of the first living entities are unknown and a cladistic approach to the origin of life is not feasible, given that all possible intermediate organisms that may have once existed have long since vanished [2,17]. Moreover, it is not possible to extend an investigation beyond the threshold that corresponds to a period of cellular evolution in which protein biosynthesis was already in operation, i.e., the RNA/protein world [3].

With these caveats, a few clues about the early evolution of autotrophy have been acquired, by searching in modern genomes. Extant beings are able to fix carbon in at least six different ways, namely: the reductive pentose phosphate cycle (Calvin–Benson cycle), the rTCA cycle, the reductive acetyl-CoA pathway, the dicarboxylate/4-hydroxybutyrate cycle, the 3-hydroxypropionate/4-hydroxybutyrate cycle, and the 3-hydroxybutyrate bicycle [13,14]. Here we critically examine the distribution and phylogenetic evolution of enzymes that are essential for the purportedly most ancient autotrophic pathways, i.e., the rTCA cycle and the reductive acetyl-CoA pathway [6, 22]. We analyzed sequences from citryl-CoA synthetase and citryl-CoA lyase, two enzymes participating in the rTCA cycle, and CO dehydrogenase/acetyl-CoA synthase, from the reductive acetyl-CoA pathway. Our results show that the early evolution of autotrophy was not free from horizontal gene transfer. Thus, these pathways cannot be invoked as proof of the LCA's metabolic abilities, nor as evidence of an autotrophic origin of life in volcanic environments rich in transition-metal sulfides.

## Materials and methods

**Sequences and genomes.** The query sequences citryl-CoA synthetase (hth:HTH\_1737), citryl-CoA lyase (hth:HTH\_0311), acetyl-CoA synthase (mta:MoH\_1202), and CO dehydrogenase subunits (mta:MoH\_1203) were retrieved, respectively, from the completely sequenced *Hydrogenobacter thermophilus* and *Moorella thermoacetica* genomes available in the Kyoto Encyclopedia of Genes and Genomes (KEGG) [16]. Reports on biochemical information were collected from KEGG, BRENDA [29], MetaCyc [7], and PDB [5] databases.

**Search for homologous genes.** Searches for homologous proteins were carried out by comparing the query sequences to the genomes database in the KEGG, using BLAST searches of its platform [<http://www.genome.jp/tools/blast/>]. The BLASTP cutoff value for homologous identification was set at  $e \leq 1 \times 10^{-7}$  and an identity  $\geq 34\%$ .

**Phylogenetic analysis.** The amino acids sequences of each enzyme were aligned using MUSCLE 3 [12] software with default parameters. Neighbor-joining (500 bootstrap replications, maximum composite likelihood distance estimation, and uniform rates among sites) and maximum-likelihood (model Kimura-2P plus gamma distribution with invariant sites, selected according to the Bayesian information criterion, 500 bootstrap replications) phylogenetic trees were constructed using MEGA5 software [33]. The root of the trees was placed using the midpoint method.

## Results and Discussion

The rTCA cycle can be defined as the (oxidative) Krebs cycle running in reverse (Fig. 1A). While the Krebs cycle is a central pathway in many organisms and is used to oxidize acetyl-CoA to  $\text{CO}_2$  and to generate intermediates for biosynthesis, the rTCA cycle allows the inverse process, i.e., the biosynthesis of acetyl-CoA from two molecules of  $\text{CO}_2$  [14]. Following its discovery in the anaerobic green sulphur photosynthetic bacterium *Chlorobium limicola*, this autotrophic pathway has been detected in strict anaerobic (and microaerobic) bacteria, such as some members of Aquificales [13]. Many of the enzymes involved in the rTCA and Krebs pathways are the same, with the exception of the key enzymes that allow the cycle to run in reverse, namely: 2-oxoglutarate synthase (2-oxoglutarate:ferredoxin oxidoreductase), fumarate reductase, and the citrate-cleaving enzymes [13,14].

The key step of the rTCA cycle is the ATP-dependent cleavage of citrate into acetyl-CoA and oxaloacetate [14]. In different species, this essential reaction is catalyzed by ATP citrate lyase (ACL; EC 2.3.3.8) in one step or by citryl-CoA synthetase (CCS; EC 6.2.1.18) and citryl-CoA lyase (CCL; EC 4.1.3.34) in two steps. Sequence analysis supports the origin of ACL through the fusion of the gene encoding CCS and CCL. Hence, ACL is a derived sequence and our study focused on

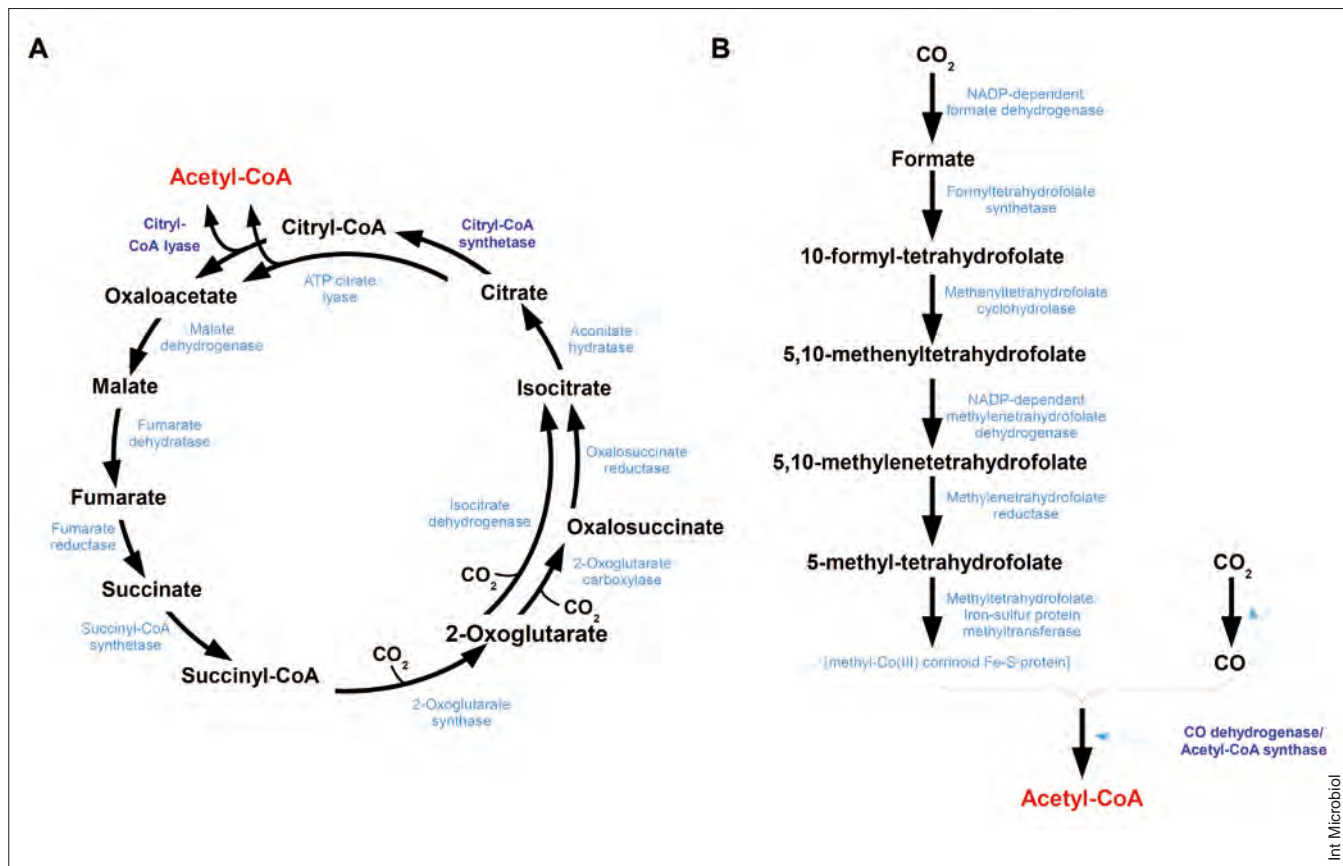


Fig. 1. Reaction schemes of (A) the rTCA cycle and (B) the Wood-Ljungdahl pathway.

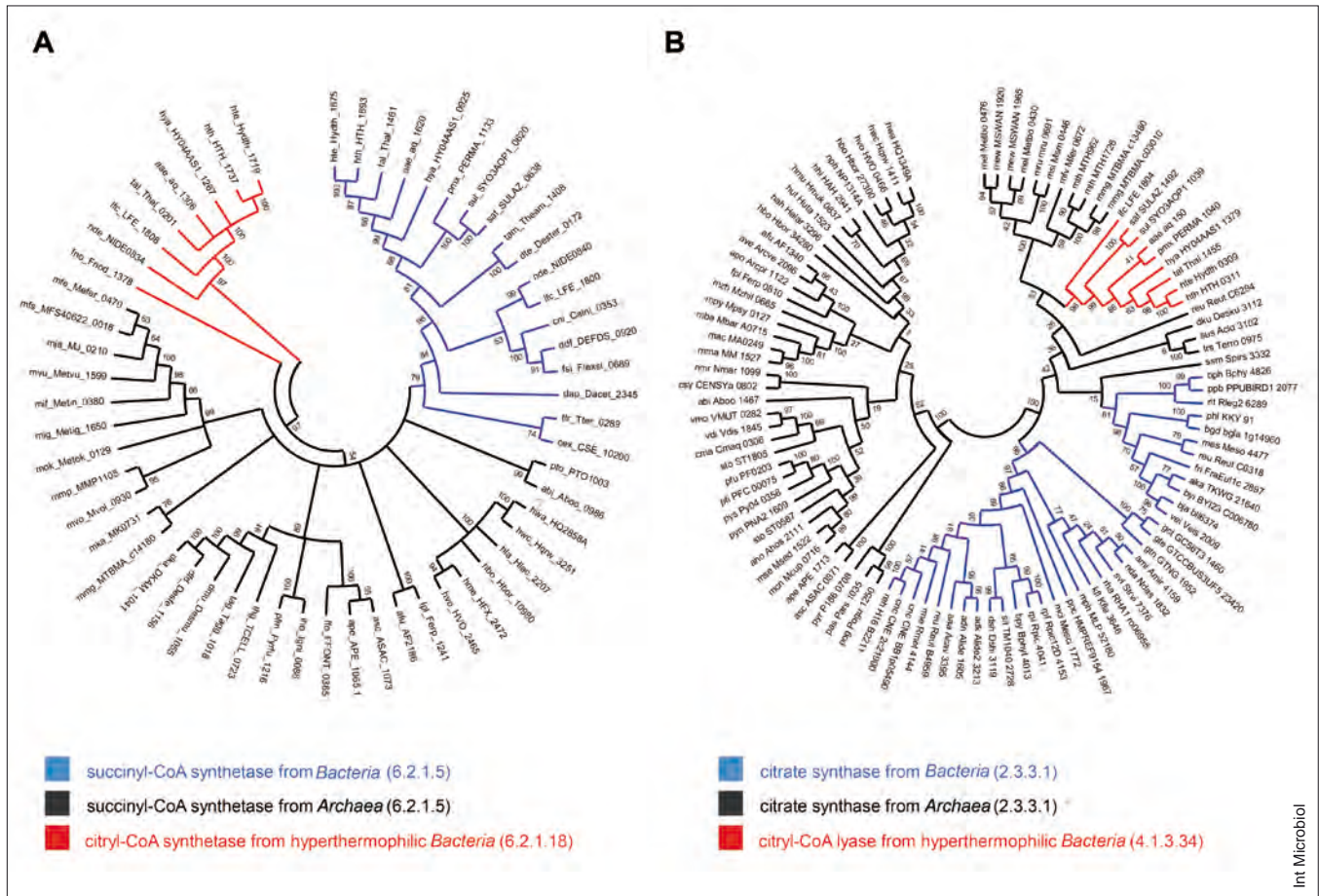
its parental enzymes (CCS/CCL). Furthermore, a common ancestry has been demonstrated between CCS and succinyl-CoA synthetase (SCS) and between CCL and citrate synthase (CS), both of which participate in the Krebs cycle [1].

A feasible evolutionary history of the reaction was proposed by Aoshima [1], who suggested that CCS and CCL activities would be the plesiomorphic state of SCS and CS, respectively. This idea is congruous with the hypothesis of an earliest appearance of a primitive version of the rTCA cycle, as proposed by Wächtershäuser [34]. However, our phylogenetic analysis based on the primary structure of the enzymes does not support Aoshima's hypothesis [1]; instead, the results indicate a horizontal gene transfer between archaea and hyperthermophilic bacteria. As shown in Fig. 2A, the CCS clade from hyperthermophilic bacteria does not appear as a sister group of the SCS bacterial clade. Similar results were obtained with the phylogenetic tree of CCL and CS (Fig. 2B).

The reductive acetyl-CoA pathway was discovered and described in acetogenic bacteria by the laboratories of Wood, Ljungdahl, Thauer, and others [4,13]. This pathway is a linear

metabolic route in which two molecules of CO<sub>2</sub> (or CO<sub>2</sub> and CO) are combined directly to form acetyl-CoA (Fig. 1B). The pathway can be divided into two branches: the methyl or "eastern" branch, in which CO<sub>2</sub> is sequentially reduced to a cofactor-bound methyl residue, and the carbonyl or "western" branch, in which another molecule of CO<sub>2</sub> is reduced to an enzyme-bound carbonyl residue [14,25]. The key enzyme in this pathway is CO dehydrogenase/acetyl-CoA synthase (CODH/ACS; EC 1.2.7.4/2.3.1.169), which has a metallic cluster integrated by iron, sulfur, nickel, zinc, and/or copper [10], although the most active form seems to be Ni-Ni-[4Fe-4S] [32]. CODH/ACS is a bifunctional catalyst that reduces CO<sub>2</sub> to carbon monoxide, forming the carbonyl group of acetyl-CoA, and catalyzes the synthesis of acetyl-CoA [10,24]. This enzyme is found both in anaerobic archaea and in chemotrophic bacteria [4,13,14].

Based on its catalytic activity and metabolic function, the acetyl-CoA synthase of CODH/ACS is a class I enzyme. Since this class is thought to be older than class II [18], CODH/ACS was the focus of our study. Moreover, it has been



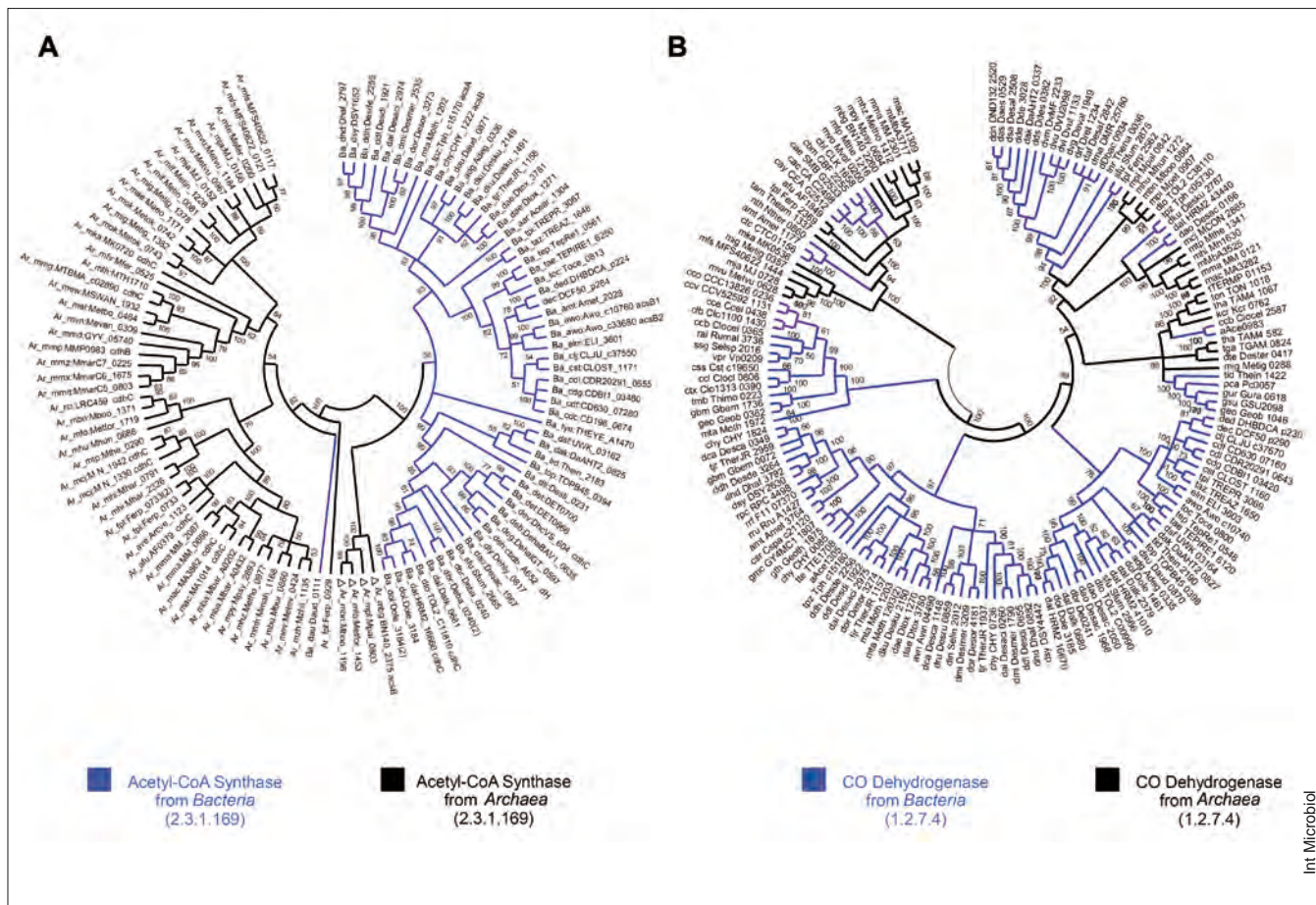
**Fig. 2.** Maximum-likelihood tree (bootstrap 500 replicates) from the amino acid sequences of (A) citryl-CoA synthetase and succinyl-CoA synthetase and (B) citryl-CoA lyase and citrate synthase. The colors of the clades describe the function and taxonomy of the OTUs. The root of the trees was placed using the midpoint method.

proposed as one of the oldest enzymes of life, responsible for the ability of early organisms to live in CO<sub>2</sub>-rich atmospheres [11].

CODH/ACS is a  $\alpha_2\beta_2$  heterotetramer that catalyzes two different reactions using seven metalloclusters. The  $\beta$ -subunit is involved in the CODH activity that generates CO from CO<sub>2</sub>, and the  $\alpha$  subunit in the ACS activity that synthesizes acetyl-CoA [9]. The subunits were found to differ in their phylogenetic distribution (Fig. 3). Also, there were more  $\beta$ -subunits and the topology of the archaeal subunit was more intricate. These findings can be explained in part by considering CODH as occurring within a large group of bacteria that includes carboxydrotrophic bacteria, species of anaerobic acetogenic bacteria, sulfate-reducing bacteria and archaea, phototrophic bacteria, hydrogenogenic bacteria, and methanogenic archaea [8]. Moreover, CODH is present in a larger number of organisms as a monofunctional enzyme than as a bifunctional activity together with the ACS subunit [15]. The topology of the

$\alpha$ -subunit tree is defined by two distinct major clades, one containing the majority of archaeal subunits and another that includes all bacteria. However, a well-supported branch of four methanoarchaea appears as the root of the bacterial clade, suggesting horizontal gene transfer from *Archaea* to *Bacteria* (Fig. 3A).

Furthermore, CODH/ACS is an enzymatic complex that requires Ni-S-Fe clusters to transfer electrons, i.e., the 2Ni-[4Fe-4S] cluster (A-cluster) of the  $\alpha$ -subunit and the [Ni-4Fe-5S] cluster (C-cluster) of the  $\beta$ -subunit. Both CODH and ACS are highly specialized enzymes that require specific chaperone systems for their assembly [21,27]. The role of the Fe-S clusters in catalysis is to funnel electrons onto an assembly line that opens and closes in order to catalyze the condensation of CO with a methyl moiety and then “pump out” acetyl-CoA [9]. This complex molecular machinery implies that the activity of CODH/ACS is an evolutionary innovation,



**Fig. 3.** Maximum-likelihood tree (bootstrap 500 replicates) from the amino acid sequences of CO dehydrogenase/acetyl-CoA synthase: (A) the  $\alpha$ -subunit ACS (EC 2.3.1.169) and (B) the  $\beta$ -subunit of CODH (EC 1.2.7.4; 1.2.99.2). The colors of the clades describe the taxonomy of the OTUs. The root of the trees was placed using the midpoint method.

involving the repurposing of a large number of pre-existing components. It also implies that CODH/ACS was not present at the very early stages of the evolution of life, but was preceded by a long evolutionary history. This possibility is supported by the phylogenetic distribution of the subunits, which suggests that the enzyme appeared after the divergence of the ancestral LCA/ LUCA population.

Our phylogenetic analysis of the citrate-cleaving enzymes participating in the rTCA cycle argue against the hypothesis of Aoshima [1], because these enzymes are not the plesiomorphic version of their homologous counterparts in the oxidative Krebs cycle. Instead, if the key enzymes that allow the Krebs cycle to run in reverse derive from older versions, then it is reasonable to propose that the enzymes required for the oxidative citric acid cycle were present on Earth earlier than those of the rTCA cycle. Phylogenetic analysis of the fundamental citrate-cleaving enzymes, CCS and CCL, do not sup-

port the presence of a complete rTCA cycle in the LCA, since both were acquired by horizontal gene transfer, from archaeal homologs participating in Krebs-cycle-like activities to bacteria. Moreover, the autotrophic growth of the strictly anaerobic Thermoproteales can be explained in full by the dicarboxylate/4-hydroxybutyrate cycle rather than by the rTCA cycle, as was initially proposed [4,13,26]. In summary, current evidence strongly suggests that the rTCA cycle is an idiosyncratic pathway of the bacterial domain that evolved after the LCA.

Similar conclusions can be derived from the phylogenetic analysis of the subunits of the bifunctional enzyme CODH/ACS, an essential component of the Wood–Ljungdahl pathway. Despite its wide distribution and diverse activities, there is no evidence that the  $\beta$ -subunit of CODH was present in the LCA. The same holds true for the  $\alpha$ -subunit (ACS), which was clearly acquired from methanoarchaea. The bifunctional

CO dehydrogenase/acetyl-CoA synthase must have emerged after its components and could have not been present during the LCA's epoch. Since both subunits are part of the Cdh-A-E complex of methanogens, then the CODH/ACS enzymes of the extant reductive acetyl-CoA pathway must have derived from those utilized by methanogens.

It is of course possible that the LCA was an autotrophic organism. However, this hypothesis is not supported by our results on the key enzymes from the oldest CO<sub>2</sub> fixation pathways. More studies, including a better phylogenetic representation of the most basal lineages in both prokaryotic domains, are needed to reveal the metabolic abilities of the LCA, and especially to determine whether it was a heterotrophic or an autotrophic organism.

As noted herein and discussed elsewhere [3,17], the origin of life is not amenable to phylogenetic analysis, since molecular cladistics and comparative genomics cannot be extended beyond a threshold that corresponds to a period of cellular evolution in which protein biosynthesis was already in operation. Accordingly, the genome distribution of enzymes from the rTCA cycle and the reductive acetyl-CoA pathway or even the presence of these metabolic routes is not evidence for an autotrophic origin of life.

**Acknowledgements.** Financial support of CONACYT (100199) to AB and (50520-Q) to AL is gratefully acknowledged. This work was completed during a sabbatical leave of absence of AB, with support from the DGAPA-UNAM and the University of Valencia, where he enjoyed the hospitality of Amparo Latorre at the Institut Cavanilles (Valencia, Spain). CGF and JP acknowledge the financial support of MINECO (Grant ref. BFU2012-39816-C02-01). We acknowledge the technical support of Sara Islas-Graciano and Ricardo Hernández-Morales.

**Competing interests.** None declared.

## References

- Aoshima M (2007) Novel enzyme reactions related to the tricarboxylic acid cycle: Phylogenetic/functional implications and biotechnological applications. *Appl Microbiol Biot* 75:249-255
- Bada J, Lazcano A (2009) The origin of life. In: Ruse M, Travis J (eds) *Evolution: The first four billion years*. Belknap/Harvard Univ. Press, Cambridge, MA, USA, pp 49-79
- Becerra A, Delaye L, Islas S, Lazcano A (2007) The very early stages of biological evolution and the nature of the last common ancestor of the three major cell domains. *Annu Rev Ecol Evol Sci* 38:361-379
- Berg IA, Kockelkorn D, Ramos-Vera WH, Say RF, Zarzycki J, Hügler M, Alber BE, Fuchs G (2010) Autotrophic carbon fixation in archaea. *Nat Rev Microbiol* 8:447-460
- Berman HM, Bhat TN, Bourne PE, Feng Z, Gilliland G, Weissig H, Westbrook J (2000) The protein data bank and the challenge of structural genomics. *Nat Struct Biol* 7:957-959
- Braakman R, Smith E (2012) The emergence and early evolution of biological carbon-fixation. *PLoS Comput Biol* 8:e1002455
- Caspi R, Altman T, Dreher K, Fulcher CA, Subhraveti P, Keseler IM (2012) The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. *Nucleic Acids Res* 40:D742-D753
- Dobbek H, Svetlitchnyi V, Gremer L, Huber R, Meyer O (2001) Crystal structure of a carbon monoxide dehydrogenase reveals a [Ni-4Fe-5S] cluster. *Science* 293:1281-1285
- Doukov TI, Blasiak LC, Seravalli J, Ragsdale SW, Drennan CL (2008) Xenon in and at the end of the tunnel of bifunctional carbon monoxide dehydrogenase/acetyl-CoA synthase. *Biochemistry* 47:3474-3483
- Doukov TI, Iverson TM, Seravalli J, Ragsdale SW, Drennan CL (2002) A Ni-Fe-Cu center in a bifunctional carbon monoxide dehydrogenase/acetyl-CoA synthase. *Science* 298:567-572
- Drennan CL, Doukov TI, Ragsdale SW (2004) The metalloclusters of carbon monoxide dehydrogenase/acetyl-CoA synthase: a story in pictures. *J Biol Inorg Chem* 9:511-515
- Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32:1792-1797
- Fuchs G (2011) Alternative pathways of carbon dioxide fixation: insights into the early evolution of life? *Annu Rev Microbiol* 65:631-658
- Hügler M, Sievert SM (2011) Beyond the Calvin cycle: autotrophic carbon fixation in the ocean. *Annu Rev Mar Sci* 3:261-289
- Jeoung JH, Dobbek H (2012) *n*-Butyl isocyanide oxidation at the [NiFe<sub>4</sub>S<sub>4</sub>OH<sub>x</sub>] cluster of CO dehydrogenase. *J Biol Inorg Chem* 17:167-173
- Kanehisa M, Goto S (2000) KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* 28:27-30
- Lazcano A (2010) Which way to life? *Origins Life Evol Biol* 40:161-167
- Lindahl PA, Chang B (2001) The evolution of acetyl-CoA synthase. *Origins Life Evol Biol* 31:403-434
- Martin W, Baross J, Kelley D, Russell MJ (2008) Hydrothermal vents and the origin of life. *Nat Rev Microbiol* 6:805-814
- Martin W, Russell MJ (2003) On the origins of cells: a hypothesis for the evolutionary transitions from abiotic geochemistry to chemoautotrophic prokaryotes, and from prokaryotes to nucleated cells. *Philos Trans R Soc Lond B Biol Sci* 358:59-83
- Meyer J (2008) Iron-sulfur protein folds, iron-sulfur chemistry, and evolution. *J Biol Inorg Chem* 13:157-170
- Peretó JG, Velasco AM, Becerra A, Lazcano A (1999) Comparative biochemistry of CO<sub>2</sub> fixation and the evolution of autotrophy. *Int Microbiol* 2:3-10
- Peretó J (2012) Out of fuzzy chemistry: from prebiotic chemistry to metabolic networks. *Chem Soc Rev* 41:5394-5403
- Ragsdale SW, Clark JE, Ljungdahl LG, Lundie LL, Drake HL (1983) Properties of purified carbon monoxide dehydrogenase from *Clostridium thermoaceticum*, a nickel, iron-sulfur protein. *J Biol Chem* 258:2364-2369
- Ragsdale SW (2008) Enzymology of the Wood-Ljungdahl pathway of acetogenesis. *Ann NY Acad Sci* 1125:129-136
- Ramos-Vera WH, Berg IA, Fuchs G (2009) Autotrophic carbon dioxide assimilation in *Thermoproteales* revisited. *J Bacteriol* 191:4286-4297
- Rees DC (2002) Great metalloclusters in enzymology. *Annu Rev Biochem* 71:221-246
- Russell MJ, Hall AJ (1997) The emergence of life from iron monosulfide bubbles at a submarine hydrothermal redox and pH front. *J Geol Soc* 154:377-402

29. Schomburg I, Chang A, Hofmann O, Ebeling C, Ehrentreich F, Schomburg D (2002) BRENDA: a resource for enzyme data and metabolic information. *Trends Biochem Sci* 27:54-56
30. Smith E, Morowitz HJ (2004) Universality in intermediary metabolism. *Proc Natl Acad Sci USA* 101:13168-13173
31. Stüeken EE, Anderson RE, Bowman JS, Brazelton WJ, Colangelo-Lillis J, Goldman AD, Som SM, Baross JA (2013) Did life originate from a global chemical reactor? *Geobiology* 11:101-126
32. Svetlitchnyi V, Dobbek H, Meyer-Klaucke W, Meins T, Thiele B, Romer P, Huber R, Meyer O (2004) A functional Ni–Ni–[4Fe–4S] cluster in the monomeric acetyl-CoA synthase from *Carboxydotherrmus hydrogenofor-mans*. *Proc Natl Acad Sci USA* 101:446-451
33. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol* 28:2731–2739
34. Wächtershäuser G (1988) Before enzymes and templates: theory of surface metabolism. *Microbiol Rev* 52:452-484
35. Wächtershäuser G (1990) Evolution of the first metabolic cycles. *Proc Natl Acad Sci USA* 87:200-204
36. Wächtershäuser G (2007) On the chemistry and evolution of the pioneer organism. *Chem Biodiver* 4:584-602





## Nanometer scale pores similar in size to the entrance of the ribosomal exit cavity are a common feature of large RNAs

Mario Rivas, Quyen Tran and George E. Fox, et al.

*RNA* 2013 19: 1349-1354 originally published online August 12, 2013  
Access the most recent version at doi:[10.1261/rna.038828.113](https://doi.org/10.1261/rna.038828.113)

---

**Supplemental Material** <http://rnajournal.cshlp.org/content/suppl/2013/08/05/rna.038828.113.DC1.html>

**References** This article cites 27 articles, 9 of which can be accessed free at:  
<http://rnajournal.cshlp.org/content/19/10/1349.full.html#ref-list-1>

**Creative Commons License** This article is distributed exclusively by the RNA Society for the first 12 months after the full-issue publication date (see <http://rnajournal.cshlp.org/site/misc/terms.xhtml>). After 12 months, it is available under a Creative Commons License (Attribution-NonCommercial 3.0 Unported), as described at <http://creativecommons.org/licenses/by-nc/3.0/>.

**Email Alerting Service** Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#).

---

---

To subscribe to *RNA* go to:  
<http://rnajournal.cshlp.org/subscriptions>

---

## REPORT

# Nanometer scale pores similar in size to the entrance of the ribosomal exit cavity are a common feature of large RNAs

MARIO RIVAS,<sup>1,3</sup> QUYEN TRAN,<sup>2,3</sup> and GEORGE E. FOX<sup>2,4</sup>

<sup>1</sup>Laboratorio de Origen de la Vida, Departamento de Biología Evolutiva, Facultad de Ciencias, Universidad Nacional Autónoma de México, Coyoacán, C.P. 0451, Mexico

<sup>2</sup>Department of Biology and Biochemistry, University of Houston, Houston, Texas 77204-5001, USA

## ABSTRACT

The highly conserved peptidyl transferase center (PTC) of the ribosome contains an RNA pore that serves as the entrance to the exit tunnel. Analysis of available ribosome crystal structures has revealed the presence of multiple additional well-defined pores of comparable size in the ribosomal (rRNA) RNAs. These typically have dimensions of 1–2 nm, with a total area of  $\sim 100 \text{ \AA}^2$  or more, and most are associated with one or more ribosomal proteins. The PTC example and the other rRNA pores result from the packing of helices. However, in the non-PTC cases the nitrogenous bases do not protrude into the pore, thereby limiting the potential for hydrogen bonding within the pore. Instead, it is the RNA backbone that largely defines the pore likely resulting in a negatively charged environment. In many but not all cases, ribosomal proteins are associated with the pores to a greater or lesser extent. With the exception of the PTC case, the large subunit pores are not found in what are thought to be the evolutionarily oldest regions of the 23S rRNA. The unusual nature of the PTC pore may reflect a history of being created by hybridization between two or more RNAs early in evolution rather than simple folding of a single RNA. An initial survey of nonribosomal RNA crystal structures revealed additional pores, thereby showing that they are likely a general feature of RNA tertiary structure.

**Keywords:** RNA tertiary structure; ribosome evolution; RNA nanopores; exit tunnel; peptidyl transferase center

## INTRODUCTION

The ribosome is responsible for coded protein synthesis (Palade et al. 1955; Watson 1963; Noller et al. 1992; Schmeing and Ramakrishnan 2009). The core transpeptidation reaction occurs in a region of the large ribosomal RNA, (rRNA), known as the peptidyl transferase center (PTC), (Nierhaus et al. 1980; Nissen et al. 2000; Sievers et al. 2004; Schmeing et al. 2005; Trobro and Aqvist 2005; Rodnina et al. 2006; Voorhees et al. 2009; Leung et al. 2011; Zaher et al. 2011). From an origins perspective it has been proposed that the PTC originated as a result of noncovalent interactions between two L-shaped RNA molecules that were ultimately ligated together (Agmon et al. 2005; Davidovich et al. 2009). One of these hypothetical RNAs would represent the core of the A-site and the other the P-site in the modern 50S subunit. When these two regions of the modern RNA interact, they create a well-defined cavity that accommodates the termini

of the A- and P-site tRNAs (Voorhees et al. 2009) and therefore forms at least part of the catalytic site.

This cavity (Fig. 1A,B) is properly regarded as a 1.5–2 nm pore. It serves as the entrance to the exit tunnel (Voss et al. 2006), which in turn is effectively a nanotube. In modern protein synthesis, the tunnel allows the nascent protein to leave the ribosome and prevents premature (mis)folding (Nakatogawa and Ito 2002). The tunnel is targeted by macrolide antibiotics that bind RNA functional groups lining its walls, with the result that peptide elongation is inhibited (Starosta et al. 2010). Two key issues that have not been explored previously are the frequency of occurrence of RNA pores and the general nature of the RNA structures that form them.

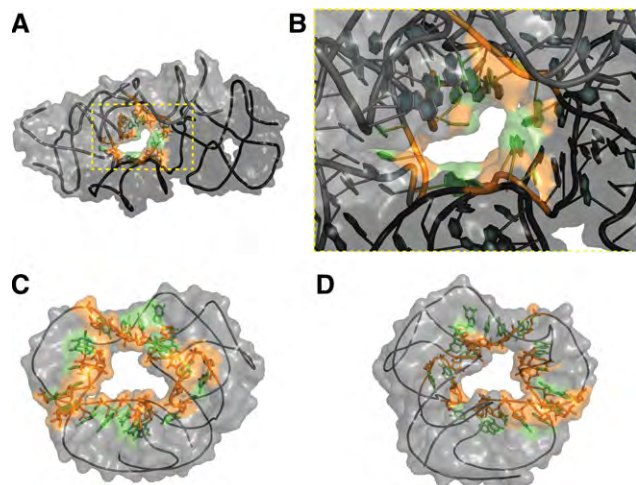
Voss et al. (2006) previously looked for cavities in the ribosome as part of their effort to characterize the exit tunnel. To accomplish this, they used a rolling ball algorithm to detect cavities in the complete ribosome structure. Because they included both proteins and RNA, their approach would necessarily overlook a pore that was formed by the RNA alone, but occluded by one or more proteins. In their effort, the largest number of cavities was found when the probe radius was in the 3–5 Å size range. A second peak occurred when the rolling ball radius was in the 8–9.5 Å size range, but the specific responsible structural features were not described.

<sup>3</sup>These authors contributed equally to this work.

<sup>4</sup>Corresponding author

E-mail [fox@uh.edu](mailto:fox@uh.edu)

Article published online ahead of print. Article and publication date are at <http://www.rnajournal.org/cgi/doi/10.1261/rna.038828.113>.



**FIGURE 1.** Although it is of similar size, the PTC Pore is distinctly different from other pores. (A) The pocket-like symmetrical region of the ribosome (Agmon et al. 2005) includes the PTC pore. Atoms that line the pore are colored in green (nitrogenous bases) and orange (phosphate backbone and sugars). (B) Close-up view of the PTC pore reveals a unique ordering of the RNA residues whereby the phosphate backbone is twisted away from the lumen of the pore, while nitrogenous bases are exposed or likely to be in contact with any molecule that occupied the space. (C) The region covering residues 2103–2186 in *Thermus thermophilus* 23S rRNA (crystal 2WDL) is shown. This region encompasses pore 10 (Table 1). In contrast with the PTC pore, the nitrogenous bases (in green) are twisted away, leaving the phosphate backbone (in orange) to line the lumen of the pore. This geometry is the same for all of the non-PTC pores found in the large and small ribosomal subunit. (D) View of pore 10 from the opposite side.

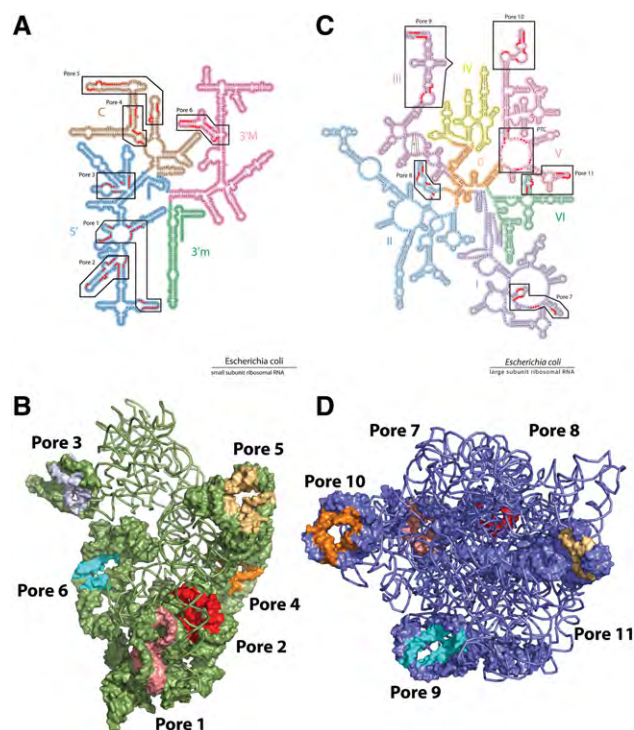
## RESULTS AND DISCUSSION

An initial examination of the *Thermus thermophilus* 16S and 23S rRNAs revealed the presence of 11 nanometer-size pores in addition to the PTC pore as listed in Table 1. Their locations in the context of the rRNA secondary structure are shown in Figure 2, A and C. With regard to the tertiary structure of the ribosome, they are mostly on the surface and avoid the interface between the two subunits. The location of the pores relative to the ribosome tertiary structure is shown in Figure 2, B and D.

It might be argued that these pores were unique features of the rRNAs and therefore would not occur in other RNAs. To address this, the crystal structures of several other RNAs were examined. No nanometer-size pores were found in a partial spliceosome structure or 4.5S RNA. However, additional pores were found in several riboswitch and RNase P structures (Table 2), bringing the total found to date to 16. For each pore, a continuous subsequence of ~54–293 residues encompasses the pore. The pores were never exactly round, so we characterized their size in terms of the smallest and largest distance across the opening from which one can estimate an “area.” Typically, these distances range from 10 to 20 Å (1–2 nm), though several of the pores seen in various riboswitch structures are somewhat smaller (7–12 Å). False pores are some-

times seen when one looks down a helical stem. The true pores are actually built, in part, on such stems, typically utilizing at least four or more residues on at least one strand of a helical element. Although it is somewhat subjective, we also sought to identify specific residues that define the boundaries of each pore. This varies between 15 and 26 residues in regions containing only one pore. A detailed summary for each pore is provided in Table 1.

Comparison of the various RNA pores suggests that there are at least two architectures. The far more common type is a phosphate-lined pore defined by the phosphodiester backbone of the RNA, as illustrated by pore 10 (Fig. 1C). Such pores will present a negative charged surface to the environment. The second type is exemplified by the single case of the PTC pore (Fig. 1B), where the pore is lined in part by the nucleic acid bases rather than just the RNA backbone. Pores of this type will present a slightly hydrophobic or positively charged



**FIGURE 2.** Location of pores relative to ribosome structure. (A) The individual RNA residues defining the pores as seen in Table 1 are highlighted in red on a secondary structure diagram obtained from the Ribovision site (<http://apollo.chemistry.gatech.edu/RiboVision>). The approximate region of RNA that bounds the pore is demarcated by a black box. (B) The distribution of pores in the 30S subunit is viewed from the 50S side. Each pore is labeled. The portion of the rRNA that bounds each pore is shown as a green surface and the residues that define the pore are in various colors. All of the remaining portions of the rRNA are shown as wire colored in green. The central medial axis of the 30S subunit is devoid of nanometer scale pores. (C) Pores in the large subunit are shown as in A. The underlying secondary structure diagram (Petrov et al. 2013) was obtained from the Ribovision website. (D) Crown view of the distribution of pores found in the 50S subunit. The pores are distributed along two planes perpendicular to the vertical axis passing through the central protuberance.

**TABLE 1.** Pores found in the *T. thermophilus* ribosome crystals PBD ID 2WDK and 2WDL

Pore number	RNA	Residues defining the pore	Residues count	Size (Å)	Protein interaction
1	16S	60–62, 101–109, 150–153, 161–169, 331–338, 348–350	36	15 to 21	S20
2	16S	119–123, 233–241, 248–254, 264–267, 276–277, 282–287	33	13 to 21	S17
3	16S	405–408, 427–428, 430–434, 498–499, 541–546	19	15 to 19	S4
4	16S	578–581, 655–659, 742–746, 754–758	19	10 to 15	S15
5	16S	689–692, 699–700, 702–710, 713–714, 775–780, 796–800	28	9 to 13	S11
6	16S	990–993, 1004–1008, 1016–1021, 1038–1041	19	11 to 18	NONE
7	23S	203–207, 217–220, 235–242, 253–257	21	16 to 18	L15, L28, L34, L35*
8	23S	1197–1199, 1213–1217, 1228–1231, 1239–1241	15	10 to 16	L20, L4, L21*
9	23S	1424–1429, 1480–1487, 1498–1503, 1559–1564	26	7 to 14	L2
10	23S	2115–2118, 2134–2138, 2148–2154, 2159–2162, 2173–2174	22	10 to 15	NONE
11	23S	2522–2526, 2531–2537, 2647–2651, 2664–2669	23	9 to 15	L6, L36*
PTC	23S	2061–2064, 2439, 2441–2442, 2450–2452, 2505–2506, 2585–2587	15	10 to 17	NONE

Instances where multiple proteins interact with the pore space are marked with an asterisk.

surface, depending on the hydration and protonation state of the bases in the solvent. In two cases, pores 1 and 2 (Table 1; Supplemental Fig. S1) and pores 4 and 5 (Table 1), a somewhat larger region of the rRNA was found to encompass two pores. As shown for the pair consisting of pores 4 and 5 (Fig. 3A), both pores are again clearly lined with backbone atoms rather than nucleotides. Single pores 3 and 7 are shown in Supplemental Figure S2.

The presence of the non-PTC rRNA pores appears to be a conserved feature of many rRNAs. An examination of the crystal structure of the *E. coli*, *Haloarcula marismortui*, and *Deinococcus radiodurans* 50S subunit structures revealed that all five non-PTC pores found in the 23S rRNA (Table 1, pores 7–11) are seen in all three structures. Although the available structures for the 30S subunit do not include a member of the Archaea, equivalent pores were again found at similar locations in *E. coli* and *Thermus thermophilus*. However, many of the pores will clearly be missing in the most minimal mitochondrial rRNAs, where large regions of the rRNAs are deleted.

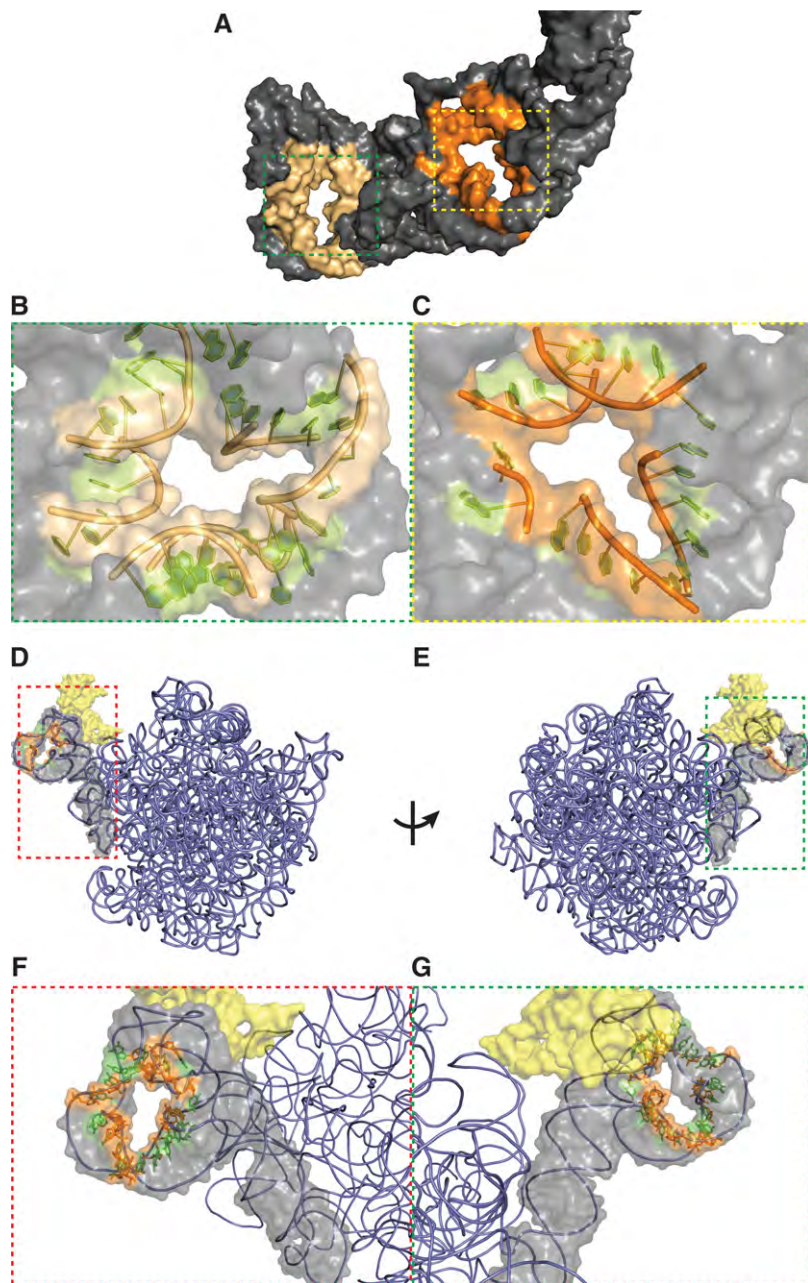
Given that the RNA secondary structures themselves are very conserved, the occurrence of equivalent pores is not unexpected. To explore the matter further, we examined the conservation of the individual residues lining each pore using preprocessed data available on the Comparative RNA web site and project (Cannone et al. 2002). The results in Supplemental Table S1 revealed that residues lining pores 1, 2, 7, and especially 5 and 11, were very conserved. The conservation, however, did not match the level of the PTC pore, where only two residues showed significant variation in all of the Bacteria Domain. In contrast, the residues that line pores 6, 8, and 9 were far less conserved, even when only the Bacterial Domain was considered.

The ribosome crystal structures are very porous, and water residues reach the interior in many places (Voss et al. 2006). It is thus unlikely that the nanometer scale pores described here

serve as channels for water flow. A more likely utility may be to stabilize the binding of various ribosomal proteins. With the possible exception of pores 6 and 10, all of the rRNA pores are associated to some extent with ribosomal proteins. An example of extensive association occurs in pore 3, where ribosomal protein S4 actually protrudes into and completely fills the pore, Supplemental Figure S3.

In order to facilitate future studies of the properties of RNA pores, it will be useful to have an effective model system. All of the rRNA pores seen to date are in the larger context of the structure of an entire rRNA. Thus, it is possible that they are stabilized directly or indirectly in part by tertiary interactions with other parts of the rRNA or ribosomal proteins and, hence, might cease to exist in the absence of such interactions. This is unlikely for at least pores 1 and 2, as it has been shown that the 5' domain of *E. coli* 16S rRNA already forms all of the expected tertiary interactions in the absence of the ribosomal proteins (Adilakshmi et al. 2005). Nevertheless, pore 10 is an interesting exception. It is located in a region of the larger ribosome structure near the ribosomal protein L1 stalk that is devoid of interactions with other parts of the RNA (Fig. 3D, E). Ribosomal protein L1 circumscribes the perimeter of the pore on the side furthest from the 30S subunit, but does not protrude into the pore as seen in the *T. thermophilus* 50S crystal structure (Fig. 3F,G). Its proximity to the pore and the E-site tRNA make it possible that the combination may play a role in the dissociation of the tRNA from the ribosome. Absent this, pore 10 may serve as a model system. It is an apparently self-contained pore that is encompassed by positions 2103–2186, and thus can be synthesized as an 85-residue sequence.

Nanometer-scale RNA pores can be regarded as a novel form of RNA tertiary structure. From the perspective of RNA design, future work may focus on efforts to predict their occurrence from primary sequence alone and as a possible construct for use in synthetic biology applications. For



**FIGURE 3.** Region of *T. thermophilus* 16S rRNA covering residues 575–810 that contains two adjacent pores and a region from the 23S rRNA showing a self-contained pore. (A) Overall surface rendering of the region with pore 4 in orange and pore 5 in yellow. (B) Close-up view of pore 4 with backbone colored in orange and nitrogenous bases in green. (C) Close-up view of pore 5 with backbone now colored in yellow. As in Figure 1, C and D, the nitrogenous bases in both cases avoid the lumen of the pore. (D) The ribosomal region that contains pore 10 that was previously highlighted in Figure 1, C and D, is shown. The 23S rRNA is displayed as a blue ribbon. Ribosomal protein L1 is rendered as a light-yellow surface, and the region containing pore 10 is highlighted by the red dotted lines. (E) Same as D but rotated 180 degrees. (F,G) Close-up view for both orientations, respectively.

example, it may be possible to create pores that facilitate catalysis. The PTC pore clearly binds RNA ligands, e.g., the termini of the A and P site tRNAs (Kim and Green 1999; Voorhees et al. 2009; Zaher et al. 2011). Most of the other

pores likely interact with peptide ligands as well, as is illustrated for ribosomal protein S4, which interacts with pore 3 (Supplemental Fig. S3).

Given that pore 10 is much more self-contained than the other pores, it could be used as a scaffold to design and test the feasibility of practical applications that utilize the aforementioned interactions. Perhaps a naïve but intuitive starting point would be to duplicate the architecture of pore 10 multiple times in a larger RNA. The primary sequence would have to be carefully designed so that, in solution, such an RNA would form many pores arrayed in a net-like macrostructure. This net-like structure would have a regular distribution of both large and small pores through which filtrates/small biomolecules could flow. The stacking and tessellation of such a structure will conceivably produce a molecular sieve that could have superior properties to those of ceramic and membrane variants. Furthermore, RNA is less toxic and its stability could be enhanced by modifications such as the addition of a 2' O-methylation on one or more residues.

Although RNA pores appear to be a common feature in many RNAs, the PTC example remains unusual in that it is not a phosphate-lined pore. It is, in fact, the only pore that is found in what are regarded as the most ancient areas of the 23S rRNA. The other 23S rRNA pores are all in regions of the RNA that are relatively recent additions to the structure (Hury et al. 2006; Bokov and Steinberg 2009; Hsiao et al. 2009; Fox 2010). The unique nature of the PTC pore may reflect a history of being created by hybridization between two or more RNAs in an RNA world rather than simple folding of a single RNA (Davidovich et al. 2009; Fox et al. 2012).

## MATERIALS AND METHODS

A heuristic search for RNA pores in the crystal structures of various large RNAs was undertaken. Crystal structures were obtained from the Protein Data Bank (PDB, <http://www.pdb.org>). Initially, the crystal structures of ribosomes from *T. thermophilus* (PDB 2WDK and 2WDL) were used to identify pores. As previously shown by Voss et al. (2006), there are numerous small cavities in the large

**TABLE 2.** Pores found in nonribosomal RNA crystals

RNA's type	PDB ID	Components	Residues defining the pore	# Residues	Size (Å)
Riboswitch	2YDH	Chain A	13, 22–23, 37–40, 60–62	10	7 to 17
	2YIF	Chain X	10–13	13	16
		Chain Z	95–103		
RNase-P	3DHS	Chain A	60–64, 77–80, 258–260, 288–290, 297–299	18	5 to 17
	3Q1Q	Chain B	223–226, 240–243, 255–263	21	10 to 18
		Chain C	83–86		
		Chain B	60–62, 223–226, 240–243, 257–263	25	11 to 20
	3Q1R	Chain C	80–86		

RNAs, which are not considered here. All of the pores found in the 16S rRNA and 23S rRNA measure at least 0.7 nm in the shorter direction with an overall area that typically exceeds 100 Å<sup>2</sup>. The location of the pore-lining residues relative to the rRNA secondary structures is shown in Figure 2, A and B. After pores were found in these structures, we verified the presence and prevalence of corresponding pores in *Escherichia coli* (PDB 3R8N, 3R8O, 3R8S, and 3R8T), *H. marismortui* (PDB 1JJ2), and *D. radiodurans* (PDB 2D3O). Currently, there are no small subunit crystals for either *H. marismortui* or *D. radiodurans*, so we couldn't look for equivalent pores in these 16S rRNAs. Subsequently, we searched for pores in crystals of riboswitches (PDB 2YDH, 3SUH, and 2YIF) the 4.5S RNA (PDB 1DUL and 1HQ1), RNase-P (PDB 3DHS, 3Q1Q, and 3Q1R), and in the partial structure of a RNA spliceosome (PDB 3SIV).

In order to find pores, the RNA portion of the crystals was visualized using Pymol (The PyMOL Molecular Graphics System, Version 1.5.0.4 Schrödinger, LLC). The whole RNA chain was first represented as a Connolly surface through PyMOL's internal surface-finding algorithm. This representation allowed easy recognition of structural pores in the RNA folding. The proteins were not removed, but rather were initially hidden from view. Once the pores had been localized, we used both the ribbon and lines representations to define their properties and nature. We also made distance measurements using PyMOL. In the future, it will be useful to develop a more formal procedure for identifying pores so that none are overlooked.

## SUPPLEMENTAL MATERIAL

Supplemental material is available for this article.

## ACKNOWLEDGMENTS

This work was supported in part by a subcontract to G.E.F. from the NASA Center for Ribosome Evolution and Adaptation at the Georgia Institute of Technology (NNA09DA78A). Q.T. was supported by a NASA Earth and Space Science Fellowship (10-Planet10R-0025). M.R. was supported by the Marine Biological Laboratory's NASA Planetary Biology Internship Program. Computational and graphics facilities were provided by the Texas Learning and Computation Center at the University of Houston. We are indebted to Dr. Steve Harvey for helpful discussion and Dr. Antonio Lazcano for his encouragement and support of M.R.

Received February 28, 2013; accepted June 19, 2013.

## REFERENCES

- Adilakshmi T, Ramaswamy P, Woodson SA. 2005. Protein-independent folding pathway of the 16S rRNA 5' domain. *J Mol Biol* **351**: 508–519.
- Agmon I, Bashan A, Zarivach R, Yonath A. 2005. Symmetry at the active site of the ribosome: Structural and functional implications. *Biol Chem* **386**: 833–844.
- Bokov K, Steinberg SV. 2009. A hierarchical model for evolution of 23S ribosomal RNA. *Nature* **457**: 977–980.
- Cannone JJ, Subramanian S, Schnare MN, Collett JR, D'Souza LM, Du Y, Feng B, Lin N, Madabusi LV, Müller KM, et al. 2002. The Comparative RNA Web (CRW) Site: An online database of comparative sequence and structure information for ribosomal, intron, and other RNAs. *BioMed Central Bioinformatics* **3**: 2. [Correction: *BioMed Central Bioinformatics*. **3**: 15.]
- Davidovich C, Belousoff M, Bashan A, Yonath A. 2009. The evolving ribosome: From non-coded peptide bond formation to sophisticated translation machinery. *Res Microbiol* **160**: 487–492.
- Fox GE. 2010. Origin and evolution of the ribosome. *Cold Spring Harb Perspect Biol* **2**: a003483.
- Fox GE, Tran Q, Yonath A. 2012. An exit cavity was crucial to the polymerase activity of the early ribosome. *Astrobiology* **12**: 57–60.
- Hsiao C, Mohan S, Kalahar BK, Williams LD. 2009. Peeling the onion: Ribosomes are ancient molecular fossils. *Mol Biol Evol* **26**: 2415–2425.
- Hury J, Nagaswamy U, Larios-Sanz M, Fox GE. 2006. Ribosome origins: The relative age of 23S rRNA domains. *Orig Life Evol Biosph* **36**: 421–429.
- Kim DF, Green R. 1999. Base-pairing between 23S rRNA and tRNA in the ribosomal A-site. *Mol Cell* **4**: 859–864.
- Leung EK, Suslov N, Tuttle N, Sengupta R, Piccirilli JA. 2011. The mechanism of peptidyl transfer catalysis by the ribosome. *Annu Rev Biochem* **80**: 527–555.
- Nakatogawa H, Ito K. 2002. The ribosomal exit tunnel functions as a discriminating gate. *Cell* **108**: 629–636.
- Nierhaus KH, Schulze H, Cooperman BS. 1980. Molecular mechanisms of the ribosomal peptidyl transferase center. *Biochem Int* **1**: 185–192.
- Nissen P, Hansen J, Ban N, Moore PB, Steitz TA. 2000. The structural basis of ribosome activity in peptide bond synthesis. *Science* **289**: 920–930.
- Noller HF, Hoffarth V, Zimniak L. 1992. Unusual resistance of peptidyl transferase to protein extraction procedures. *Science* **256**: 1416–1419.
- Palade GE. 1955. A small particulate component of the cytoplasm. *J Biophys Biochem Cytol* **1**: 59–68.
- Petrov AS, Bernier CR, Hershkovits E, Xue Y, Waterbury CC, Hsiao C, Stepanov VG, Gaucher EA, Grover MA, Harvey SC, et al. 2013. Secondary structure and domain architecture of the 23S rRNA. *Nucleic Acids Res* (in press). doi: 10.1093/nar/gkt513.

- Rodnina MV, Beringer M, Wintermeyer W. 2006. Mechanism of peptide bond formation on the ribosome. *Q Rev Biophys* **39**: 203–225.
- Schmeing TM, Ramakrishnan V. 2009. What recent ribosome structures have revealed about the mechanism of translation. *Nature* **461**: 1234–1242.
- Schmeing TM, Huang KS, Kitchen DE, Strobel SA, Steitz TA. 2005. Structural insights into the roles of water and the 2' hydroxyl of the P site tRNA in the peptidyl transferase reaction. *Mol Cell* **20**: 437–448.
- Sievers A, Beringer M, Rodnina MV, Wolfenden R. 2004. The ribosome as an entropy trap. *Proc Natl Acad Sci* **101**: 7897–7901.
- Starosta AL, Karpenko VV, Shishkina AV, Mikolajka A, Sumbatyan NV, Schluenzen F, Korshunova GA, Bogdanov AA, Wilson DN. 2010. Interplay between the ribosomal tunnel, nascent chain, and macrolides influences drug inhibition. *Chem Biol* **16**: 1180–1189.
- Trobro S, Aqvist J. 2005. Mechanism of peptide bond synthesis on the ribosome. *Proc Natl Acad Sci* **102**: 12395–12400.
- Voorhees RM, Weixlbaumer A, Loakes D, Kelley AC, Ramakrishnan V. 2009. Insights into substrate stabilization from snapshots of the peptidyl transferase center of the intact 70S ribosome. *Nat Struct Mol Biol* **16**: 528–533.
- Voss NR, Gerstein M, Steitz TA, Moore PB. 2006. The geometry of the ribosomal polypeptide exit tunnel. *J Mol Biol* **360**: 893–906.
- Watson JD. 1963. Involvement of RNA in the synthesis of proteins. *Science* **140**: 17–26.
- Zaher HS, Shaw JJ, Strobel SA, Green R. 2011. The 2'-OH group of the peptidyl-tRNA stabilizes an active conformation of the ribosomal PTC. *EMBO J* **30**: 2445–2453.

### Conclusiones del capítulo 3

*A phylogenetic approach to the early evolution of autotrophy: the case of the reverse TCA and the reductive acetyl-CoA pathways.*

La propuesta de que el último ancestro común (LCA) de los seres vivos contemporáneos era un organismo quimiolitautótrofo anaerobio se ha sustentado en gran medida en la especulación sobre la antigüedad de las rutas de fijación de carbono como el ciclo reductivo del ácido cítrico o el ciclo reductivo del acetil-CoA. En favor de esta hipótesis también se han invocado por un lado la posición en la base del árbol de la vida de algunos organismos autotróficos y por otro las propiedades catalíticas de los minerales como la pirita ( $\text{FeS}_2$ ) que parecen similares a las de los centros hierro-azufre de las proteínas contemporáneas.

Basados en la distribución filogenética y el análisis de la evolución molecular de las enzimas citril-CoA sintetasa (EC 6.2.1.18), citril-CoA liasa (EC 4.1.3.34), ambas esenciales para el ciclo reductivo del ácido cítrico, y del heterodímero CO deshidrogenasa/acetil-CoA sintasa (EC 1.2.7.4/2.3.1.169) indispensable para el ciclo reductivo del acetil-CoA, nuestro equipo concluyó que estas enzimas han sufrido una cantidad importante de eventos de transferencia horizontal y por lo tanto es poco probable que fueran parte del inventario de enzimas del LCA, por lo que no debieran ser utilizadas como prueba de las habilidades metabólicas del LCA y mucho menos en favor de un origen autotrófico de la vida.



*Nanometer scale pores similar in size to the entrance of the ribosomal exit cavity are a common feature of large RNAs.*

El centro peptidil-transferasa (PTC) del ribosoma contemporáneo es en términos estructurales un poro en el RNA que sirve como entrada al túnel de salida por donde el péptido naciente, durante la síntesis de proteínas, deja el ribosoma. El análisis de las estructuras cristalográficas del ribosoma disponibles para diversos organismos, reveló la presencia de múltiples estructuras bien definidas similares en tamaño al poro del PTC. Se ha llegado a la conclusión de que estas estructuras son el resultado de del empaquetamiento de varias hélices de RNA. Lo que caracteriza el poro del PTC es la exposición hacia la luz, es decir, el interior del poro de algunas bases (altamente conservadas) que le confieren sus capacidades catalíticas, mientras que en el resto de los poros del ribosoma es el esqueleto de fosfatos de las hélices de RNA el que define el poro lo que resulta en un ambiente cargado negativamente en la luz de esos poros. La búsqueda de este tipo de poros en estructuras cristalográficas de RNAs distintos al ribosoma ha revelado que la formación de este tipo de estructuras en los plegamientos del RNA representa una característica inherente de la estructura terciaria de las cadenas de RNA.

Desde la perspectiva del origen del ribosoma, este tipo de estructuras parecen ser el resultado de la interacción no covalente entre dos estructuras similares a los tRNAs de los organismos contemporáneos, que esencialmente están formadas de hélices de RNA y que podrían ser los ancestros que dieron origen al “PTC” de una entidad hipotética que hoy llamamos el protoribosoma.

# Referencias

## *Referencias de la Introducción General*

- Becerra, A., Delaye, L., Islas, S., & Lazcano, A. (2007). The very early stages of biological evolution and the nature of the last common ancestor of the three major cell domains. *Annual Review of Ecology, Evolution, and Systematics*, 38(1), 361–379.
- Caetano-Anollés, G., Kim, H. S., & Mittenthal, J. E. (2007). The origin of modern metabolic networks inferred from phylogenomic analysis of protein architecture. *Proceedings of the National Academy of Sciences*, 104(22), 9358–63.
- Caetano-Anollés, K., & Caetano-Anollés, G. (2013). Structural phylogenomics reveals gradual evolutionary replacement of abiotic chemistries by protein enzymes in purine metabolism. *PloS One*, 8(3), e59300.
- Canfield, D. E. (2005). The early history of atmospheric oxygen: homage to Robert M. Garrels. *Annual Review of Earth and Planetary Sciences*, 33(1), 1–36.
- Clarke, P. H., & Elsdén, S. R. (1980). The earliest catabolic pathways. *Journal of Molecular Evolution*, 15(4), 333–338.
- Cunchillos, C., & Lecointre, G. (2002). Early steps of metabolism evolution inferred by cladistic analysis of amino acid catabolic pathways. *Comptes Rendus Biologies*, 325(2), 119–29.
- Goldfine, H. (1965). The evolution of oxygen as a biosynthetic reagent. *The Journal of General Physiology*, 49(1), Suppl:253-74.
- Keefe, A. D., Lazcano, A., & Miller, S. L. (1995). Evolution of the biosynthesis of the branched-chain amino acids. *Origins of Life and Evolution of the Biosphere*, 25(1–3), 99–110.
- Keller, M. A., Piedrafita, G., & Ralser, M. (2015). The widespread role of non-enzymatic reactions in cellular metabolism. *Current Opinion in Biotechnology*, 34C, 153–161.
- Krebs, H. A. (1981). The Evolution of Metabolic Pathways. In *Molecular and Cellular Aspects of Microbial Evolution* (pp. 215–228). Great Britain: Cambridge University

Press.

- Krebs, H. A., & Kornberg, H. L. (1957). *Energy Transformations in Living Matter*. Berlin, Heidelberg: Springer Berlin Heidelberg.
- Lazcano, A. (2016). Alexandr I. Oparin and the origin of life: a historical reassessment of the heterotrophic theory. *Journal of Molecular Evolution*, 83(5–6), 214–222.
- Lazcano, A., & Miller, S. L. (1999). On the origin of metabolic pathways. *Journal of Molecular Evolution*, 49(4), 424–31.
- Lyons, T. W., Reinhard, C. T., & Planavsky, N. J. (2014). The rise of oxygen in Earth's early ocean and atmosphere. *Nature*, 506(7488), 307–15.
- Michal, G., & Schomburg, D. (Eds.). (2012). *Biochemical pathways: an atlas of biochemistry and molecular biology* (Second). Hoboken, NJ, USA: John Wiley & Sons, Inc.
- Oparin, A. I. (1924). *The Origin of Life (Russian)*. Moscow: Moscow Worker.
- Oparin, A. I. (1938). *The Origin of Life (English)* (First Engl). New York: The Macmillan Company.
- Schönheit, P., Buckel, W., & Martin, W. F. (2016). On the origin of heterotrophy. *Trends in Microbiology*, 24(1), 12–25.

### *Referencias de la Metodología General*

- Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, 215(3), 403–10.
- Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., ... Bourne, P. E. (2000). The Protein Data Bank. *Nucleic Acids Research*, 28(1), 235–242.
- Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., ... Schäffer, A. (2009). BLAST+: architecture and applications. *BMC Bioinformatics*, 10(1), 421.
- Caspi, R., Altman, T., Billington, R., Dreher, K., Foerster, H., Fulcher, C. A., ... Karp, P. D. (2014). The MetaCyc database of metabolic pathways and enzymes and the

- BioCyc collection of Pathway/Genome Databases. *Nucleic Acids Research*, 42(Database issue), D459-71.
- Chothia, C., & Lesk, A. M. (1986). The relation between the divergence of sequence and structure in proteins. *The EMBO Journal*, 5(4), 823–6.
- Edgar, R. C. (2004a). MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics*, 5, 113.
- Edgar, R. C. (2004b). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, 32(5), 1792–7.
- Embley, T. M., & Williams, T. A. (2015). Steps on the road to eukaryotes. *Nature*, 521(7551), 169–170.
- Finn, R. D., Mistry, J., Schuster-Böckler, B., Griffiths-Jones, S., Hollich, V., Lassmann, T., ... Bateman, A. (2006). Pfam: clans, web tools and services. *Nucleic Acids Research*, 34(Database issue), D247-51.
- Forterre, P., & Gribaldo, S. (2010). Bacteria with a eukaryotic touch: a glimpse of ancient evolution? *Proceedings of the National Academy of Sciences*, 107(29), 12739–12740.
- Fu, L., Niu, B., Zhu, Z., Wu, S., & Li, W. (2012). CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics (Oxford, England)*,
- Geer, L. Y., Marchler-Bauer, A., Geer, R. C., Han, L., He, J., He, S., ... Bryant, S. H. (2010). The NCBI BioSystems database. *Nucleic Acids Research*, 38(Database issue), D492-6.
- Irving, J. A., Whisstock, J. C., & Lesk, A. M. (2001). Protein structural alignments and functional genomics. *Proteins*, 42(3), 378–82.
- Kanehisa, M., & Goto, S. (2000). KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Research*, 28(1), 27–30.
- Kanehisa, M., Goto, S., Sato, Y., Kawashima, M., Furumichi, M., & Tanabe, M. (2014). Data, information, knowledge and principle: back to metabolism in KEGG. *Nucleic Acids Research*, 42(Database issue), D199-205.
- Magrane, M., & Consortium, U. (2011). UniProt Knowledgebase: a hub of integrated protein data. *Database : The Journal of Biological Databases and Curation*, 2011(0), bar009.

- McMurry, J., & Begley, T. P. (2005). *The Organic Chemistry of Biological Pathways*. Roberts and Company Publishers.
- Michal, G., & Schomburg, D. (Eds.). (2012). *Biochemical pathways: an atlas of biochemistry and molecular biology* (Second). Hoboken, NJ, USA: John Wiley & Sons, Inc.
- Mistry, J., Finn, R. D., Eddy, S. R., Bateman, A., & Punta, M. (2013). Challenges in homology search: HMMER3 and convergent evolution of coiled-coil regions. *Nucleic Acids Research*, gkt263-.
- Prlic, A., Bliven, S., Rose, P. W., Bluhm, W. F., Bizon, C., Godzik, A., & Bourne, P. E. (2010). Pre-calculated protein structure alignments at the RCSB PDB website. *Bioinformatics (Oxford, England)*, 26(23), 2983–5.
- Punta, M., Coggill, P. C., Eberhardt, R. Y., Mistry, J., Tate, J., Boursnell, C., ... Finn, R. D. (2012). The Pfam protein families database. *Nucleic Acids Research*, 40(Database issue), D290-301.
- Schomburg, I., Chang, A., Placzek, S., Söhngen, C., Rother, M., Lang, M., ... Schomburg, D. (2013). BRENDA in 2013: integrated reactions, kinetic data, enzyme function data, improved disease classification: new options and contents in BRENDA. *Nucleic Acids Research*, 41(Database issue).
- Subbiah, S., Laurents, D. V., & Levitt, M. (1993). Structural similarity of DNA-binding domains of bacteriophage repressors and the globin core. *Current Biology*, 3(3), 141–8.
- The UniProt Consortium. (2012). Reorganizing the protein space at the universal protein resource (UniProt). *Nucleic Acids Research*, 40(Database issue), D71-5.
- The UniProt Consortium. (2014). Activities at the universal protein resource (UniProt). *Nucleic Acids Research*, 42(Database issue), D191-8.
- Ye, Y., & Godzik, A. (2003). Flexible structure alignment by chaining aligned fragment pairs allowing twists. *Bioinformatics (Oxford, England)*, 19 Suppl 2, ii246-55.

## Referencias del Capítulo 2

- Alexandrova, A. N., & Jorgensen, W. L. (2011). On the mechanism and rate of spontaneous decomposition of amino acids. *Journal of Physical Chemistry B*, 115(46), 13624–13632.
- Burton, A. S., Stern, J. C., Elsila, J. E., Glavin, D. P., & Dworkin, J. P. (2012). Understanding prebiotic chemistry through the analysis of extraterrestrial amino acids and nucleobases in meteorites. *Chemical Society Reviews*, 41(16), 5459–72.
- Canfield, D. E. (2005). The early history of atmospheric oxygen: homage to Robert M. Garrels. *Annual Review of Earth and Planetary Sciences*, 33(1), 1–36.
- Caspi, R., Altman, T., Billington, R., Dreher, K., Foerster, H., Fulcher, C. A., ... Karp, P. D. (2014). The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of Pathway/Genome Databases. *Nucleic Acids Research*, 42(Database issue), D459-71.
- Clarke, P. H., & Elsdén, S. R. (1980). The earliest catabolic pathways. *Journal of Molecular Evolution*, 15(4), 333–338.
- Ferris, J. P., Joshi, P. C., Edelson, E. H., & Lawless, J. G. (1978). HCN: A plausible source of purines, pyrimidines and amino acids on the primitive earth. *Journal of Molecular Evolution*, 11(4), 293–311.
- Johnson, A. P., Cleaves, H. J., Dworkin, J. P., Glavin, D. P., Lazcano, A., & Bada, J. L. (2008). The Miller volcanic spark discharge experiment. *Science (New York, N.Y.)*, 322(5900), 404.
- Lyons, T. W., Reinhard, C. T., & Planavsky, N. J. (2014). The rise of oxygen in Earth's early ocean and atmosphere. *Nature*, 506(7488), 307–15.
- McMurry, J., & Begley, T. P. (2005). *The Organic Chemistry of Biological Pathways*. Roberts and Company Publishers.
- Michal, G., & Schomburg, D. (Eds.). (2012). *Biochemical pathways: an atlas of biochemistry and molecular biology* (Second). Hoboken, NJ, USA: John Wiley & Sons, Inc.
- Miller, S. L. (1953). A production of amino acids under possible primitive earth conditions. *Science (New York, N.Y.)*, 117(3046), 528–9.

- Miller, S. L., & Orgel, L. E. (1974). *The origins of life on the Earth*. United States: Englewood Cliffs, N. J., Prentice-Hall [c1974].
- Parker, E. T., Cleaves, H. J., Callahan, M. P., Dworkin, J. P., Glavin, D. P., Lazcano, A., & Bada, J. L. (2011). Prebiotic synthesis of methionine and other sulfur-containing organic compounds on the primitive Earth: a contemporary reassessment based on an unpublished 1958 Stanley Miller experiment. *Origins of Life and Evolution of the Biosphere*, 41(3), 201–12.
- Parker, E. T., Cleaves, H. J., Dworkin, J. P., Glavin, D. P., Callahan, M. P., Aubrey, A., ... Bada, J. L. (2011). Primordial synthesis of amines and amino acids in a 1958 Miller H<sub>2</sub>S-rich spark discharge experiment. *Proceedings of the National Academy of Sciences*, 108(14), 5526.
- Pizzarello, S., Huang, Y., & Fuller, M. (2004). The carbon isotopic distribution of Murchison amino acids. *Geochimica et Cosmochimica Acta*, 68(23), 4963–4969.
- Schönheit, P., Buckel, W., & Martin, W. F. (2016). On the origin of heterotrophy. *Trends in Microbiology*, 24(1), 12–25.

## Anexo 1: Catálogo del Catabolismo

El catálogo contiene los detalles de las rutas catabólicas analizadas en presente trabajo. La información de cada enzima es presentada en el siguiente orden: a) su número enzimático (EC), b) su nombre común o nombre sistemático, c) el tipo de enzima según su clasificación en función de su actividad catalítica específica, d) un resumen de la reacción que cataliza y e) los cofactores y/o las coenzimas necesarias para su actividad catalítica que han sido descritas en la literatura especializada. El código de colores se corresponde con la siguiente descripción:

- Purpura → Contiene el nombre del sustrato y el producto de la vía metabólica, así como el código asignado para la ruta.
- Azul → Contiene los títulos sobre la información que se presenta para cada enzima.
- Verde → Las filas verdes indican las reacciones de cada ruta. El color verde intenso y claro se alterna para indicar pasos consecutivos en la vía metabólica. Cuando más de una fila se presenta con el mismo tono de verde, esto indica que existe más de una enzima capaz de producir el metabolito de ese paso específico, ya sea implementando una reacción equivalente o una diferente.
- Amarillo → Las filas de color amarillo indican enzimas capaces de catalizar la reacción si hipotéticamente se invierte el sentido de la reacción que se ha descrito para ellas o si representan enzimas que no son indispensables para completar la ruta, es decir alternativas a un mecanismo previamente descrito. De igual forma el amarillo intenso y el claro se alternan para indicar reacciones contiguas y redundancia enzimática.
- Rojo → Las filas con fondo rojo representa enzimas y/o pasos metabólicos poco caracterizados y a los cuales no se les ha relacionado con una sola secuencia que permitiera buscar la distribución de genes homólogos.
- Letras negras → El texto resaltado en letras negras indican la reacción donde uno de los metabolitos de encrucijada es obtenido.
- Letras rojas → El texto resaltado en letras rojas indica una enzima dependiente de oxígeno molecular.
- La frase “non enzymatic” aparece en la columna designada para el número EC cuando existe información de que la reacción ocurre en ausencia de enzimas, es decir, se considera espontánea.



Enzymes of Glycolysis (from glucose to pyruvate)				
EC Number	Enzyme Name	Class	Reaction	Coenzyme/Cofactor
2.7.1.1	Hexokinase	Transferases	ATP + D-glucose = ADP + D-glucose 6-phosphate	Co2+, Mg2+, Mn2+, Ni2+
2.7.1.2	Glucokinase	Transferases	ATP + D-glucose = ADP + D-glucose 6-phosphate	Co2+, Fe2+, Mg2+, Mn2+, Ni2+, Zn2+
2.7.1.63	polyphosphate Glucokinase	Transferases	polyphosphate + D-glucose = polyphosphate + D-glucose 6-phosphate	Ca2+, Co2+, Cu2+, Fe2+, Mg2+, Mn2+, Zn2+
2.7.1.147	ADP-specific Glucokinase	Transferases	ADP + D-glucose = AMP + D-glucose 6-phosphate	Ca2+, Co2+, Cu2+, Mg2+, Mn2+, Zn2+
5.3.1.9	Glucose-6-phosphate isomerase	Isomerases	D-glucose 6-phosphate = D-fructose 6-phosphate	Co2+, Cu2+, Fe2+, Mg2+, Mn2+, Ni2+, Zn2+
2.7.1.11	6-phosphofructokinase	Transferases	ATP + D-fructose 6-phosphate = ADP + D-fructose 1,6-bisphosphate	Ca2+, Co2+, Fe2+, Mg2+, Mn2+, Ni2+, Zn2+
2.7.1.146	ADP phosphofructokinase	Transferases	ADP + D-fructose 6-phosphate = AMP + D-fructose 1,6-bisphosphate	Ca2+, Co2+, Fe2+, Mg2+, Mn2+, Ni2+
4.1.2.13	fructose-bisphosphate aldolase	Lyases	D-fructose 1,6-bisphosphate = glyceralone P + D-glyceraldehyde 3-P	Ca2+, Cd2+, Co2+, Cu2+, Fe2+, Mg2+, Mn2+, Ni2+, Zn2+
5.3.1.1	Triose-phosphate isomerase	Isomerases	glyceralone phosphate = D-glyceraldehyde 3-P	Mg2+
1.2.1.12	Triosephosphate dehydrogenase	Oxidoreductases	D-glyceraldehyde 3-P + PO4,3- + NAD+ = 1,3-bisphospho-D-glycerate + NADH + H+	NAD
1.2.1.59	Glyceraldehyde-3-phosphate dehydrogenase	Oxidoreductases	D-glyceraldehyde 3-P + PO4,3- + NAD(P)+ = 1,3-bisphospho-D-glycerate + NAD(P)H + H+	NAD, NADP
2.7.2.3	Phosphoglycerate kinase	Transferases	ADP + 1,3-bisphospho-D-glycerate = ATP + 3-phospho-D-glycerate	Ca2+, Co2+, Cd2+, Mg2+, Mn2+, Ni2+, Zn2+
5.4.2.11	Phosphoglycerate mutase	Isomerases	3-phospho-D-glycerate = 2-phospho-D-glycerate	2,3-diphosphoglycerate
5.4.2.12	Phosphoglycerate mutase	Isomerases	3-phospho-D-glycerate = 2-phospho-D-glycerate	Ca2+, Cd2+, Co2+, Mg2+, Mn2+
4.2.1.11	Enolase	Lyases	2-phospho-D-glycerate = phosphoenolpyruvate + H2O	Al3+, Cd2+, Co2+, Cu2+, Fe2+, Mg2+, Mn2+, Ni2+, Zn2+
2.7.1.40	Pyruvate kinase	Transferases	<b>ADP + phosphoenolpyruvate = ATP + pyruvate</b>	Ca2+, Co2+, Cu2+, Fe2+, K+, Mg2+, Mn2+, Na+, Zn2+

**Enzymes of the Ribose Catabolic Pathway (from ribose to D-glyceraldehyde 3-P)**

EC Number	Enzyme Name	Class	Reaction	Coenzyme/Cofactor
2.7.1.15	Ribokinase	Transferases	ATP + D-ribose = ADP + D-ribose 5-phosphate	Cd <sup>2+</sup> , Co <sup>2+</sup> , Cs <sup>2+</sup> , Cu <sup>2+</sup> , Mg <sup>2+</sup> , Mn <sup>2+</sup> , Ni <sup>2+</sup> , Rb <sup>2+</sup> , Zn <sup>2+</sup>
2.2.1.1	Transketolase	Transferases	<b>D-ribose 5-P + D-xylulose 5-P = sedoheptulose 7-P + D-glyceraldehyde 3-P</b>	Thiamine diphosphate
5.3.1.6	Phosphoriboisomerase	Isomerases	D-ribose 5-phosphate = D-ribulose 5-phosphate	Co <sup>2+</sup> , Mg <sup>2+</sup>
5.1.3.1	Phosphoribulose epimerase	Isomerases	D-ribulose 5-phosphate = D-xylulose 5-phosphate	Co <sup>2+</sup> , Fe <sup>2+</sup> , Mg <sup>2+</sup> , Mn <sup>2+</sup> , Zn <sup>2+</sup>

**Enzymes of the Ribose Catabolic Pathway (from ribose to D-fructose 6-P and D-glyceraldehyde 3-P)**

EC Number	Enzyme Name	Class	Reaction	Coenzyme/Cofactor
2.7.1.15	Ribokinase	Transferases	ATP + D-ribose = ADP + D-ribose 5-phosphate	Cd <sup>2+</sup> , Co <sup>2+</sup> , Cs <sup>2+</sup> , Cu <sup>2+</sup> , Mg <sup>2+</sup> , Mn <sup>2+</sup> , Ni <sup>2+</sup> , Rb <sup>2+</sup> , Zn <sup>2+</sup>
2.2.1.1	Transketolase	Transferases	D-ribose 5-P + D-xylulose 5-P = sedoheptulose 7-P + D-glyceraldehyde 3-P	Thiamine diphosphate
5.3.1.6	Phosphoriboisomerase	Isomerases	D-ribose 5-phosphate = D-ribulose 5-phosphate	Co <sup>2+</sup> , Mg <sup>2+</sup>
5.1.3.1	Phosphoribulose epimerase	Isomerases	D-ribulose 5-phosphate = D-xylulose 5-phosphate	Co <sup>2+</sup> , Fe <sup>2+</sup> , Mg <sup>2+</sup> , Mn <sup>2+</sup> , Zn <sup>2+</sup>
2.2.1.2	Transaldolase	Transferases	sedoheptulose 7-P + D-glyceraldehyde 3-P = D-erythrose 4-P + D-fructose 6-P	--
5.1.3.1	Phosphoribulose epimerase	Isomerases	D-ribulose 5-phosphate = D-xylulose 5-phosphate	Co <sup>2+</sup> , Fe <sup>2+</sup> , Mg <sup>2+</sup> , Mn <sup>2+</sup> , Zn <sup>2+</sup>
2.2.1.1	Transketolase	Transferases	<b>D-erythrose 4-P + D-xylulose 5-P = D-fructose 6-P + D-glyceraldehyde 3-P</b>	Thiamine diphosphate

Adenine to Urate (Phase 1)				
EC Number	Enzyme Name	Class	Reaction	Coenzyme/Cofactor
3.5.4.2	adenine deaminase	Hydrolases	adenine + H <sub>2</sub> O = hypoxanthine + NH <sub>3</sub>	Co <sup>2+</sup> , Fe <sup>2+</sup> , Mn <sup>2+</sup> , Zn <sup>2+</sup>
1.17.1.4	xanthine dehydrogenase	Oxidoreductases	hypoxanthine + NAD <sup>+</sup> + H <sub>2</sub> O = xanthine + NADH + H <sup>+</sup>	NAD, FAD
<b>1.17.3.2</b>	<b>xanthine oxidase</b>	<b>Oxidoreductases</b>	<b>hypoxanthine + Oxygen + H<sub>2</sub>O = xanthine + H<sub>2</sub>O<sub>2</sub></b>	<b>FAD</b>
1.17.1.4	xanthine dehydrogenase	Oxidoreductases	xanthine + NAD <sup>+</sup> + H <sub>2</sub> O = urate + NADH + H <sup>+</sup>	NAD, FAD
<b>1.17.3.2</b>	<b>xanthine oxidase</b>	<b>Oxidoreductases</b>	<b>xanthine + H<sub>2</sub>O + O<sub>2</sub> = urate + H<sub>2</sub>O<sub>2</sub></b>	<b>FAD</b>

Urate to allantoin (Phase 2a)				
EC Number	Enzyme Name	Class	Reaction	Coenzyme/Cofactor
<b>1.7.3.3</b>	<b>uric acid oxidase</b>	<b>Oxidoreductases</b>	<b>urate + O<sub>2</sub> + H<sub>2</sub>O = 5-hydroxyisourate + H<sub>2</sub>O<sub>2</sub></b>	<b>Co<sup>2+</sup>, Cu<sup>2+</sup>, Fe<sup>2+</sup>, Mg<sup>2+</sup>, Mn<sup>2+</sup></b>
<b>1.14.13.113</b>	<b>FAD-dependent urate oxidase</b>	<b>Oxidoreductases</b>	<b>urate + FADH + H<sup>+</sup> + O<sub>2</sub> = 5-hydroxyisourate + FAD<sup>+</sup> + H<sub>2</sub>O</b>	<b>FAD, NAD</b>
3.5.2.17	hydroxyisourate hydrolase	Hydrolases	5-hydroxyisourate + H <sub>2</sub> O = 5-hydroxy-2-oxo-4-ureido-2,5-dihydro-1H-imidazole-5-carboxylate	--
4.1.1.97	2-oxo-4-hydroxy-4-carboxy-5-ureidoimidazole decarboxylase	Lyases	5-hydroxy-2-oxo-4-ureido-2,5-dihydro-1H-imidazole-5-carboxylate = (S)(+)-allantoin + CO <sub>2</sub>	--

Urate to allantoin (Phase 2b)				
EC Number	Enzyme Name	Class	Reaction	Coenzyme/Cofactor
<b>1.7.3.3</b>	<b>uric acid oxidase</b>	<b>Oxidoreductases</b>	<b>urate + O<sub>2</sub> + H<sub>2</sub>O = 5-hydroxyisourate + H<sub>2</sub>O<sub>2</sub></b>	<b>Co<sup>2+</sup>, Cu<sup>2+</sup>, Fe<sup>2+</sup>, Mg<sup>2+</sup>, Mn<sup>2+</sup></b>
<b>1.14.13.113</b>	<b>FAD-dependent urate oxidase</b>	<b>Oxidoreductases</b>	<b>urate + FADH + H<sup>+</sup> + O<sub>2</sub> = 5-hydroxyisourate + FAD<sup>+</sup> + H<sub>2</sub>O</b>	<b>FAD, NAD</b>
3.5.2.17	hydroxyisourate hydrolase	Hydrolases	5-hydroxyisourate + H <sub>2</sub> O = 5-hydroxy-2-oxo-4-ureido-2,5-dihydro-1H-imidazole-5-carboxylate	--
NONE	Spontaneous reaction	--	5-hydroxy-2-oxo-4-ureido-2,5-dihydro-1H-imidazole-5-carboxylate + H <sup>+</sup> = (R)(-)-allantoin + CO <sub>2</sub>	--
5.1.99.3	allantoin racemase	Isomerases	(R)(-)-allantoin = (S)(+)-allantoin	--

allantoin to allantoate (Phase 3)				
EC Number	Enzyme Name	Class	Reaction	Coenzyme/Cofactor
3.5.2.5	allantoinase	Hydrolases	(S)-allantoin + H <sub>2</sub> O = allantoate	Mn <sup>2+</sup> , Zn <sup>2+</sup>

allantoate to Urea (Phase 4)				
EC Number	Enzyme Name	Class	Reaction	Coenzyme/Cofactor
3.5.3.4	allantoicase	Hydrolases	allantoate + H <sub>2</sub> O = (S)-ureidoglycolate + urea	Cd <sup>2+</sup> , Mg <sup>2+</sup> , Mn <sup>2+</sup>
4.3.2.3	ureidoglycolate lyase	Lyases	(S)-ureidoglycolate = glyoxylate + urea	Ca <sup>2+</sup> , Co <sup>2+</sup> , Fe <sup>2+</sup> , Mg <sup>2+</sup> , Mn <sup>2+</sup> , Ni <sup>2+</sup> Zn <sup>2+</sup>

Urea to CO <sub>2</sub> and NH <sub>3</sub> (Phase 5)				
EC Number	Enzyme Name	Class	Reaction	Coenzyme/Cofactor
3.5.1.5	urease	Hydrolases	urea + H <sub>2</sub> O = CO <sub>2</sub> + 2 NH <sub>3</sub>	Ca <sup>2+</sup> , Mg <sup>2+</sup> , Mn <sup>2+</sup> , Ni <sup>2+</sup>

Guanine to Urate (Phase 1)				
EC Number	Enzyme Name	Class	Reaction	Coenzyme/Cofactor
3.5.4.3	guanine deaminase	Hydrolases	guanine + H <sub>2</sub> O = xanthine + NH <sub>3</sub>	Ca <sup>2+</sup> , Cu <sup>2+</sup> , Mg <sup>2+</sup> , Mn <sup>2+</sup> , Zn <sup>2+</sup>
1.17.1.4	xanthine dehydrogenase	Oxidoreductases	xanthine + NAD <sup>+</sup> + H <sub>2</sub> O = urate + NADH + H <sup>+</sup>	NAD, FAD
1.17.3.2	<b>xanthine oxidase</b>	<b>Oxidoreductases</b>	<b>xanthine + H<sub>2</sub>O + O<sub>2</sub> = urate + H<sub>2</sub>O<sub>2</sub></b>	<b>FAD</b>

Urate to allantoin (Phase 2a)				
EC Number	Enzyme Name	Class	Reaction	Coenzyme/Cofactor
1.7.3.3	<b>uric acid oxidase</b>	<b>Oxidoreductases</b>	<b>urate + O<sub>2</sub> + H<sub>2</sub>O = 5-hydroxyisourate + H<sub>2</sub>O<sub>2</sub></b>	<b>Co<sup>2+</sup>, Cu<sup>2+</sup>, Fe<sup>2+</sup>, Mg<sup>2+</sup>, Mn<sup>2+</sup></b>
1.14.13.113	<b>FAD-dependent urate oxidase</b>	<b>Oxidoreductases</b>	<b>urate + FADH + H<sup>+</sup> + O<sub>2</sub> = 5-hydroxyisourate + FAD<sup>+</sup> + H<sub>2</sub>O</b>	<b>FAD, NAD</b>
3.5.2.17	hydroxyisourate hydrolase	Hydrolases	5-hydroxyisourate + H <sub>2</sub> O = 5-hydroxy-2-oxo-4-ureido-2,5-dihydro-1H-imidazole-5-carboxylate	--
4.1.1.97	2-oxo-4-hydroxy-4-carboxy-5-ureidoimidazoline decarboxylase	Lyases	<b>5-hydroxy-2-oxo-4-ureido-2,5-dihydro-1H-imidazole-5-carboxylate = (S)(+)-allantoin + CO<sub>2</sub></b>	--

Urate to allantoin (Phase 2b)				
EC Number	Enzyme Name	Class	Reaction	Coenzyme/Cofactor
1.7.3.3	<b>uric acid oxidase</b>	<b>Oxidoreductases</b>	<b>urate + O<sub>2</sub> + H<sub>2</sub>O = 5-hydroxyisourate + H<sub>2</sub>O<sub>2</sub></b>	<b>Co<sup>2+</sup>, Cu<sup>2+</sup>, Fe<sup>2+</sup>, Mg<sup>2+</sup>, Mn<sup>2+</sup></b>
1.14.13.113	<b>FAD-dependent urate oxidase</b>	<b>Oxidoreductases</b>	<b>urate + FADH + H<sup>+</sup> + O<sub>2</sub> = 5-hydroxyisourate + FAD<sup>+</sup> + H<sub>2</sub>O</b>	<b>FAD, NAD</b>
3.5.2.17	hydroxyisourate hydrolase	Hydrolases	5-hydroxyisourate + H <sub>2</sub> O = 5-hydroxy-2-oxo-4-ureido-2,5-dihydro-1H-imidazole-5-carboxylate	--
NONE	Spontaneous reaction	--	5-hydroxy-2-oxo-4-ureido-2,5-dihydro-1H-imidazole-5-carboxylate + H <sup>+</sup> = (R)(-)-allantoin + CO <sub>2</sub>	--
5.1.99.3	allantoin racemase	Isomerases	<b>(R)(-)-allantoin = (S)(+)-allantoin</b>	--

allantoin to allantoate (Phase 3)				
EC Number	Enzyme Name	Class	Reaction	Coenzyme/Cofactor
3.5.2.5	allantoinase	Hydrolases	<b>(S)-allantoin + H<sub>2</sub>O = allantoate</b>	Mn <sup>2+</sup> , Zn <sup>2+</sup>

allantoate to Urea (Phase 4)				
EC Number	Enzyme Name	Class	Reaction	Coenzyme/Cofactor
3.5.3.4	allantoicase	Hydrolases	allantoate + H <sub>2</sub> O = (S)-ureidoglycolate + urea	Cd <sup>2+</sup> , Mg <sup>2+</sup> , Mn <sup>2+</sup>
4.3.2.3	ureidoglycolate lyase	Lyases	<b>(S)-ureidoglycolate = glyoxylate + urea</b>	Ca <sup>2+</sup> , Co <sup>2+</sup> , Fe <sup>2+</sup> , Mg <sup>2+</sup> , Mn <sup>2+</sup> , Ni <sup>2+</sup> , Zn <sup>2+</sup>

Urea to CO <sub>2</sub> and NH <sub>3</sub> (Phase 5)				
EC Number	Enzyme Name	Class	Reaction	Coenzyme/Cofactor
3.5.1.5	urease	Hydrolases	urea + H <sub>2</sub> O = CO <sub>2</sub> + 2 NH <sub>3</sub>	Ca <sup>2+</sup> , Mg <sup>2+</sup> , Mn <sup>2+</sup> , Ni <sup>2+</sup>

**Enzymes of the Cytosine Catabolic Pathway (from cytosine to beta-alanine)**

EC Number	Enzyme Name	Class	Reaction	Coenzyme/Cofactor
3.5.4.1	cytosine aminohydrolase	Hydrolases	cytosine + H <sub>2</sub> O = uracil + NH <sub>3</sub>	Fe <sup>2+</sup> , Zn <sup>2+</sup>
1.3.1.1	5,6-dihydropyrimidine:NAD <sup>+</sup> oxidoreductase	Oxidoreductases	uracil + NADH + H <sup>+</sup> = 5,6-dihydrouracil + NAD <sup>+</sup>	NAD
1.3.1.2	5,6-dihydrouracil:NADP <sup>+</sup> 5-oxidoreductase	Oxidoreductases	uracil + NADPH + H <sup>+</sup> = 5,6-dihydrouracil + NADP <sup>+</sup>	NADP
3.5.2.2	5,6-dihydropyrimidine amidohydrolase	Hydrolases	5,6-dihydrouracil + H <sub>2</sub> O = 3-ureidopropanoate	Co <sup>2+</sup> , Cu <sup>2+</sup> , Fe <sup>2+</sup> , Mg <sup>2+</sup> , Mn <sup>2+</sup> , Zn <sup>2+</sup>
3.5.1.6	N-carbamoyl-beta-alanine amidohydrolase	Hydrolases	<b>3-ureidopropionate + H<sub>2</sub>O = beta-alanine + CO<sub>2</sub> + NH<sub>3</sub></b>	Co <sup>2+</sup> , Mg <sup>2+</sup> , Mn <sup>2+</sup> , Ni <sup>2+</sup> , Zn <sup>2+</sup>

**Enzymes of the Uracil Catabolic Pathway (from uracil to beta-alanine)**

EC Number	Enzyme Name	Class	Reaction	Coenzyme/Cofactor
1.3.1.1	5,6-dihydropyrimidine:NAD <sup>+</sup> oxidoreductase	Oxidoreductases	uracil + NADH + H <sup>+</sup> = 5,6-dihydrouracil + NAD <sup>+</sup>	NAD
1.3.1.2	5,6-dihydrouracil:NADP <sup>+</sup> 5-oxidoreductase	Oxidoreductases	uracil + NADPH + H <sup>+</sup> = 5,6-dihydrouracil + NADP <sup>+</sup>	NADP
3.5.2.2	5,6-dihydropyrimidine amidohydrolase	Hydrolases	5,6-dihydrouracil + H <sub>2</sub> O = 3-ureidopropanoate	Co <sup>2+</sup> , Cu <sup>2+</sup> , Fe <sup>2+</sup> , Mg <sup>2+</sup> , Mn <sup>2+</sup> , Zn <sup>2+</sup>
3.5.1.6	N-carbamoyl-beta-alanine amidohydrolase	Hydrolases	<b>3-ureidopropionate + H<sub>2</sub>O = beta-alanine + CO<sub>2</sub> + NH<sub>3</sub></b>	Co <sup>2+</sup> , Mg <sup>2+</sup> , Mn <sup>2+</sup> , Ni <sup>2+</sup> , Zn <sup>2+</sup>

**Enzymes of the Thymine Catabolic Pathway (from thymine to 3-aminoisobutyrate)**

EC Number	Enzyme Name	Class	Reaction	Coenzyme/Cofactor
1.3.1.1	5,6-dihydropyrimidine:NAD <sup>+</sup> oxidoreductase	Oxidoreductases	thymine + NADH + H <sup>+</sup> = 5,6-dihydrothymine + NAD <sup>+</sup>	NAD
1.3.1.2	5,6-dihydrouracil:NADP <sup>+</sup> 5-oxidoreductase	Oxidoreductases	thymine + NADPH + H <sup>+</sup> = 5,6-dihydrothymine + NADP <sup>+</sup>	NADP
3.5.2.2	5,6-dihydropyrimidine amidohydrolase	Hydrolases	5,6-dihydrothymine + H <sub>2</sub> O = 3-ureidoisobutyrate	Co <sup>2+</sup> , Cu <sup>2+</sup> , Fe <sup>2+</sup> , Mg <sup>2+</sup> , Mn <sup>2+</sup> , Zn <sup>2+</sup>
3.5.1.6	N-carbamoyl-beta-alanine amidohydrolase	Hydrolases	<b>3-ureidoisobutyrate + H<sub>2</sub>O = 3-aminoisobutyrate + CO<sub>2</sub> + NH<sub>3</sub></b>	Co <sup>2+</sup> , Mg <sup>2+</sup> , Mn <sup>2+</sup> , Ni <sup>2+</sup> , Zn <sup>2+</sup>

**Alanine to Pyruvate (Pathway A1)**

EC Number	Enzyme Name	Class	Reaction	Coenzyme/Cofactor
2.6.1.44	L-alanine:glyoxylate aminotransferase	Transferases	<b>L-alanine + glyoxylate = pyruvate + glycine</b>	pyridoxal 5'-phosphate

**Alanine to Pyruvate (Pathway A2)**

EC Number	Enzyme Name	Class	Reaction	Coenzyme/Cofactor
2.6.1.2	L-alanine:2-oxoglutarate aminotransferase	Transferases	<b>L-alanine + 2-oxoglutarate = pyruvate + L-glutamate</b>	pyridoxal 5'-phosphate

**Alanine to Pyruvate (Pathway A3)**

EC Number	Enzyme Name	Class	Reaction	Coenzyme/Cofactor
1.4.1.1	L-alanine:NAD <sup>+</sup> oxidoreductase	Oxidoreductases	<b>L-alanine + H<sub>2</sub>O + NAD<sup>+</sup> = pyruvate + NH<sub>3</sub> + NADH + H<sup>+</sup></b>	NAD, NADP

**Alanine to Pyruvate (Pathway A4)**

EC Number	Enzyme Name	Class	Reaction	Coenzyme/Cofactor
5.1.1.1	alanine racemase	Isomerases	L-alanine = D-alanine	pyridoxal 5'-phosphate
2.6.1.21	D-alanine aminotransferase	Transferases	<b>D-alanine + 2-oxoglutarate = pyruvate + D-glutamate</b>	pyridoxal 5'-phosphate

**Arginine to Glutamate (Pathway R1)**

EC Number	Enzyme Name	Class	Reaction	Cofactor
3.5.3.1	arginase	Hydrolases	L-arginine + H <sub>2</sub> O = L-ornithine + urea	Ca <sup>2+</sup> , Co <sup>2+</sup> , Fe <sup>2+</sup> , Mg <sup>2+</sup> , Mn <sup>2+</sup> , Ni <sup>2+</sup> Zn <sup>2+</sup>
2.6.1.13	ornithine aminotransferase	Transferases	L-ornithine + 2-oxoglutarate = L-glutamate 5-semialdehyde + L-glutamate	pyridoxal 5'-phosphate
non enzymatic	spontaneous reaction	--	L-glutamate 5-semialdehyde = (S)-1-pyrroline-5-carboxylate + H <sub>2</sub> O	--
1.2.1.88	L-pyrroline-5-carboxylate-NAD <sup>+</sup> oxidoreductase	Oxidoreductases	<b>(S)-1-pyrroline-5-carboxylate + NAD<sup>+</sup> + 2 H<sub>2</sub>O = L-glutamate + NADH + H<sup>+</sup></b>	NAD, NADP

**Arginine to Proline (Pathway R2)**

EC Number	Enzyme Name	Class	Reaction	Cofactor
3.5.3.1	arginase	Hydrolases	L-arginine + H <sub>2</sub> O = L-ornithine + urea	Ca <sup>2+</sup> , Co <sup>2+</sup> , Fe <sup>2+</sup> , Mg <sup>2+</sup> , Mn <sup>2+</sup> , Ni <sup>2+</sup> Zn <sup>2+</sup>
4.3.1.12	ornithine cyclase	Lyases	<b>L-ornithine = L-proline + NH<sub>3</sub></b>	NAD

**Asparagine to Oxaloacetate (Pathway N1)**

EC Number	Enzyme Name	Class	Reaction	Cofactor
2.6.1.14	asparagine-oxo-acid transaminase	Transferases	L-asparagine + glyoxylate* = 2-oxosuccinamate + glycine*	pyridoxal 5'-phosphate
3.5.1.3	omega-amidase	Hydrolases	<b>2-oxosuccinamate + H2O = oxaloacetate + NH3</b>	--

**Asparagine to Aspartate (Pathway N2)**

EC Number	Enzyme Name	Class	Reaction	Cofactor
3.5.1.1	asparaginase	Hydrolases	<b>L-asparagine + H2O = L-aspartate + NH3</b>	Ca2+, Co2+, Cu2+, Mg2+, Zn2+

**Asparagine to Aspartate (Pathway N3)**

EC Number	Enzyme Name	Class	Reaction	Cofactor
3.5.1.38	glutaminase-asparaginase	Hydrolases	<b>L-asparagine + H2O = L-aspartate + NH3</b>	Fe3+, Ni2+

**Asparagine to Aspartate (Pathway N4)**

EC Number	Enzyme Name	Class	Reaction	Cofactor
6.3.5.4	asparagine synthase (glutamine-hydrolysing)	Ligases	<b>AMP + diphosphate + L-asparagine = ATP + L-aspartate + NH3</b>	Br-, Ca+, Cl-, Co2+, Cu2+, F-, I-, Mg2+, Mn2+, Na+, Ni2+

**Asparagine to Aspartate (Pathway N5)**

EC Number	Enzyme Name	Class	Reaction	Cofactor
6.3.1.1	aspartate-ammonia ligase	Ligases	<b>AMP + diphosphate + L-asparagine = ATP + L-aspartate + NH3</b>	Mg2+, Mn2+



Aspartate to Oxaloacetate (Pathway D1)				
EC Number	Enzyme Name	Class	Reaction	Cofactor
2.6.1.1	aspartate transaminase	Transferases	L-aspartate + 2-oxoglutarate = oxaloacetate + L-glutamate	pyridoxal 5'-phosphate

Aspartate to Oxaloacetate (Pathway D2)				
EC Number	Enzyme Name	Class	Reaction	Cofactor
1.4.3.2	L-amino-acid oxidase	Oxidoreductases	L-aspartate + H <sub>2</sub> O + O <sub>2</sub> = oxaloacetate + NH <sub>3</sub> + H <sub>2</sub> O <sub>2</sub>	FAD, FMN

Aspartate to Oxaloacetate (Pathway D3)				
EC Number	Enzyme Name	Class	Reaction	Cofactor
1.4.3.16	L-aspartate oxidase	Oxidoreductases	L-aspartate + H <sub>2</sub> O + O <sub>2</sub> = oxaloacetate + NH <sub>3</sub> + H <sub>2</sub> O <sub>2</sub>	FAD

Aspartate to Oxaloacetate (Pathway D4)				
EC Number	Enzyme Name	Class	Reaction	Cofactor
5.1.1.13	aspartate racemase	Isomerases	L-aspartate = D-aspartate	pyridoxal 5'-phosphate
1.4.3.1	D-aspartate oxidase	Oxidoreductases	D-aspartate + H <sub>2</sub> O + O <sub>2</sub> = oxaloacetate + NH <sub>3</sub> + H <sub>2</sub> O <sub>2</sub>	FAD, FMN

Aspartate to Fumarate (Pathway D5)				
EC Number	Enzyme Name	Class	Reaction	Cofactor
4.3.1.1	aspartate ammonia-lyase	Lyases	L-aspartate = fumarate + NH <sub>3</sub>	Ca <sup>2+</sup> , Cd <sup>2+</sup> , Co <sup>2+</sup> , Mg <sup>2+</sup> , Mn <sup>2+</sup> , Zn <sup>2+</sup>

Aspartate to Fumarate (Pathway D6)				
EC Number	Enzyme Name	Class	Reaction	Cofactor
6.3.4.4	adenylosuccinate synthase	Ligases	GTP + IMP + L-aspartate = GDP + orthophosphate + N <sub>6</sub> -(1,2-dicarboxyethyl)-AMP	Ca <sup>2+</sup> , Co <sup>2+</sup> , Mg <sup>2+</sup> , Mn <sup>2+</sup> , Ni <sup>2+</sup>
4.3.2.2	adenylosuccinate lyase	Lyases	N <sub>6</sub> -(1,2-dicarboxyethyl)AMP = fumarate + AMP	--

Aspartate to Fumarate (Pathway D7)				
EC Number	Enzyme Name	Class	Reaction	Cofactor
6.3.4.5	argininosuccinate synthase	Ligases	ATP + L-citrulline + L-aspartate = AMP + diphosphate + 2-(Nomega-L-arginino)succinate	Mg <sup>2+</sup>
4.3.2.1	argininosuccinate lyase	Lyases	2-(Nomega-L-arginino)succinate = fumarate + L-arginine	--

Cysteine to Pyruvate (Pathway C1)				
EC Number	Enzyme Name	Class	Reaction	Cofactor
4.4.1.1	cystathionine gamma-lyase	Lyases	<b>L-cysteine + H<sub>2</sub>O = SH<sub>2</sub> + pyruvate + NH<sub>3</sub></b>	pyridoxal 5'-phosphate

Cysteine to Pyruvate (Pathway C2)				
EC Number	Enzyme Name	Class	Reaction	Cofactor
4.4.1.8	cystathionine beta-lyase	Lyases	<b>L-cysteine + H<sub>2</sub>O = SH<sub>2</sub> + pyruvate + NH<sub>3</sub></b>	pyridoxal 5'-phosphate

Cysteine to Pyruvate (Pathway C3)				
EC Number	Enzyme Name	Class	Reaction	Cofactor
1.8.1.6	cystine reductase	Oxidoreductases	2 L-cysteine + NAD <sup>+</sup> = L-cystine + NADH + H <sup>+</sup>	NAD
4.4.1.1	cystathionine gamma-lyase	Lyases	<b>L-Cystine + H<sub>2</sub>O = pyruvate + NH<sub>3</sub> + thiocysteine</b>	pyridoxal 5'-phosphate
4.4.1.8	cystathionine beta-lyase	Lyases	<b>L-Cystine + H<sub>2</sub>O = pyruvate + NH<sub>3</sub> + thiocysteine</b>	pyridoxal 5'-phosphate

Cysteine to Pyruvate (Pathway C4)				
EC Number	Enzyme Name	Class	Reaction	Cofactor
5.1.1.10	L-amino acid racemase	Isomerases	L-cysteine = D-Cysteine	pyridoxal 5'-phosphate
4.4.1.15	D-cysteine desulfhydrase	Lyases	<b>D-cysteine + H<sub>2</sub>O = SH<sub>2</sub> + NH<sub>3</sub> + pyruvate</b>	pyridoxal 5'-phosphate

Cysteine to Pyruvate (Pathway C5)				
EC Number	Enzyme Name	Class	Reaction	Cofactor
<b>1.13.11.20</b>	<b>cysteine oxidase</b>	<b>Oxidoreductases</b>	<b>L-cysteine + O<sub>2</sub> = L-cysteine-sulfinat</b>	<b>Fe<sup>2+</sup>, NAD, FAD</b>
2.6.1.1	transaminase A	Transferases	L-cysteine-sulfinat + 2-oxoglutarate = 3-sulfinylpyruvate + L-glutamate	pyridoxal 5'-phosphate
non enzymatic	spontaneous reaction	--	<b>3-sulfinylpyruvate + H<sub>2</sub>O = sulfite + pyruvate</b>	--

Cysteine to Pyruvate (Pathway C6)				
EC Number	Enzyme Name	Class	Reaction	Cofactor
2.6.1.1	transaminase A	Transferases	L-cysteine + 2-oxoglutarate = mercaptopyruvate + L-glutamate	pyridoxal 5'-phosphate
2.6.1.3	cysteine transaminase	Transferases	L-cysteine + 2-oxoglutarate = mercaptopyruvate + L-glutamate	pyridoxal 5'-phosphate
2.8.1.2	3-mercaptopyruvate sulfurtransferase	Transferases	<b>mercaptopyruvate + cyanide = pyruvate + thiocyanate</b>	Cu <sup>2+</sup> , Zn <sup>2+</sup>

Cysteine to Pyruvate (Pathway C7)				
EC Number	Enzyme Name	Class	Reaction	Cofactor
4.4.1.10	cysteine lyase	Lyases	L-cysteine + sulfite = SH <sub>2</sub> + L-cysteate	pyridoxal 5'-phosphate
4.4.1.25	L-cysteate sulfo-lyase	Lyases	<b>L-cysteate + H<sub>2</sub>O = pyruvate + bisulfite + NH<sub>3</sub></b>	pyridoxal 5'-phosphate

Cysteine to Pyruvate (Pathway C8)				
EC Number	Enzyme Name	Class	Reaction	Cofactor
4.4.1.10	cysteine lyase	Lyases	L-cysteine + sulfite = SH <sub>2</sub> + L-cysteate	pyridoxal 5'-phosphate
2.6.1.1	aspartate transaminase	Transferases	L-cysteate + 2-oxoglutarate = 3-sulfo-pyruvate + L-glutamate	pyridoxal 5'-phosphate
1.1.1.272 (1.1.1.37)	D-2-hydroxyacid dehydrogenase (NADP <sup>+</sup> )	Oxidoreductases	3-sulfo-pyruvate + NADH + H <sup>+</sup> = (2R)-3-sulfolactate + NAD <sup>+</sup>	NAD, NADP
1.1.1.338	(2R)-3-sulfolactate dehydrogenase (NADP <sup>+</sup> )	Oxidoreductases	3-sulfo-pyruvate + NADPH + H <sup>+</sup> = (2R)-3-sulfolactate + NADP <sup>+</sup>	NADP
4.4.1.24	(2R)-sulfolactate sulfo-lyase	Lyases	<b>(2R)-3-sulfolactate = pyruvate + bisulfite</b>	Co <sup>2+</sup> , Fe <sup>2+</sup> , Mn <sup>2+</sup>

Glutamate to 2-oxoglutarate (Pathway E1)				
EC Number	Enzyme Name	Class	Reaction	Cofactor
1.4.1.2	glutamate dehydrogenase	Oxidoreductases	<b>L-glutamate + H<sub>2</sub>O + NAD<sup>+</sup> = 2-oxoglutarate + NH<sub>3</sub> + NADH + H<sup>+</sup></b>	NAD <sup>+</sup> , NADP <sup>+</sup>

Glutamate to 2-oxoglutarate (Pathway E2)				
EC Number	Enzyme Name	Class	Reaction	Cofactor
1.4.1.3	glutamate dehydrogenase [NAD(P) <sup>+</sup> ]	Oxidoreductases	<b>L-glutamate + H<sub>2</sub>O + NAD(P)<sup>+</sup> = 2-oxoglutarate + NH<sub>3</sub> + NAD(P)H + H<sup>+</sup></b>	NAD <sup>+</sup> , NADP <sup>+</sup>

Glutamate to 2-oxoglutarate (Pathway E3)				
EC Number	Enzyme Name	Class	Reaction	Cofactor
1.4.1.4	glutamate dehydrogenase (NADP <sup>+</sup> )	Oxidoreductases	<b>L-glutamate + H<sub>2</sub>O + NADP<sup>+</sup> = 2-oxoglutarate + NH<sub>3</sub> + NADPH + H<sup>+</sup></b>	NAD <sup>+</sup> , NADP <sup>+</sup>

Glutamate to Glutamine and 2-oxoglutarate (Pathway E4)				
EC Number	Enzyme Name	Class	Reaction	Cofactor
1.4.1.13	glutamate synthase (NADPH)	Oxidoreductases	<b>2 L-glutamate + NADP<sup>+</sup> = L-glutamine + 2-oxoglutarate + NADPH + H<sup>+</sup></b>	Fe <sup>2+</sup> , Mg <sup>2+</sup> +FAD, FMN, NADP <sup>+</sup>

Glutamate to Glutamine and 2-oxoglutarate (Pathway E5)				
EC Number	Enzyme Name	Class	Reaction	Cofactor
1.4.1.14	glutamate synthase (NADH)	Oxidoreductases	<b>2 L-glutamate + NAD<sup>+</sup> = L-glutamine + 2-oxoglutarate + NADH + H<sup>+</sup></b>	Cd <sup>2+</sup> , Mg <sup>2+</sup> , FMN, NAD <sup>+</sup>

**Glutamine to 2-oxoglutarate (Pathway Q1)**

EC Number	Enzyme Name	Class	Reaction	Cofactor
2.6.1.15	L-glutamine transaminase	Transferases	L-glutamine + pyruvate = 2-oxoglutarate + L-alanine	--
3.5.1.3	2-oxoglutarate amidohydrolase	Hydrolases	<b>2-oxoglutarate + H<sub>2</sub>O = 2-oxoglutarate + NH<sub>3</sub></b>	--

**Glutamine to Glutamate (Pathway Q2)**

EC Number	Enzyme Name	Class	Reaction	Cofactor
3.5.1.2	glutaminase	Hydrolases	<b>L-glutamine + H<sub>2</sub>O = L-glutamate + NH<sub>3</sub></b>	Ca <sup>2+</sup>

**Glutamine to Glutamate (Pathway Q3)**

EC Number	Enzyme Name	Class	Reaction	Cofactor
3.5.1.38	glutaminase-asparaginase	Hydrolases	<b>L-glutamine + H<sub>2</sub>O = L-glutamate + NH<sub>3</sub></b>	Fe <sup>3+</sup> , Ni <sup>2+</sup>

**Glutamine to Glutamate (Pathway Q4)**

EC Number	Enzyme Name	Class	Reaction	Cofactor
1.4.1.13	glutamate synthase (NADPH)	Oxidoreductases	<b>L-glutamine + 2-oxoglutarate + NADPH + H<sup>+</sup> = 2 L-glutamate + NADP<sup>+</sup></b>	NAD, FAD, FMN

**Glutamine to Glutamate (Pathway Q5)**

EC Number	Enzyme Name	Class	Reaction	Cofactor
1.4.1.14	glutamate synthase (NADH)	Oxidoreductases	<b>L-glutamine + 2-oxoglutarate + NADH + H<sup>+</sup> = 2 L-glutamate + NAD<sup>+</sup></b>	NAD, FMN

**Glutamine to Glutamate (Pathway Q6)**

EC Number	Enzyme Name	Class	Reaction	Cofactor
6.3.1.2	glutamine synthetase	Ligase	<b>ADP + orthophosphate + L-glutamine = ATP + L-glutamate + NH<sub>3</sub></b>	Ca <sup>2+</sup> , Cd <sup>2+</sup> , Cu <sup>2+</sup> , Fe <sup>2+</sup> , Mg <sup>2+</sup> , Mn <sup>2+</sup> , Ni <sup>2+</sup> , Zn <sup>2+</sup>

Glycine to Glyoxylate (Pathway G1)				
EC Number	Enzyme Name	Class	Reaction	Cofactor
1.4.3.2	L-amino-acid oxidase	Oxidoreductases	glycine + H2O + O2 = glyoxylate + NH3 + H2O2	FAD, FMN

Glycine to Glyoxylate (Pathway G2)				
EC Number	Enzyme Name	Class	Reaction	Cofactor
1.4.3.19	glycine oxidase	Oxidoreductases	glycine + H2O + O2 = glyoxylate + NH3 + H2O2	FAD

Glycine to Serine (Pathway G3)				
EC Number	Enzyme Name	Class	Reaction	Cofactor
2.6.1.45	Serine-glyoxylate transaminase	Transferases	3-hydroxypyruvate + Glycine = L-Serine + Glyoxylate	pyridoxal 5'-phosphate

Glycine to Serine (Pathway G4)				
EC Number	Enzyme Name	Class	Reaction	Cofactor
2.1.2.1	Glycine hydroxymethyltransferase	Transferases	5,10-methylenetetrahydrofolate + Glycine + H2O = tetrahydrofolate + L-Serine	pyridoxal 5'-phosphate

Glycine to Glyoxylate				
EC Number	Enzyme Name	Class	Reaction	Cofactor
1.4.1.10	glycine NAD oxidoreductase	Oxidoreductases	glycine + H2O + NAD+ = glyoxylate + NH3 + NADH + H+	NAD

**Histidine to 2-oxoglutarate (Pathway H1)**

EC Number	Enzyme Name	Class	Reaction	Cofactor
4.3.1.3	histidine ammonia-lyase	Lyases	L-histidine = urocanate + NH <sub>3</sub>	Ca <sup>2+</sup> , Cd <sup>2+</sup> , Fe <sup>2+</sup> , Mg <sup>2+</sup> , Mn <sup>2+</sup> , Ni <sup>2+</sup> , Zn <sup>2+</sup>
4.2.1.49	urocanase	Lyases	urocanate + H <sub>2</sub> O = 4-imidazolone-5-propanoate	NAD
non enzymatic	spontaneous reaction	--	4-imidazolone-5-propanoate + H <sub>2</sub> O = 4-oxoglutaramate + formate + NH <sub>3</sub>	--
non enzymatic	spontaneous reaction	--	<b>4-oxoglutaramate + H<sub>2</sub>O = 2-oxoglutarate + NH<sub>3</sub></b>	--

**Histidine to Glutamate (Pathway H2)**

EC Number	Enzyme Name	Class	Reaction	Cofactor
4.3.1.3	histidine ammonia-lyase	Lyases	L-histidine = urocanate + NH <sub>3</sub>	Ca <sup>2+</sup> , Cd <sup>2+</sup> , Fe <sup>2+</sup> , Mg <sup>2+</sup> , Mn <sup>2+</sup> , Ni <sup>2+</sup> , Zn <sup>2+</sup>
4.2.1.49	urocanase	Lyases	urocanate + H <sub>2</sub> O = 4-imidazolone-5-propanoate	NAD
3.5.2.7	imidazolonepropionase	Hydrolases	4-imidazolone-5-propanoate + H <sub>2</sub> O = N-formimino-L-glutamate + H <sup>+</sup>	Zn <sup>2+</sup>
3.5.3.8	formimidoylglutamase	Hydrolases	<b>N-formimino-L-glutamate + H<sub>2</sub>O = L-glutamate + formamide</b>	glutathione, tetrahydrofolate, Ca <sup>2+</sup> , Cd <sup>2+</sup> , Fe <sup>2+</sup> , Mg <sup>2+</sup> , Mn <sup>2+</sup> , Ni <sup>2+</sup> , Zn <sup>2+</sup>
2.1.2.5	glutamate formimidoyltransferase	Transferases	<b>tetrahydrofolate + N-formimino-L-glutamate = 5-formiminotetrahydrofolate + L-glutamate</b>	tetrahydrofolate

Isoleucine to Acetyl-CoA (Pathway I1)				
EC Number	Enzyme Name	Class	Reaction	Cofactor
1.4.3.2	L-amino-acid oxidase	Oxidoreductases	L-isoleucine + H <sub>2</sub> O + O <sub>2</sub> = (S)-3-methyl-2-oxopentanoate + NH <sub>3</sub> + H <sub>2</sub> O <sub>2</sub>	FAD
2.6.1.42	branched-chain-amino-acid transaminase	Transferases	L-isoleucine + 2-oxoglutarate = (S)-3-methyl-2-oxopentanoate + L-glutamate	pyridoxal 5'-phosphate
1.2.7.7	3-methyl-2-oxobutanoate dehydrogenase	Oxidoreductases	(S)-3-methyl-2-oxopentanoic acid + CoA + 2 oxidized ferredoxin = (S)-2-methylbutanoyl-CoA + CO <sub>2</sub> + 2 reduced ferredoxin + H <sup>+</sup>	ferredoxin
1.3.8.1	butyryl-CoA dehydrogenase	Oxidoreductases	(S)-2-methylbutanoyl-CoA + acceptor = 2-methylbut-2-enoyl-CoA + reduced acceptor	FAD
1.3.99.12	2-methylacyl-CoA dehydrogenase	Oxidoreductases	(S)-2-methylbutanoyl-CoA + acceptor = 2-methylbut-2-enoyl-CoA + reduced acceptor	Ag <sup>+</sup> , Cu <sup>2+</sup> , Hg <sup>2+</sup>
1.3.8.7	medium-chain acyl-CoA dehydrogenase	Oxidoreductases	(S)-2-methylbutanoyl-CoA + acceptor = 2-methylbut-2-enoyl-CoA + reduced acceptor	FAD
4.2.1.17	enoyl-CoA hydratase	Lyases	2-methylbut-2-enoyl-CoA + H <sub>2</sub> O = (S)-3-hydroxy-2-methylbutyryl-CoA	FAD
1.1.1.35	beta-hydroxyacyl dehydrogenase	Oxidoreductases	(S)-3-hydroxy-2-methylbutyryl-CoA + NAD <sup>+</sup> = 2-methylacetoacetyl-CoA + NADH + H <sup>+</sup>	NAD
1.1.1.178	3-hydroxy-2-methylbutyryl-CoA dehydrogenase	Oxidoreductases	(S)-3-hydroxy-2-methylbutyryl-CoA + NAD <sup>+</sup> = 2-methylacetoacetyl-CoA + NADH + H <sup>+</sup>	NAD
2.3.1.16	acetyl-CoA C-acyltransferase	Transferases	CoA + 2-methylacetoacetyl-CoA = propanoyl-CoA + acetyl-CoA	Ca <sup>2+</sup> , Mg <sup>2+</sup>

Isoleucine to Acetyl-CoA (Pathway I2)				
EC Number	Enzyme Name	Class	Reaction	Cofactor
1.4.3.2	L-amino-acid oxidase	Oxidoreductases	L-isoleucine + H <sub>2</sub> O + O <sub>2</sub> = (S)-3-methyl-2-oxopentanoate + NH <sub>3</sub> + H <sub>2</sub> O <sub>2</sub>	FAD
2.6.1.42	branched-chain-amino-acid transaminase	Transferases	L-isoleucine + 2-oxoglutarate = (S)-3-methyl-2-oxopentanoate + L-glutamate	pyridoxal 5'-phosphate
1.2.4.4	branched chain keto acid dehydrogenase	Oxidoreductases	(S)-3-methyl-2-oxopentanoate + thiamin diphosphate = 2-methyl-1-hydroxybutyl-ThPP + CO <sub>2</sub>	CoA, NAD, ThPP
1.2.4.4	branched chain keto acid dehydrogenase	Oxidoreductases	2-methyl-1-hydroxybutyl-ThPP + lipoamide-E = S-(2-methylbutanoyl)-dihydrolipoamide-E + thiamin diphosphate	CoA, NAD, ThPP
2.3.1.168	dihydrolipoyl transacylase	Transferases	CoA + S-(2-methylbutanoyl)-dihydrolipoamide-E = (S)-2-methylbutanoyl-CoA + dihydrolipoamide-E	--
1.8.1.4	dihydrolipoyl dehydrogenase	Oxidoreductases	dihydrolipoamide-E + NAD <sup>+</sup> = lipoamide-E + NADH + H <sup>+</sup>	FAD, NAD
1.3.8.1	butyryl-CoA dehydrogenase	Oxidoreductases	(S)-2-methylbutanoyl-CoA + acceptor = 2-methylbut-2-enoyl-CoA + reduced acceptor	FAD
1.3.99.12	2-methylacyl-CoA dehydrogenase	Oxidoreductases	(S)-2-methylbutanoyl-CoA + acceptor = 2-methylbut-2-enoyl-CoA + reduced acceptor	Ag <sup>+</sup> , Cu <sup>2+</sup> , Hg <sup>2+</sup>
1.3.8.7	medium-chain acyl-CoA dehydrogenase	Oxidoreductases	(S)-2-methylbutanoyl-CoA + acceptor = 2-methylbut-2-enoyl-CoA + reduced acceptor	FAD
4.2.1.17	enoyl-CoA hydratase	Lyases	2-methylbut-2-enoyl-CoA + H <sub>2</sub> O = (S)-3-hydroxy-2-methylbutyryl-CoA	FAD
1.1.1.35	beta-hydroxyacyl dehydrogenase	Oxidoreductases	(S)-3-hydroxy-2-methylbutyryl-CoA + NAD <sup>+</sup> = 2-methylacetoacetyl-CoA + NADH + H <sup>+</sup>	NAD
1.1.1.178	3-hydroxy-2-methylbutyryl-CoA dehydrogenase	Oxidoreductases	(S)-3-hydroxy-2-methylbutyryl-CoA + NAD <sup>+</sup> = 2-methylacetoacetyl-CoA + NADH + H <sup>+</sup>	NAD
2.3.1.16	acetyl-CoA C-acyltransferase	Transferases	CoA + 2-methylacetoacetyl-CoA = propanoyl-CoA + acetyl-CoA	Ca <sup>2+</sup> , Mg <sup>2+</sup>



Isoleucine to Acetyl-CoA to Succinyl-CoA (Pathway I3)				
EC Number	Enzyme Name	Class	Reaction	Cofactor
1.4.3.2	L-amino-acid oxidase	Oxidoreductases	L-isoleucine + H <sub>2</sub> O + O <sub>2</sub> = (S)-3-methyl-2-oxopentanoate + NH <sub>3</sub> + H <sub>2</sub> O <sub>2</sub>	FAD
2.6.1.42	branched-chain-amino-acid transaminase	Transferases	L-isoleucine + 2-oxoglutarate = (S)-3-methyl-2-oxopentanoate + L-glutamate	pyridoxal 5'-phosphate
1.2.7.7	3-methyl-2-oxobutanoate dehydrogenase	Oxidoreductases	(S)-3-methyl-2-oxopentanoic acid + CoA + 2 oxidized ferredoxin = (S)-2-methylbutanoyl-CoA + CO <sub>2</sub> + 2 reduced ferredoxin + H <sup>+</sup>	ferredoxin
1.3.8.1	butyryl-CoA dehydrogenase	Oxidoreductases	(S)-2-methylbutanoyl-CoA + acceptor = 2-methylbut-2-enoyl-CoA + reduced acceptor	FAD
1.3.99.12	2-methylacyl-CoA dehydrogenase	Oxidoreductases	(S)-2-methylbutanoyl-CoA + acceptor = 2-methylbut-2-enoyl-CoA + reduced acceptor	Ag <sup>+</sup> , Cu <sup>2+</sup> , Hg <sup>2+</sup>
1.3.8.7	medium-chain acyl-CoA dehydrogenase	Oxidoreductases	(S)-2-methylbutanoyl-CoA + acceptor = 2-methylbut-2-enoyl-CoA + reduced acceptor	FAD
4.2.1.17	enoyl-CoA hydratase	Lyases	2-methylbut-2-enoyl-CoA + H <sub>2</sub> O = (S)-3-hydroxy-2-methylbutyryl-CoA	FAD
1.1.1.35	beta-hydroxyacyl dehydrogenase	Oxidoreductases	(S)-3-hydroxy-2-methylbutyryl-CoA + NAD <sup>+</sup> = 2-methylacetoacetyl-CoA + NADH + H <sup>+</sup>	NAD
1.1.1.178	3-hydroxy-2-methylbutyryl-CoA dehydrogenase	Oxidoreductases	(S)-3-hydroxy-2-methylbutyryl-CoA + NAD <sup>+</sup> = 2-methylacetoacetyl-CoA + NADH + H <sup>+</sup>	NAD
2.3.1.16	acetyl-CoA C-acyltransferase	Transferases	CoA + 2-methylacetoacetyl-CoA = propanoyl-CoA + acetyl-CoA	Ca <sup>2+</sup> , Mg <sup>2+</sup>
6.4.1.3	propionyl-CoA carboxylase	Ligases	ATP + propanoyl-CoA + HCO <sub>3</sub> <sup>-</sup> = ADP + phosphate + (S)-methylmalonyl-CoA	biotin
5.1.99.1	methylmalonyl-CoA epimerase	Isomerases	(S)-methylmalonyl-CoA = (R)-methylmalonyl-CoA	Co <sup>2+</sup> , Fe <sup>2+</sup> , Mn <sup>2+</sup> , Ni <sup>2+</sup> , Zn <sup>2+</sup>
5.4.99.2	methylmalonyl-CoA mutase	Isomerases	(R)-methylmalonyl-CoA = succinyl-CoA	adenosylcobalamin

Isoleucine to Acetyl-CoA to Succinyl-CoA (Pathway I4)				
EC Number	Enzyme Name	Class	Reaction	Cofactor
1.4.3.2	L-amino-acid oxidase	Oxidoreductases	L-isoleucine + H <sub>2</sub> O + O <sub>2</sub> = (S)-3-methyl-2-oxopentanoate + NH <sub>3</sub> + H <sub>2</sub> O <sub>2</sub>	FAD
2.6.1.42	branched-chain-amino-acid transaminase	Transferases	L-isoleucine + 2-oxoglutarate = (S)-3-methyl-2-oxopentanoate + L-glutamate	pyridoxal 5'-phosphate
1.2.4.4	branched chain keto acid dehydrogenase	Oxidoreductases	(S)-3-methyl-2-oxopentanoate + thiamin diphosphate = 2-methyl-1-hydroxybutyl-ThPP + CO <sub>2</sub>	CoA, NAD, ThPP
1.2.4.4	branched chain keto acid dehydrogenase	Oxidoreductases	2-methyl-1-hydroxybutyl-ThPP + lipoamide-E = S-(2-methylbutanoyl)-dihydrolipoamide-E + thiamin diphosphate	CoA, NAD, ThPP
2.3.1.168	dihydrolipoyl transacylase	Transferases	CoA + S-(2-methylbutanoyl)-dihydrolipoamide-E = (S)-2-methylbutanoyl-CoA + dihydrolipoamide-E	--
1.8.1.4	dihydrolipoyl dehydrogenase	Oxidoreductases	dihydrolipoamide-E + NAD <sup>+</sup> = lipoamide-E + NADH + H <sup>+</sup>	FAD, NAD
1.3.8.1	butyryl-CoA dehydrogenase	Oxidoreductases	(S)-2-methylbutanoyl-CoA + acceptor = 2-methylbut-2-enoyl-CoA + reduced acceptor	FAD
1.3.99.12	2-methylacyl-CoA dehydrogenase	Oxidoreductases	(S)-2-methylbutanoyl-CoA + acceptor = 2-methylbut-2-enoyl-CoA + reduced acceptor	Ag <sup>+</sup> , Cu <sup>2+</sup> , Hg <sup>2+</sup>
1.3.8.7	medium-chain acyl-CoA dehydrogenase	Oxidoreductases	(S)-2-methylbutanoyl-CoA + acceptor = 2-methylbut-2-enoyl-CoA + reduced acceptor	FAD
4.2.1.17	enoyl-CoA hydratase	Lyases	2-methylbut-2-enoyl-CoA + H <sub>2</sub> O = (S)-3-hydroxy-2-methylbutyryl-CoA	FAD
1.1.1.35	beta-hydroxyacyl dehydrogenase	Oxidoreductases	(S)-3-hydroxy-2-methylbutyryl-CoA + NAD <sup>+</sup> = 2-methylacetoacetyl-CoA + NADH + H <sup>+</sup>	NAD
1.1.1.178	3-hydroxy-2-methylbutyryl-CoA dehydrogenase	Oxidoreductases	(S)-3-hydroxy-2-methylbutyryl-CoA + NAD <sup>+</sup> = 2-methylacetoacetyl-CoA + NADH + H <sup>+</sup>	NAD
2.3.1.16	acetyl-CoA C-acyltransferase	Transferases	CoA + 2-methylacetoacetyl-CoA = propanoyl-CoA + acetyl-CoA	Ca <sup>2+</sup> , Mg <sup>2+</sup>
6.4.1.3	propionyl-CoA carboxylase	Ligases	ATP + propanoyl-CoA + HCO <sub>3</sub> <sup>-</sup> = ADP + phosphate + (S)-methylmalonyl-CoA	biotin
5.1.99.1	methylmalonyl-CoA epimerase	Isomerases	(S)-methylmalonyl-CoA = (R)-methylmalonyl-CoA	Co <sup>2+</sup> , Fe <sup>2+</sup> , Mn <sup>2+</sup> , Ni <sup>2+</sup> , Zn <sup>2+</sup>
5.4.99.2	methylmalonyl-CoA mutase	Isomerases	(R)-methylmalonyl-CoA = succinyl-CoA	adenosylcobalamin

Leucine to Acetyl-CoA and Acetoacetate (Pathway L1)				
EC Number	Enzyme Name	Class	Reaction	Cofactor
1.4.1.9	leucine dehydrogenase	Oxidoreductases	L-leucine + H <sub>2</sub> O + NAD <sup>+</sup> = 4-methyl-2-oxopentanoate + NH <sub>3</sub> + NADH + H <sup>+</sup>	NAD
2.6.1.42	branched-chain-amino-acid transaminase	Transferases	L-leucine + 2-oxoglutarate = 4-methyl-2-oxopentanoate + L-glutamate	pyridoxal 5'-phosphate
1.2.7.7	3-methyl-2-oxobutanoate dehydrogenase	Oxidoreductases	4-methyl-2-oxopentanoate + CoA + 2 oxidized ferredoxin = 3-methylbutanoyl-CoA + CO <sub>2</sub> + 2 reduced ferredoxin + H <sup>+</sup>	ferredoxin
1.3.8.4	isovaleryl-CoA dehydrogenase	Oxidoreductases	3-methylbutanoyl-CoA + FAD = 3-methylbut-2-enoyl-CoA + FADH <sub>2</sub>	FAD
1.3.8.7	medium-chain acyl-CoA dehydrogenase	Oxidoreductases	3-methylbutanoyl-CoA + FAD = 3-methylbut-2-enoyl-CoA + FADH <sub>2</sub>	FAD
6.4.1.4	methylcrotonoyl-CoA carboxylase	Ligases	ATP + 3-methylbut-2-enoyl-CoA + HCO <sub>3</sub> <sup>-</sup> = ADP + phosphate + 3-methylglutaconyl-CoA	biotin
4.2.1.18	methylglutaconyl-CoA hydratase	Lyases	3-methylglutaconyl-CoA + H <sub>2</sub> O = (S)-3-hydroxy-3-methylglutaryl-CoA	--
4.1.3.4	hydroxymethylglutaryl-CoA lyase	Lyases	<b>(S)-3-hydroxy-3-methylglutaryl-CoA = acetyl-CoA + acetoacetate</b>	Ca <sup>2+</sup> , Cu <sup>2+</sup> , Mg <sup>2+</sup> , Mn <sup>2+</sup>

Leucine to Acetyl-CoA and Acetoacetate (Pathway L2)				
EC Number	Enzyme Name	Class	Reaction	Cofactor
1.4.1.9	leucine dehydrogenase	Oxidoreductases	L-leucine + H <sub>2</sub> O + NAD <sup>+</sup> = 4-methyl-2-oxopentanoate + NH <sub>3</sub> + NADH + H <sup>+</sup>	NAD
2.6.1.42	branched-chain-amino-acid transaminase	Transferases	L-leucine + 2-oxoglutarate = 4-methyl-2-oxopentanoate + L-glutamate	pyridoxal 5'-phosphate
1.2.4.4	branched chain keto acid dehydrogenase	Oxidoreductases	4-methyl-2-oxopentanoate + thiamin diphosphate = 3-methyl-1-hydroxybutyl-ThPP + CO <sub>2</sub>	CoA, NAD, ThPP
1.2.4.4	branched chain keto acid dehydrogenase	Oxidoreductases	3-methyl-1-hydroxybutyl-ThPP + lipoamide-E = S-(3-methylbutanoyl)-dihydrolipoamide-E + thiamin diphosphate	CoA, NAD, ThPP
2.3.1.168	dihydrolipoyl transacylase	Transferases	CoA + S-(3-methylbutanoyl)-dihydrolipoamide-E = 3-methylbutanoyl-CoA + dihydrolipoamide-E	--
1.8.1.4	dihydrolipoyl dehydrogenase	Oxidoreductases	dihydrolipoamide-E + NAD <sup>+</sup> = lipoamide-E + NADH + H <sup>+</sup>	FAD, NAD
1.3.8.4	isovaleryl-CoA dehydrogenase	Oxidoreductases	3-methylbutanoyl-CoA + FAD = 3-methylbut-2-enoyl-CoA + FADH <sub>2</sub>	FAD
1.3.8.7	medium-chain acyl-CoA dehydrogenase	Oxidoreductases	3-methylbutanoyl-CoA + FAD = 3-methylbut-2-enoyl-CoA + FADH <sub>2</sub>	FAD
6.4.1.4	methylcrotonoyl-CoA carboxylase	Ligases	ATP + 3-methylbut-2-enoyl-CoA + HCO <sub>3</sub> <sup>-</sup> = ADP + phosphate + 3-methylglutaconyl-CoA	biotin
4.2.1.18	methylglutaconyl-CoA hydratase	Lyases	3-methylglutaconyl-CoA + H <sub>2</sub> O = (S)-3-hydroxy-3-methylglutaryl-CoA	--
4.1.3.4	hydroxymethylglutaryl-CoA lyase	Lyases	<b>(S)-3-hydroxy-3-methylglutaryl-CoA = acetyl-CoA + acetoacetate</b>	Ca <sup>2+</sup> , Cu <sup>2+</sup> , Mg <sup>2+</sup> , Mn <sup>2+</sup>

Lysine to Acetyl-CoA (Pathway K1)				
EC Number	Enzyme Name	Class	Reaction	Cofactor
1.5.1.7	lysine-2-oxoglutarate reductase	Oxidoreductases	L-lysine + 2-oxoglutarate + NADH + H <sup>+</sup> = saccharopine + NAD <sup>+</sup> + H <sub>2</sub> O	NAD
1.5.1.8	lysine-ketoglutarate reductase	Oxidoreductases	L-lysine + 2-oxoglutarate + NADPH + H <sup>+</sup> = saccharopine + NADP <sup>+</sup> + H <sub>2</sub> O	NADP
1.5.1.9	saccharopine dehydrogenase	Oxidoreductases	saccharopine + NAD <sup>+</sup> + H <sub>2</sub> O = L-glutamate + L-2-aminoadipate 6-semialdehyde + NADH + H <sup>+</sup>	NAD
1.5.1.10	saccharopine reductase	Oxidoreductases	saccharopine + NADP <sup>+</sup> + H <sub>2</sub> O = L-glutamate + L-2-aminoadipate 6-semialdehyde + NADPH + H <sup>+</sup>	NADP
1.2.1.31	aminoadipate semialdehyde dehydrogenase	Oxidoreductases	L-2-aminoadipate 6-semialdehyde + NAD(P) <sup>+</sup> + H <sub>2</sub> O = L-2-aminoadipate + NAD(P)H + H <sup>+</sup>	NAD, NADP
2.6.1.39	2-aminoadipate transaminase	Transferases	L-2-aminoadipate + 2-oxoglutarate = 2-oxoadipate + L-glutamate	pyridoxal 5'-phosphate
1.2.4.2	2-oxoadipate dehydrogenase	Oxidoreductases	2-oxoadipate + CoA + NAD <sup>+</sup> = glutaryl-CoA + CO <sub>2</sub> + NADH + H <sup>+</sup>	thiamine diphosphate
2.3.1.61	dihydrolipoyl transsuccinylase	Transferases		--
1.3.8.6	glutaryl-CoA dehydrogenase	Oxidoreductases	glutaryl-CoA + electron-transferring flavoprotein = crotonoyl-CoA + reduced electron-transferring flavoprotein + CO <sub>2</sub>	FAD
4.2.1.17	enoyl-CoA hydratase	Lyases	crotonoyl-CoA + H <sub>2</sub> O = (S)-3-hydroxybutanoyl-CoA	FAD
1.1.1.35	3-hydroxyacyl-CoA dehydrogenase	Oxidoreductases	(S)-3-hydroxybutanoyl-CoA + NAD <sup>+</sup> = acetoacetyl-CoA + NADH + H <sup>+</sup>	NAD
2.3.1.9	acetyl-CoA C-acetyltransferase	Transferases	<b>CoA + acetoacetyl-CoA = 2 acetyl-CoA</b>	Ca <sup>2+</sup> , Mg <sup>2+</sup> , Mn <sup>2+</sup>

Lysine to Acetyl-CoA (Pathway K2)				
EC Number	Enzyme Name	Class	Reaction	Cofactor
<b>1.13.12.2</b>	<b>lysine monooxygenase</b>	<b>Oxidoreductases</b>	<b>L-lysine + O<sub>2</sub> = 5-aminopentanamide + CO<sub>2</sub> + H<sub>2</sub>O</b>	<b>FAD</b>
3.5.1.30	5-aminopentanamidase	Hydrolases	5-aminopentanamide + H <sub>2</sub> O = 5-aminopentanoate + NH <sub>3</sub>	Al <sup>3+</sup> , Fe <sup>3+</sup>
2.6.1.48	5-aminovalerate transaminase	Transferases	5-aminopentanoate + 2-oxoglutarate = glutarate semialdehyde + L-glutamate	pyridoxal 5'-phosphate
1.2.1.20	glutarate-semialdehyde dehydrogenase	Oxidoreductases	glutarate semialdehyde + NAD(P) <sup>+</sup> + H <sub>2</sub> O = glutarate + NAD(P)H + 2 H <sup>+</sup>	NAD, NADP
6.2.1.6	glutaryl-CoA synthetase	Ligases	ATP + glutarate + CoA = ADP + phosphate + glutaryl-CoA	--
1.3.8.6	glutaryl-CoA dehydrogenase	Oxidoreductases	glutaryl-CoA + electron-transferring flavoprotein = crotonoyl-CoA + reduced electron-transferring flavoprotein + CO <sub>2</sub>	FAD
4.2.1.17	crotonyl hydratase	Lyases	crotonoyl-CoA + H <sub>2</sub> O = (S)-3-hydroxybutanoyl-CoA	FAD
1.1.1.35	3-hydroxyacyl-CoA dehydrogenase	Oxidoreductases	(S)-3-hydroxybutanoyl-CoA + NAD <sup>+</sup> = acetoacetyl-CoA + NADH + H <sup>+</sup>	NAD
2.3.1.9	acetyl-CoA C-acetyltransferase	Transferases	<b>CoA + acetoacetyl-CoA = 2 acetyl-CoA</b>	Ca <sup>2+</sup> , Mg <sup>2+</sup> , Mn <sup>2+</sup>

Lysine to Acetoacetate (Pathway K3)				
EC Number	Enzyme Name	Class	Reaction	Cofactor
5.4.3.2	lysine 2,3-aminomutase	Isomerases	L-lysine = L-beta-lysine	pyridoxal 5'-phosphate -- S-adenosyl-L-methionine
5.4.3.3	beta-lysine mutase	Isomerases	L-beta-lysine = (3S,5S)-3,5-diaminohexanoate	adenosylcobalamin -- pyridoxal 5'-phosphate
1.4.1.11	L-3,5-diaminohexanoate dehydrogenase	Oxidoreductases	(3S,5S)-3,5-diaminohexanoate + H <sub>2</sub> O + NAD <sup>+</sup> = (S)-5-amino-3-oxohexanoate + NH <sub>3</sub> + NADH + H <sup>+</sup>	NAD
2.3.1.247/Kce	3-keto-5-amino-hexanoate cleavage enzyme	Transferases	(S)-5-amino-3-oxohexanoate + acetyl-CoA = L-3-aminobutyryl-CoA + acetoacetate	Co <sup>2+</sup> , Mn <sup>2+</sup> , Zn <sup>2+</sup>
4.3.1.14	3-aminobutyryl-CoA ammonia-lyase	Lyases	L-3-aminobutyryl-CoA = crotonoyl-CoA + NH <sub>3</sub>	--

Methionine to Cysteine (Pathway M1)				
EC Number	Enzyme Name	Class	Reaction	Cofactor
2.1.1.5	betaine-homocysteine S-methyltransferase	Transferases	dimethylglycine + L-methionine = betaine + L-homocysteine	Zn <sup>2+</sup>
2.1.1.10	homocysteine S-methyltransferase	Transferases	2 L-methionine = S-methyl-L-methionine + L-homocysteine	Cd <sup>2+</sup> , Zn <sup>2+</sup>
2.1.1.13	methionine synthase	Transferases	tetrahydrofolate + L-methionine = 5-methyltetrahydrofolate + L-homocysteine	cobalamin, S-adenosyl-L-methionine
2.1.1.14	cobalamin-independent methionine synthase	Transferases	tetrahydropteroyltri-L-glutamate + L-methionine = 5-methyltetrahydropteroyltri-L-glutamate + L-homocysteine	Ca <sup>2+</sup> , Mg <sup>2+</sup> , Mn <sup>2+</sup> , Zn <sup>2+</sup>
4.2.1.22	cystathionine beta-synthase	Lyases	L-serine + L-homocysteine = L-cystathionine + H <sub>2</sub> O	pyridoxal 5'-phosphate, S-adenosyl-L-methionine
4.4.1.1	cystathionine gamma-lyase	Lyases	<b>L-cystathionine + H<sub>2</sub>O = L-cysteine + NH<sub>3</sub> + 2-oxobutanoate</b>	pyridoxal 5'-phosphate

Methionine to Succinyl-CoA (Pathway M2)				
EC Number	Enzyme Name	Class	Reaction	Cofactor
2.1.1.5	betaine-homocysteine S-methyltransferase	Transferases	dimethylglycine + L-methionine = betaine + L-homocysteine	Zn <sup>2+</sup>
2.1.1.10	homocysteine S-methyltransferase	Transferases	2 L-methionine = S-methyl-L-methionine + L-homocysteine	Cd <sup>2+</sup> , Zn <sup>2+</sup>
2.1.1.13	methionine synthase	Transferases	tetrahydrofolate + L-methionine = 5-methyltetrahydrofolate + L-homocysteine	cobalamin, S-adenosyl-L-methionine
2.1.1.14	cobalamin-independent methionine synthase	Transferases	tetrahydropteroyltri-L-glutamate + L-methionine = 5-methyltetrahydropteroyltri-L-glutamate + L-homocysteine	Ca <sup>2+</sup> , Mg <sup>2+</sup> , Mn <sup>2+</sup> , Zn <sup>2+</sup>
4.2.1.22	cystathionine beta-synthase	Lyases	L-serine + L-homocysteine = L-cystathionine + H <sub>2</sub> O	pyridoxal 5'-phosphate, S-adenosyl-L-methionine
4.4.1.1	cystathionine gamma-lyase	Lyases	<b>L-cystathionine + H<sub>2</sub>O = L-cysteine + NH<sub>3</sub> + 2-oxobutanoate</b>	pyridoxal 5'-phosphate
2.3.1.54	formate C-acetyltransferase	Transferases	2-oxobutanoate + CoA = propanoyl-CoA + formate	Mg <sup>2+</sup>
1.2.7.1	pyruvate synthase	Oxidoreductases	2 oxidized ferredoxin + 2-oxobutanoate + CoA = 2 reduced ferredoxin + propanoyl-CoA + CO <sub>2</sub> + 2 H <sup>+</sup>	Mg <sup>2+</sup>
2.1.3.1	transcarboxylase	Transferases	propanoyl-CoA + oxaloacetate = (S)-methylmalonyl-CoA + pyruvate	biotin
4.1.1.41	methylmalonyl-CoA decarboxylase	Lyases	propanoyl-CoA + CO <sub>2</sub> = (S)-methylmalonyl-CoA	biotin
6.4.1.3	propionyl-CoA carboxylase	Ligases	ATP + propanoyl-CoA + HCO <sub>3</sub> <sup>-</sup> = ADP + phosphate + (S)-methylmalonyl-CoA	biotin
5.1.99.1	methylmalonyl-CoA epimerase	Isomerases	(S)-methylmalonyl-CoA = (R)-methylmalonyl-CoA	Co <sup>2+</sup> , Ni <sup>2+</sup> , Mn <sup>2+</sup> , Zn <sup>2+</sup>
5.4.99.2	methylmalonyl-CoA mutase	Isomerases	<b>(R)-methylmalonyl-CoA = succinyl-CoA</b>	adenosylcobalamin

Phenylalanine to Fumarate (Pathway F1)				
EC Number	Enzyme Name	Class	Reaction	Cofactor
4.3.1.24	phenylalanine deaminase	Lyases	L-phenylalanine = trans-cinnamate + NH <sub>3</sub>	MIO
4.3.1.25	phenylalanine/Tyrosine ammonia-lyase	Lyases	L-phenylalanine = trans-cinnamate + NH <sub>3</sub>	--
non enzymatic	spontaneous reaction	--	trans-cinnamate = trans-3-hydroxycinnamate	--
<b>1.14.13.127</b>	<b>3-(3-hydroxyphenyl)propanoate hydroxylase</b>	<b>Oxidoreductases</b>	<b>trans-3-hydroxycinnamate + O<sub>2</sub> + NADH + H<sup>+</sup> = trans-2,3-dihydroxycinnamate + H<sub>2</sub>O + NAD<sup>+</sup></b>	<b>NAD</b>
<b>1.13.11.16</b>	<b>3-carboxyethylcatechol 2,3-dioxygenase</b>	<b>Oxidoreductases</b>	<b>trans-2,3-dihydroxycinnamate + O<sub>2</sub> = 2-hydroxy-6-ketononatrienedioate</b>	<b>Fe<sup>2+</sup></b>
3.7.1.14	2-hydroxy-6-oxonona-2,4-dienedioate hydrolase	Hydrolases	<b>2-hydroxy-6-ketononatrienedioate + H<sub>2</sub>O = 2-hydroxy-2,4-pentadienoate + fumarate</b>	--

Phenylalanine to Fumarate (Pathway F2)				
EC Number	Enzyme Name	Class	Reaction	Cofactor
4.3.1.24	phenylalanine deaminase	Lyases	L-phenylalanine = trans-cinnamate + NH <sub>3</sub>	MIO
4.3.1.25	phenylalanine/Tyrosine ammonia-lyase	Lyases	L-phenylalanine = trans-cinnamate + NH <sub>3</sub>	--
<b>1.14.12.19</b>	<b>3-phenylpropanoate dioxygenase</b>	<b>Oxidoreductases</b>	<b>trans-cinnamate + O<sub>2</sub> + NADH + H<sup>+</sup> = cis-3-(3-carboxyethenyl)-3,5-cyclohexadiene-1,2-diol + NAD<sup>+</sup></b>	<b>NAD</b>
1.3.1.87	3-(cis-5,6-dihydroxycyclohexa-1,3-dien-1-yl)propanoate dehydrogenase	Oxidoreductases	cis-3-(3-carboxyethenyl)-3,5-cyclohexadiene-1,2-diol + NAD <sup>+</sup> = trans-2,3-dihydroxycinnamate + NADH + H <sup>+</sup>	NAD
<b>1.13.11.16</b>	<b>3-carboxyethylcatechol 2,3-dioxygenase</b>	<b>Oxidoreductases</b>	<b>trans-2,3-dihydroxycinnamate + O<sub>2</sub> = 2-hydroxy-6-ketononatrienedioate</b>	<b>Fe<sup>2+</sup></b>
3.7.1.14	2-hydroxy-6-oxonona-2,4-dienedioate hydrolase	Hydrolases	<b>2-hydroxy-6-ketononatrienedioate + H<sub>2</sub>O = 2-hydroxy-2,4-pentadienoate + fumarate</b>	--

**Proline to Glutamate (Pathway P1)**

EC Number	Enzyme Name	Class	Reaction	Cofactor
1.5.1.2	proline oxidase	Oxidoreductases	L-proline + NAD(P) <sup>+</sup> = 1-pyrroline-5-carboxylate + NAD(P)H + H <sup>+</sup>	NAD, NADP
1.5.5.2	proline dehydrogenase	Oxidoreductases	L-proline + quinone = (S)-1-pyrroline-5-carboxylate + quinol	FAD
1.2.1.88	L-pyrroline-5-carboxylate-NAD <sup>+</sup> oxidoreductase	Oxidoreductases	<b>(S)-1-pyrroline-5-carboxylate + NAD<sup>+</sup> + 2 H<sub>2</sub>O = L-glutamate + NADH + H<sup>+</sup></b>	NAD, NADP

**Proline to 2-oxoglutarate (Pathway P2)**

EC Number	Enzyme Name	Class	Reaction	Cofactor
<b>1.14.11.2</b>	<b>procollagen-proline dioxygenase</b>	<b>Oxidoreductases</b>	<b>L-proline + 2-oxoglutarate + oxygen = hydroxyproline + succinate + CO<sub>2</sub></b>	<b>ascorbate</b>
5.1.1.8	4-hydroxyproline epimerase	Isomerases	trans-4-hydroxy-L-proline = cis-4-hydroxy-D-proline	--
<b>1.4.3.3</b>	<b>D-amino-acid oxidase</b>	<b>Oxidoreductases</b>	<b>cis-4-hydroxy-D-proline + O<sub>2</sub> = 1-pyrroline-4-hydroxy-2-carboxylate + H<sub>2</sub>O<sub>2</sub></b>	<b>FAD</b>
3.5.4.22	1-pyrroline-4-hydroxy-2-carboxylate deaminase	Hydrolases	1-pyrroline-4-hydroxy-2-carboxylate + H <sub>2</sub> O = 2,5-dioxopentanoate + NH <sub>3</sub>	--
1.2.1.3	NAD-aldehyde dehydrogenase	Oxidoreductases	<b>2,5-dioxopentanoate + NAD<sup>+</sup> + H<sub>2</sub>O = 2-oxoglutarate + NADH + H<sup>+</sup></b>	NAD
1.2.1.26	NADP-aldehyde dehydrogenase	Oxidoreductases	<b>2,5-dioxopentanoate + NADP<sup>+</sup> + H<sub>2</sub>O = 2-oxoglutarate + NADPH + H<sup>+</sup></b>	NADP

Serine to Pyruvate - (Pathway S1)				
EC Number	Enzyme Name	Class	Reaction	Cofactor
4.3.1.19	serine deaminase	Lyases	<b>L-serine = pyruvate + NH3</b>	pyridoxal 5'-phosphate

Serine to Pyruvate - (Pathway S2)				
EC Number	Enzyme Name	Class	Reaction	Cofactor
4.3.1.17	L-serine ammonia-lyase	Lyases	<b>L-serine = pyruvate + NH3</b>	pyridoxal 5'-phosphate

Serine to Pyruvate (Pathway S3)				
EC Number	Enzyme Name	Class	Reaction	Cofactor
5.1.1.10	amino-acid racemase	Isomerases	L-serine = D-serine	pyridoxal 5'-phosphate
5.1.1.18	serine racemase	Isomerases	L-serine = D-serine	pyridoxal 5'-phosphate
4.3.1.18	D-serine ammonia-lyase	Lyases	<b>D-serine = pyruvate + NH3</b>	pyridoxal 5'-phosphate

Threonine to Acetyl-CoA and glycine (Pathway T1)				
EC Number	Enzyme Name	Class	Reaction	Cofactor
1.1.1.103	L-threonine dehydrogenase	Oxidoreductases	L-threonine + NAD <sup>+</sup> = L-2-amino-acetoacetate + NADH + H <sup>+</sup>	NAD
2.3.1.29	L-2-amino-acetoacetate CoA ligase	Transferases	<b>L-2-amino-acetoacetate + CoA = acetyl-CoA + glycine</b>	pyridoxal 5'-phosphate

Threonine to glycine (Pathway T2)				
EC Number	Enzyme Name	Class	Reaction	Cofactor
4.1.2.5	L-threonine aldolase	Lyases	<b>L-threonine = glycine + acetaldehyde</b>	pyridoxal 5'-phosphate

Threonine to Acetaldehyde to Acetyl-CoA (Pathway T3)				
EC Number	Enzyme Name	Class	Reaction	Cofactor
4.1.2.5	L-threonine aldolase	Lyases	<b>L-threonine = glycine + acetaldehyde</b>	pyridoxal 5'-phosphate
1.2.1.3	aldehyde dehydrogenase (NAD <sup>+</sup> )	Oxidoreductases	acetaldehyde + NAD <sup>+</sup> + H <sub>2</sub> O = acetate + NADH + H <sup>+</sup>	NAD
1.2.1.4	aldehyde dehydrogenase (NADP <sup>+</sup> )	Oxidoreductases	acetaldehyde + NADP <sup>+</sup> + H <sub>2</sub> O = acetate + NADPH + H <sup>+</sup>	NADP
1.2.1.5	aldehyde dehydrogenase [NAD(P) <sup>+</sup> ]	Oxidoreductases	acetaldehyde + NAD(P) <sup>+</sup> + H <sub>2</sub> O = acetate + NAD(P)H + H <sup>+</sup>	NAD/NADP
1.2.99.3	aldehyde dehydrogenase (PQQ)	Oxidoreductases	acetaldehyde + PQQ + H <sub>2</sub> O = acetate + PQQH <sub>2</sub>	pyrroloquinoline quinone
1.2.99.6	carboxylate reductase	Oxidoreductases	acetaldehyde + acceptor + H <sub>2</sub> O = acetate + reduced acceptor	molybdopterin/NADP
3.1.2.1	acetyl-CoA hydrolase	Hydrolases	<b>acetate + CoA = acetyl-CoA + H<sub>2</sub>O</b>	Ca <sup>2+</sup> , Mg <sup>2+</sup>
6.2.1.1	acetyl-CoA synthetase	Ligases	<b>ATP + acetate + CoA = AMP + diphosphate + acetyl-CoA</b>	--
6.2.1.13	acetyl-CoA synthetase (ADP-forming)	Ligases	<b>ATP + acetate + CoA = ADP + phosphate + acetyl-CoA</b>	Ca <sup>2+</sup> , Co <sup>2+</sup> , Mg <sup>2+</sup> , Mn <sup>2+</sup>
2.8.3.1	propionate CoA-transferase	Transferases	<b>acetate + propanoyl-CoA = acetyl-CoA + propanoate</b>	--
2.8.3.8	acetate CoA-transferase	Transferases	<b>acetoacetyl-CoA + acetate = acetoacetate + acetyl-CoA</b>	Mg <sup>2+</sup>
2.8.3.10	citrate CoA-transferase	Transferases	<b>acetate + (3S)-citryl-CoA = acetyl-CoA + citrate</b>	--

Threonine to Succinyl-CoA (Pathway T4)				
EC Number	Enzyme Name	Class	Reaction	Cofactor
4.3.1.19	threonine deaminase	Lyases	L-threonine = 2-oxobutanoate + NH <sub>3</sub>	pyridoxal 5'-phosphate
2.3.1.54	formate C-acetyltransferase	Transferases	2-oxobutanoate + CoA = propanoyl-CoA + formate	Mg <sup>2+</sup>
1.2.7.1	pyruvate synthase	Oxidoreductases	2 oxidized ferredoxin + 2-oxobutanoate + CoA = 2 reduced ferredoxin + propanoyl-CoA + CO <sub>2</sub> + 2 H <sup>+</sup>	Mg <sup>2+</sup>
2.1.3.1	transcarboxylase	Transferases	propanoyl-CoA + oxaloacetate = (S)-methylmalonyl-CoA + pyruvate	biotin
4.1.1.41	methylmalonyl-CoA decarboxylase	Lyases	propanoyl-CoA + CO <sub>2</sub> = (S)-methylmalonyl-CoA	biotin
6.4.1.3	propionyl-CoA carboxylase	Ligases	ATP + propanoyl-CoA + HCO <sub>3</sub> <sup>-</sup> = ADP + phosphate + (S)-methylmalonyl-CoA	biotin
5.1.99.1	methylmalonyl-CoA epimerase	Isomerases	(S)-methylmalonyl-CoA = (R)-methylmalonyl-CoA	Co <sup>2+</sup> , Ni <sup>2+</sup> , Mn <sup>2+</sup> , Zn <sup>2+</sup>
5.4.99.2	methylmalonyl-CoA mutase	Isomerases	<b>(R)-methylmalonyl-CoA = succinyl-CoA</b>	adenosylcobalamin



Tryptophan to Pyruvate (Pathway W1)				
EC Number	Enzyme Name	Class	Reaction	Cofactor
4.1.99.1	tryptophanase	Lyases	L-tryptophan = indole + 2-aminoprop-2-enoate	pyridoxal 5'-phosphate
non enzymatic	spontaneous reaction	--	2-aminoprop-2-enoate = 2-iminopropanoate	--
non enzymatic	spontaneous reaction	--	<b>2-iminopropanoate + H2O = pyruvate + NH3</b>	--

Tryptophan to Acetyl-CoA (Pathway W2)				
EC Number	Enzyme Name	Class	Reaction	Cofactor
<b>1.13.11.11</b>	<b>tryptophan oxygenase</b>	<b>Oxidoreductases</b>	<b>L-tryptophan + O2 = N-formyl-L-kynurenine</b>	<b>heme-group</b>
<b>1.13.11.52</b>	<b>tryptophan pyrrolase</b>	<b>Oxidoreductases</b>	<b>L-tryptophan + O2 = N-formyl-L-kynurenine</b>	<b>heme-group</b>
3.5.1.9	kynurenine formamidase	Hydrolases	N-formyl-L-kynurenine + H2O = formate + L-kynurenine	pyridoxal 5'-phosphate
<b>1.14.13.9</b>	<b>kynurenine 3-monooxygenase</b>	<b>Oxidoreductases</b>	<b>L-kynurenine + NADPH + H+ + O2 = 3-hydroxy-L-kynurenine + NADP+ + H2O</b>	<b>FAD, NAD, NADP</b>
3.7.1.3	kynureninase	Hydrolases	3-hydroxy-L-kynurenine + H2O = 3-hydroxyanthranilate + L-alanine	pyridoxal 5'-phosphate
<b>1.13.11.6</b>	<b>3-hydroxyanthranilate oxygenase</b>	<b>Oxidoreductases</b>	<b>3-hydroxyanthranilate + O2 = 2-amino-3-carboxyruconate semialdehyde</b>	<b>Fe2+, Ni2+</b>
4.1.1.45	aminocarboxyruconate-semialdehyde decarboxylase	Lyases	2-amino-3-carboxyruconate semialdehyde = 2-aminomuconate semialdehyde + CO2	Cd2+, Co2+, Fe2+, Mn2+, Zn2+
1.2.1.32	aminomuconate-semialdehyde dehydrogenase	Oxidoreductases	2-aminomuconate semialdehyde + NAD+ + H2O = 2-aminomuconate + NADH + 2 H+	NAD
1.5.1.X	--	--	2-aminomuconate + NADH + H+ + H2O = 2-oxoadipate + ammonia + NAD+	NAD, NADP
1.2.4.2	oxoglutarate dehydrogenase	Oxidoreductases	2-oxoadipate + CoA + NAD(P)+ = glutaryl-CoA + CO2 + NAD(P)H + H+	thiamine diphosphate
1.3.8.6	glutaryl-CoA dehydrogenase	Oxidoreductases	glutaryl-CoA + FAD+ = (E)-glutaconyl-CoA + FADH (E)-glutaconyl-CoA = crotonyl-CoA + CO2	FAD
4.2.1.17	enoyl-CoA hydratase	Lyases	crotonyl-CoA + H2O = (S)-3-hydroxybutanoyl-CoA	FAD
1.1.1.35	3-hydroxyacyl-CoA dehydrogenase	Oxidoreductases	(S)-3-hydroxybutanoyl-CoA + NAD(P)+ = acetoacetyl-CoA + NAD(P)H + H+	NAD, NADP
2.3.1.9	acetyl-CoA C-acetyltransferase	Transferases	<b>CoA + acetoacetyl-CoA = 2 acetyl-CoA</b>	Ca2+, Mg2+, Mn2+

Tryptophan to Pyruvate to Acetyl-CoA (Pathway W3)				
EC Number	Enzyme Name	Class	Reaction	Cofactor
<b>1.13.11.11</b>	<b>tryptophan oxygenase</b>	<b>Oxidoreductases</b>	<b>L-tryptophan + O2 = N-formyl-L-kynurenine</b>	<b>heme-group</b>
<b>1.13.11.52</b>	<b>tryptophan pyrrolase</b>	<b>Oxidoreductases</b>	<b>L-tryptophan + O2 = N-formyl-L-kynurenine</b>	<b>heme-group</b>
3.5.1.9	kynurenine formamidase	Hydrolases	N-formyl-L-kynurenine + H2O = formate + L-kynurenine	pyridoxal 5'-phosphate
<b>1.14.13.9</b>	<b>kynurenine 3-monooxygenase</b>	<b>Oxidoreductases</b>	<b>L-kynurenine + NADPH + H+ + O2 = 3-hydroxy-L-kynurenine + NADP+ + H2O</b>	<b>FAD, NAD, NADP</b>
3.7.1.3	kynureninase	Hydrolases	3-hydroxy-L-kynurenine + H2O = 3-hydroxyanthranilate + L-alanine	pyridoxal 5'-phosphate
<b>1.13.11.6</b>	<b>3-hydroxyanthranilate oxygenase</b>	<b>Oxidoreductases</b>	<b>3-hydroxyanthranilate + O2 = 2-amino-3-carboxyruconate semialdehyde</b>	<b>Fe2+, Ni2+</b>
4.1.1.45	aminocarboxyruconate-semialdehyde decarboxylase	Lyases	2-amino-3-carboxyruconate semialdehyde = 2-aminomuconate semialdehyde + CO2	Cd2+, Co2+, Fe2+, Mn2+, Zn2+
1.2.1.32	aminomuconate-semialdehyde dehydrogenase	Oxidoreductases	2-aminomuconate semialdehyde + NAD+ + H2O = 2-aminomuconate + NADH + 2 H+	NAD
3.5.99.5	2-aminomuconate deaminase	Hydrolases	2-aminomuconate + H2O = (3E)-2-oxohex-3-enedioate + NH3	--
4.1.1.77	2-oxo-3-hexenedioate decarboxylase	Lyases	(3E)-2-oxohex-3-enedioate = 2-oxopent-4-enoate + CO2	Mg2+
4.2.1.80	2-oxopent-4-enoate hydratase	Lyases	2-oxopent-4-enoate + H2O = 4-hydroxy-2-oxopentanoate	Ca2+, Co2+, Mg2+, Mn2+, Zn2+
4.1.3.39	4-hydroxy-2-oxovalerate aldolase	Lyases	<b>4-hydroxy-2-oxopentanoate = acetaldehyde + pyruvate</b>	NAD
1.2.1.10	acetaldehyde dehydrogenase	Oxidoreductases	<b>acetaldehyde + CoA + NAD+ = acetyl-CoA + NADH + H+</b>	NAD

**Tyrosine to Acetoacetate and Fumarate (Pathway Y1)**

EC Number	Enzyme Name	Class	Reaction	Cofactor
1.4.1.20	phenylalanine dehydrogenase	Oxidoreductases	L-tyrosine + H <sub>2</sub> O + NAD <sup>+</sup> = 3-(4-hydroxyphenyl)pyruvate + NH <sub>3</sub> + NADH + H <sup>+</sup>	NAD
<b>1.4.3.2</b>	<b>L-amino-acid oxidase</b>	<b>Oxidoreductases</b>	<b>L-tyrosine + H<sub>2</sub>O + O<sub>2</sub> = 3-(4-hydroxyphenyl)pyruvate + NH<sub>3</sub> + H<sub>2</sub>O<sub>2</sub></b>	<b>FAD, FMN</b>
<b>1.13.11.27</b>	<b>4-hydroxyphenylpyruvate dioxygenase</b>	<b>Oxidoreductases</b>	<b>3-(4-hydroxyphenyl)pyruvate + O<sub>2</sub> = homogentisate + CO<sub>2</sub></b>	<b>Cu<sup>2+</sup>, Fe<sup>2+</sup>, Zn<sup>2+</sup></b>
<b>1.13.11.5</b>	<b>homogentisate 1,2-dioxygenase</b>	<b>Oxidoreductases</b>	<b>homogentisate + O<sub>2</sub> = 4-maleylacetoacetate</b>	<b>Fe<sup>2+</sup></b>
5.2.1.2	maleylacetoacetate isomerase	Isomerases	4-maleylacetoacetate = 4-fumarylacetoacetate	--
3.7.1.2	fumarylacetoacetase	Hydrolases	<b>4-fumarylacetoacetate + H<sub>2</sub>O = acetoacetate + fumarate</b>	Ca <sup>2+</sup> , Cu <sup>2+</sup> , Mg <sup>2+</sup>

Valine to Succinyl-CoA (Pathway V1)				
EC Number	Enzyme Name	Class	Reaction	Cofactor
1.4.1.23	valine dehydrogenase (NAD <sup>+</sup> )	Oxidoreductases	L-valine + H <sub>2</sub> O + NAD <sup>+</sup> = 3-methyl-2-oxobutanoate + NH <sub>3</sub> + NADH + H <sup>+</sup>	NAD
2.6.1.42	branched-chain-amino-acid transaminase	Transferases	L-valine + 2-oxoglutarate = 3-methyl-2-oxobutanoate + L-glutamate	pyridoxal 5'-phosphate
1.2.7.7	3-methyl-2-oxobutanoate dehydrogenase	Oxidoreductases	3-methyl-2-oxobutanoic acid + CoA + 2 oxidized ferredoxin = isobutyryl-CoA + CO <sub>2</sub> + 2 reduced ferredoxin + H <sup>+</sup>	ferredoxin
1.3.8.7	acyl-CoA dehydrogenase	Oxidoreductases	isobutyryl-CoA + acceptor = methacrylyl-CoA + reduced acceptor	FAD
1.3.8.1	butyryl-CoA dehydrogenase	Oxidoreductases	isobutyryl-CoA + acceptor = methacrylyl-CoA + reduced acceptor	FAD
1.3.99.12	2-methylacyl-CoA dehydrogenase	Oxidoreductases	isobutyryl-CoA + acceptor = methacrylyl-CoA + reduced acceptor	Ag <sup>+</sup> , Cu <sup>2+</sup> , Hg <sup>2+</sup>
1.3.99.X	2-methylpropanoyl-CoA: 2,3-oxidoreductase	Oxidoreductases	isobutyryl-CoA + acceptor = methacrylyl-CoA + reduced acceptor	--
4.2.1.17	enoyl-CoA hydratase	Lyases	methacrylyl-CoA + H <sub>2</sub> O = (S)-3-hydroxyisobutyryl-CoA	FAD
3.1.2.4	3-hydroxyisobutyryl-CoA hydrolase	Hydrolases	(S)-3-hydroxyisobutyryl-CoA + H <sub>2</sub> O = CoA + (S)-3-hydroxyisobutyrate	--
1.1.1.31	3-hydroxyisobutyrate dehydrogenase	Oxidoreductases	(S)-3-hydroxyisobutyrate + NAD <sup>+</sup> = (S)-methylmalonate semialdehyde + NADH + H <sup>+</sup>	NAD
1.1.1.35	beta-hydroxyacyl dehydrogenase	Oxidoreductases	(S)-3-hydroxyisobutyrate + NAD <sup>+</sup> = (S)-methylmalonate semialdehyde + NADH + H <sup>+</sup>	NAD
1.2.1.3	aldehyde dehydrogenase (NAD <sup>+</sup> )	Oxidoreductases	(S)-methylmalonate semialdehyde + NAD <sup>+</sup> + H <sub>2</sub> O = methylmalonate + NADH + H <sup>+</sup>	NAD
<b>1.2.3.1</b>	<b>aldehyde oxidase</b>	<b>Oxidoreductases</b>	<b>(S)-methylmalonate semialdehyde + O<sub>2</sub> + H<sub>2</sub>O = methylmalonate + H<sub>2</sub>O<sub>2</sub></b>	<b>FAD</b>
6.2.1.-	methylmalonate:CoA ligase (AMP-forming)	Ligases	methylmalonate + CoA + ATP = (R)-methylmalonyl-CoA + AMP + diphosphate	--
5.4.99.2	methylmalonyl-CoA mutase	Isomerases	<b>(R)-methylmalonyl-CoA = succinyl-CoA</b>	adenosylcobalamin

Valine to Succinyl-CoA (Pathway V2)				
EC Number	Enzyme Name	Class	Reaction	Cofactor
1.4.1.23	valine dehydrogenase (NAD <sup>+</sup> )	Oxidoreductases	L-valine + H <sub>2</sub> O + NAD <sup>+</sup> = 3-methyl-2-oxobutanoate + NH <sub>3</sub> + NADH + H <sup>+</sup>	NAD
2.6.1.42	branched-chain-amino-acid transaminase	Transferases	L-valine + 2-oxoglutarate = 3-methyl-2-oxobutanoate + L-glutamate	pyridoxal 5'-phosphate
1.2.4.4	branched chain keto acid dehydrogenase	Oxidoreductases	3-methyl-2-oxobutanoic acid + thiamin diphosphate = 2-methyl-1-hydroxypropyl-ThPP + CO <sub>2</sub>	CoA, NAD, ThPP
1.2.4.4	branched chain keto acid dehydrogenase	Oxidoreductases	2-methyl-1-hydroxypropyl-ThPP + lipoamide-E = S-(2-methylpropanoyl)-dihydrolipoamide-E + thiamin diphosphate	CoA, NAD, ThPP
2.3.1.168	dihydrolipoyl transacylase	Transferases	CoA + S-(2-methylpropionyl)-dihydrolipoamide-E = isobutyryl-CoA + dihydrolipoamide-E	--
1.8.1.4	dihydrolipoyl dehydrogenase	Oxidoreductases	dihydrolipoamide-E + NAD <sup>+</sup> = lipoamide-E + NADH + H <sup>+</sup>	FAD, NAD
1.3.8.7	acyl-CoA dehydrogenase	Oxidoreductases	isobutyryl-CoA + acceptor = methacrylyl-CoA + reduced acceptor	FAD
1.3.8.1	butyryl-CoA dehydrogenase	Oxidoreductases	isobutyryl-CoA + acceptor = methacrylyl-CoA + reduced acceptor	FAD
1.3.99.12	2-methylacyl-CoA dehydrogenase	Oxidoreductases	isobutyryl-CoA + acceptor = methacrylyl-CoA + reduced acceptor	Ag <sup>+</sup> , Cu <sup>2+</sup> , Hg <sup>2+</sup>
1.3.99.X	2-methylpropanoyl-CoA: 2,3-oxidoreductase	Oxidoreductases	isobutyryl-CoA + acceptor = methacrylyl-CoA + reduced acceptor	--
4.2.1.17	enoyl-CoA hydratase	Lyases	methacrylyl-CoA + H <sub>2</sub> O = (S)-3-hydroxyisobutyryl-CoA	FAD
3.1.2.4	3-hydroxyisobutyryl-CoA hydrolase	Hydrolases	(S)-3-hydroxyisobutyryl-CoA + H <sub>2</sub> O = CoA + (S)-3-hydroxyisobutyrate	--
1.1.1.31	3-hydroxyisobutyrate dehydrogenase	Oxidoreductases	(S)-3-hydroxyisobutyrate + NAD <sup>+</sup> = (S)-methylmalonate semialdehyde + NADH + H <sup>+</sup>	NAD
1.1.1.35	beta-hydroxyacyl dehydrogenase	Oxidoreductases	(S)-3-hydroxyisobutyrate + NAD <sup>+</sup> = (S)-methylmalonate semialdehyde + NADH + H <sup>+</sup>	NAD
1.2.1.3	aldehyde dehydrogenase (NAD <sup>+</sup> )	Oxidoreductases	(S)-methylmalonate semialdehyde + NAD <sup>+</sup> + H <sub>2</sub> O = methylmalonate + NADH + H <sup>+</sup>	NAD
<b>1.2.3.1</b>	<b>aldehyde oxidase</b>	<b>Oxidoreductases</b>	<b>(S)-methylmalonate semialdehyde + O<sub>2</sub> + H<sub>2</sub>O = methylmalonate + H<sub>2</sub>O<sub>2</sub></b>	<b>FAD</b>
6.2.1.-	methylmalonate:CoA ligase (AMP-forming)	Ligases	methylmalonate + CoA + ATP = (R)-methylmalonyl-CoA + AMP + diphosphate	--
5.4.99.2	methylmalonyl-CoA mutase	Isomerases	<b>(R)-methylmalonyl-CoA = succinyl-CoA</b>	adenosylcobalamin

Valine to Succinyl-CoA (Pathway V3)				
EC Number	Enzyme Name	Class	Reaction	Cofactor
1.4.1.23	valine dehydrogenase (NAD <sup>+</sup> )	Oxidoreductases	L-valine + H <sub>2</sub> O + NAD <sup>+</sup> = 3-methyl-2-oxobutanoate + NH <sub>3</sub> + NADH + H <sup>+</sup>	NAD
2.6.1.42	branched-chain-amino-acid transaminase	Transferases	L-valine + 2-oxoglutarate = 3-methyl-2-oxobutanoate + L-glutamate	pyridoxal 5'-phosphate
1.2.7.7	3-methyl-2-oxobutanoate dehydrogenase	Oxidoreductases	3-methyl-2-oxobutanoic acid + CoA + 2 oxidized ferredoxin = isobutyryl-CoA + CO <sub>2</sub> + 2 reduced ferredoxin + H <sup>+</sup>	ferredoxin
1.3.8.7	acyl-CoA dehydrogenase	Oxidoreductases	isobutyryl-CoA + acceptor = methacrylyl-CoA + reduced acceptor	FAD
1.3.8.1	butyryl-CoA dehydrogenase	Oxidoreductases	isobutyryl-CoA + acceptor = methacrylyl-CoA + reduced acceptor	FAD
1.3.99.12	2-methylacyl-CoA dehydrogenase	Oxidoreductases	isobutyryl-CoA + acceptor = methacrylyl-CoA + reduced acceptor	Ag <sup>+</sup> , Cu <sup>2+</sup> , Hg <sup>2+</sup>
1.3.99.X	2-methylpropanoyl-CoA: 2,3-oxidoreductase	Oxidoreductases	isobutyryl-CoA + acceptor = methacrylyl-CoA + reduced acceptor	--
4.2.1.17	enoyl-CoA hydratase	Lyases	methacrylyl-CoA + H <sub>2</sub> O = (S)-3-hydroxyisobutyryl-CoA	FAD
3.1.2.4	3-hydroxyisobutyryl-CoA hydrolase	Hydrolases	(S)-3-hydroxyisobutyryl-CoA + H <sub>2</sub> O = CoA + (S)-3-hydroxyisobutyrate	--
1.1.1.31	3-hydroxyisobutyrate dehydrogenase	Oxidoreductases	(S)-3-hydroxyisobutyrate + NAD <sup>+</sup> = (S)-methylmalonate semialdehyde + NADH + H <sup>+</sup>	NAD
1.1.1.35	beta-hydroxyacyl dehydrogenase	Oxidoreductases	(S)-3-hydroxyisobutyrate + NAD <sup>+</sup> = (S)-methylmalonate semialdehyde + NADH + H <sup>+</sup>	NAD
1.2.1.27	methylmalonate-semialdehyde dehydrogenase (CoA-acylating)	Oxidoreductases	(S)-methylmalonate semialdehyde + CoA + NAD <sup>+</sup> = propanoyl-CoA + CO <sub>2</sub> + NADH + H <sup>+</sup>	NAD
6.4.1.3	propionyl-CoA carboxylase	Ligases	ATP + propanoyl-CoA + HCO <sub>3</sub> <sup>-</sup> = ADP + phosphate + (S)-methylmalonyl-CoA	biotin
5.1.99.1	methylmalonyl-CoA epimerase	Isomerases	(S)-methylmalonyl-CoA = (R)-methylmalonyl-CoA	Co <sup>2+</sup> , Fe <sup>2+</sup> , Mn <sup>2+</sup> , Ni <sup>2+</sup> , Zn <sup>2+</sup>
5.4.99.2	methylmalonyl-CoA mutase	Isomerases	<b>(R)-methylmalonyl-CoA = succinyl-CoA</b>	adenosylcobalamin

Valine to Succinyl-CoA (Pathway V4)				
EC Number	Enzyme Name	Class	Reaction	Cofactor
1.4.1.23	valine dehydrogenase (NAD <sup>+</sup> )	Oxidoreductases	L-valine + H <sub>2</sub> O + NAD <sup>+</sup> = 3-methyl-2-oxobutanoate + NH <sub>3</sub> + NADH + H <sup>+</sup>	NAD
2.6.1.42	branched-chain-amino-acid transaminase	Transferases	L-valine + 2-oxoglutarate = 3-methyl-2-oxobutanoate + L-glutamate	pyridoxal 5'-phosphate
1.2.4.4	branched chain keto acid dehydrogenase	Oxidoreductases	3-methyl-2-oxobutanoic acid + thiamin diphosphate = 2-methyl-1-hydroxypropyl-ThPP + CO <sub>2</sub>	CoA, NAD, ThPP
1.2.4.4	branched chain keto acid dehydrogenase	Oxidoreductases	2-methyl-1-hydroxypropyl-ThPP + lipoamide-E = S-(2-methylpropanoyl)-dihydrolipoamide-E + thiamin diphosphate	CoA, NAD, ThPP
2.3.1.168	dihydrolipoyl transacylase	Transferases	CoA + S-(2-methylpropionyl)-dihydrolipoamide-E = isobutyryl-CoA + dihydrolipoamide-E	--
1.8.1.4	dihydrolipoyl dehydrogenase	Oxidoreductases	dihydrolipoamide-E + NAD <sup>+</sup> = lipoamide-E + NADH + H <sup>+</sup>	FAD, NAD
1.3.8.7	acyl-CoA dehydrogenase	Oxidoreductases	isobutyryl-CoA + acceptor = methacrylyl-CoA + reduced acceptor	FAD
1.3.8.1	butyryl-CoA dehydrogenase	Oxidoreductases	isobutyryl-CoA + acceptor = methacrylyl-CoA + reduced acceptor	FAD
1.3.99.12	2-methylacyl-CoA dehydrogenase	Oxidoreductases	isobutyryl-CoA + acceptor = methacrylyl-CoA + reduced acceptor	Ag <sup>+</sup> , Cu <sup>2+</sup> , Hg <sup>2+</sup>
1.3.99.X	2-methylpropanoyl-CoA: 2,3-oxidoreductase	Oxidoreductases	isobutyryl-CoA + acceptor = methacrylyl-CoA + reduced acceptor	--
4.2.1.17	enoyl-CoA hydratase	Lyases	methacrylyl-CoA + H <sub>2</sub> O = (S)-3-hydroxyisobutyryl-CoA	FAD
3.1.2.4	3-hydroxyisobutyryl-CoA hydrolase	Hydrolases	(S)-3-hydroxyisobutyryl-CoA + H <sub>2</sub> O = CoA + (S)-3-hydroxyisobutyrate	--
1.1.1.31	3-hydroxyisobutyrate dehydrogenase	Oxidoreductases	(S)-3-hydroxyisobutyrate + NAD <sup>+</sup> = (S)-methylmalonate semialdehyde + NADH + H <sup>+</sup>	NAD
1.1.1.35	beta-hydroxyacyl dehydrogenase	Oxidoreductases	(S)-3-hydroxyisobutyrate + NAD <sup>+</sup> = (S)-methylmalonate semialdehyde + NADH + H <sup>+</sup>	NAD
1.2.1.27	methylmalonate-semialdehyde dehydrogenase (CoA-acylating)	Oxidoreductases	(S)-methylmalonate semialdehyde + CoA + NAD <sup>+</sup> = propanoyl-CoA + CO <sub>2</sub> + NADH + H <sup>+</sup>	NAD
6.4.1.3	propionyl-CoA carboxylase	Ligases	ATP + propanoyl-CoA + HCO <sub>3</sub> <sup>-</sup> = ADP + phosphate + (S)-methylmalonyl-CoA	biotin
5.1.99.1	methylmalonyl-CoA epimerase	Isomerases	(S)-methylmalonyl-CoA = (R)-methylmalonyl-CoA	Co <sup>2+</sup> , Fe <sup>2+</sup> , Mn <sup>2+</sup> , Ni <sup>2+</sup> , Zn <sup>2+</sup>
5.4.99.2	methylmalonyl-CoA mutase	Isomerases	<b>(R)-methylmalonyl-CoA = succinyl-CoA</b>	adenosylcobalamin