



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

INSTITUTO DE BIOTECNOLOGÍA

**Promiscuidad Catalítica y Adaptabilidad Evolutiva:
Evolución *in vitro* y Reclutamiento Contingente
de un Intermediario Generalista en la Ruta
de Biosíntesis de la Biotina**

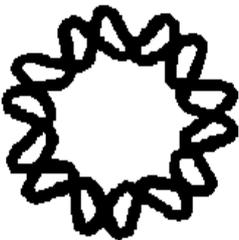
TESIS

QUE PARA OBTENER EL GRADO DE
MAESTRO EN CIENCIAS BIOQUÍMICAS

PRESENTA

BIOL. LUIS GABRIEL CONTRERAS FERRAT

TUTOR: DR. JUAN ENRIQUE MORETT SÁNCHEZ



CUERNAVACA, MORELOS

2009



Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

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

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

Este trabajo se realizó bajo la dirección del Dr. Juan Enrique Morett Sánchez, en el Departamento de Ingeniería Celular y Biocatálisis del Instituto de Biotecnología, UNAM. Durante la realización de este trabajo se contó con la beca de maestría otorgada por el Consejo Nacional de Ciencia y Tecnología.

Agradecimientos

Agradezco al Dr. Enrique Morett la oportunidad de trabajar bajo su dirección.

Al Dr. Humberto Flores Soto, por las críticas y la orientación en todos los aspectos del desarrollo de este trabajo.

Al Dr. Ricardo Grande Cano, al M. en C. Alfredo Mendoza Vargas, a la M. en Admon. Leticia Olvera Rodríguez y a la Biol. Maricela Olvera Rodríguez, por sus contribuciones al trabajo experimental.

Al Dr. Paul Gaytán Colín, al Q. I. Santiago Becerra Ramírez, al Quím. Jorge Arturo Yáñez Ponce de León y al M. en C. Eugenio López Bustos, por su valiosa colaboración.

A la Sra. Juana Ferrer y al Sr. Francisco Reyes Reyes, por el apoyo técnico.

Agradezco la revisión de esta tesis a los sinodales, el Dr. Lorenzo Segovia Forcella, el Dr. Humberto Flores Soto, el Dr. Miguel Ángel Cevallos Gaos y el Dr. David Romero Camarena.

Índice

Resumen.....	6
Introducción.....	7
Antecedentes.....	9
Procesos evolutivos fundamentales en la diversificación enzimática.....	9
Mutación y selección natural.....	9
Duplicación génica.....	11
Tendencias en la evolución estructural.....	13
Tendencias en la evolución de las actividades enzimáticas.....	16
Constreñimientos evolutivos intrínsecos.....	18
La diversificación de las actividades enzimáticas en el contexto metabólico.....	21
Promiscuidad catalítica.....	26
Las determinantes del poder catalítico como base de la promiscuidad.....	26
Distribución de la promiscuidad catalítica.....	27
Promiscuidad de enzimas que evolucionaron recientemente en la naturaleza.....	27
Evolución <i>in vitro</i>	29
Enzimas dependientes de PLP.....	30
La biosíntesis de la biotina.....	32
Planteamiento del problema.....	36
Objetivo general.....	36
Objetivos particulares.....	36
Estrategia experimental.....	37
Materiales y métodos.....	38
Resultados y discusión.....	45
Mutagénesis al azar en la secuencia génica <i>hemA</i> de <i>Bradyrhizobium japonicum</i>	45

Selección <i>in vivo</i> de un alelo <i>hemA</i>^{mut} que revierte de auxotrofia de biotina en una cepa Δ<i>bioF</i> de <i>Escherichia coli</i>.....	45
Detección <i>in vitro</i> de las actividades de HemA^{mut1}	48
El camino evolutivo del alelo <i>hemA</i>^{mut1} sugiere el hallazgo de posiciones de plasticidad.....	49
Exploración de la plasticidad de las posiciones de HemA^{mut1}	55
HemA^{mut1} revierte la auxotrofia asociada a la remoción de un segundo gen de la biosíntesis de la biotina.....	64
Otras variantes exhiben un repertorio de actividades promiscuas.....	66
La promiscuidad catalítica como una componente de la adaptabilidad evolutiva de las enzimas.....	67
Conclusiones	76
Perspectivas.....	77
Referencias.....	78
Protocolos.....	89

Resumen

Las enzimas catalizan reacciones específicas, pero algunas pueden mostrar bajos niveles de promiscuidad catalítica. Las enzimas homólogas suelen compartir mecanismos de catálisis. Y muchas enzimas homólogas han surgido a partir de duplicaciones intragenómicas. A estas enzimas se les denomina parálogas. De manera interesante, las actividades promiscuas de algunas enzimas parálogas corresponden a las actividades principales de otras enzimas parálogas. Dado lo anterior, la promiscuidad catalítica podría ser un vestigio de las actividades principales de los ancestros comunes de las enzimas. Por otro lado, experimentos de evolución *in vitro* muestran que el nivel de una actividad promiscua puede aumentarse mediante mutaciones discretas y sin afectar gravemente la actividad principal. Asimismo, se ha observado que el aumento de una actividad promiscua suele acompañarse por el aumento de otras actividades adicionales. Dado lo anterior, se ha propuesto que el cambio gradual de actividades promiscuas podría permitir la evolución de nuevas actividades biológicamente relevantes. Este cambio gradual incluiría intermediarios evolutivos generalistas.

Si la evolución explota la promiscuidad catalítica, quizá un intermediario generalista podría contribuir a la adecuación celular con más de una actividad. Esta contribución podría incluir el ensamble de pasos de una rta metabólica. En este trabajo se llevaron a cabo experimentos de evolución *in vitro* y selección de variantes enzimáticas capaces de restablecer un paso metabólico esencial de *Escherichia coli*. Como blanco de la evolución *in vitro* se escogió a HemA de *Bradyrhizobium japonicum*. HemA es una enzima asociada a la ruta de biosíntesis de los tetrapirroles. El blanco de la selección fue la actividad llevada a cabo por BioF, una enzima de la ruta de biosíntesis de la biotina. Se evaluó si se alteró la actividad principal de las variantes HemA seleccionadas. Asimismo, se evaluó la capacidad de las variantes enzimáticas para restablecer el paso metabólico llevado a cabo por BioA. De manera importante, BioF y BioA actúan consecutivamente en la ruta de biosíntesis de la biotina. HemA, BioF y BioA catalizan reacciones comparables y son parálogas. Se provee evidencia de la obtención de variantes de HemA capaces de restablecer las actividades de los parálogos BioF y BioA. Los resultados sugieren que una enzima con una promiscuidad catalítica aumentada podría ser útil para la evolución de nuevas actividades biológicamente relevantes. Asimismo, sugieren un escenario evolutivo donde un intermediario generalista podría sostener varias reacciones metabólicas.

Introducción

Muchos aspectos de los sistemas biológicos son entendibles como el resultado de tendencias evolutivas. Una de estas tendencias corresponde a la divergencia de las actividades enzimáticas. Dicha divergencia está asociada a la variación del contenido informacional del DNA. El DNA puede ser afectado por varias formas de mutaciones, entre las que se destacan las duplicaciones génicas y las sustituciones nucleotídicas. Las duplicaciones pueden generar secuencias génicas redundantes, mientras que las sustituciones alteran la estabilidad y la función de las proteínas codificadas. Un efecto adicional concebible es el surgimiento de una nueva actividad enzimática.

El impacto de la divergencia enzimática es apreciable en la complejidad del metabolismo: Cientos de reacciones catalizadas por enzimas canalizan materia y energía hacia las estructuras y funciones celulares. Esta complejidad parece inevitable si la retención de las novedades evolutivas ocurre por un mecanismo determinista. En la síntesis moderna de la teoría evolutiva, tal mecanismo corresponde casi indefectiblemente a la selección natural. En este mismo marco teórico, la evolución es un proceso histórico en que importan la variación heredable, la descendencia compartida y la supervivencia diferencial de los organismos de una población. Dicho lo anterior, se infiere que presiones de selección emergen de los distintos niveles de organización de los sistemas biológicos. Algunas de estas presiones son ambientales, mientras que otras son orgánicas e inclusive intrínsecas de los caracteres.

El surgimiento de una nueva actividad enzimática parece estar constreñido estructuralmente y por el poder catalítico. Por un lado, las enzimas exhiben una estabilidad estructural marginal. Por el otro, la funcionalidad de los sitios activos requiere ambientes químicos particulares y grupos funcionales en posiciones precisas. Ambos factores están asociados a las propiedades fisicoquímicas de la cadena polipeptídica. Así que la exploración evolutiva parece posible a través de mutaciones que preservan el equilibrio entre factores intrínsecos. A esto debe sumarse que la enzima actúa en macroprocesos. Constreñimientos emergen de las interacciones funcionales dentro de la célula. Y estas interacciones deben mantener su cohesión pese a las perturbaciones ambientales. El escenario anterior permite ver porqué se dice que la contingencia es una característica de los procesos evolutivos. Tales procesos dependen de la concomitancia espacio-temporal de muchos factores. Aunque cada factor tiene un peso estadístico *per se*, son tantos que las consecuencias evolutivas resultan impredecibles (Luisi, 2003).

Los estudios de estructuras proteínicas y de genómica presentan un panorama en que muchas enzimas son caracteres parálogos, *i. e.*, descendientes de eventos de duplicación. Ciertas estructuras se han reutilizado de manera repetida, aunque han divergido extensivamente a nivel de secuencia. El repertorio de secuencias compatibles con una estructura supone un potencial evolutivo. Sin embargo, resulta difícil compaginar la divergencia de secuencia, la conservación estructural y la innovación enzimática. Esto es que no puede

determinarse *a priori* la cantidad de variación necesaria para el surgimiento de una nueva actividad. Tal incapacidad de predicción se ve agravada por el desconocimiento de las determinantes de la estabilidad estructural, la cual se ve afectada por la mayoría de las mutaciones. Por otro lado, los estudios de los mecanismos catalíticos y de evolución *in vitro* han permitido una aproximación al potencial evolutivo de las enzimas.

Se ha observado que las enzimas homólogas comparten mecanismos catalíticos. De hecho, esta conservación favorece hipótesis de ancestría común entre secuencias muy divergentes que adoptan una misma estructura. Por la abundancia en la naturaleza y la química compartida de las enzimas parálogas, parece posible que alguna propiedad intrínseca facilite la divergencia catalítica. Evidencias experimentales sugieren que esta propiedad es la promiscuidad catalítica. Algunas enzimas muestran promiscuidad en niveles apenas detectables. La preexistencia de reacciones secundarias podría ser un punto de partida para el surgimiento de actividades biológicamente ventajosas. Experimentos de evolución *in vitro* han demostrado la posibilidad de elevar las reacciones promiscuas hasta niveles importantes para los sistemas biológicos.

Antecedentes

Las enzimas llevan a cabo reacciones específicas con factores de aceleración en el intervalo de 10^7 a 10^{19} . Para poner estos valores en contexto, las mismas reacciones en disolución tendrían tasas espontáneas con vidas medias de 5 s a 1×10^9 años (Benkovic & Hammes-Schiffer, 2003; Wolfenden & Snider, 2001). Por lo anterior, resulta comprensible que las enzimas participen en la mayoría de los procesos celulares, y que la catálisis sea sustantiva en las definiciones operativas de la vida (Tsokolov, 2009; Ruiz-Mirazo *et al.*, 2004). Asimismo, se entiende que los genes de las enzimas sean fracciones significativas del contenido informacional de los genomas. Por ejemplo, más de la tercera parte de los ~ 4500 genes de *E. coli* es codifican enzimas (Riley *et al.*, 2006).

Una interrogante fundamental es el surgimiento de nuevas actividades enzimáticas. El problema es de tipo molecular, pero sus implicaciones afectan a la evolución de los organismos. Esto se debe a que las enzimas constituyen el metabolismo, el macroproceso que sostiene a las estructuras y funciones celulares. Asimismo, se debe a que la expansión del repertorio enzimático ha permitido la divergencia biológica. Ambos factores se ven representados en los microorganismos, que extraen energía y sintetizan compuestos esenciales a partir de lo disponible en ambientes heterogéneos o empobrecidos. En otras palabras, el problema es el surgimiento de unidades funcionales de un sistema cuya subsistencia depende de la interacción con el ambiente.

Considérese a las enzimas como caracteres codificados a nivel genético. La abundancia de estos caracteres en la naturaleza sugiere dos modos de divergencia evolutiva. El primero ocurre por eventos de especiación y resulta en enzimas con funciones comparables en especies diferentes. A estos caracteres se les denomina ortólogos. Por otro lado, también es posible la divergencia funcional tras la duplicación intragenómica de una secuencia codificante. En este caso, los caracteres resultantes son llamados parálogos (Gogarten & Olendzensky, 1999; Sonnhammer & Koonin, 2002). Los caracteres parálogos suelen tener papeles distintos en la célula y son un recipiente principal de la divergencia a nivel catalítico. ¿Qué procesos derivan en estas consecuencias evolutivas?

Procesos evolutivos fundamentales en la divergencia enzimática: Mutación y selección natural

La codificación genética y la replicación permiten la transmisión generacional de los caracteres. Pero éstos pueden variar por efecto de las mutaciones. De los tipos de mutación conocidos, nos resultan interesantes las substituciones debidas a errores de la DNA polimerasa. La baja tasa de error de esta enzima sugiere que un gen podría acumular substituciones de manera gradual. Que el producto de la expresión génica se altere por una substitución depende de la degeneración del código genético. Si el producto llega alterarse en su

estructura, entonces pueden esperarse efectos funcionales. Desde este punto, procesos en escala orgánica determinan la permanencia a largo plazo de las variantes de un carácter.

La existencia de los genotipos está limitada. No obstante, esta transitoriedad puede evitarse por la reproducción dentro de las poblaciones. Las poblaciones sirven como reservorios de genotipos, cuya abundancia depende del efecto de ciertos mecanismos. Uno de estos mecanismos emerge de las consecuencias funcionales de los caracteres codificados, lo que supone una contribución a la capacidad orgánica para sobrevivir y dejar descendencia. El peso cuantitativo de esta capacidad se conoce como adecuación (Orr, 2005), pero el mecanismo subyacente es la selección natural.

La selección natural es un proceso físico de exclusión competitiva proyectado sobre individuos que crecen diferencialmente (Hoelzer *et al.*, 2006). Esto es que algunos individuos sobreviven y se reproducen más que otros, debido a la adecuación que confieren sus caracteres. Para la síntesis moderna de la teoría evolutiva, el resultado de la selección es un sesgo estadístico en la tasa de supervivencia de los genotipos. En una escala temporal de generaciones, los genotipos más frecuentes son los que aportan más adecuación. En una escala evolutiva, otra consecuencia de la selección corresponde a las adaptaciones.

La narrativa biológica abunda en descripciones de cómo ciertos atributos orgánicos están diseñados para responder a aspectos particulares del ambiente. Estos atributos confieren mayor adecuación y pueden definirse como adaptaciones, si su surgimiento es explicable por la selección natural (Winther, 2008; Shanahan, 2008). Esta conexión causal puede implicar la construcción de fenotipos y caracteres complejos, como procesos bioquímicos y estructuras anatómicas. Sin embargo, los procesos evolutivos no están subordinados al aumento o mejoramiento de las partes de un sistema (O'Grady, 1984). Por el contrario, la selección puede causar simplemente el mantenimiento del *status quo*.

Cuando la selección se considera en términos genéticos, las mutaciones pueden ser positivas, neutrales ó negativas por su efecto en la adecuación. Se dice que la selección actúa de manera direccional al retener mutaciones positivas, y que es purificadora al eliminar las negativas. El *status quo* implica que la función génica está en un óptimo, y que son raras las mutaciones capaces de aumentarla (DePristo *et al.*, 2005). De acuerdo con la teoría neutral, la mayoría de las diferencias interespecíficas a nivel de secuencias de DNA y de proteínas son indistintas para la adecuación (Nei, 2005). Esto sugiere que la selección purificadora ha prevalecido en la constitución de los genomas, y que las mutaciones positivas son muy raras. Lo anterior atrae un problema interpretativo, pues la mayoría de las mutaciones tienen un efecto negativo en la estructura y la función de las proteínas (como se explica más adelante). Una generalización teórica más reciente considera que substituciones ligeramente negativas pueden ser *cuasi* neutrales. La presencia de estas substituciones se ve afectada por la selección o la deriva aleatoria en forma dependiente del tamaño de la población en que ocurren (Ohta, 1992). De manera evidente, los efectos de la selección tienen explicaciones a

nivel de individuos y de poblacionales. Ahora debe establecerse si a nivel molecular hay unidades de selección natural.

Las unidades de selección satisfacen una condición de individualidad (Winther, 2008; Mitton, 2001; Gould & Lloyd, 1999). En los organismos, dicha condición se identifica con la delimitación física y una interdependencia de partes. Los genes no se ajustan a estos factores. No obstante, al igual que los organismos, tienen puntos de nacimiento y de muerte separados por un período de estabilidad más o menos prolongado. Asimismo, puede llamarse unidad de selección a toda entidad que: (i) interacciona con un ambiente; (ii) muestra caracteres que imparten una reproducción diferencial, y; (iii) altera su base hereditaria de manera generacional. Puede inferirse que los genes codifican caracteres y que representan unidades de selección (Gould & Lloyd, 1999). Muchas innovaciones evolutivas han ocurrido a través de la duplicación génica, pues ésta provee material hereditario para una dinámica evolutiva sometida al escrutinio de la selección natural.

Procesos evolutivos fundamentales en la diversificación enzimática: Duplicación génica

La duplicación puede deberse al deslizamiento retrógrado de las cadenas de DNA en la replicación, así como a la recombinación de sitios homólogos entrecruzados de forma asimétrica (Patthy, 1999). Estos eventos pueden abarcar regiones intragénicas, secuencias codificantes e inclusive segmentos cromosómicos (Long, 2001). Nuevos genes funcionales han emergido por la combinación de exones y la iteración de motivos de secuencia. (Long *et al.*, 2003). La redundancia génica también ocurre por transferencias horizontales, transposiciones y segregaciones cromosómicas incompletas (Hughes, 2005; Sankoff, 2001). Los genes funcionales generados por duplicaciones arrostran alguno de los destinos representados en la Figura 1.

Un destino posible es la persistencia de las copias redundantes y un aumento concomitante de la transcripción de la secuencia ancestral. Se supone que este aumento suele ser neutral (Massingham *et al.*, 2001). De manera alternativa, sustituciones en distintos elementos reguladores pueden dar lugar a la partición de funciones ancestrales (Force *et al.*, 1999). Esta hipótesis propone que los duplicados llegan a expresarse de manera diferencial en el tiempo y el espacio. Por otro lado, la selección puede actuar de forma asimétrica entre los duplicados de un gen ancestral. En este caso, una copia puede mantenerse bajo el efecto purificador de la selección, mientras que la otra acumula mutaciones. Se supone que la acumulación de mutaciones puede construir pseudogenes, pero también puede iniciar la divergencia funcional (Taylor & Raes, 2004). El impacto evolutivo de la duplicación puede medirse en la configuración de los genomas eucariontes y procariontes.

Eventos masivos de duplicación se asocian a la aparición de las células eucariontes y el *phylum* Chordata (Holland, 1999; Makarova *et al.*, 2005). Asimismo, los genomas de *Saccharomyces cerevisiae* y *Arabidopsis thaliana* parecen descender de duplicaciones cromosómicas (Wolfe & Shields, 1997; Kellis *et*

al., 2004; Sankoff, 2001). Puede suponerse que dichas duplicaciones genómicas han tenido un papel, tanto en la divergencia de linajes como en el aumento de la complejidad orgánica. Otros eventos de duplicación encontrados en eucariontes implican la redundancia de secuencias génicas particulares. Por ejemplo, hay variaciones del número de genes para la gliceraldehído-3-fosfato deshidrogenasa de *Drosophila melanogaster*, así como para los pigmentos visuales de humanos (Taylor & Raes, 2004; Takano *et al.*, 1989; Neitz & Keitz, 1995).

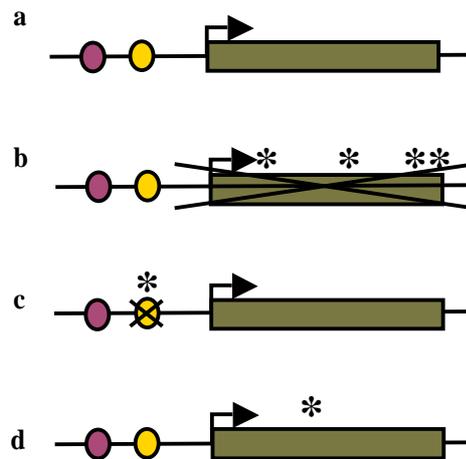


Figura 1. Alternativas de pérdida y diversificación funcional en genes duplicados. La duplicación génica seguida de mutaciones (asteriscos) puede afectar a los descendientes de un gen (rectángulo verde) con elementos de regulación en *cis* (elipses morada y amarilla). (a) Un duplicado puede quedar bajo selección purificadora, mientras que en el otro se relajan los contraejemplos sobre la evolución (b-d). En (b), la acumulación de mutaciones conduce a la pseudogenización. En (c), la incidencia de mutaciones en un elemento de regulación puede conducir a la expresión diferencial del duplicado. En (d), las mutaciones pueden resultar en el surgimiento de una nueva función.

En el análisis de 106 genomas procariontes, entre 7 % y 41 % de los genes descienden de duplicaciones. En *E. coli*, alrededor del 50 % de los genes representan caracteres parálogos (Labedan & Riley 1995). Estas observaciones sugieren que las consecuencias evolutivas de la duplicación pueden rastrearse hasta el establecimiento de los procesos metabólicos fundamentales. De acuerdo con esta hipótesis, ciertas duplicaciones están conservadas de manera casi universal, *e. g.*, aminoacil-tRNA sintetasas y helicasas (Gogarten & Olendzenski, 1999). El nexo entre la duplicación y la divergencia enzimática es apreciable en *E. coli*, donde 60 % de las enzimas del metabolismo de moléculas pequeñas pertenece a pares o grupos de parálogos (Labedan & Riley, 1999).

La evolución suele ser considerada como un proceso histórico en el que exploraciones aleatorias y efectos de difusión resultan en tendencias (Witting, 2003). De acuerdo con esto, los ciclos de duplicación,

mutación y selección han dado lugar a una enorme cantidad de genes. Pero el significado biológico de estos procesos se extiende a la divergencia funcional de las proteínas. La historia natural de las enzimas muestra que la tendencia de divergencia funcional puede expresarse como una tendencia de reutilización.

Tendencias en la evolución estructural

Las enzimas alcanzan una estructura tridimensional por un proceso de autoensamblaje conocido como plegamiento. En este proceso, interacciones débiles cooperan para causar una transición en la cadena polipeptídica, desde un estado desordenado y lineal, hasta uno ordenado y topológicamente único (Honig, 1999; Creighton, 1990). El resultado es la conformación nativa, en la que combinaciones de configuraciones periódicas exhiben orientaciones distintivas. Si la divergencia de las actividades enzimáticas es un problema central de la evolución, el del origen de las estructuras que la sostienen es uno consubstancial e insoslayable.

Algunos aspectos de la codificación estructural son predecibles, como la propensión de patrones de secuencia a adoptar conformaciones periódicas locales (Honig, 1999; Creighton, 1990). Las contribuciones de los residuos de una secuencia son de tipo cooperativo y contextual. No obstante, se desconocen los elementos de la secuencia que especifican al autoensamblaje y a la topología (Honig, 1999). Más aún, resulta enigmático que una misma topología corresponda a secuencias cuya identidad es casi la esperada al azar. Lo anterior conduce al concepto del espacio de secuencia, definido como todas las combinaciones posibles en una sucesión de residuos de aminoácidos. Partiendo de una cadena con N residuos, y de la disponibilidad de 20 aminoácidos, el número de combinaciones en la sucesión es 20^N . Esta numerología suele usarse con trivialidad, pero en ella hay un mensaje valioso (Dryden *et al.*, 2008; Luisi, 2003). A saber, el espacio de secuencia ocupado por las proteínas es una fracción del universo de las secuencias posibles. Aunque las estructuras eluden todavía una predicción *ab initio*, la aproximación a su evolución puede lograrse por la simple observación de cómo se distribuyen en la naturaleza.

La intuición acierta en que la conservación estructural constriñe la divergencia a patrones no aleatorios. En la suposición de un espacio de secuencia inconmensurable, la identidad significativa entre dos secuencias parece una evidencia de ancestría común. Metodologías automatizadas como PsiBLAST detectan semejanzas entre las secuencias. La complementación de las semejanzas con evidencias estructurales y funcionales permite la agrupación de las proteínas homólogas en familias y superfamilias. La fracción de residuos idénticos entre secuencias de una familia suele ser $> 30\%$. Los valores correspondientes a nivel de superfamilia son $< 50\%$, aunque pueden estar por abajo de 20% (Glasner *et al.*, 2007). Bases de información estructural como SCOP y CATH incorporan familias y superfamilias en esquemas de clasificación jerárquica, los cuales están basados en el concepto de dominios.

Los dominios son unidades de estructura y función delimitables a nivel de secuencia. Los análisis de asignación estructural a escala genómica indican que ~ 60 % de las proteínas procariontes, y al menos 80 % de las eucariontes, consisten en más de un dominio (Orengo & Thornton, 2005). Ciertos dominios existen en combinaciones o como proteínas independientes. Por lo anterior, los dominios también son considerados como unidades evolutivas. La fusión de estas unidades ha originado muchísimas proteínas; el intercambio e invasión de partes de éstas también han contribuido a la divergencia estructural (Thornton *et al.*, 1999; Aroul-Selvam *et al.*, 2004). Una pregunta fundamental es cuántos dominios existen y cómo se han usado.

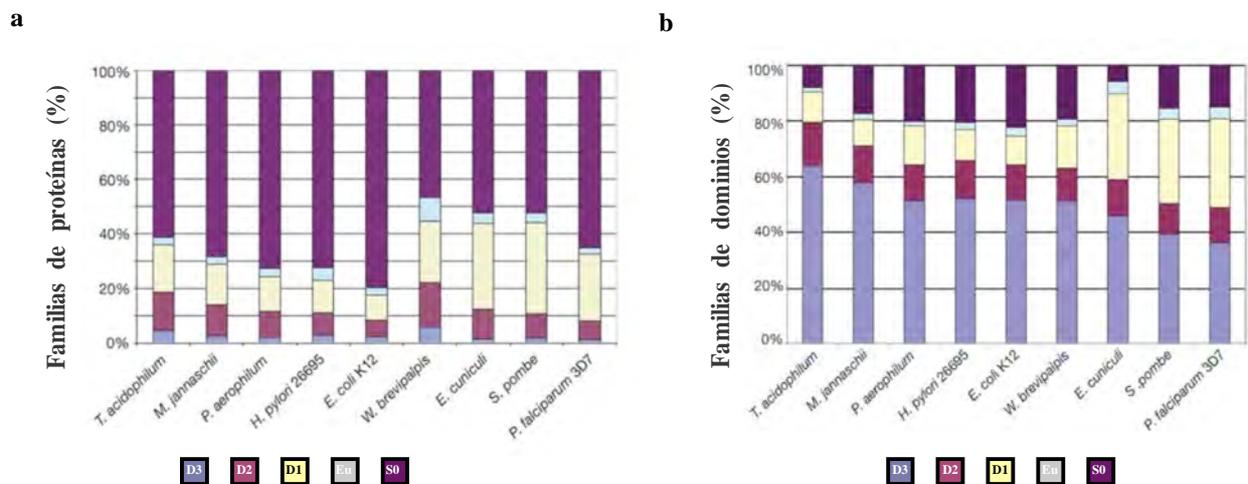


Figura 2. Distribución de los dominios proteínicos en 9 especies representativas de los tres dominios de la vida. (a) Como porcentaje de todas las combinaciones de dominios proteínicos asignadas a un genoma, se muestran las familias de proteínas distribuidas en un dominio (D1), dos dominios (D2) o tres dominios de la vida (D3), en la especie de origen únicamente (Eu), o como singletons (S0). **(b)** Como porcentaje de todos los dominios asignados a un genoma, se muestran las familias de dominios comunes a un dominio (D1), dos dominios (D2), tres dominios (D3) en el organismo de origen únicamente (Eu), o como singletons (S0). (Tomada de Orengo & Thornton. 2005.)

Aproximadamente 50,000 familias se han definido con base en la composición de dominios de $> 1 \times 10^6$ proteínas predichas en 120 genomas de los tres dominios de la vida (Orengo & Thornton, 2005, Lee *et al.*, 2005). En general, menos de 15 % de las combinaciones de dominios en un genoma aparecen en otro genoma o dominio de la vida (Fig. 2a). En este sentido, la mayoría de las proteínas son invenciones específicas (de las especies o de ciertos linajes biológicos). No obstante, ~ 70 % de las secuencias de dominios pueden agruparse en tan sólo 1,400 familias. Más aún, casi la mitad de las secuencias de dominios pertenecen a alguna de 213 familias comunes a los tres dominios de la vida (Fig. 2b) (Lee *et al.*, 2003; Lee *et al.*, 2005). De manera sorprendente, la distribución de las unidades estructurales puede colapsarse más si se consideran sólo el tipo y conectividad de las conformaciones secundarias.

La coincidencia estructural en ausencia virtual de identidad de secuencia o función constituye una ‘zona de incertidumbre’, donde no es posible discernir entre la ancestría común y la evolución convergente. No obstante, las configuraciones secundarias tienen pocas alternativas de conectividad y disposición espacial. Así que las proteínas pueden agruparse con base en prototipos topológicos, o grupos de plegamiento (Liu *et al.*, 2004; Thornton *et al.*, 1999; Aroul-Selvam *et al.*, 2004). En la actualidad se conocen poco más de 850 grupos de plegamiento; se estima que el número total de topologías en la naturaleza está en el intervalo de 10^3 a 10^4 (Orengo *et al.*, 2005; Liu *et al.*, 2004; Chothia, 1992). Menos de 10 plegamientos reúnen a ~ 40 % de los tipos de estructura de la base PDB (Fig. 3). Entre estos plegamientos recurrentes están los de barril (β/α)₈, asa-P NTPasa, Rossmann y HUP (Orengo & Thornton, 2005; Copley & Bork, 2000; Winstanley *et al.*, 2005). Dichas topologías son numerosas en las superfamilias CATH y son adoptadas por secuencias muy divergentes (Thornton *et al.*, 1999). Por ejemplo, el plegamiento de barril (β/α)₈ es compartido por más de 25 superfamilias funcionalmente diversas, que incluyen al menos 76 familias de secuencias homólogas (Nagano *et al.*, 2002).

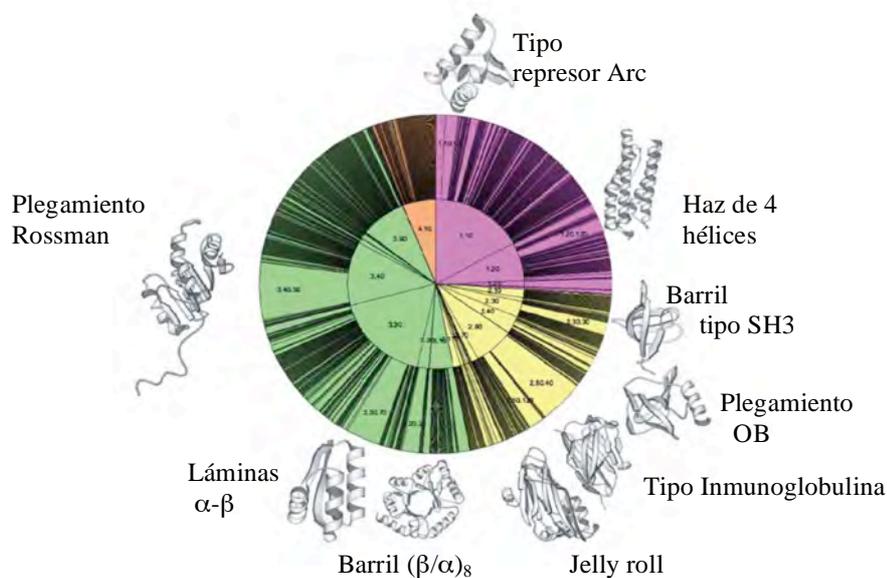


Figura 3. Distribución de las estructuras de dominios de la base PDB entre los niveles de clasificación jerárquica CATH. En colores se muestran las arquitecturas siguientes: principalmente α (rosa), principalmente β (amarillo) y α - β (verde). El círculo interno corresponde a diferentes arquitecturas, mientras que el externo está ocupado por los grupos de plegamiento. Cada grupo está dividido de acuerdo al número y poblaciones de diferentes superfamilias de homólogos que lo adoptan. (Tomado de Orengo & Thornton, 2005.)

Una fracción menor de los dominios proteínicos aparecen decenas y hasta centenas de veces en cada genoma o dominio de la vida. Estos dominios recurrentes constituyen casi la mitad de las anotaciones de estructuras en los genomas procariontes (Orengo & Thornton, 2005). Y muchos de estos dominios adoptan alguno de diez

plegamientos altamente representados en las bases de información estructural. En general, estas aproximaciones sugieren que una parte importante de la diversidad de las proteínas es consecuencia de la reutilización de un número reducido de estructuras ancestrales (Winstanley *et al.*, 2005). La cuestión que sigue es si puede encontrarse una tendencia similar en la divergencia de las actividades enzimáticas.

Tendencias en la evolución de las actividades enzimáticas

Análisis estructurales han revelado los mecanismos subyacentes a la actividad de muchas enzimas (Ghanem & Raushel, 2007). Un rasgo invariable es la presencia de residuos capaces de polarizar al sustrato y de estabilizar estados de transición. Estos residuos catalíticos muestran cadenas laterales polares o cargadas, cuya orientación espacial facilita la alteración química de átomos y enlaces del sustrato. De manera importante, los residuos catalíticos pueden ampliar sus capacidades al interactuar con residuos aledaños (Gutteridge & Thornton, 2005; Bartlett *et al.*, 2003).

La información de secuencia indica que los residuos catalíticos suelen conservarse entre las enzimas homólogas. A partir de esta observación pueden buscarse tendencias evolutivas. Se ha observado que la química de reacción está compartida en 80 % de las superfamilias de enzimas de la base SCOP (Schmidt *et al.*, 2003b). Casi la mitad de estas superfamilias albergan enzimas que catalizan variaciones de una reacción con diferentes especificidades de sustrato (Fig 4). Esta diversificación de especificidad es mostrada por enzimas cuyas secuencias comparten una identidad promedio de 38 % (Schmidt *et al.*, 2003b). A este nivel de divergencia pueden observarse residuos catalíticos conservados en términos de identidad y posición de secuencia. Por otro lado, las enzimas homólogas más divergentes llevan a cabo actividades diferentes en términos de la química de reacción (Glasner *et al.*, 2006; Glasner *et al.*, 2007).

Alrededor de 20 % de las superfamilias de la base SCOP agrupan enzimas que catalizan transformaciones químicamente distintas (Fig. 4; Schmidt *et al.*, 2003b). No obstante, se espera que al menos una tercera parte de estas agrupaciones albergue rasgos catalíticos compartidos. Todd *et al.* (2001) encontraron evidencia de una conservación mecánica parcial en 22 (de 27) superfamilias catalíticamente diversas de la base CATH (Todd *et al.*, 2001). Se han caracterizado algunas de estas superfamilias, y se sabe que cada una alberga más de 11 reacciones distintas (Gerlt & Babbitt, 2001). Estas reacciones son catalizadas por enzimas cuya identidad de secuencia está entre 25 % y 13 % (Schmidt *et al.*, 2003b). A este nivel de divergencia hay residuos catalíticos conservados, aunque con identidades y posiciones variables (Schmidt *et al.*, 2003b; Gerlt & Babbitt, 2001). En realidad, la idea de la consistencia catalítica es persuasiva cuando se observa cómo se conducen las reacciones.

Pese a su divergencia, los homólogos catalíticamente diversos suelen compartir alguna estrategia catalítica, como una reacción parcial, un intermediario o estado de transición (Glasner *et al.*, 2006; Glasner *et*

al., 2007; Gerlt & Babbitt, 2001). Esto puede verse en las reacciones de la superfamilia enolasas, que incluyen racemizaciones, cicloisomerizaciones y eliminaciones β . Estas reacciones necesitan la abstracción de un protón α por un catalista ácido/base, al igual que la estabilización de un intermediario enólico por un ión metálico divalente. Los residuos que coordinan al ión metálico están conservados, pero no así la identidad y posición del catalista ácido/base (Glasner *et al.*, 2007).

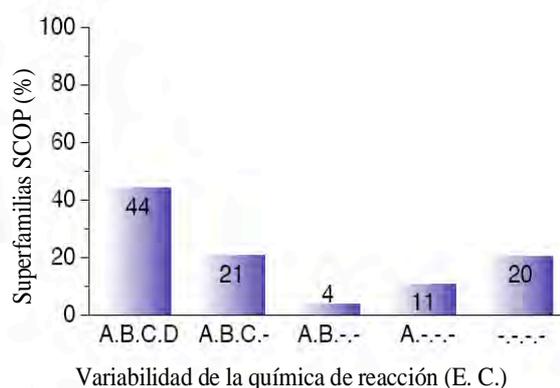


Figura 4. Variabilidad de la química de reacción entre las superfamilias de proteínas de la base SCOP. Alrededor de 40 % de las superfamilias alberga homólogos que catalizan reacciones idénticas y consecuentemente comparten dígitos E. C. idénticos. Muchos de estos homólogos posiblemente son ortólogos. Una proporción de 36 % corresponde a superfamilias cuyos miembros catalizan reacciones que coinciden en la química de reacción, pero que varían en el uso de sustratos. La proporción restante corresponde a superfamilias en las que puede apreciarse cierta diversificación en la química de reacción. (Editada de Schmidt *et al.*, 2003.)

El ejemplo anterior muestra factores mecánicos que compiten en importancia con los residuos catalíticos. Tales factores constituyen ambientes químicos locales y pueden tener *status* de firmas. Por ejemplo, ciertos grupos amida y carbonilo de la cadena polipeptídica forman espacios oxianiónicos, que estabilizan intermediarios de reacción en superfamilias como las de serín hidrolasas y crotonasas (Glasner *et al.*, 2007; Perona & Craik, 1995). Asimismo, iones metálicos y cofactores son parte de la definición operativa de las enolasas y las enzimas dependientes de piridoxal-5'-fosfato (PLP), respetivamente (Glasner *et al.*, 2007; Eliot & Kirsch, 2004; Mehta & Christen, 2000). Por otro lado, efectos electrostáticos dispersos en el sitio activo generan dipolos fijos, que estabilizan configuraciones del sustrato semejantes al estado de transición de la reacción (Warshel, 2006). A esta preorganización electrostática se atribuye el poder catalítico de las enzimas (Benkovic & Hammes-Schiffer, 2003; Smith *et al.*, 2008).

La complejidad del sitio activo impide el mapeo cabal de sus elementos constituyentes, aunque éstos dan base a una afinidad por especies químicas del sustrato diferentes al estado basal. Esta afinidad diferencial constriñe la divergencia catalítica, pues aspectos del mecanismo de catálisis deben conservarse mientras se posicionan diferencialmente residuos catalíticos secundarios. Para redondear esta idea,

considérese que pocas superfamilias representan casos de verdadera versatilidad catalítica (Glasner *et al.*, 2007). No sorprende que tales superfamilias pertenezcan a grupos de plegamiento modulares, como los barriles (β/α)₈ (Glasner *et al.*, 2006). Por otro lado, las enzimas muestran gran especificidad al discriminar entre sustratos potenciales. Esta afinidad absoluta se alcanza mediante interacciones no covalentes entre las paredes del sitio activo y los sustratos cognados (Hedstrom, 2002; Perona & Craik, 1995).

El establecimiento de nuevas interacciones parece una solución evolutiva más accesible que la invención de un sitio activo. Frecuentemente se encuentran enzimas homólogas con diferentes especificidades de sustrato. En parte, esto se debe a nuevas interacciones establecidas en asas que presentan libertad conformacional (Glasner *et al.*, 2006). Lo anterior no supone restricciones evolutivas infranqueables, pero las fuentes de las afinidades diferencial y absoluta están asociadas. Es probable que esta situación imponga constreñimientos al uso de nuevos sustratos. En coincidencia con lo anterior, más de la mitad de las 87 superfamilias CATH presentes en *E. coli* muestran sustratos principales con conservación estructural (Nobeli *et al.*, 2005).

Evidentemente, la evolución representa una búsqueda de combinaciones de residuos que satisfacen factores estructurales y funcionales. La insatisfacción de cualquiera de estos factores intrínsecos puede cancelar la exploración de nuevas alternativas evolutivas. Aparentemente, un constreñimiento importante emerge de la incapacidad de las proteínas para soportar mutaciones sin ver alterada su estabilidad estructural.

Constreñimientos evolutivos intrínsecos

Pese a la infinidad de conformaciones alternativas, las proteínas en disolución pueden plegarse en milisegundos a segundos (Creighton, 1990). Dicha transición puede satisfacer condiciones cinéticas y termodinámicas sin la intervención de agentes externos. Esto sugiere la exploración del espacio conformacional por rutas definidas (Levinthal, 1968; Pande *et al.*, 1998). Muchos modelos intentan explicar estas rutas a través de reacciones, intermediarios y estados de transición (Dill & Chan., 1997; Fersht, 1997; Dill *et al.*, 1995). Por otro lado, el proceso *in toto* puede representarse termodinámicamente como la discriminación entre ensambles probabilísticos (Onuchic, 1997; Onuchic & Wolynes, 2004). Para que un ensamble sea poblado, la pérdida de entropía implícita en el orden estructural debe balancearse con una disminución de energía libre (Honig, 1999; Dill & Chan, 1997; Dobson & Karplus, 1999). En otras palabras, se favorece la conformación nativa porque corresponde a un mínimo local de energía. Esta situación no significa que el estado plegado sea estático. Al contrario, las proteínas exhiben un comportamiento dinámico.

La dinámica interna de las proteínas abarca distintas magnitudes y escalas temporales. En una escala mínima se encuentran las energías cinética, rotacional y vibratoria de átomos individuales (Cooper, 1999). Después están desplazamientos de conformaciones secundarias y dominios. En termodinámica, los

desplazamientos estructurales se reducen a una distribución estadística entre dos estados (Anfinsen, 1973; Cooper, 1999). En coincidencia con lo anterior, hay evidencia experimental de que las proteínas alternan entre la conformación nativa y un estado desordenado (Zwanzig, 1997). La probabilidad de encontrar a la proteína en estado plegado es finita. Esta probabilidad suele representarse con un parámetro de estabilidad termodinámica, o de cambio de energía libre (ΔG) (Cooper, 1999). La intuición dicta que la conformación nativa debe predominar si es que han de satisfacer demandas biológicas. Los valores de ΔG de muchas proteínas están en el intervalo de -5 a -10 kcal mol⁻¹. Para poner estos valores en contexto, la energía de un puente de hidrógeno es de 2 a 10 kcal mol⁻¹ (DePristo *et al.*, 2005). De aquí se infiere que la conformación nativa sólo es marginalmente más estable que el estado desordenado. Esto sugiere que muchas proteínas están desplegándose en cualquier momento dado.

La conformación nativa es un estado metaestable determinado de manera biológica (Anfinsen, 1973). Evidencia de lo anterior es que las proteínas de especies termófilas muestran más estabilidad que sus ortólogos mesófilos (Sterner & Liebl, 2001). Sin duda, las diferencias de estabilidad son consecuencia de un proceso de adaptación. El margen de estabilidad de las proteínas tiene implicaciones para la regulación y la prevención de fenómenos de citotoxicidad (De Pristo *et al.*, 2005). Una proteína inestable tiene una concentración efectiva disminuida, mientras que un exceso de estabilidad implica una baja proteólisis y dificultades para regular el nivel de proteína. Asimismo, la rigidez estructural es desventajosa para las enzimas, pues la catálisis requiere cierta flexibilidad mecánica (Fenton, 2008). Curiosamente, las características intramoleculares favorables a la estructura tienen tendencias opuestas a las que confieren actividad.

La contribución de las interacciones intramoleculares a la estabilidad termodinámica es un problema controvertido (Cooper, 1999). Sin embargo, hay acuerdo en que el plegamiento se conduce por el confinamiento de regiones de secuencia hidrofóbicas. Asimismo, se reconoce que al interior de las estructuras quedan satisfechas casi todas las interacciones posibles. Puentes de hidrógeno, fuerzas de van der Waals y puentes salinos estabilizan conformaciones secundarias, al tiempo que contribuyen con el empacamiento estructural. En contraste, la catálisis ocurre en una cavidad con superficies hidrofóbicas expuestas al disolvente, grupos con carga secuestrados lejos del disolvente, cargas semejantes enfrentadas y puentes de hidrógeno insatisfechos (Creighton, 1990). Parece que esta organización significa un compromiso para la estabilidad termodinámica.

De acuerdo con la hipótesis del compromiso entre estabilidad y actividad, los residuos catalíticos y de unión de sustratos no están optimizados para hacer contribuciones estructurales (Shoichet *et al.*, 1995). En la β -lactamasa AmpC, la substitución de estos residuos por otros más afines a la estructura significan aumentos de estabilidad hasta de 4.7 kcal mol⁻¹ (Beadle & Shoichet, 2002). Esto muestra que la funcionalidad del sitio activo tiene un costo negativo en términos de estabilidad estructural. Pero un costo análogo se impone sobre la

catálisis, por las alteraciones estructurales sufridas ante fluctuaciones energéticas del ambiente. En el intervalo de temperatura de crecimiento de *E. coli*, las enzimas toleran variaciones de $\sim 1.0 \text{ kcal mol}^{-1}$ sin pérdidas de estabilidad (y actividad) significativas. Fuera de este intervalo se requiere la asistencia de chaperonas (De Pristo *et al.*, 2005).

El balance entre estabilidad y actividad es una ventana termodinámica estrecha, situada dentro de otra apenas más amplia y correspondiente a la estabilidad *per se*. En contra de la intuición, las enzimas homólogas divergentes sugieren que la topología y la función no obedecen tanto al detalle químico como a una correlación de geometría macromolecular. En realidad, la conformación nativa implica la minimización energética de todo un sistema. Es decir, la estructura debe entenderse como consecuencia de la totalidad de interacciones atómicas asociadas a cada secuencia (Anfinsen, 1973). Desde luego, el balance determinado de forma intrínseca para cada enzima debe representar constreñimientos a la divergencia evolutiva.

Dos observaciones dan cuenta de la susceptibilidad de la estabilidad termodinámica frente a las sustituciones de residuos. La primera es que la diferencia de estabilidad causada por las sustituciones ($\Delta\Delta G$) suele ser de 0.5 a 5 kcal mol^{-1} (DePristo *et al.*, 2005). En una aproximación *in silico* se estimaron los $\Delta\Delta G$ de todas las mutaciones posibles de 16 proteínas. Los resultados sugieren que $\sim 70\%$ de las sustituciones causan desestabilizaciones menores, mientras que $\sim 20\%$ resultan muy desestabilizadoras (Tokuriki *et al.*, 2007). La segunda observación es que las proteínas de especies extremófilas toleran más mutaciones que los ortólogos mesófilos. Los ortólogos corismato mutasa de *Methanococcus janaschii* y *E. coli* fueron sometidos a una mutagénesis extensiva. En el contexto de *E. coli*, la detección *in vivo* de variantes activas mostró que el ortólogo termoestable tolera 10 veces más mutaciones (Besenmatter *et al.*, 2007).

La susceptibilidad ante las sustituciones sigue una distribución acorde con la accesibilidad del disolvente. Se ha observado esto en la búsqueda de variantes enzimáticas con actividad suficiente para la supervivencia de *E. coli*: Guo *et al.* (2004) calcularon un índice de substitucionalidad con base en 920 mutaciones toleradas por la 3-metiladenina DNA glicosilasa. Según dicho índice, las superficies accesibles al disolvente son más tolerantes que los residuos enterrados. En una estimación *in silico*, el efecto desestabilizador promedio en posiciones superficiales es de $\sim 0.6 \text{ kcal mol}^{-1}$, mientras que en el núcleo hidrofóbico corresponde a $\sim 1.4 \text{ kcal mol}^{-1}$ (Tokuriki *et al.*, 2007). Obviamente, el peor escenario es la pérdida de la actividad enzimática. En estudios independientes se ha observado que la probabilidad de desactivación por la incidencia de una sola substitución es $\sim 33\%$ (Guo *et al.*, 2004; Loh *et al.*, 2007). Es decir, al menos un tercio de las sustituciones podrían afectar de forma severa la actividad, aunque no hayan incidido en el sitio activo (Camps *et al.*, 2007).

Por otra parte, las tasas de mutación varían de forma significativa entre los genes de las proteínas. La dinámica evolutiva de las proteínas también parece estar constreñida por el grado de integración funcional dentro de la célula (Pál *et al.*, 2006). Entre los factores de integración que resultan relevantes están la

dispensabilidad (Jordan *et al.*, 2002; Wall *et al.*, 2005), el nivel de expresión (Pal *et al.*, 2001; Pal *et al.*, 2003; Rocha & Danchin, 2004; Drummond *et al.*, 2005) y las interacciones proteína-proteína (Fraser, 2005; Fraser *et al.*, 2002). Las diferencias entre las tasas evolutivas también sugieren que mutaciones potencialmente dañinas logran fijarse como consecuencia del balance entre mutación, selección y adecuación. Si la mayoría de las sustituciones pueden implicar una penalización sobre la adecuación, quizá la divergencia a nivel de secuencia está determinada por una forma de selección que no produce innovaciones. Entonces, hay posibilidad de que la divergencia esté dominada por mutaciones compensadoras.

En un escenario de compensación, el efecto negativo de una mutación puede ser resarcido desde sitios que interaccionan epistáticamente. Kondrashov *et al.* (2002) encontraron evidencias de compensación entre los ortólogos de tres enzimas humanas, que en conjunto exhiben 10 sustituciones negativas. En los ortólogos no humanos, la presencia de estas sustituciones se correlaciona con la conservación de residuos cercanos estructuralmente. Los residuos que acompañan siempre a una mutación potencialmente negativa quizá son sustituciones compensadoras. Se ha estimado que por cada mutación negativa, hasta 12 mutaciones podrían compensar la pérdida de adecuación (Poon *et al.*, 2005).

Si sustituciones potencialmente desfavorables pueden enmascarar el efecto negativo de otras, entonces el efecto último de las mutaciones depende del contexto de secuencia y del balance preexistente entre los parámetros de estabilidad y actividad (DePristo *et al.*, 2005). En términos de la adecuación, muchas sustituciones pueden ser negativas o *cuasi* neutrales de manera condicional (Poon & Chao, 2005). El surgimiento de combinaciones de residuos satisfactorias para ambos parámetros parece poco probable. Asimismo, el surgimiento de una innovación evolutiva parece posible, pero puede tener un costo para la estabilidad termodinámica. Constreñimientos estructurales y funcionales limitan las opciones de variación de las proteínas. Las demandas de un factor pueden ser inconsistentes con las del otro, así que la divergencia de la catálisis debe ocurrir a través de combinaciones de mutaciones que mantienen (o restablecen) un equilibrio intrínseco. Por otro lado, para entender las consecuencias de los constreñimientos y la evolución en la divergencia de las actividades enzimáticas, es necesario entender a las funciones catalíticas en un contexto de colectividad. Esto es porque el surgimiento y retención evolutiva de las actividades novedosas sirve a la constitución del metabolismo.

La diversificación de las actividades enzimáticas en el contexto metabólico

La célula se sustenta como un todo, pero sus moléculas no son partes permanentes; continuamente son tomadas del exterior, transformadas químicamente y excretadas por el metabolismo. Este direccionamiento de materia y energía ocurre a través de múltiples reacciones químicas, la mayoría catalizadas por enzimas. Los productos de las reacciones pueden comportarse como intermediarios al ser sustratos en otras reacciones.

Asimismo, ciertas reacciones tienen un orden de sucesión, que desemboca en la formación de compuestos esenciales. El acoplamiento entre enzimas es mayor cuando se consideran todos los metabolitos participantes en las reacciones. Por lo anterior, las reacciones metabólicas pueden representarse como un conjunto de rutas ó como una red densamente entrelazada.

El metabolismo posiblemente evolucionó desde un conjunto pequeño de reacciones. A partir de tal punto, un aumento de complejidad ha sido más o menos proporcional a la aparición de nuevas enzimas. La aparición de nuevas enzimas depende de la duplicación y la divergencia génicas. Pero la retención de las nuevas enzimas ha estado circunscrita tanto por factores estructura/función como por la integración funcional. Diversos modelos intentan explicar la evolución de procesos metabólicos (Lazcano & Miller, 1999). De estos modelos, los de retroevolución y de evolución en mosaico incluyen dos factores que probablemente influyen en la construcción de rutas metabólicas (Figs. 5 y 6). Estos factores son la composición química del ambiente y el potencial catalítico de las enzimas.

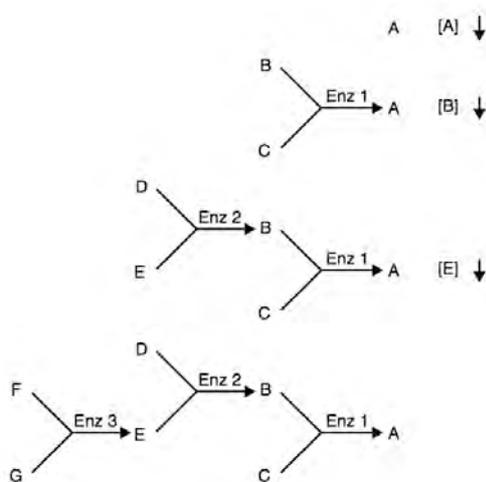


Figura 5. Surgimiento de las rutas metabólicas de acuerdo con la hipótesis de retroevolución. Un organismo agota las reservas ambientales de un compuesto esencial A. El surgimiento de una enzima (Enz1) capaz de usar los intermediarios B y C para restablecer la disponibilidad de A confiere una ventaja de supervivencia. Las concentraciones de B y E también disminuyen, pero el reclutamiento de las enzimas Enz2 y Enz3 puede compensar esta disminución, respectivamente. (Tomado de Rison & Thornton, 2002.)

El modelo de retroevolución propone que las rutas se ensamblaron en sentido opuesto al flujo metabólico hacia un producto final (Fig 5; Horowitz, 1945; Lazcano & Miller, 1999; Rison & Thornton, 2002). La hipótesis asume un ambiente con compuestos esenciales e intermediarios potenciales. Una célula consume un compuesto esencial hasta que se limita su crecimiento. La disponibilidad de este compuesto puede restablecerse por el reclutamiento de una enzima capaz de generarlo a partir de lo disponible en el ambiente. La presión de selección que conduce al reclutamiento se presenta de nuevo, con el agotamiento de los

intermediarios que sirven como precursores del compuesto esencial. La repetición de este proceso implica la derivación del compuesto esencial desde intermediarios cada vez más distantes. En la retroevolución, la limitación de compuestos condiciona la evolución de enzimas capaces de producirlos.

En el modelo de evolución en mosaico, las rutas se ensamblaron a partir de un número reducido de enzimas ineficientes (Fig. 6; Ycas, 1974; Jensen, 1976; Lazcano & Miller, 1999; Rison & Thornton, 2002; Jensen, 1996). Cada una de estas enzimas cataliza de manera inespecífica reacciones diversas con compuestos similares. Esta promiscuidad catalítica resulta en cadenas de reacciones, donde se sintetizan compuestos esenciales en bajas concentraciones. Pero la duplicación y la divergencia dan lugar a enzimas más eficientes, de manera que van configurándose rutas en repetidas etapas de selección y reclutamiento. Como resultado, enzimas parálogas que catalizan reacciones similares se reclutan en diferentes rutas; es decir, se forma un mosaico enzimático. En este modelo, la evolución es conducida por un potencial catalítico.

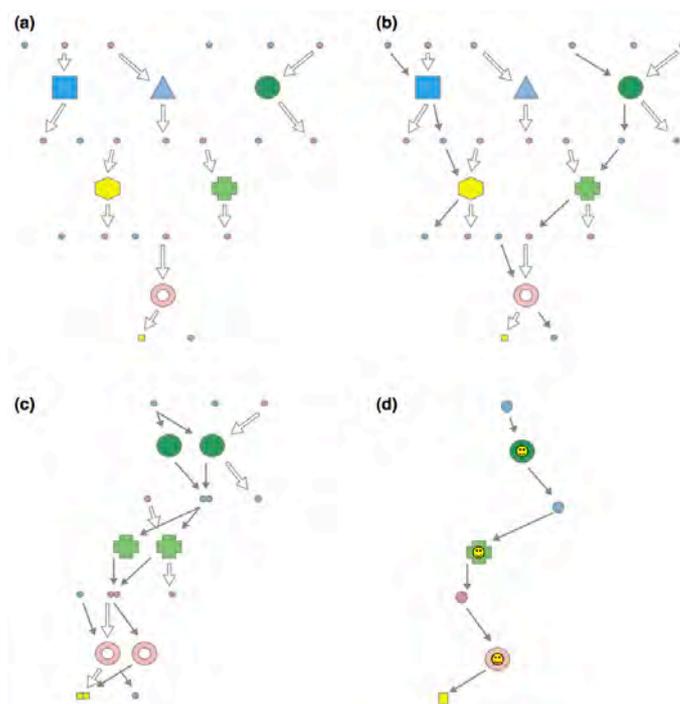


Figura 6. Modelo de evolución en mosaico para el surgimiento de las rutas metabólicas. (a) En un sistema de enzimas pueden verse favorecidas ciertas transformaciones químicas. (b) La existencia de promiscuidad catalítica permite la constitución de cadenas, como la que incluye a las enzimas representadas por un círculo (verde), una cruz (verde) y un toroide (rosa). (c y d) La duplicación génica y la divergencia pueden conducir a la especialización de las enzimas y concomitantemente al establecimiento de rutas similares a las observadas en la actualidad. (Tomado de Rison & Thornton, 2002.)

La integración de información bioquímica y genómica permite la aproximación al diseño y evolución del metabolismo. Se han construido diversas matrices de interacción entre enzimas parálogas y metabolitos. Tales

abstracciones difieren en el grado de acoplamiento permitido entre reacciones. Algunas abstracciones se basan en la idea de que el metabolismo es un conjunto de rutas interconectadas. Bajo esta idea, se ha encontrado que los dominios homólogos se distribuyen ampliamente entre las rutas metabólicas de *E. coli*. Entre homólogos, aspectos de la mecánica catalítica están conservados con más frecuencia que el reconocimiento de los sustratos (Tsoka & Ouzounis, 2001; Teichmann *et al.*, 2001; Rison *et al.*, 2002; Orengo & Thornton, 2005). Estas observaciones concuerdan con el modelo de evolución en mosaico, pero no permiten la inferencia de patrones evolutivos. Esto es porque la consideración de las rutas metabólicas como unidades funcionales independientes es arbitraria. Las interacciones observables entre las enzimas se multiplican cuando se consideran todos los metabolitos asociados a sus reacciones, *i. e.*, sustratos, productos, iones y cofactores. Desde esta perspectiva, cualquier compuesto es teóricamente sintetizable a partir de otro. La consideración del acoplamiento entre reacciones de distintas rutas es importante, porque puede reflejar constreñimientos evolutivos intrínsecos (de las enzimas) y sistémicos (del metabolismo). Dicha consideración requiere una abstracción en forma de red.

El metabolismo puede representarse como una red constituida por nodos y líneas de conexión. Tal abstracción suele organizarse alrededor de los metabolitos, de manera que éstos constituyen nodos conectados por las reacciones que los interconvierten (Barabási & Oltvai, 2004). De manera alternativa, las enzimas pueden ser nodos conectados por los metabolitos que se producen o se consumen. Las redes metabólicas sugieren patrones de interacción entre reacciones enzimáticas. La suma de estos patrones podría significar la compartimentación de las enzimas en bloques funcionales (Ravasz *et al.*, 2002). Estos bloques podrían dar lugar a propiedades de autorregulación y tolerancia a fallas en escala global (Albert *et al.*, 2000; Jeong *et al.*, 2000). Dicha organización tendría origen en la duplicación y el acoplamiento preferencial entre enzimas parálogas.

Se ha encontrado una topología libre de escala en 43 redes metabólicas de los tres dominios de la vida. Tal topología implica que la distribución de conexiones por nodo es heterogénea y semejante a una ley de potencias (Jeong *et al.*, 2000; Wagner & Fell, 2001): Muchos nodos están conectados de forma escasa, mientras que unos cuantos establecen múltiples conexiones y son centrales para la cohesión de la red. Lo anterior supone que la mayoría de los metabolitos participan en unas cuantas reacciones y que sólo algunos tienen una participación amplia, *e. g.*, piruvato y CoA. Las redes metabólicas también describen un efecto de mundo pequeño: La separación entre dos nodos cualesquiera suele consistir en pocos pasos. Quizá esto permite que las variaciones de concentración de los metabolitos afecten con rapidez a todo el metabolismo (Wagner & Fell, 2001).

Cuando se integran criterios de homología se obtienen aproximaciones al modo en que la duplicación y la divergencia han participado en la evolución metabólica. El surgimiento de las redes con topología libre de escala es explicable por un proceso de acoplamiento preferencial. Las redes libres de escala crecen por

adición de nodos, que tienden a unirse a otros preexistentes y muy conectados (Barabási & Oltvai, 2004; Barabási & Albert, 1999). Si esto ha ocurrido en la evolución metabólica, entonces las enzimas ortólogas más ancestrales deberían exhibir un número mayor de conexiones. En *E. coli*, los ortólogos exclusivamente procariontes tienen menos conexiones que los ortólogos presentes en los tres dominios de la vida. El anclaje preferencial podría estar asociado a la conservación del uso de metabolitos por las enzimas parálogas (Light *et al.*, 2005). Un nodo gana conexiones si se duplica alguno de los vecinos con los que interacciona. En este sentido, una enzima muy conectada tiene mayor probabilidad de ganar las conexiones (Barabási & Oltvai, 2004).

¿Puede integrarse el concepto de reclutamiento en una red metabólica? La respuesta es afirmativa. Pero si la reticularidad es un aspecto *bona fide* de la evolución, entonces conviene tomar en cuenta el concepto de distancias metabólicas. Es decir, deben considerarse los pasos entre dos enzimas parálogas, siendo cada paso la transformación enzimática de uno o más sustratos (Rison *et al.*, 2002). Así, el reclutamiento tiene dos formas, local y a larga distancia, dependiendo de si la separación entre dos homólogos es menor ó mayor que 3 pasos, respectivamente. En esta definición no importan las rutas a las que son asignadas las enzimas. El reclutamiento a larga distancia es tan frecuente como el local en 12 redes metabólicas de los tres dominios de la vida (Alves *et al.*, 2002). No obstante, la observación trascendente de éste y otros análisis concierne al reclutamiento local. El porcentaje de pares de enzimas homólogas separadas por 1 y 2 pasos es significativamente mayor que el observado cuando la red se reconecta al azar (Rison *et al.*, 2002; Alves *et al.*, 2002; Light & Kraulis, 2004; Díaz-Mejía *et al.*, 2007). En *E. coli*, las enzimas homólogas separadas por 1 paso representan 3.8 % de los pares a esta distancia, o 58 % del total de pares homólogos en la red. Esto implica una sobre-representación de 3 veces respecto a lo esperado al azar (Light & Kraulis, 2004). La correlación entre distancia metabólica y homología se extiende al tipo de reacción catalizada. También en *E. coli*, la mayoría de los homólogos contiguos comparten los primeros dos dígitos E.C. (Light & Kraulis, 2004; Díaz-Mejía *et al.*, 2007).

Consideraciones genéticas, estructurales y de mecánica catalítica indican que la evolución de nuevas actividades enzimáticas está condicionada por estreñimientos. Pero el metabolismo es un atributo contextual medible en miles de interacciones enzimáticas. Tales interacciones suponen el reclutamiento de enzimas en procesos de direccionamiento. Y las abstracciones del metabolismo sugieren que el reclutamiento incluye de forma principal a enzimas parálogas. Entonces, las coincidencias catalíticas y en el uso de sustratos de las enzimas parálogas podrían significar la organización del metabolismo en bloques funcionales. Propiedades emergentes como ésta debilitan al argumento de los estreñimientos, pues algunos de éstos también son identificables con el potencial evolutivo de las enzimas. Esto lleva a pensar que alguna propiedad intrínseca de las enzimas podría facilitar el surgimiento de nuevas actividades y el ensamblaje de rutas metabólicas. Recientemente se ha sugerido que parte del potencial evolutivo está contenido en la

promiscuidad catalítica de los sitios activos. Esta sugerencia contraviene la idea de que las enzimas han evolucionado para catalizar reacciones específicas. Sin embargo, hay evidencia persuasiva del papel evolutivo de la promiscuidad.

Promiscuidad catalítica

Una noción tradicional en bioquímica es que las actividades enzimáticas muestran una alta especificidad de sustrato y de reacción. Según esta misma noción, un comportamiento inespecífico podría ser un artefacto experimental o un fenómeno fortuito. Las enzimas exhiben actividades adicionales a su actividad principal en condiciones *ex vivo*. Por ejemplo, la lipasa B de *Candida antarctica* es capaz de catalizar una adición aldólica con propanal o hexanal (Bornscheuer & Kaslauskas, 2004). Esta actividad consume sustratos no fisiológicos, ocurre en disolvente orgánico y es 10^5 veces menos eficiente que la hidrólisis de triglicéridos (la actividad principal). Por otro lado, numerosas observaciones indican que algunas enzimas también pueden catalizar más de una reacción en condiciones fisiológicas. La capacidad de un sitio activo para acelerar reacciones adicionales a la principal es denominada promiscuidad catalítica.

Las determinantes del poder catalítico como base de la promiscuidad

En el ambiente dipolar de los sitios activos parece albergar al poder catalítico. Se dice esto porque dicho ambiente parece orientarse parcialmente hacia la estabilización de configuraciones del sustrato semejantes al estado de transición (Benkovic & Hammes-Schiffer, 2003; Warshel *et al.*, 2006). Esta orientación es parcial porque se espera que los sustratos contribuyan a la distribución de los dipolos. Lo anterior indica que hay grados de afinidad. En coincidencia con lo anterior, las enzimas muestran una afinidad absoluta al discriminar moléculas incapaces de satisfacer interacciones no covalentes que emergen desde las paredes del sitio activo. Ésta es la base de lo que se denomina especificidad de sustrato.

En el escenario anterior, la especificidad y eficiencia catalíticas necesitan grados de correspondencia entre el sitio activo y los sustratos. No obstante, una correspondencia íntima no es absolutamente necesaria para la catálisis. Esto es porque el poder catalítico está almacenado en forma de energía de plegamiento, no de interacciones de especificidad. En un sentido amplio, la promiscuidad catalítica es inevitable por la reactividad inherente del sitio activo. Esto no significa una oposición a las restricciones químicas impuestas por la arquitectura del sitio activo. Por el contrario, las reacciones promiscuas suelen mostrar características reconocibles en la actividad principal.

Las analogías entre reacciones alternativas de un sitio activo pueden ser obvias, como en el uso de gluconato y galactonato por la gluconato deshidratasa de *Sulfolobus solfataricus* (Lamble *et al.*, 2004). Esta enzima cataliza versiones de una sola reacción con sustratos semejantes. Las semejanzas tienen que buscarse

en los mecanismos catalíticos. Considérese a las serín hidrolasas. La actividad nativa de la quimiotripsina es la hidrólisis de un enlace C-N. Esta reacción incluye un estado de transición tetrahédrico y el ataque nucleofílico por el residuo Ser de una tríada catalítica Ser/His/Asp. Asimismo, un sitio oxianiónico polariza al enlace carbonilo del sustrato en el estado de transición. La quimiotripsina también es capaz de romper amidas, ésteres, ésteres-tiol y cloruros. La estructura de los sustratos y enlaces afectados difieren, pero todas las reacciones son procesos hidrolíticos con intermediarios y mecanismos comunes.

Las eficiencias de las actividades promiscuas suelen ser órdenes de magnitud menores que la eficiencia de la actividad principal. Por ejemplo, la β -glucuronidasa de *E. coli* hidroliza de forma promiscua β -galáctosidos, pero con una eficiencia 10^6 veces menor a la del rompimiento de β -glucurónidos (Bornscheuer & Kaslauskas, 2004). ¿Acaso hay constreñimientos en la evolución de la multifuncionalidad? Aunque resulta posible la promiscuidad, quizá la eficiencia de sus reacciones no logra satisfacer a la célula. Esto podría deberse a una imposibilidad física para mejorar y mantener más de una reacción. Por otro lado, la competencia de los sustratos alternativos por participar en el recambio catalítico quizá dificulta la regulación de las actividades. Es decir, la multifuncionalidad podría verse desfavorecida por razones estructurales y de homeorhesis. No obstante, las analogías entre las reacciones de un sitio activo son similares a las que dan sustento a hipótesis de ancestría común entre las enzimas. Esto sugiere que la promiscuidad refleja aspectos de la divergencia enzimática.

Distribución de la promiscuidad catalítica

La tabla 1 muestra las coincidencias entre las reacciones promiscuas y nativas de enzimas homólogas agrupadas en superfamilias. En ciertos casos, una actividad promiscua es compartida por pares de enzimas homólogas, como la hidratación de 2-oxo-3-pentinoato por la malonato semialdehído descarboxilasa y la trans-3-cloroacrilato deshalogenasa (Khersonky *et al.*, 2006). También hay casos de reciprocidad entre las actividades principales y promiscuas. Por ejemplo, las enzimas 3' ceto L-gulonato 6-fosfato descarboxilasa y la D-arabino-hex-3-ulosa 6 fosfato sintasa catalizan de forma promiscua la reacción principal de la otra (Yew *et al.*, 2005). La promiscuidad catalítica de algunas enzimas lleva a la interrogante de si la preexistencia de actividades secundarias no facilita la evolución de nuevas actividades principales. Esta hipótesis se apoya en la observación de enzimas de aparición reciente en la naturaleza y por experimentos de evolución *in vitro*.

Promiscuidad de enzimas recientemente evolucionadas en la naturaleza

Variantes naturales que degradan compuestos antropogénicos (Copley, 2009) proveen evidencia de la participación de la promiscuidad en la evolución enzimática. Un ejemplo favorito es la enzima fosfotriesterasa

(FTE), que transforma compuestos liberados en la naturaleza durante el siglo XX. En apariencia, la selección natural ha permitido el surgimiento de una actividad biológicamente ventajosa en un lapso brevísimo. La evolución de dicha actividad podría haberse facilitado por su preexistencia en el repertorio de actividades promiscuas de los ancestros de la FTE.

Familia/ Superfamilia	Enzimas	Enzima: Actividad nativa*	Enzima: Actividades promiscuas [†]	Refs.
Enolesas (subgrupo de la enzima lactonizante de mucronato)	<i>o</i> -Succinilbenzoato sintasa	OSBS: deshidratación de SHCHC (2-succinil-6R-hidroxil-2,4-ciclohexadieno-1R-carboxilato (2.5×10^5))	OSBS: <i>N</i> -acilaminocácido racemasa con isómeros de la <i>N</i> -acetil metionina ($4.9 - 5.9 \times 10^5$)	Palmer <i>et al.</i> , 1999
Tautomerazas	Malonato semialdehído descarboxilasa	MSAD: descarboxilación de malonato semialdehído (2.2×10^7)	MSAD: hidratación de 2-oxo-3-pentincato (6×10^5)	Reukens <i>et al.</i> , 2004
	4-Oxalocrotonato tautomeraza	4OI: isomerización de 2-oxo-4E-hexadienocato a 2-oxo-3E-hexadienocato a través de 2-hidroxi-2,4E-hexadienocato (2.0×10^7 , segunda reacción)	4OI: actividad CasD, hidratación de 3E-clorocacrilato (2.6×10^5)	
	YwhB tautomeraza	YwhB, análoga a 4OI (2.8×10^6)	YwhB: actividad CasD, hidratación de 3E-clorocacrilato (4.4×10^5)	
Orotidina 5' monofosfato desaminasa	<i>lys</i> -3-clorocacrilato deshalogenasa	CaaD: deshalogenación hidrolítica (hidatasa) de 3E-halocacrilatos (1.2×10^7 , 3E-clorocacrilato)	CaaD: hidratación de 2-oxo-pentincato (6.4×10^5)	
	3' oxo-L-gliconato 6-fosfato descarboxilasa	KGFD5: Descarboxilación de 3' oxo-L-gliconato 6-fosfato (7.7×10^6)	HPS: KGFD5 (2.3×10^5)	Yew <i>et al.</i> , 2005
	D-arabino-hex-3-ulosa 6-fosfato sintasa	HPS: condensación abiótica de D-ribulosa 5-fosfato y formaldehído (1.6×10^6)	KGFD5: HPS (8.2×10^5)	
ROK (repressor open reading frame, kinase)	NanK	N-acetil-D-mansarina cirasa (2.2×10^5)	Glucosa cirasa (5.1×10^5)	Miller & Rames, 2004; 2005
	YajP	Fructosa cirasa (1.1×10^6)	(2×10^5)	
	YcEX	Desconocida	(2.4×10^5)	
	AlaK	Alcota cirasa (6.5×10^4)	(15)	

Tabla 1. Algunas actividades promiscuas compartidas por enzimas parálogas. [†] k_{cat}/K_M ($M^{-1} S^{-1}$). (Con base en Khersonsky *et al.*, 2006.)

La FTE hidroliza organofosfatos sintéticos introducidos en la naturaleza desde 1950 (Raushel & Holden, 2000; Scalan & Reid, 1995). Su mejor sustrato es el insecticida paraoxon, con el que exhibe una eficiencia próxima al límite controlado de difusión ($k_{cat} = 104 s^{-1}$, $k_{cat}/K_m = 4 \times 10^7 M^{-1} s^{-1}$; Omburo *et al.*, 1992). En apariencia, la FTE constituye una respuesta biológica a la limitación de fosfatos en el suelo. De acuerdo con esto, la enzima existe en especies diversas, que incluyen a *Pseudomonas diminuta*, *Flavobacterium* sp. y *Agrobacterium radiobacter*. Curiosamente, la FTE hidroliza algunas lactonas en forma significativa (Roodveldt & Tawfik, 2005). Dicha reacción puede ser vestigio de un ancestro evolutivo, perteneciente a un grupo de lactonazas microbianas caracterizadas recientemente. El papel fisiológico de estas lactonazas podría

ser el procesamiento de señales extracelulares de *quorum*, pues hidrolizan con eficiencia *N*-acil homoserín lactonas (k_{cat}/K_m de 6×10^4 a $1.7 \times 10^6 \text{ M}^{-1}\text{s}^{-1}$). De manera sorprendente, estas enzimas muestran actividad hidrolítica con paraoxon, aunque en forma vestigial (k_{cat}/K_m de 0.5 a $4000 \text{ M}^{-1}\text{s}^{-1}$) (Afriat *et al.*, 2006).

Evolución *in vitro*

Matsumura y Ellington (2001) estudiaron la especificidad de sustrato de la β -glucuronidasa de *E. coli*. Variantes con mayor actividad β -galactosidasa se detectaron con una cepa *lacZ* y el cromóforo X-Gal. Las mejores variantes de una librería de PCR mutagénica se sometieron a *shuffling*. Al final se obtuvo una mutante triple con actividad galactosidasa 21 veces mayor, y una actividad principal 10 veces disminuida (k_{cat}/K_m de $500 \text{ M}^{-1}\text{min}^{-1}$ y $7 \times 10^6 \text{ M}^{-1}\text{min}^{-1}$, respectivamente). De manera inesperada, la variante mostró una mayor eficiencia con glucósidos (sustratos ausentes en el medio de selección). Los autores propusieron que la evolución de la especificidad de sustrato ocurre a través de intermediarios evolutivos ‘generalistas’.

En el ejemplo anterior, la evolución *in vitro* requirió la preexistencia de una actividad promiscua. Además, se experimentó con una sola transformación química. Por razones de arquitectura y mecánica catalítica, la promiscuidad podría ser una capacidad latente. Esto fue sugerido por la manipulación experimental de miembros de la superfamilia enolasas: Las enzimas lactonizante de muconato II (ELM II) y *o*-succinilbenzoato sintasa (OSBS) catalizan una cicloisomerización y una eliminación β , respectivamente. La ELM II no muestra reacción OSBS *in vitro*. No obstante, mediante *shuffling* y complementación *in vivo* se obtuvo una variante, cuya $k_{\text{cat}}/k_{\text{uncat}}$ para la producción de *o*-succinilbenzoato es sólo 10 veces menor a la de una OSBS *bona fide* (Schmidt *et al.*, 2003). Dicha variante exhibe una sola sustitución y una actividad principal comparable con la promiscua OSBS (k_{cat}/K_m de $1.3 \times 10^3 \text{ M}^{-1}\text{s}^{-1}$ y $1.9 \times 10^3 \text{ M}^{-1}\text{s}^{-1}$).

Aharoni *et al.* (2005) sometieron a las enzimas anhidrasa carbónica, FTE y paraoxonasa sérica a evolución *in vitro* para incrementar seis actividades promiscuas. Sus observaciones coinciden con las de numerosos estudios de evolución *in vitro*: (i) Entre 1 y 4 sustituciones suelen ser suficientes para alterar la preferencia de sustrato y aún la de reacción; (ii) *grosso modo*, las actividades promiscuas se aumentan entre 10 y 1000 veces; (iii) las actividades seleccionadas están acompañadas por otras para las que no se implementaron criterios de selección, y; (iiii) las variantes enzimáticas retienen la actividad principal con alteraciones menores (Khersonsky *et al.*, 2006).

Desde una perspectiva evolutiva, las actividades promiscuas podrían existir libres del efecto de la selección en tanto no comprometan a la actividad principal. De manera interesante, la capacidad de catalizar el recambio de sustratos diferentes en reacciones análogas podría dar poder a la evolución para generar funciones novedosas y biológicamente ventajosas. El propósito de estudiar el papel evolutivo de la promiscuidad catalítica hace deseable contar con enzimas bien caracterizadas. Las enzimas dependientes del

cofactor piridoxal-5'-fosfato (PLP) son un grupo de catalistas versátiles y atractivos por su participación en procesos celulares diversos. En lo particular, estas enzimas catalizan transformaciones útiles en el metabolismo de los aminoácidos, por lo que pueden considerarse como parte fundamental del metabolismo. Asimismo, su papel se extiende a procesos como la biosíntesis de cofactores.

Enzimas dependientes del cofactor PLP

Las enzimas dependientes del cofactor PLP constituyen un grupo ubicuo y funcionalmente diverso. Las secuencias de estas enzimas representan hasta 1.5 % del contenido génico total en genomas procariontes. Las secuencias son más numerosas en eucariontes, aunque representan una fracción genómica mucho menor (Percudani & Peracchi, 2003). Por otro lado, las reacciones catalizadas por las enzimas dependientes de PLP pertenecen a 5 de las 6 clases definidas por el Comité de Nomenclatura Enzimática. Más de 140 actividades distintas, o ~ 4 % del total clasificado, son llevadas a cabo por este tipo de enzimas. La actividad aminotransferasa (E.C. 2.6.1.) reúne a la mayoría de las reacciones dependientes de PLP.

La mayoría de las enzimas dependientes de PLP intervienen en la síntesis, interconversión y degradación de aminoácidos. Dado lo anterior, resultan esenciales para la conexión de los metabolismos de nitrógeno y de carbono, para el reaprovisionamiento de unidades de carbono y para la formación de aminas biogénicas. Las enzimas que transforman aminoácidos catalizan reacciones de transaminación, descarboxilación, racemización y sustitución de grupos funcionales unidos a los carbonos β y γ de los aminoácidos (Mehta & Christen, 2000). Un conjunto de características compartidas hacen posible todas estas reacciones (Jansonius, 1998).

En la figura 7 se muestran los intermediarios críticos en las reacciones catalizadas por las enzimas dependientes de PLP. En ausencia de sustratos se mantiene una unión covalente entre el C4' del PLP y el grupo ϵ -amino de un residuo Lys enzimático. En el inicio de la catálisis, el grupo NH_2 del sustrato aminoácido entrante ataca nucleofílicamente al C4', y substituye al NH_2 del residuo Lys activo. El resultado de esta transaminación es un intermediario PLP-sustrato aminoácido, *i. e.*, la aldimina externa. En el siguiente paso, uno de los enlaces hacia el C α de la aldimina externa se rompe, con lo que se forma un intermediario quinoide. Este proceso es facilitado por el PLP. Éste actúa como hundidero de electrones derivados de la ruptura de enlace, y estabiliza la carga negativa de intermediarios carbaniónicos. La interacción entre el PLP y la arquitectura del sitio activo permite la discriminación entre rutas catalíticas. En otras palabras, la especificidad de reacción depende del control enzimático hacia alguno de los intermediarios tipo quinoide.

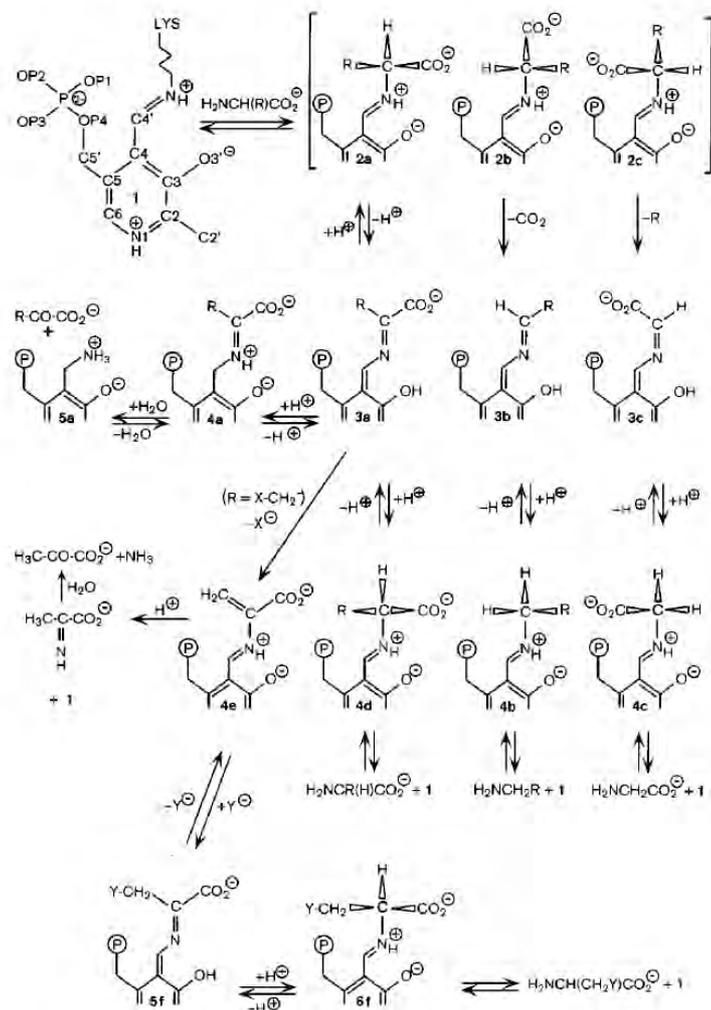


Figura 7. Intermediarios de las reacciones catalizadas por las enzimas dependientes de PLP. La estructura **1** muestra la aldimina interna protonada entre el PLP y el residuo Lys del sitio catalítico. Las aldiminas externas **2a-c** se forman por transiminación con el substrato aminoácido entrante. De acuerdo con Dunathan, en las **reacciones de transaminación** la desprotonación del C α en **2a** produce el quinoide **3a**, que por protonación en C4' se convierte en la cetimina **4a**. Por hidrólisis se obtienen el **oxoácido** y **PMP 5a**. La reversión de esta reacción parcial con un substrato oxo-ácido entrante (como aceptor de grupos amino) completa la transaminación. Alternativamente, en las **reacciones de racemización**, **3a** puede reprotonarse desde el lado opuesto y dar la aldimina externa **4d**, que por transiminación produce el **D-aminoácido** y **1**. Las **eliminaciones β y γ** de un grupo X $^\circ$ también se conducen desde **3a**; en este ejemplo, una cadena lateral X-CH $_2$ elimina a X $^\circ$, y produce el aminoacrilato **4e**. La reprotonación y transiminación reforman **1** y un **producto imina**, que se hidroliza en piruvato y amoníaco. En las **reacciones de remplazamiento β** , un grupo Y $^\circ$ es sumado al C β ; esto conduce al quinoide **5f**, que es reprotonado en el C α para dar la aldimina externa **6f**. La transiminación libera el producto y **1**. Las **descarboxilaciones α** se conducen desde **2b** a través del quinoide **3b**; luego, por reprotonación en el C α hacia la aldimina externa **4b**, y hacia el **producto amino** y **1**. Alternativamente, la reprotonación en el C4' da la cetimina y finalmente un oxo-producto y PMP (**3a** \rightarrow **4a** \rightarrow **5a**). La ruptura del enlace C α -C β en **2c** produce, con el tetrahidrofolato como aceptor, 5,10-metilentetrahidrofolato y, vía **3c** y la aldimina externa **4c**, glicina y **1**. esta reacción es catalizada por la glicina hidroximetiltransferasa y otras enzimas que cortan el enlace C α -C β (Tomado de Jansonius, 1998).

Dunathan propuso el primer modelo concerniente al control de la especificidad de reacción de las enzimas dependientes de PLP (Eliot & Kirsch, 2003; 2004). En este modelo, la unión de sustratos supone un posicionamiento tal que el enlace que va a romperse queda alineado de manera perpendicular respecto del anillo del PLP. La superposición de este enlace con el orbital π del PLP minimiza la energía del estado de transición para la ruptura de enlace. El sistema π del cofactor se extiende en un carbono. El par electrónico libre puede deslocalizarse sobre estos orbitales moleculares π extendidos. Este modelo ha sido apoyado por diversos estudios. En todas las estructuras cristalográficas reportadas, el enlace C4'-N de la aldimina externa adopta una conformación *cisoides*. En esta conformación, las cargas negativas de los grupos fosfato (del PLP) y α -carboxilato (del aminoácido) están a una distancia máxima. Resulta razonable pensar que la conformación *cisoides* es energéticamente favorable para mantener la unión de la aldimina a la apoenzima (Christen *et al.*, 1996).

Pese a la uniformidad de sus mecanismos catalíticos, las enzimas dependientes de PLP se distribuyen en 4 familias estructuralmente diferentes: *i*) la familia α , con la aspartato aminotransferasa como prototipo; *ii*) la familia β , representada por la subunidad β de la triptofano sintasa; *iii*) la familia de la D-alanina aminotransferasa, y; *iv*) la familia de la alanina racemasa. La familia α es la más grande y diversa en términos de función (Mehta & Christen, 2000; Jansonius, 1998; Eliot & Kirsch, 2004). En muchos de sus miembros, la regioespecificidad de los cambios covalentes en el sustrato aminoácido están restringidos al C α (Alexander *et al.*, 1994). La familia α reúne diversas transaminasas y descarboxilasas, pero también enzimas que catalizan sustituciones y eliminaciones de tipos β y γ . Inclusive, las actividades del resto de las familias se sobrelapan con la observadas en la familia α . Posiblemente, la familia α emergió primero en la evolución y por ello ocupó mayor espacio catalítico. De manera relevante para este trabajo, esta familia incluye enzimas que participan en la ruta de biosíntesis de la biotina.

La biosíntesis de la biotina

La biotina es un cofactor en enzimas carboxilativas, como la piruvato carboxilasa y la metilcrotonil-CoA carboxilasa. En *E. coli*, este cofactor interacciona con dos proteínas: la acetil CoA carboxilasa, que produce malonil-CoA en el primer paso de la biosíntesis de ácidos grasos, y; BirA, que actúa como holocarboxilasa sintasa y represor transcripcional del operón para la biosíntesis de la biotina (Streit & Entcheva, 2003). En la figura 8 se muestran los intermediarios y las enzimas que participan en la biosíntesis de la biotina (Fig. 8; Streit & Entcheva, 2003; Ifuku *et al.*, 1994). De estas interconversiones nos resultan interesantes las que afectan al pimeloil-CoA y al 7-cetoaminopelargonato, pues son llevadas a cabo por enzimas dependientes de PLP. En *E. coli*, las actividades 7-ceto-aminopelargonato sintasa (CAPS) y 7,8 diaminopelargonato sintasa (DAPS) dependen de los productos de los genes *bioF* y *bioA*, respectivamente (Fig. 9a).

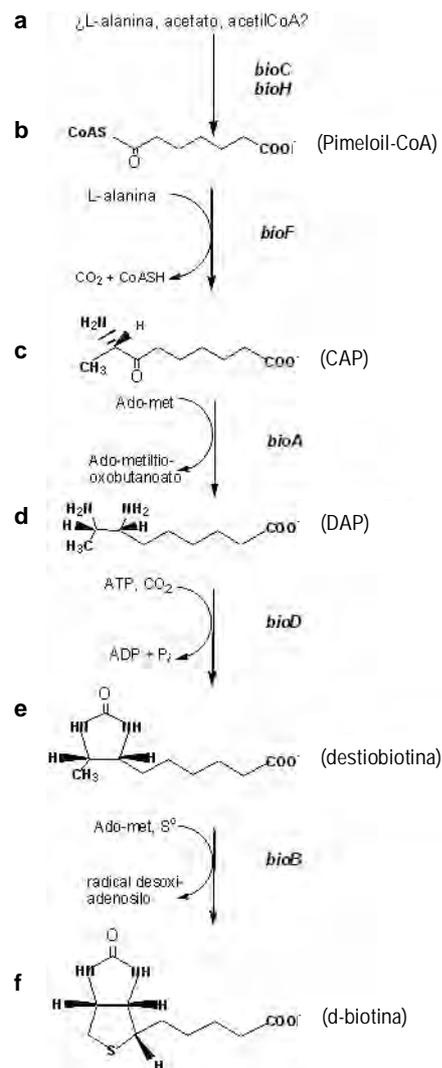
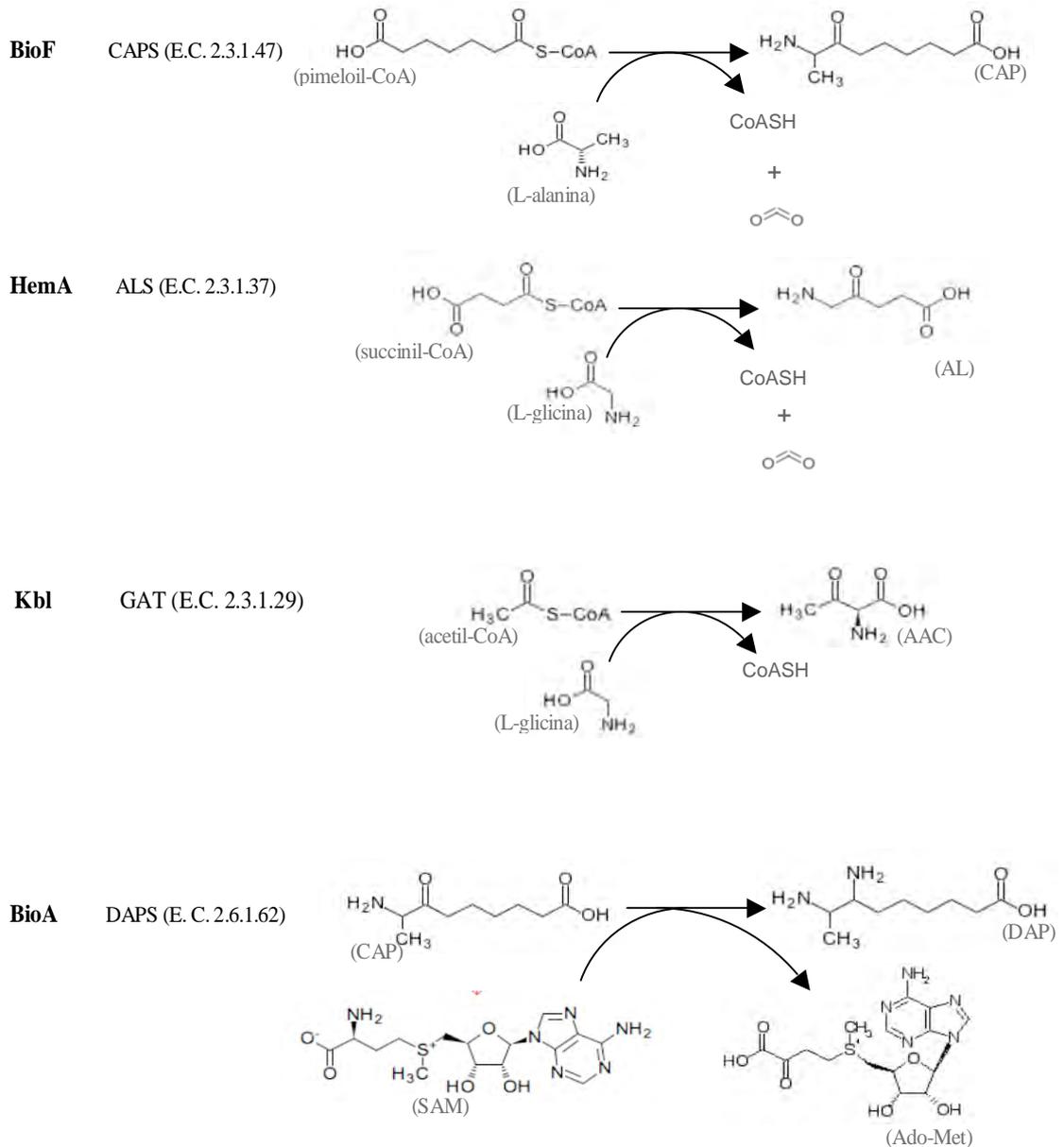


Figura 8. Ruta de síntesis de la biotina en *E. coli*. Se muestran los intermediarios y genes necesarios para la biosíntesis de la biotina. En *E. coli*, la síntesis de pimeloil-CoA depende de los genes *bioC* y *bioH*, pero no se conocen claramente los precursores y las reacciones que lo originan (a-b). Después se forman los intermediarios pimeloil CoA → 7-ceto-aminogelargonato (CAP) → 7,8 diaminogelargonato (DAP) → d-destiobiotina → d-biotina. Las reacciones b-c y c-d dependen de la CAP sintasa y la DAP aminotransferasa, codificadas por *bioF* y *bioA*, respectivamente. En la cuarta reacción, la destiobiotina sintasa, BioD, actúa sobre DAP (d-e). La biotina sintasa, BioB, transforma la destiobiotina en d-biotina, la forma biológicamente activa del cofactor (e-f).

BioF forma parte de la familia α de las enzimas dependientes de PLP. Esta enzima forma una pequeña subfamilia junto con Kbl y HemA (Fig. 9). Kbl participa en la degradación de la L-treonina para formar L-glicina. Por otro lado, HemA genera al precursor universal de los tetrapirroles en eucariontes no fotosintéticos y α -proteobacterias (Ploux & Marquet, 1996; Schneider & Linqvist, 2001; Alexeev *et al.*, 1998; Webster *et al.*, 2000; Panek & O'Brian, 2002). La figura 9a muestra que BioF, Kbl y HemA catalizan aminotransferencias con substratos muy parecidos. Las tres enzimas usan un aminoácido pequeño y apolar como donador de grupos NH₂, y un ácido relativamente débil unido a CoASH como aceptor. BioA también

es una aminotransferasa de la familia α . Sin embargo, esta enzima no usa un aminoácido como donador de grupos NH_2 , sino *S*-adenosilmetionina. En la figura 9b se muestran los porcentajes de identidad compartidos por BioF, HemA, Kbl y BioA. Puede verse que estas enzimas han divergido entre sí de manera muy significativa. Debido a las semejanzas catalíticas entre BioF y HemA, se supuso que podrían lograrse experimentos para evaluar la importancia de la promiscuidad catalítica en la evolución de actividades biológicamente relevantes.

a**b**

	BioF	HemA	Kbl	BioA
BioF	100,00	29,88	29,59	14,41
HemA		100,00	27,30	17,26
Kbl			100,00	12,75
BioA				100,00

Figura 9. Actividades e identidades de las enzimas BioF, HemA, Kbl y BioA. En (a), números E. C. y reacciones catalizadas por los parálogos BioF (7-ceto-aminopelargonato sintasa, CAPS); HemA (5-aminolevulinato sintasa, ALS), Kbl (glicina acetiltransferasa, GAT) y; BioA (7,8 diaminopelargonato sintasa, DAPS). En (b), porcentajes de identidad entre las secuencias de los parálogos. (Con base en Streit & Entcheva (2003) y en la base de datos Kyoto Encyclopedia of Genes and Genomes; www.genome.jp/kegg/.)

Planteamiento del problema

La promiscuidad catalítica es abundante en la naturaleza y se ha aumentado de manera discreta por evolución *in vitro*. También son discretas las mutaciones que dan lugar a dichos aumentos, así que el cambio gradual de actividades preexistentes podría dar base a la divergencia de nuevas actividades en la naturaleza. En esta búsqueda de papeles biológicos para la promiscuidad catalítica, el siguiente paso es el reclutamiento contingente de un intermediario generalista en diferentes regiones del metabolismo. Este escenario permitiría la valoración empírica de hipótesis generales sobre la evolución metabólica. Asimismo, daría pie a la especulación sobre la importancia de la promiscuidad para la incursión orgánica en nichos inusitados. Con base en lo anterior pueden plantearse objetivos y una estrategia experimental.

Objetivo general

Ampliar la comprensión del papel evolutivo de la promiscuidad catalítica en el surgimiento de nuevas actividades enzimáticas y en la constitución de rutas metabólicas.

Objetivos particulares

La obtención de variantes de HemA con actividad 7-ceto-aminopelargonato sintasa por evolución dirigida y selección *in vivo*, en una cepa $\Delta bioF$ de la bacteria *Escherichia coli*.

La determinación del potencial evolutivo de las variantes seleccionadas para restaurar la actividad 7,8 diaminopelargonato sintasa, en mutantes $\Delta bioA$ y $\Delta bioA \Delta bioF$ de la bacteria *Escherichia coli*.

Estrategia experimental

Los principios subyacentes a la estrategia experimental son simples, *i. e.*, la variación sistemática del gen codificante de una enzima y la detección de variantes enzimáticas con la actividad deseada. La noción de mutagenizar una secuencia génica acepta implícitamente la imposibilidad de explorar todas las combinaciones posibles de la cadena polipeptídica codificada. Sin embargo, puede esperarse una exploración productiva si la mutagénesis resulta en una librería de entre 10^5 y 10^9 secuencias mutantes. Este número no deja de ser vasto, por lo que se necesita un sistema que revele de manera eficaz la presencia de la actividad deseada.

La metodología circunviene problemas prácticos de manera darwiniana. El primer paso recrea la generación de variantes alélicas mediante PCR en condiciones que aumentan la tasa de error de la DNA polimerasa. En el segundo, los ciclos recursivos de mutación se amplifican selectivamente. Esto es que las variantes de la librería se conservan ó se eliminan por un sistema biológico. El sistema es una mutante de *E. coli* que arrostra opciones de vida o muerte ante la ausencia de un nutrimento esencial. El poder de esta aproximación está en que un fenotipo de crecimiento puede extrapolarse al resurgimiento de una actividad enzimática eliminada con premeditación. Pero las actividades generadas *in vitro* son predeciblemente débiles y difícilmente detectables. Por lo anterior, como sistema de selección se usó una mutante $\Delta bioF$, que por definición genética y fenotípica es auxótrofa de biotina. La biotina es esencial para el crecimiento, pero se requiere en cantidades mínimas. En principio, este requerimiento mínimo debería permitir la detección de variantes enzimáticas que catalizan la actividad deseada, aunque ésta sea poco robusta.

El blanco de la variación fue HemA de *B. japonicum*, un parálogo muy divergente de BioF de *E. coli* (~30 % de residuos idénticos compartidos; Fig. 9b). Pese a su baja identidad, ambas enzimas catalizan una aminotransferencia con substratos muy parecidos. Un parálogo más distante está representado por BioA (17 % de identidad compartida con HemA; Fig. 9b). Las reacciones catalizadas por BioF y BioA constituyen pasos consecutivos en la biosíntesis de biotina. Las mutantes $\Delta bioA$ y $\Delta bioA \Delta bioF$ se usaron para evaluar la capacidad de las variantes de HemA para restaurar una segunda actividad necesaria en la biosíntesis de biotina. Una derivación metodológica interesante consistió en la selección de variantes generadas por mutagénesis al azar de las posiciones correspondientes a codones selectos.

El significado biológico de los hallazgos puede extenderse a un contexto orgánico por la naturaleza del sistema de selección de las variantes enzimáticas. En otras palabras, la estrategia experimental recrea principios de la evolución natural. Además, dado que el escrutinio de la variabilidad genética tiene impacto en un nivel de organización definitivo de un organismo, puede valorarse la evolución de nuevas actividades enzimáticas en el contexto de la creación de complejidad y la adaptación.

Materiales y métodos

Cepas, plásmidos y condiciones de crecimiento. Las cepas usadas en este trabajo fueron *Bradyrhizobium japonicum* 110spc4 (Sp^f), *Escherichia coli* MC1061 [Sm^f Δ(*lacIPOZYA*) X74 *hsdR*], *E. coli* BL21 [F⁻ *dcm ompT hsdS*(r_B- m_B-) *gal*] y *E. coli* ER2566 F- λ- *fhuA2* [lon] *ompT lacZ::T7 gene 1 gal sulA11* Δ(*mcrC-mrr*)114::IS10 R(*mcr-73::miniTn10-TetS*)2 R(*zgb-210::Tn10*)(*TetS*) *endA1* [*dcm*] (New England Biolabs, Inc.). El plásmido pUC18 (Ap^r), un derivado multicopia del plásmido pBR322, se usó como vector de expresión (Yanisch-Perron *et al.*, 1985). El plásmido pKO3 (Cm^r) es un vector de remplazamiento útil para una metodología de remoción génica. El plásmido pET28 (Km^r) se usó como vector para la purificación de proteína. *E. coli* se cultivó en medio rico LB (Lysogeny Broth) y en medio mínimo M9, a 37 °C. Los cultivos líquidos se incubaron en condiciones aeróbicas (con agitación a 250 rpm). Para la selección se usó medio mínimo M9: glucosa (27 mM); MgSO₄ (1 mM); disolución de sales M9 [NH₄Cl (18 mM), Na₂HPO₄ (47 mM), KH₂PO₄ (22 mM), NaCl (8mM)]; L-leucina (0.4 mM); Bacto-Agar (1.5%). A menos de que se indique lo contrario, este medio se complementó con casaminoácidos libres de vitaminas. Todos los cultivos se complementaron con los antibióticos apropiados en las concentraciones siguientes: ampicilina (200 μg mL⁻¹); estreptomycin (200 μg mL⁻¹), cloranfenicol (35 μg mL⁻¹); kanamicina (25 μg mL⁻¹). Todos los químicos se obtuvieron de J.T. Baker, Sigma-Aldrich o Merck.

Manipulación de DNA. Las manipulaciones rutinarias se hicieron de acuerdo con procedimientos estándar. Las descripciones detalladas se presentan en la sección final de protocolos. El DNA cromosómico de *E. coli* se aisló de acuerdo con Sambrook *et al.* (1989). Los plásmidos se purificaron con el método ‘High Pure Plasmid Isolation Kit’ de Roche. Las transformaciones se llevaron a cabo por electroporación (Tung & Chow, 1995), con un equipo Gene PulserTM de BioRad. Las reacciones de digestión y de ligación se hicieron con enzimas de New England Biolabs Inc., siguiendo las instrucciones del fabricante. La DNA-polimerasa Taq se purificó previamente en el laboratorio, a partir de una cepa de *E. coli* en que se sobreexpresa el gen *pol* de *Thermus aquaticus*. También se empleó la DNA-polimerasa Vent_R[®] de New England Biolabs, Inc.

Las amplificaciones por PCR se realizaron con un termociclador Robocycler 40 de Stragene. Las reacciones no mutagénicas se prepararon en volúmenes de 100 μL, con las siguientes proporciones: 20 ng de DNA templado; 30 pmoles de cada oligonucleótido; MgCl₂ (1.5 mM); dTTP, dCTP, dGTP y dATP en concentraciones equimolares (0.2 mM); 2.5 u de DNA-Polimerasa Taq; 10 μL de amortiguador de reacción 10x [(NH₄)₂SO₄ (0.166 M), Tris-HCl (0.67 M, pH 8.8), MgCl₂ (15 mM), EDTA (67 mM, pH 8.0), BSA (1.7 mg mL⁻¹)]. Las reacciones de amplificación ocurrieron bajo los parámetros siguientes: 1 ciclo inicial de desnaturalización, a 94 °C por 5 min; 30 ciclos de amplificación (94 °C por 5min, 55 °C por 1 min y 72 °C

por 2 min); 1 ciclo terminal de extensión, a 72 °C por 5 min. Las condiciones de PCR mutagénica se mencionan más adelante. Todos los productos de amplificación y de digestión se purificaron a partir de geles de agarosa con el método comercial 'High Pure PCR Isolation Kit' de Roche. La secuenciación de DNA se llevó a cabo en la unidad de secuencia del IBT, con un secuenciador ABI Prism 3700 automatic DNA sequencer (Perkin-Elmer).

Construcción de los plásmidos pUC18/*hemA*, pUC18/*bioA*, pUC18/*bioF* y pUC18/*bioA bioF*. El plásmido pUC18/*hemA* se construyó en este trabajo. Los plásmidos pUC18/*bioA*, pUC18/*bioF* y pUC18/*bioA bioF* se construyeron previamente en el laboratorio. Preparaciones de DNA cromosómico se usaron como templado para amplificar las secuencias silvestres de los genes *hemA*, *bioF* y *bioA*. Los oligonucleótidos incluyeron sitios de digestión *EcoRI* (5'-gaattc-3') y *BamHI* (5'-ggatcc-3') para aprovechar los sitios de clonación de pUC18. Los oligonucleótidos 5'-gggaattccatgatggattacagccagttcttaattccg-3' y 5'-cgggatccctactccgccgagcacttttgcttgag-3' sirvieron para amplificar la secuencia *hemA* de *B. japonicum*. Los oligonucleótidos 5'-ccggaattcttaaaggaggagcactatgagctggcaggagaaaatc-3' y 5'-cgggatcccttaaccgttgccatgcagcacctccaacag-3' se usaron para la secuencia *bioF* de *E. coli*. De manera similar, la secuencia *bioA* de *E. coli* se obtuvo con el par 5'-ccggaattcttaaaggaggagcactatgacaacggacgatc-3' y 5'-ttagtggatccccgctttgcgcgagcactcc-3'. Con el fin de construir los vectores pUC18/*hemA*, pUC18/*bioA* y pUC18/*bioF*, los productos de amplificación se purificaron y se digirieron. Los fragmentos de DNA de las secuencias génicas fueron purificados y ligados a un vector pUC18 previamente digerido. Para construir el plásmido pUC18/*bioA bioF*, se generaron dos fragmentos con los siguientes pares de oligonucleótidos: 5'-ccggaattcttaaaggaggagcactatgacaacggacgatc-3' y 5'-tctagagctcctcctctgaatcttaaccgttgccatgcagcacc-3'; 5'-attcggagctcatgatgacaacggacgatcttgctttg-3' y 5'-gactaagctttattggcaaaaaatgtttcatcc-3'. Los productos de amplificación se trataron como se describió antes. Digestiones con *EcoRI* y *SacI* (5'-gagctc-3'), o con *SacI* y *HindIII* (5'-aagctt-3'), permitieron la clonación de ambas secuencias en posición contigua en pUC18.

Construcción de las cepas mutantes de *E. coli* Δ *bioA*, Δ *bioF* y Δ *bioA* Δ *bioF*. Las mutantes se contruyeron previamente en el laboratorio. Los genes de las enzimas de la biosíntesis de la biotina de *E. coli* están organizados en dos operones que se transcriben de manera divergente (Streit & Entcheva, 2003). La primer unidad transcripcional se sitúa en la cadena *plus* del cromosoma y consiste en los genes *bioB*, *bioF*, *bioH*, *bioC* y *bioD*. La segunda unidad se encuentra en la cadena *minus* y está constituida por los genes *bioA* y *ybhB* (éste último no parece no asociarse a la síntesis de la biotina). El extremo 5' de la secuencia génica *bioF* se sobrelapa con 3 bp de la de *bioB*, mientras que el extremo 3' comparte 14 bp con *bioC*. Las secuencias *bioA* y *bioB* están separadas por 87 bp. Remociones precisas se construyeron mediante la metodología de Link *et al.* (1997). Esta metodología depende de la recombinación entre una región cromosómica y el

plásmido pKO3. El plásmido pKO3 contiene un origen de replicación termosensible; asimismo, confiere capacidad para hidrolizar sacarosa y para resistir al cloranfenicol. Estas características sirven como marcadores de selección positiva ó negativa, según sea necesario seleccionar la integración ó escisión de pKO3 respecto del cromosoma.

Para construir la secuencia de remplazamiento de *bioF*, las regiones aledañas al gen se amplificaron con los siguientes pares de oligonucleótidos: *bioB*-for (5'-cgcgatccccacgctggacattgtcgaagtc-3') y *bioB*-rev (5'-gccggaattctcatagtctgccgcgttgaatattc-3'); *bioC*-for (5'-cactatgagaattccggcctctcttgagggtgctgcatggcaacg-3') y *bioC*-rev (5'-ccgggatccttactcacgaacaatcactccc-3'). Los oligonucleótidos *bioB*-rev y *bioC*-for se sobrelapan parcialmente. Los productos de la amplificación se alinearon y se amplificaron como un solo fragmento para obtener la secuencia de remplazamiento, la cual se trató con *Bam*HI para ligarse a un vector pKO3 cortado con la misma enzima. La secuencia de remplazamiento para la remoción de la secuencia génica *bioA* se construyó de manera similar, con los siguientes pares de oligonucleótidos: *ybhB*-for (5'-acagggatccaat-tgccgcatctaccgccgctgaatcg-3') y *ybhB*-rev (5'-gatgccgcctatattctcccgaacag-3'); *bioB*-for (5'-aatatagggcgcatccgacataccgcaaccaggcgtctg-3') y *bioB*-rev (5'-tagtggatccgaacagttgcaggtttatcttc-3'). Los plásmidos resultantes, pKO3/ Δ *bioF* y pKO3/ Δ *bioA*, se introdujeron en *E. coli*. Las células transformantes se sembraron en LB-cloranfenicol y se incubaron a 43 °C, por 12 h. A esta temperatura, pKO3 es incapaz de replicarse y el crecimiento queda imposibilitado, a menos que el plásmido se integre en el cromosoma. Para hacer más eficiente la recuperación de células con plásmidos integrados, se cultivaron colonias en medio líquido LB-cloranfenicol a 43 °C, por 6 h. Diluciones seriales de estos cultivos fueron sembradas en medio sólido con sacarosa (5%) y estreptomycin, e incubadas a 30 °C. Se seleccionaron colonias para ser sembradas por duplicado, en presencia de sacarosa y kanamicina, a 30 °C, y en LB-cloranfenicol, a 45 °C. Las colonias que lograron crecer exclusivamente a la temperatura permisiva, se cosecharon para su caracterización ulterior. Para construir la cepa mutante Δ *bioA* Δ *bioF*, células Δ *bioF* se transformaron con el plásmido pUC18/*bioF*. Esta cepa fue tratada para eliminar la copia cromosómica de *bioA*. Para eliminar el plásmido pUC18 *bioF*, la doble mutante se cultivó de manera continua en medio líquido LB sin ampicilina.

El análisis de fenotipo de las cepas Δ *bioF*, Δ *bioA* y Δ *bioA* Δ *bioF* reveló la conservación de la funcionalidad de los genes restantes de la biosíntesis de la biotina. Las cepas fueron incapaces de crecer en medio mínimo M9 sin biotina; el crecimiento se restableció por la expresión extracromosómica de las secuencias silvestres correspondientes. Asimismo, análisis genéticos revelaron la remoción precisa de los genes *bioF* y *bioA*. Reacciones de PCR a partir de colonia produjeron fragmentos de longitud predecible. La secuenciación de DNA cromosómico reveló que tras la remoción génica, el marco de lectura del operón permaneció intacto en cada mutante. La delección de *bioF* se verificó con los oligonucleótidos 5'-cagtcgctaacggcaggattc-3' y 5'-gccggacgcgagcagatgaacg-3'. La delección de *bioA* se revisó con los oligonucleótidos *ybhB*-for y *bioB*-rev.

Construcción y exploración de librerías de variantes *hemA* (*hemA^{mut}*) por mutagénesis al azar. Se llevaron a cabo reacciones de PCR mutagénica con una versión modificada del protocolo de Leung *et al.* (1989). Las reacciones fueron preparadas en volúmenes de 100 μ L, con lo siguiente: 20 ng de templado pUC18 *hemA*; 30 pmoles de los oligonucleótidos *hemA*-for (5'-cgggaattcttaaaggaggagcacatatggattacagccagtctttaaattccgcc-3') y *hemA*-rev (5'-cacctaaagcttctactccgcgccagcagcttttgc-3'); $MnCl_2$ (0.5 mM); dGTP y dATP (0.2 mM); dCTP y dTTP (1 mM); 10 μ L de DMSO; β -mercaptoetanol (0.01 mM); 5 u de DNA-polimerasa Taq; 10 μ L de amortiguador de reacción 10x [(NH_4)₂SO₄ (0.166 M), Tris-HCl (0.67 M, pH 8.8), MgCl₂ (61 mM), EDTA (67 mM, pH 8.0), BSA (1.7 mg mL⁻¹)]. Los parámetros de reacción fueron los siguientes: 1 ciclo inicial de desnaturalización, a 94 °C por 5 min; 30 ciclos de amplificación (94 °C por 5min, 55 °C por 1 min y 72 °C por 2 min); 1 ciclo terminal de extensión, a 72 °C por 5 min. El producto de la amplificación mutagénica se purificó, se digirió con *Hind*III y *Eco*RI, se purificó nuevamente y se ligó a pUC18. La librería resultante se introdujo en células competentes Δ *bioF* recién preparadas. Las transformantes se sembraron en cajas con medio sólido LB-estreptomicina-ampicilina. Las cajas se incubaron por 24 h para permitir el aumento de la abundancia de cada secuencia mutante. Después se aisló el plásmido. El tamaño de la librería se calculó mediante diluciones seriales y cuantificación de células provenientes de la electroporación. La distribución de las mutaciones en la librería se estimó por secuenciación.

La librería se introdujo en células Δ *bioF* para la selección de variantes *hemA^{mut}* capaces de revertir la auxotrofia de biotina. La selección se llevó a cabo en dos rondas. En la primera, células Δ *bioF* recién transformadas con la librería se lavaron 4 veces. Cada lavado consistió en una centrifugación a 3,000 rpm, por 10 min, seguida de la decantación del sobrenadante y la resuspensión de las células con 5 mL de disolución de sales M9. (La resuspensión se logró con micropipeta.) Las células se sembraron en medio mínimo M9 complementado con cantidades picomolares de biotina (2.5, 5 y 10 pM). El número de células por caja varió entre ~300 y ~3000. Las colonias que aparecieron en el transcurso de 9 días de incubación se cosecharon y se mezclaron en un cultivo líquido de LB-estreptomicina-ampicilina. El plásmido se aisló y se introdujo nuevamente en células Δ *bioF*. Para la segunda ronda de selección, las transformantes se lavaron y se sembraron en números de ~300. Las colonias más conspicuas al término de 5 días se cosecharon individualmente para aislar plásmido. Las variantes *hemA^{mut}* seleccionadas se reclonaron en pUC18, para asegurar que los fenotipos observados fuesen atribuibles exclusivamente a la secuencia clonada. La aparición de falsos positivos y el crecimiento de fondo obligaron a controlar estrictamente la presencia de biotina en los medios de cultivo. La presencia de trazas de vitamina, y de otros restos de materia orgánica, se evitó con el uso de vidriería lavada con mezcla de H₂SO₄-K₂Cr₂O₇. En el mismo sentido, la disoluciones fueron esterilizadas sucesivamente por filtración y autoclave (excepto los antibióticos).

Construcción y exploración de una librería de variantes *hemA*^{mut} por mutagénesis dirigida a sitios específicos. Tres rondas de PCR sirvieron para construir variantes *hemA*^{mut} con distintos nucleótidos en los codones 165 y 288. En cada ronda se usó pUC18/*hemA* como templado. Cada producto de amplificación sirvió como ‘megaprimer’ en la ronda siguiente. La ronda 1 incluyó los oligonucleótidos 5’-gggaattccatatggattacagccagttcttaattccg-3’ y 5’-caacagcgcttcgagG/CNNggcgagatcggttg-3’. El ‘megaprimer’ se purificó y se usó en la ronda 2 junto con el oligonucleótido 5’ gcttgatcgggcG/CNNcgggccgagcgat-3’. El producto de esta amplificación se usó con el oligonucleótido 5’-cgcggatccctactccgcccagcgacttttgcttgag-3’. El producto de la ronda 3 se reamplificó con los oligonucleótidos *hemA*-for y *hemA*-rev. La secuencia se purificó, se digirió con *Hind*III y *Eco*RI, se purificó nuevamente, y se ligó a pUC18. El tamaño de la librería fue calculado como se describió antes. Todas las amplificaciones se llevaron a cabo en condiciones no mutagénicas. Sin embargo, se observó la aparición de mutaciones adicionales a las introducidas deliberadamente. Este factor se minimizó con el uso combinado de las DNA-Polimerasas Taq y Vent_R[®] (proveniente de *Thermococcus litoralis*). La tasa de error de la DNA-Polimerasa Taq es 2.2×10^{-5} por nucleótido por ciclo. Debido en parte a su actividad exonucleasa 3’-5’, la enzima Vent_R[®] exhibe una fidelidad 5 a 15 veces mayor a la de la enzima Taq.

La librería se usó para transformar la cepa Δ *bioF*. La selección de variantes capaces de revertir la auxotrofia de biotina se realizó en una sola ronda. Las transformantes se lavaron y se sembraron en medio mínimo complementado con biotina 10 pM. El número de células por caja fue de ~1000. Veinte colonias se cosecharon individualmente al término de 9 días de incubación. Los plásmidos se aislaron para retransformar a la cepa Δ *bioF*, con el fin de verificar los fenotipos observados. Después se reclonaron cinco alelos mutantes para realizar experimentos de complementación en los fondos de mutación Δ *bioF*, Δ *bioA* y Δ *bioA* Δ *bioF*. Asimismo, se obtuvo la secuencia de las variantes. El fenotipo de complementación conferido por las secuencias *hemA*^{mut} se evaluó por comparación de crecimiento en medio mínimo M9 complementado con concentraciones picomolares de biotina. Células auxótrofas de biotina, transformadas con los plásmidos pUC18, pUC18/*hemA* ó pUC18/*hemA*^{mut}, se cultivaron en medio líquido LB-estreptomicina-ampicilina hasta la fase logarítmica ($DO_{540} \approx 0.4$). Estas transformantes se cosecharon por centrifugación y se lavaron 4 veces. Se ajustó la $DO_{600} \approx 0.6$ de cada mezcla celular para preparar diluciones seriales. Se inocularon volúmenes de 10 μ L de las diluciones en cajas con medio mínimo M9, complementado con antibióticos y trazas de biotina. El crecimiento celular se monitoreó a 37 °C. Este crecimiento fue evidente en forma de manchas (en las diluciones más concentradas) o de colonias separadas (en las más diluidas).

Expresión y purificación de proteína HemA. Las secuencias génicas de *hemA* y *hemA*^{mut1} se amplificaron y se clonaron en el plásmido pET28. Se construyó una cepa de *E. coli* ER2566 Δ *bioF* para la producción de proteína. En la cepa ER2566, el gen de la RNA polimerasa del bacteriofago T7 se expresa bajo el control del

promotor *plac*. Se prepararon cultivos de 500 mL de medio líquido LB mediante la inoculación 1:100 de un precultivo incubado por 12 h. Una incubación a 37° C, por 3.5 h, antecedió a la inducción de la expresión de las secuencias *hemA*. La inducción se logró por la adición de isopropil- β -D-1-tiogalactósido (IPTG, 1 mM) y la incubación a 20° C, por 16 h. A continuación, los cultivos se centrifugaron (5,000 rpm) a 4° C, por 10 min. Las células se resuspendieron en 20 mL de amortiguador A [KH_2PO_4 (50 mM), NaCl (300 mM, pH 7.63)] y se sonicaron en un equipo Branson 450. Tras una centrifugación (10,000 rpm) a 4° C, por 20 min, se hizo atravesar el sobrenadante por un filtro de membrana (diámetro de poro de 0.2 μm). La proteína se separó del sobrenadante mediante cromatografía de afinidad a Níquel (se empleó una 'HiTrap Ni²⁺ affinity column' de Amersham-Pharmacia Biotech). La purificación se logró en un equipo de FPLC ÄKTA (Amersham-Pharmacia Biotech). La columna se lavó con amortiguador A repetidamente, y una vez con Amortiguador B [KH_2PO_4 (50 mM), NaCl (300 mM)]. La elución de la proteína se logró con un gradiente de imidazol 10-500 (mM) en amortiguador B. Para remover el imidazol, se llevó a cabo una doble diálisis con amortiguador C [KH_2PO_4 (50 mM), EDTA (1mM), DTT (0.1mM, pH 7.80)], a 4° C. La pureza de la proteína se determinó por electroforesis en gel de poliacrilamida (10 %) en condiciones de desnaturalización. La proteína era inestable y se observó degradación tras la purificación.

Detección de las actividades enzimáticas. Las actividades 7-ceto-8-aminopelargonato sintasa (CAPS) y 5-aminolevulinato sintasa (ALS) se midieron mediante la detección espectrométrica de la liberación de CoASH a partir de pimeloil-CoA ó succinil-CoA. La detección se logró con el reactivo de Ellman, DTNB (Bhor *et al.*, 2006). El pimeloil-CoA fue producido en la Unidad de Síntesis del IBT, mientras que el succinil-CoA se compró a Sigma-Aldrich. Las mezclas de reacción consistieron en volúmenes de 100 μL , que contuvieron: 0.25 μg de enzima; DTNB (0.1 mM; reactivo de Ellman) en KH_2PO_4 (50 mM; pH 7.7); diferentes concentraciones de succinil-CoA ó pimeloil-CoA (6.25 - 400 μM); diferentes concentraciones de L-glicina ó L-alanina (31.5 - 8000 μM). Los cambios de absorbancia ($\lambda= 412 \text{ nm}$) se monitorearon a intervalos de 30 s, por 1 h, en el equipo Safire Quad4 Monochromator Microplate Reader (Tecan Group Ltd., Switzerland). Las constantes V_{max} and K_m se determinaron mediante análisis de regresión no-lineal con el programa SigmaPlot2001.

Modelaje estructural y análisis de secuencias

Para obtener un modelo *in silico* de HemA de *Bradyrhizobium japonicum* se empleó el software PromodII del servidor www.expasy.ch/swissmod/swiss-model/html. El modelo se basó en la estructura cristalográfica de HemA de *Rhodobacter capsulatus* en ausencia de sustratos. También se usaron las estructuras cristalográficas obtenidas con la enzima en presencia de sustratos (Astner *et al.*, 2005). Los modelos del

ortólogo de *R. capsulatus* están disponibles en la base de información estructural Protein Data Bank (www.rcsb.org), bajo las claves 2BWN, 2BWO y 2BWP. Se obtuvieron imágenes del ortólogo de *B. japonicum* con el programa MacPyMol del servidor <http://www.pymol.org> (DeLano Scientific LLC, 2006). Los sustratos pueden observarse en este modelo de *B. japonicum* gracias a la siguiente manipulación. Tras superponer las imágenes de los dos ortólogos, se eliminó la correspondiente a la enzima de *R. capsulatus*. La imagen de los sustratos fue conservada en una posición fija y como objeto independiente. Esto permitió ver interacciones entre partes de los sustratos y residuos de la enzima de *B. japonicum*. Tales interacciones coincidieron con las observadas originalmente en el análisis estructural de Astner *et al.* (2005). Los análisis a nivel de secuencia incluyeron alineamientos múltiples y la estimación de porcentajes de identidad. Para esto se utilizó el software CLC main workbench (<http://www.clcbio.com>).

Resultados y discusión

La promiscuidad catalítica podría constituir un potencial para la evolución de nuevas reacciones y procesos metabólicos. Con el fin de evaluar esta posibilidad se generaron variantes de HemaA por mutagénesis al azar; después se seleccionaron las variantes capaces de proveer pasos faltantes en la ruta de biosíntesis de la biotina de *E. coli*:

Mutagénesis aleatoria en la secuencia génica *hemA* de *B. japonicum*

Se generaron mutaciones en la secuencia *hemA* mediante PCR mutagénica (Leung et al., 1989). La incorporación de nucleótidos erróneos dependió de la presencia de MgCl₂ en exceso y de dNTPs en concentraciones desiguales; las condiciones mutagénicas también incluyeron DMSO y MnCl₂. Se obtuvo una librería de $\sim 1.12 \times 10^6$ alelos mutantes de *hemA* (*hemA*^{mut}). La frecuencia de las mutaciones se estimó a partir de 8 alelos *hemA*^{mut} aislados al azar. En un espacio de secuencia de 6,740 nucleótidos, el porcentaje de posiciones mutadas fue 1.76 %. Las transiciones y transversiones constituyeron 46.2 % y 53.8 % de los cambios, respectivamente. El número medio de mutaciones por secuencia fue ~ 15 . No se observó un sesgo significativo en la frecuencia de las mutaciones. En conclusión, la mutagénesis de la secuencia *hemA* satisface una premisa fundamental de la evolución natural: La acumulación sistemática y aleatoria de variación genética.

Selección *in vivo* de un alelo *hemA*^{mut} que revierte la auxotrofia de biotina en una cepa $\Delta bioF$

La selección de alelos *hemA*^{mut} que restauran la biosíntesis de biotina se llevó a cabo en dos rondas. En la primera, células $\Delta bioF$ transformadas con la librería se lavaron y se sembraron en medio de selección complementado con trazas de biotina. Se evaluaron $\sim 3.9 \times 10^6$ transformantes, *i. e.*, 3.5 veces el total de secuencias en la librería. En el transcurso de 9 días se observó el crecimiento de 17 colonias, las cuales se cosecharon y se cultivaron en medio rico. Una mezcla de plásmidos se recuperó y se introdujo en células $\Delta bioF$ para la segunda ronda. De nuevo, las transformantes se lavaron y se sembraron en medio con trazas de biotina. Las colonias más conspicuas tras 5 días se cosecharon por separado. Se caracterizaron los plásmidos de cinco colonias elegidas al azar. Todos los plásmidos mostraron una misma secuencia, que se denominó *hemA*^{mut1}. En comparación con el alelo silvestre, *hemA*^{mut1} mostró substituciones en las posiciones 494 (A→T), 862 (A→G) y 864 (C→A). Estas substituciones provocaron los cambios de codón GAC→GTC (en la posición 165) y ACC→GCA (en la posición 288). Los cambios esperados en la estructura primaria de

HemA^{mut1} son D165V y T288A. Resulta razonable suponer que la secuencia *hemA*^{mut1} se generó una sola vez durante la amplificación mutagénica. Así que el aislamiento repetido de *hemA*^{mut1} puede atribuirse a la estrategia de selección: En la segunda ronda se eliminaron falsos positivos, así como los alelos mutantes con la menor capacidad para estimular el crecimiento de las células $\Delta bioF$. La secuencia *hemA*^{mut1} se reclonó en pUC18 para verificar el fenotipo observado.

La figura 10 muestra el fenotipo asociado a la expresión extracromosómica del alelo *hemA*^{mut1}. Este fenotipo se comparó con el asociado al vector y al alelo *hemA* silvestre. (No se incluyeron células con el alelo *bioF* silvestre; éstas liberan biotina que captan las células auxótrofas, las cuales crecen aún si se elimina el cofactor de forma exógena.) Se sembraron diluciones seriales en medio con biotina en exceso. Se observaron números similares de UFCs entre puntos de crecimiento equivalentes. En apariencia, la presencia del vector y de los alelos *hemA* es indistinta genéticamente ante un suministro ilimitado de vitamina. Por otro lado, las mismas diluciones se sembraron en medio con trazas de biotina. La auxotrofia asociada a la delección de *bioF* impide el crecimiento de las células transformadas con el vector y el alelo *hemA* silvestre. En contraste, la presencia de *hemA*^{mut1} se correlacionó con el desarrollo de colonias desde el quinto día de incubación. Las dimensiones de las colonias aumentaron en todos los puntos de crecimiento hasta el final del experimento, a los 9 días.



Figura 10. Fenotipo de crecimiento conferido a las células $\Delta bioF$ por el alelo *hemA*^{mut1}. Células $\Delta bioF$ transformadas con los plásmidos pUC18 (carril 1), pUC18/*hemA* (carril 2) y pUC18/*hemA*^{mut1} (carril 3) se lavaron extensivamente y se sembraron en medio de selección complementado con biotina en exceso (panel a) ó en trazas (panel b). En el panel b se presenta el fenotipo observado tras 7 días de incubación. (Resultados representativos de al menos 5 experimentos independientes.)

El sistema de selección permitió la detección de una nueva actividad enzimática. Asimismo, permitió la evaluación *a priori* de dicha actividad en un contexto biológico. Esto es porque el sistema muestra sensiblemente que las consecuencias de la mutación para el cambio fenotípico están condicionadas por componentes del propio fenotipo (Kirschner & Gerhart, 1998). Aunque la auxotrofia de biotina supone una incapacidad para crecer, las mutantes transformadas con el vector desarrollaron UFCs en los puntos de mayor concentración celular. Este fenotipo es reproducible y distinto al asociado a *hemA*. Las diferencias observadas suscitan el cuestionamiento sobre los fenómenos poblacionales y celulares representados por el fenotipo. Lo anterior conlleva la interrogante acerca de la influencia de la actividad seleccionada en dichos fenómenos.

Aun en condiciones constantes, las células exhiben diferencias en parámetros como el *status* metabólico, la integridad de membrana y la viabilidad (Ericsson *et al.*, 2000; Makinoshima *et al.*, 2002). Es decir, el fenotipo es una representación a nivel poblacional de procesos celulares inherentemente heterogéneos. Esta heterogeneidad se exagera ante el conflicto de supervivencia causado por una presión de selección. El desenlace del conflicto podría ser el mantenimiento de la viabilidad a costa de la extinción de subpoblaciones celulares. A partir de lo anterior puede proponerse una explicación para el fenotipo de las células $\Delta bioF/pUC18$. Se sabe que la auxotrofia asociada a la mutación de *bioF* provoca la disminución de la viabilidad. Esta disminución podría afectar anticipadamente a un número de células, las cuales perderían la integridad de membrana. Biotina residual podría liberarse al medio. Una subpoblación que conserva el control de la permeabilidad podría incorporar esta vitamina. En los puntos de mayor densidad de UFCs habría suficiente biotina para mantener la división celular, de manera transitoria. En los puntos de menor densidad celular, la biotina residual sería insuficiente para que las células supervivientes reanuden la división. Lo anterior ocurriría por el requerimiento del cofactor para iniciar la síntesis de ácidos grasos (Berkovitch *et al.*, 2004), y consecuentemente de las envolturas celulares.

Las células $\Delta bioF pUC18/hemA$ no mostraron crecimiento, ni siquiera tras 20 días. Resulta evidente que *hemA* afecta al fenotipo de la auxotrofia. Esto podría deberse a un fenómeno de citotoxicidad. En el contexto de *E. coli*, la expresión de *hemA* está desregulada y su producto podría ser inestable estructuralmente. Asimismo, HemA lleva a cabo una actividad redundante (Panek & O'Brian, 2002). La expresión de un gen espurio representa una carga energética adicional a la del mantenimiento del plásmido. El crecimiento quedaría imposibilitado si la carga resulta insostenible para la mayoría de las células. Además, esta actividad implica la interconversión de intermediarios cuyo consumo, direccionamiento o acumulación fortuitos podrían conducir al colapso del metabolismo en una situación de estrés.

La expresión extracromosómica de *hemA*^{mut1} permite el restablecimiento del crecimiento. De manera relevante, el fenotipo parece independiente de la dinámica de supervivencia asociable a la heterogeneidad poblacional: Pueden compararse los números de UFCs en los puntos de menor concentración celular de las

condiciones de exceso y limitación de biotina. Esto indica que la mayoría de las células mantuvo la viabilidad ante un suministro de biotina escaso. El producto de la expresión de *hemA^{mut1}* no se ajusta al contexto celular de referencia, pero el costo de su mantenimiento parece compensarse por el beneficio que supone la actividad seleccionada. Puede concluirse que se ha reproducido una segunda premisa de la evolución natural: La supervivencia diferencial por la selección de un nuevo genotipo.

Los resultados indican el hallazgo de un alelo mutante que complementa una función celular suprimida premeditadamente. Esta función es un proceso metabólico en el se eliminó una actividad enzimática. ¿Es inferible que la expresión del alelo mutante restaura la integridad del proceso al proveer la actividad faltante? Sí, pero la evidencia apenas permite la lectura de una función génica sin el conocimiento de la actividad enzimática *per se*.

Detección *in vitro* de las actividades de Hema^{mut1}

En un ensayo invitro se determinaron las actividades específicas CAPS (de BioF) y ALS (la actividad principal de Hema). El ensayo se basó en la detección espectrométrica de la liberación de CoASH (Bhor *et al.*, 2006). En la figura 11a se muestra el comportamiento de Hema y Hema^{mut1} en presencia de los sustratos pimeloil-CoA y L-alanina (actividad CAPS). Hema parece exhibir un nivel vestigial de actividad CAPS. En comparación, parece que Hema^{mut1} cataliza la actividad CAPS con mayor rapidez. Se sugiere que la evolución de Hema^{mut1} significó el incremento discreto de una actividad promiscua preexistente. En la figura 11b se muestran los valores de K_m para los sustratos de la actividad principal ALS (succinil-CoA y L-glicina). En Hema^{mut1}, la K_m de la glicina disminuyó 26 veces, lo que sugiere un aumento de afinidad por el sustrato aminoácido. En contraste, la K_m del succinil-CoA aumentó 9 veces. Lo anterior supone que en Hema^{mut1} disminuyó la afinidad por el sustrato succinil-CoA. Aparentemente, en Hema^{mut1} no se alteró significativamente la actividad principal ALS.

El ensayo de la actividad CAPS se realizó sólo una vez, mientras que el ensayo de la actividad ALS se llevó a cabo por triplicado. La confirmación de las actividades se vio imposibilitada por la dificultad para manipular los sustratos. Tanto el pimeloil-CoA como el succinil-CoA parecen degradarse rápidamente en disolución. Esto coincide con el conocimiento de que los enlaces tioéster son lábiles en medios acuosos. La degradación del succinil-CoA fue confirmada por HPLC en la Unidad de Síntesis del Instituto. Se analizaron los picos de absorbancia generados en una preparación fresca de succinil-CoA. Se observó más de un pico, y se determinó que dos corresponden al CoASH y el ácido succínico por separado. A pesar de lo anterior, se supone que los resultados de los ensayos *in vitro* son fiables. Por un lado, se realizaron ensayos preliminares con el fin de encontrar la mejor concentración de sustratos para medir cada actividad. Y en todos estos

ensayos se observaron cambios de absorbancia que sugieren un proceso catalizado enzimáticamente. En HemA^{mut1}, el cambio de los parámetros catalíticos para la actividad principal ALS podría permitir que substratos alternativos participen en el recambio catalítico. Asimismo, con base en los experimentos de complementación genética, puede inferirse que dichos cambios dieron como resultado una actividad promiscua suficientemente robusta para restablecer una función metabólica y, consecuentemente, para permitir el crecimiento celular.

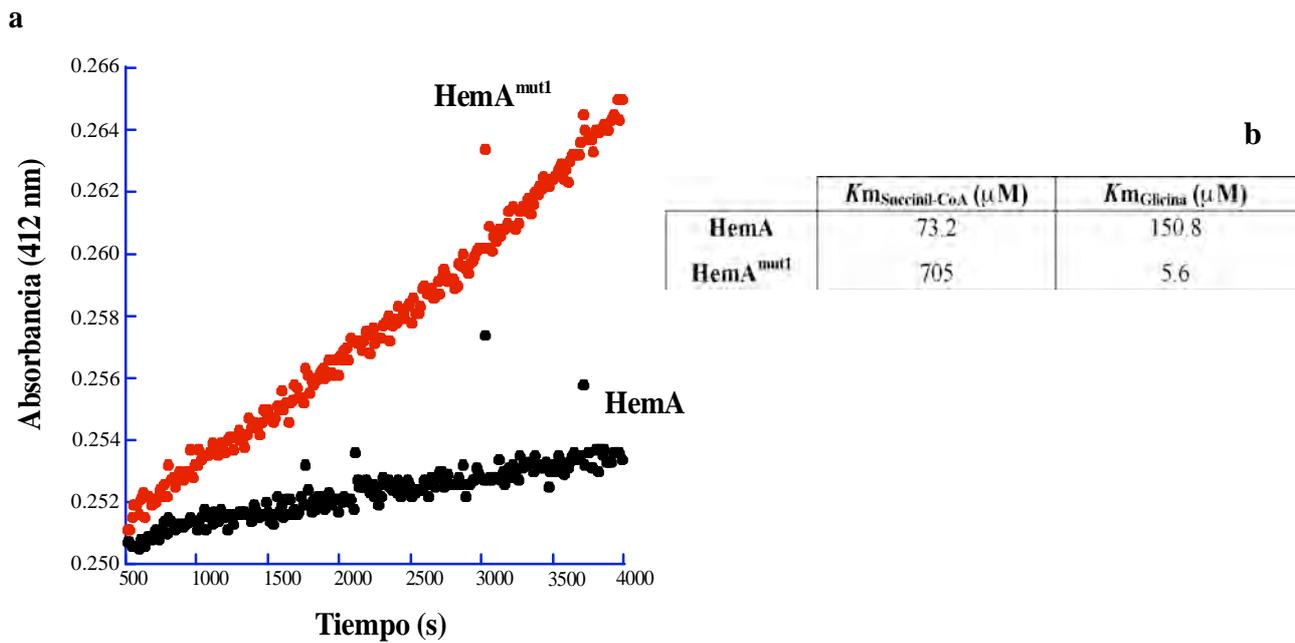


Figura 11. Caracterización bioquímica de la variante HemA^{mut1}. (a) Actividad CAPS de las proteínas HemA and HemA^{mut1}. La liberación de CoASH a partir de los substratos fue medida espectrométricamente con el reactivo de Ellman. (b) Afinidad por los substratos cognados de la actividad ALS de HemA y HemA^{mut1}. (Resultados de un experimento único realizado por triplicado.)

El camino evolutivo de HemA^{mut1} sugiere el hallazgo de posiciones de plasticidad

Los cambios esperados en la estructura primaria de HemA^{mut1} son D165V y T288A. La cuestión es cómo dos cambios discretos tienen efecto en la actividad de una entidad tridimensional. Los cambios de residuos se mapearon en un modelo *in silico* de HemA^{mut1} (Fig. 12). El modelo se basa en la estructura cristalográfica del ortólogo de *Rhodobacter capsulatus* (Astner *et al.*, 2005). El alto porcentaje de identidad de secuencia entre estos homólogos es 55.4 %, lo cual permite hacer extrapolaciones fiables. Por lo pronto, el cambio D165V se sitúa en un elemento secundario contiguo a la pared del sitio de unión del substrato acil-CoA. El cambio T288A está situado distantemente en la superficie. Ninguna de las posiciones afectadas parece implicarse en la actividad enzimática de manera directa.

La comparación de secuencias homólogas permite la identificación de elementos de secuencia importantes. La ausencia de sustituciones, y la presencia de cambios conservativos, sugieren la necesidad de ciertos residuos para la estructura y la función. Se construyó un alineamiento de 22 ortólogos de HemA provenientes de α -proteobacterias y eucariontes (Fig. 13). Los residuos D165 y T288 están representados en 18 % y 20 % de las secuencias, respectivamente. En la mayoría de los ortólogos, la posición 165 exhibe un residuo His en lugar de Asp. La posición 288 muestra más variabilidad, aunque Leu y Thr se repiten varias veces. La presencia repetida de ciertos residuos podría indicar la conservación de propiedades físicoquímicas. De acuerdo con esto, la posición 165 está ocupada por residuos muy hidrofílicos, aunque éstos difieren en volumen molecular y polaridad. Por otro lado, la posición 288 puede exhibir mayor variación. En general, puede suponerse una relajación local de los constreñimientos evolutivos. En contraposición, un alto nivel de conservación global es observable entre las secuencias HemA. Los ortólogos de especies diversas muestran entre 46 % y 68 % con HemA de *B. japonicum*.

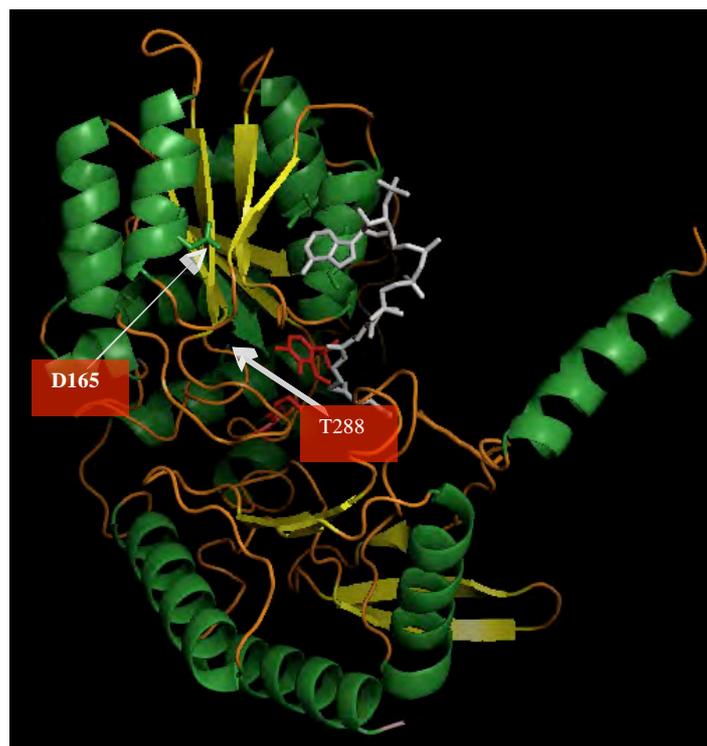


Figura 12. Mapeo de las posiciones 165 y 288 en el monómero de HemA de *B. japonicum*. La posición 165 se sitúa en una región aledaña a las paredes del bolso de unión del sustrato sCoA. La posición 288 es parte de una hélice distante respecto del sitio activo. El modelo fue obtenido a partir de la estructura cristalográfica del ortólogo de *R. capsulatus* (Astner *et al.*, 2005). El PLP y el sCoA están representados en rojo y blanco, respectivamente.

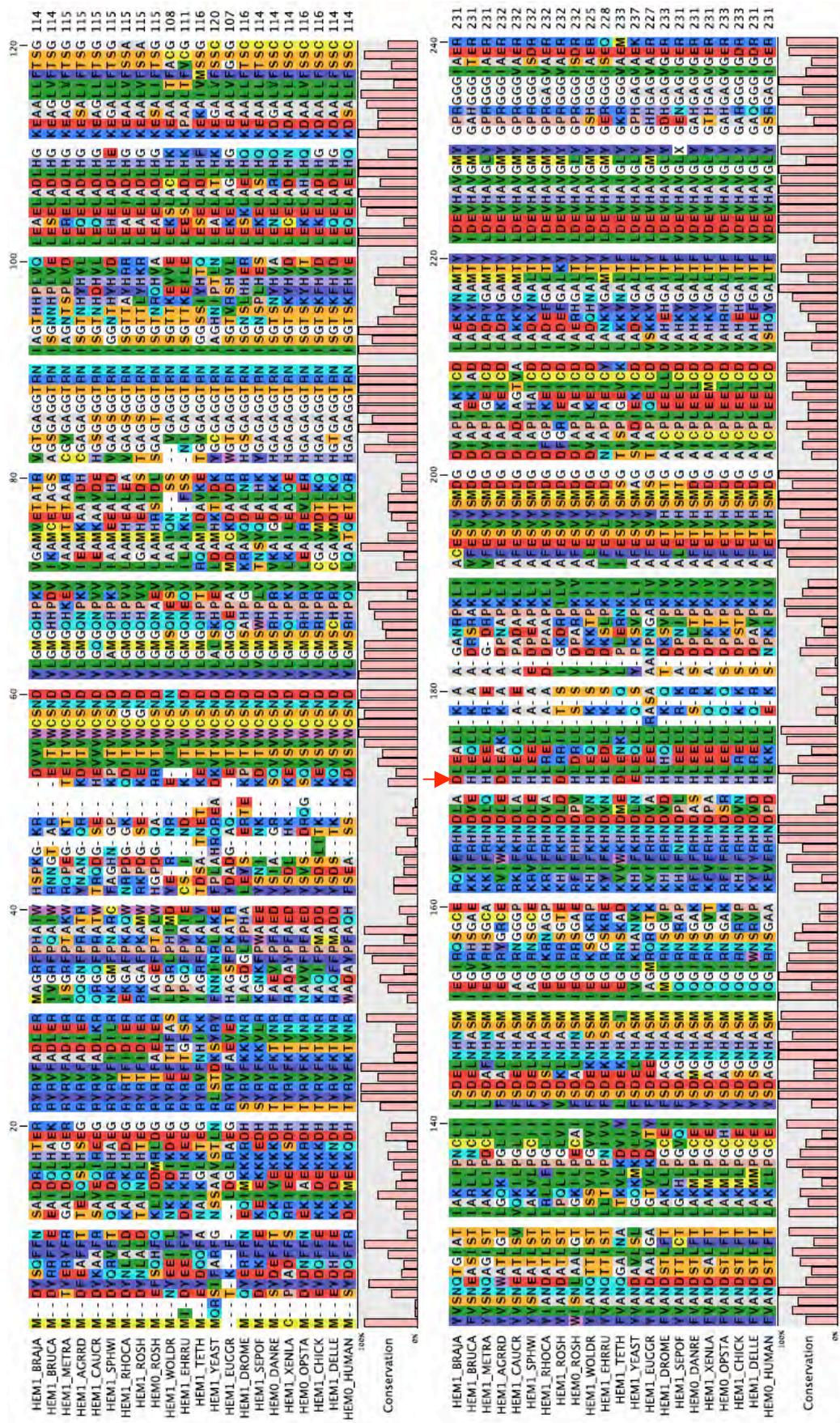
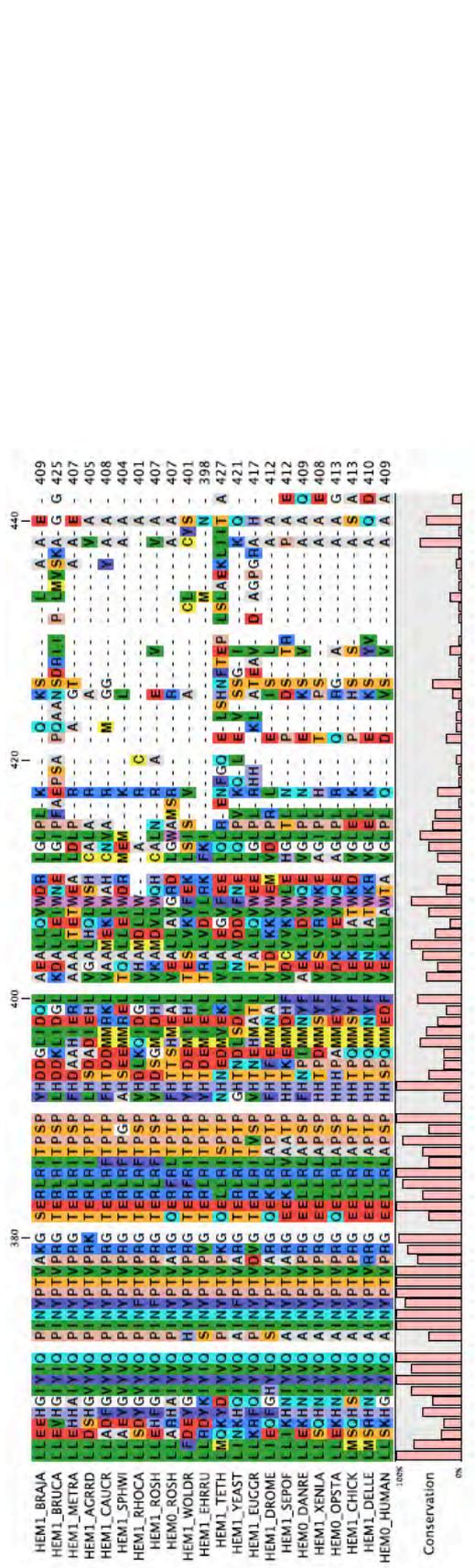
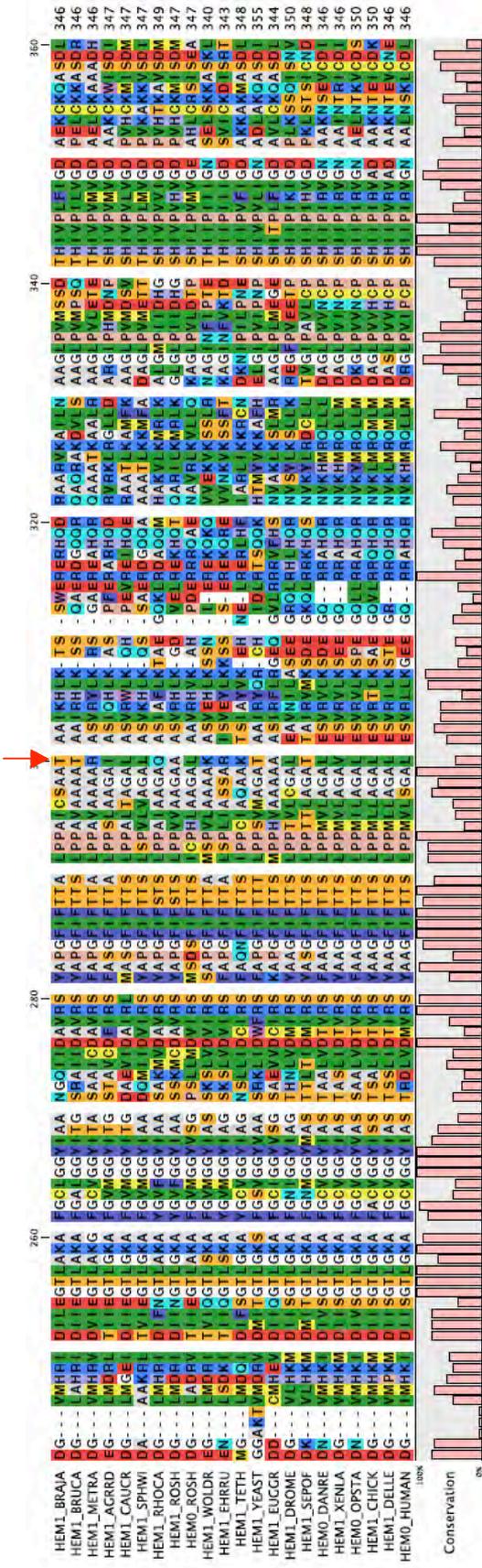


Figura 13. Comparación de secuencias ortólogas de HemA. El alineamiento muestra 11 secuencias provenientes de especies de la rama α de proteobacterias; asimismo, se incluyen 11 especies representativas de la diversidad de linajes eucariotes. BRAJA, *Bradyrhizobium japonicum*; BRUCA, *Brucella canis*; METRA, *Methylobacterium radiotolerans*; AGRRD, *Agrobacterium radiobacter*; CAUCR, *Caulobacter crescentis*; SPHWJ, *Sphingomonas wittichii*; RHOCA, *Rhodobacter capsulatus*; RHOSH, *Rhodobacter sphaeroides*; WOLDR, *Wolbachia* (endosimbionte de *Drosophila melanogaster*); EHRRU, *Ehrlichia ruminantium*; TETH, *Tetrahymena thermophila*; YEAST, *Saccharomyces cerevisiae*; EUGGR, *Euglena gracilis*; DROME, *D. melanogaster*; SPOF, *Sepia officinalis*; DANRE, *Danio rerio*; XENLA, *Xenopus laevis*; OPSTA, *Opsanus tau*; CHICK, *Gallus gallus*; DELLE, *Delphinapterus leucas*; HUMAN, *Homo sapiens*. (Continúa en la página siguiente.)



(Continuación.) En *R. sphaerooides*, Hem1 y 0 representan distintas copias de la enzima. Las flechas señalan las posiciones 165 y 288 del ortólogo HemA de *B. japonicum*- En eucariotes, Hem1 representa la forma mitocondrial inespecífica, mientras que Hem0 es una isoenzima específica eritroide.

Posiblemente, los residuos en las posiciones de interés reflejan necesidades particulares (para la secuencia de *B. japonicum*), más que universales (para el conjunto de ortólogos). La evolución de una actividad promiscua se ha asociado a la variación de posiciones alejadas del sitio de unión de sustratos y catálisis. Por esto parece paradójica su implicación en una innovación enzimática. Con el fin de ofrecer una explicación razonable a la evolución de HemA^{mut1}, es necesario cuestionar si la conservación refleja más que necesidades para los parámetros intrínsecos de estabilidad estructural y catálisis. Asimismo, es necesario considerar a la enzima como una suma de interacciones dinámicas, no como un cuerpo rígido.

Pese a la divergencia de los linajes que los albergan, los ortólogos de HemA han permanecido esencialmente inalterados en amplias regiones de secuencia. Esto podría ser resultado de constreñimientos sobre la variación. Las mutaciones provocan fluctuaciones en los parámetros intrínsecos de la enzima, y algunas fluctuaciones tienen consecuencias inadmisibles. Pero los constreñimientos no sólo son imposiciones intrínsecas. Si lo fuesen, la conservación entre parálogos con actividades comparables sería tan significativa como entre ortólogos. Las enzimas parálogas BioF, HemA y Kbl catalizan reacciones muy parecidas (Fig. 9a). Sin embargo, sus secuencias comparten identidades de ~30 % (Fig. 9b). Esta conservación es menor que la observada entre grupos de ortólogos de cualquiera de las enzimas. Por ejemplo, los ortólogos HemA provenientes de humano y *X. laevis* muestran ~ 49 % de identidad respecto del de *B. japonicum*.

Los constreñimientos también emergen de forma extrínseca, en parte porque la enzima participa en una red de interacciones. La alteración de los parámetros intrínsecos significa la perturbación de un flujo metabólico. Este flujo implica el direccionamiento de intermediarios dentro de una ruta, pero también entre regiones del metabolismo. Es decir, ciertos fenotipos son inaccesibles por las leyes físicas del plegamiento y la catálisis, pero también por el contexto biológico en que ocurren. Dentro de la célula, la enzima interactúa con proteínas, otras componentes celulares e inclusive agentes ambientales. Un número de estas interacciones podrían participar en mecanismos de regulación. El mapeo de interacciones a escala proteómica da evidencia en favor de esta suposición (Collins & Chouhary, 2008; Goll & Uetz, 2006). El uso de metodologías de purificación por afinidad en tándem y de espectrometría de masas ha permitido la identificación de cientos de complejos proteicos en *S. cerevisiae* (Gavin *et al.*, 2006; Krogan *et al.*, 2006). HemA aparece asociada con proteínas diversas, lo que sugiere un papel en la comunicación entre procesos celulares diferentes (Tabla 2). Entonces, el número y exactitud de las interacciones podrían determinar una diversidad funcional adicional a la actividad catalítica. Esta diversidad funcional podría restringir las alternativas de cambio en posiciones (y regiones) de la estructura enzimática.

El escenario anterior permite considerar la importancia de las posiciones 165 y 288. Parece posible que la variación de estas posiciones provoque alteraciones menores en la funcionalidad. Inclusive es predecible que la actividad principal de HemA^{mut1} haya permanecido inalterada. Sin embargo, dicha variación

puede influir en los parámetros de la actividad enzimática. Las proteínas son ensamblajes dinámicos donde coinciden desplazamientos atómicos con amplitudes y escalas temporales diversas. La perturbación de los subestados conformacionales de la estructura podría traducirse en la elevación discreta de una actividad promiscua. El cambio D165V podría significar fluctuaciones conformacionales en las paredes del sitio de unión de sustratos. Asimismo, el cambio T288A podría incidir en la transmisión vectorial de vibraciones hasta el sitio activo. Tales perturbaciones podrían facilitar el acceso o el recambio de sustratos alternativos. El efecto de las mutaciones se reconsiderará más adelante. El planteamiento importante es que las posiciones 165 y 288 tienen un carácter de plasticidad: Probablemente albergan un potencial evolutivo para modular actividades diferentes a la principal, o sobre la que normalmente actúa la selección.

Proteína	Función celular	Referencia
PSA1	GDP-manosa-1-fosfato guaniltransferasa; actividad necesaria para la síntesis de pared celular.	Gavin <i>et al.</i> , 2006.
HSP60	Chaperonina; media el replegamiento proteínico en respuesta al estrés por calor; necesaria para prevenir la agregación de proteínas; papel en la transmisión del mtDNA; similar a GroEL.	
Pro3	δ -1-pirrolina-5-carboxilato reductasa; catálisis del último paso de la biosíntesis de prolina.	Krogen <i>et al.</i> , 2006.
GSY2	Glucógeno sintasa; expresión inducida ante la limitación de glucosa, la inanición por nitrógeno, el estrés de calor y la fase estacionaria.	
RFA1	Subunidad del factor de replicación RF-A; papel en la replicación, reparación y recombinación del DNA.	
BDF1	Papel en la iniciación de la transcripción en promotores con secuencias TATA; se asocia con el factor de transcripción basal TFIIID.	
GCD11	Subunidad y del factor de traducción eIF2; papel en la identificación del codón de inicio.	
RRP5	Proteína requerida para la síntesis de los rRNAs 18 S y 5.8 S.	

Tabla 2. Proteínas purificadas en complejo junto con HemA en *Saccharomyces cerevisiae*.

Se mencionó que cierto nivel de promiscuidad catalítica es inevitable. La variación en posiciones de plasticidad podría suponer la fluctuación de reacciones promiscuas entre niveles que no compiten con el nivel de la actividad principal. Mutaciones discretas podrían dar relevancia fisiológica a una actividad promiscua. Obviamente, esto depende de que dicha actividad facilite la supervivencia celular. En otras palabras, la importancia de las posiciones de plasticidad es la facilitación de la innovación evolutiva. Pero para que este

potencial adquiriera relevancia es indispensable una presión de selección. Asimismo, se necesita un contexto y una oportunidad: Contexto, porque las posiciones de plasticidad son elementos de secuencia cuya variación no compromete críticamente la estabilidad o la actividad principal. Oportunidad, porque las actividades promiscuas asociadas a la variación sólo quedan atrapadas por la selección en un momento y ambiente apropiados.

Si la generación de una actividad promiscua seleccionable dependió de mutaciones en posiciones de plasticidad, la exploración de las posibilidades de cambio en estas posiciones podría conducir a una variante HemA^{mut} más eficiente. Esta exploración necesita que combinaciones de mutaciones incidan en los nucleótidos de los codones correspondientes a las posiciones 165 y 288. Dado este requerimiento, la exploración de la variabilidad por PCR mutagénica hubiese sido inasequible. Se eligió una aproximación de mutagénesis dirigida a sitios específicos.

Exploración de la plasticidad de las posiciones de HemA^{mut1}

La aproximación consistió en generar variantes enzimáticas con una distribución aleatoria de residuos en las posiciones de interés. Se construyó una librería de alelos *hemA*^{mut} que exhiben G o C en el sitio 3, ó cualquier nucleótido en los sitios 1 y 2, de los codones 165 y 288. A partir de una librería con $\sim 1.0 \times 10^6$ secuencias se seleccionaron alelos que revierten la auxotrofia de la biotina. Células $\Delta bioF$ transformadas con la librería se lavaron y se sembraron en medio con trazas de biotina. Se seleccionaron 19 colonias al término de 9 días de incubación. Los plásmidos se aislaron y luego se reintrodujeron en la cepa $\Delta bioF$. El fenotipo conferido por los alelos seleccionados se comparó con el asociable a pUC18, *hemA* y *hemA*^{mut1} (Figs. 14 y 15). El fenotipo de complementación fue reproducible en diecisiete casos. Sin embargo, ningún alelo confirió un fenotipo más robusto que el asociado a *hemA*^{mut1}.

La tabla 3 muestra los cambios observados en la estructura primaria de las variantes. Se encontraron distintos residuos en las posiciones 165 y 288. Esto apoya la hipótesis del hallazgo de posiciones de plasticidad. La selección *in vivo* permitió segregar la variación de cada posición. Cambios individuales por residuos Cys, His o Arg se encontraron en la posición 165. El cambio por un residuo Ser apareció por separado en la posición 288. Estas observaciones sugieren que ambas posiciones modulan la actividad promiscua deseada. No obstante, las opciones de variación parecen diferir entre las posiciones. Llama la atención que se hayan encontrado más cambios para la posición 165, pues ésta es la que exhibe una mayor conservación entre los ortólogos. Esta posición podría tener mayor efecto en la actividad. Esto parece posible, dada su cercanía respecto del sitio de unión de substratos. Por otro lado, parece que sólo algunos residuos en la posición provocan fluctuaciones en la actividad deseada. Lo anterior lleva a considerar el papel de la

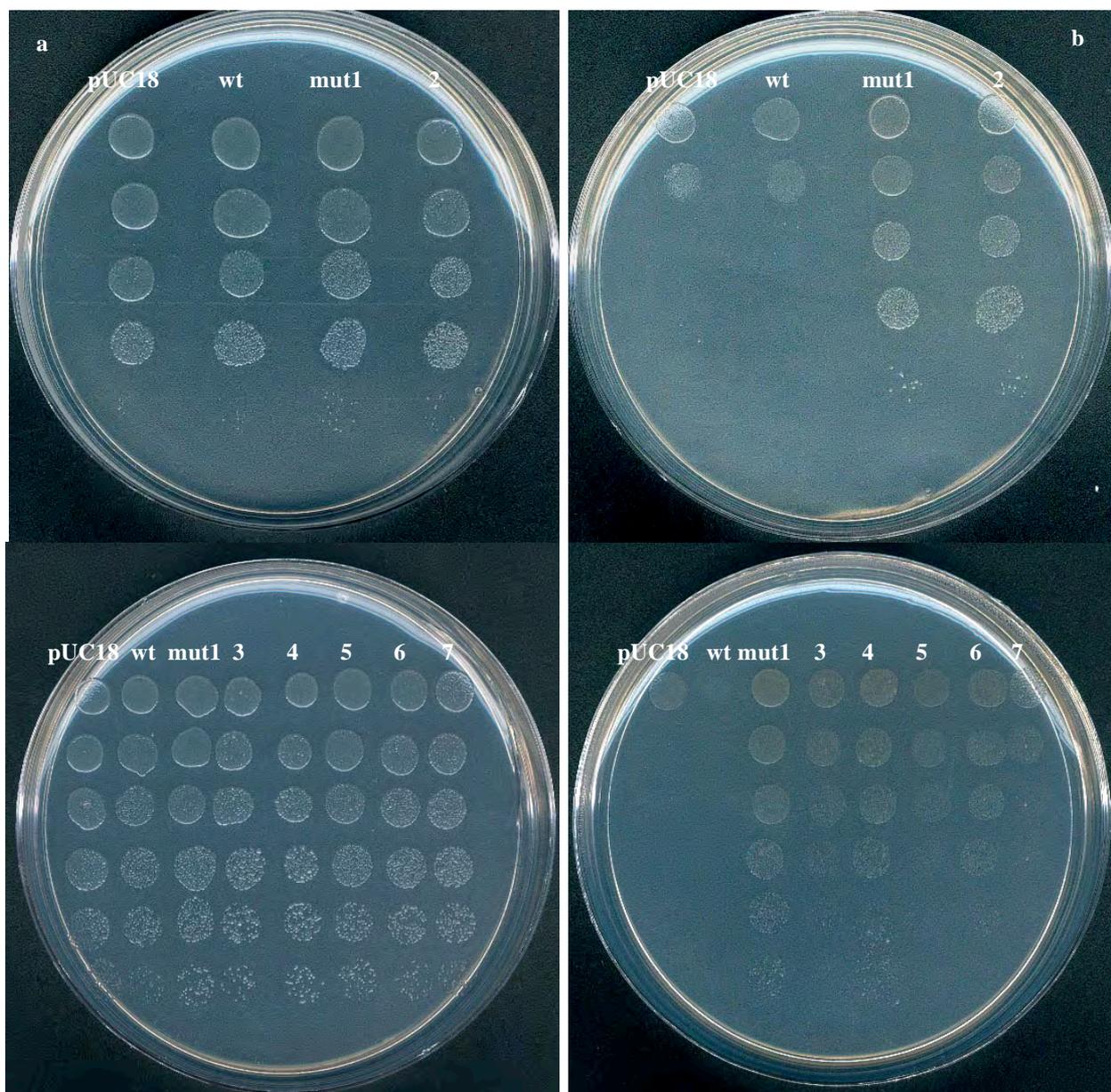


Figura 14. Selección de alelos *hema*^{mut} generados por mutagénesis sitioespecífica. Células $\Delta bioF$ se transformaron, se lavaron y se sembraron en medio complementado con biotina en exceso (panel **a**) ó en trazas (panel **b**). De izquierda a derecha, fenotipos asociables al vector (**pUC18**), *hema* (**wt**), *hema*^{mut1} (**mut1**), *hema*^{mut2} (**2**), *hema*^{mut3} (**3**), *hema*^{mut4} (**4**), *hema*^{mut5} (**5**), *hema*^{mut6} (**6**), y *hema*^{mut7} (**7**). En el panel **b** se observa el crecimiento tras 9 días de incubación. (Resultados de un experimento único.)

epistasia en la evolución de las variantes Hema^{mut}. Posiblemente, la mayoría de las mutaciones de la posición 288 producen actividades seleccionables sólo si en la 165 hay ciertos cambios. Las mutaciones de Hema^{mut1} podrían ser un caso de sinergia positiva. Curiosamente, los cambios D165V y T288A no fueron encontrados. Además, la substitución simultánea de las posiciones se observó sólo en dos casos. En el primero (Hema^{mut2}),

los residuos S165 y K288 fueron acompañados por un cambio adicional (D326G). En el segundo ($HemA^{mut22}$), los residuos Q165 y R288 resultaron irrelevantes para conseguir la complementación genética.

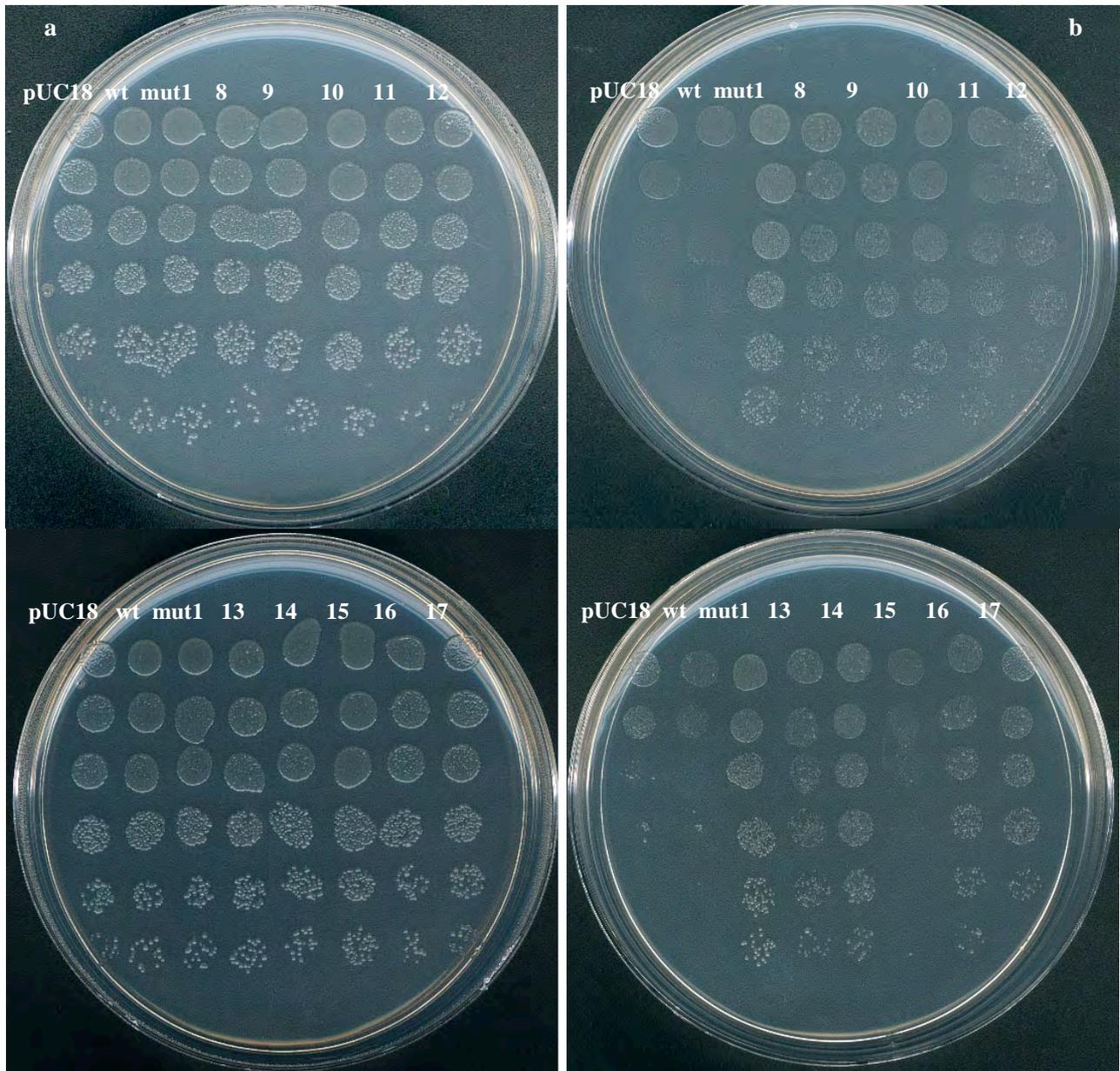


Figura 15. Selección de alelos $hemA^{mut}$ generados por mutagénesis dirigida. Células $\Delta bioF$ se transformaron, se lavaron y se sembraron en medio complementado con biotina en exceso (panel **a**) ó en trazas (panel **b**). De izquierda a derecha, el fenotipo asociable al vector (**pUC18**), $hemA$ (**wt**), $hemA^{mut1}$ (**mut1**), $hemA^{mut8}$ (**8**), $hemA^{mut9}$ (**9**), $hemA^{mut10}$ (**10**), $hemA^{mut11}$ (**11**), $hemA^{mut12}$ (**12**), $hemA^{mut13}$ (**13**), $hemA^{mut14}$ (**14**), $hemA^{mut15}$ (**15**), $hemA^{mut16}$ (**16**) y $hemA^{mut17}$ (**17**). En el panel **b** se observa el crecimiento tras 9 días de incubación. (Resultados de un experimento único.)

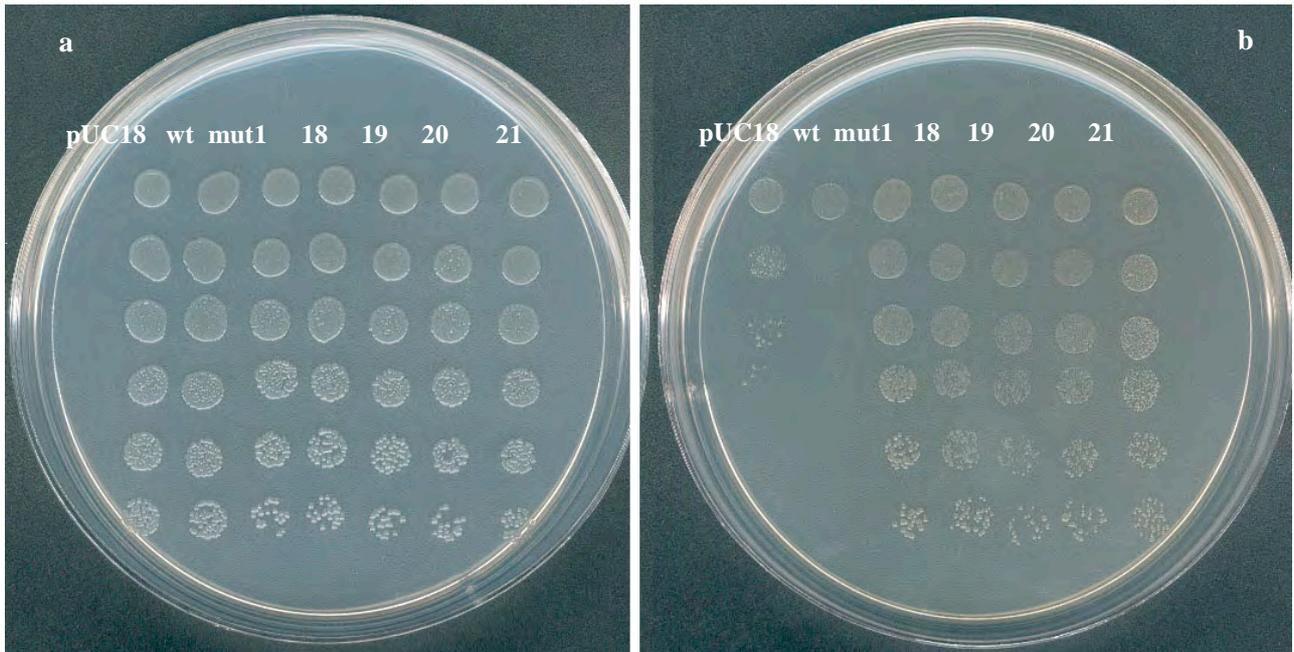


Figura 15 (continuación). Selección de alelos *hemA^{mut}* generados por mutagénesis sitioespecífica. Células $\Delta bioF$ se transformaron, se lavaron y se sembraron en medio complementado con biotina en exceso (panel a) ó en trazas (panel b). De izquierda a derecha el fenotipo asociable al vector (pUC18), *hemA* (wt), *hemA^{mut1}* (mut1), *hemA^{mut18}* (18), *hemA^{mut19}* (19), *hemA^{mut20}* (20) y *hemA^{mut121}* (21). En el panel b se observa el crecimiento tras 9 días de incubación. (Resultados de un experimento único.)

Variante	Posición																	Fenotipo	
	22	30	62	161	165	168	171	174	176	200	223	231	241	262	280	288	326		381
Wt	V	A	P	N	D	A	K	A	N	C	P	R	L	I	P	T	D	G	-
1					V											A			+
2					S											K	G		+
3																		S	+
4					C														+
5																		S	+
6					H														+
7											C	H							+
8					T										R				+
9																			+
10																		S	+
11	M	V			N		E							V					+
12					T														+
13			L																+
14	M																	S	+
15									D	S									-
16																		S	+
17											Q								+
18				S	S														+
19					G			(Z)											+
20					L	V													+
21					R														+
22					Q											R			-

Tabla 3. Distribución de mutaciones en las variantes Hema^{mut}

Se encontraron variantes con cambios en posiciones distintas a las esperadas. Ciertos cambios ocurrieron de manera individual, en secuencias que de otra manera serían silvestres: P62L (HemA^{mut13}), G381S (HemA^{mut3, 5 y 10}) y P223Q (HemA^{mut17}). Un cambio interesante (V22M) se encontró en secuencias que difieren por la variación de las posiciones 165 y 288 (HemA^{mut11}, HemA^{mut14}). Los hallazgos sugieren la existencia de múltiples caminos evolutivos que convergen en una actividad promiscua biológicamente relevante. Sin embargo, es necesario discernir entre los cambios que resultaron de un proceso de selección de los que se fijaron al azar. El mapeo de los cambios en el modelo *in silico* indica que las posiciones afectadas se distribuyen principalmente en la interfase y la superficie de la enzima. Resulta intrigante que sólo se haya encontrado un cambio en la zona estructural intermedia. Las mutaciones en la interfase son interesantes, dado que coinciden con el espacio donde se aloja el sustrato succinil-CoA.

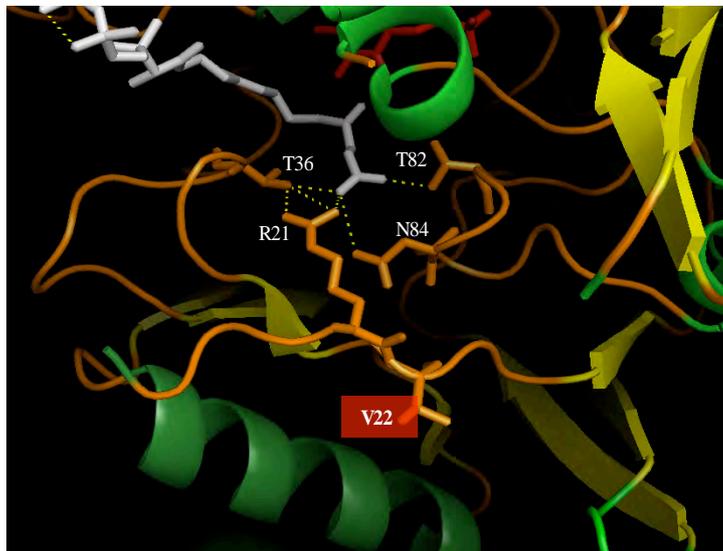


Figura 16. Modelo del canal de acceso al bolsillo de unión del sustrato succinil-CoA. La mutación de la posición V22 podría alterar la capacidad del residuo aledaño (R21) para establecer un puente de hidrógeno importante para el posicionamiento de la porción succinato del sustrato sCoA.

Los cambios V22M, A30V, P62L y P280R se sitúan en la interfase, cerca de residuos clave para la estabilidad estructural y la unión del sustrato succinil-CoA. La estructura de HemA consiste en dos monómeros entrelazados íntimamente (Astner *et al.*, 2005). El sitio de unión a sustratos está delimitado por los dominios catalíticos de los dos monómeros y por el C-terminal del monómero que une al PLP. Los sustratos ingresan por un canal que da alojamiento a la porción succinato del succinil-CoA. El cambio V22M se sitúa cerca de la entrada del canal, junto al residuo conservado R21 (Fig. 16). La evidencia estructural indica que R21 establece un puente de sal con la función carboxilo del succinato; se trata de un residuo necesario para el posicionamiento del sustrato succinil-CoA (Astner *et al.*, 2005).

Siguiendo por el canal se encuentra el asa $\alpha 3-\alpha 4$, que incluye una secuencia rica en residuos Gly (Fig. 16). Dentro de esta secuencia están residuos que comparten el papel de R21 en el posicionamiento del sustrato: T82, N84 e I85. En posición opuesta están Q60 y G31; de acuerdo con la evidencia estructural, estos residuos comparten un puente de hidrógeno relevante para la interacción de los dominios catalítico y N-terminal (Astner *et al.*, 2005). La figura 17 muestra otro puente de hidrógeno relevante entre R28 y D54. Los cambios A30V, P62L y P280R se sitúan en la inmediación del asa $\alpha 3-\alpha 4$ y de las interacciones mencionadas. Evidentemente, estas mutaciones inciden en una región estructural fácilmente perturbable.

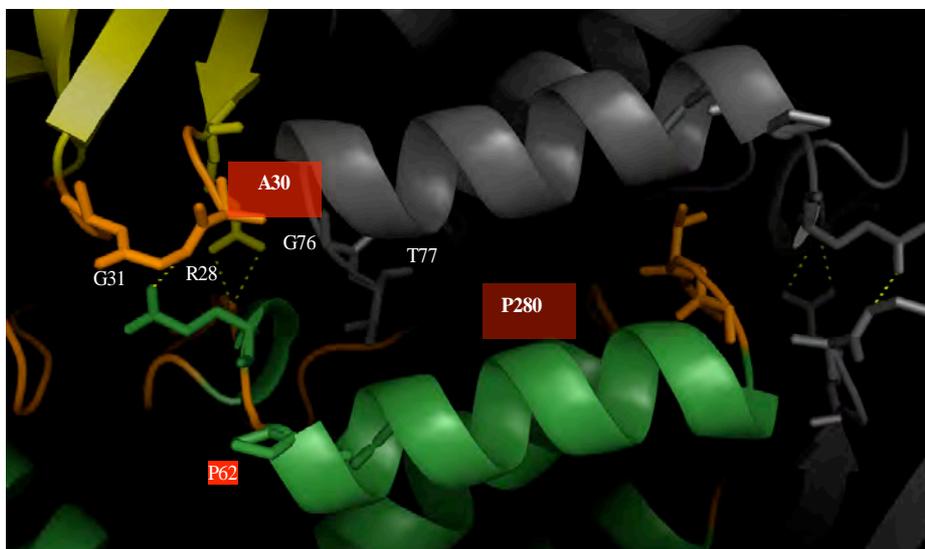


Figura 17. Modelo de la interfase de los monómeros de HemA. La mutación de los residuos V22, A30, P62 y P280 afecta una región estructural fácilmente perturbable. En la inmediación de dichas posiciones se encuentran interacciones y residuos clave para la estabilidad estructural y la especificidad de sustrato.

Los cambios D165V y N161S se sitúan en elementos de estructura contiguos al sitio de unión del sustrato succinil-CoA. En el modelo estructural de HemA, la hélice $\alpha 6$ y la lámina $\beta 6$ se exponen al disolvente. Las interacciones originadas desde el asa $\alpha 6-\beta 6$ podrían restar libertad conformacional a dichos segmentos estructurales (Fig. 18). De manera importante, la lámina $\beta 6$ alberga al residuo V157. Este residuo forma la base del sitio hidrofóbico en el que se posiciona la función pirimidina del CoA. El cambio D165V podría alterar la libertad conformacional de la lámina $\beta 6$. Se propone que esta alteración se debe a la eliminación de una interacción comparable a las observadas en el asa $\alpha 6-\beta 6$ de la estructura de ortólogo de *R. capsulatus*. En el modelo de *B. japonicum*, la separación entre D165 y R159 permite la predicción de un

puede de sal. Dicha interacción no sería establecida por un residuo Val, ni por el resto de los residuos encontrados en la posición 165. De acuerdo con esta especulación, ninguna variante exhibió el cambio D165Q, que podría significar la conservación de las propiedades físicoquímicas y la posibilidad de alguna interacción de atracción. La mutagénesis experimental de la posición R159 daría evidencia de la importancia de una interacción para la evolución de la actividad promiscua seleccionada.

El cambio N161S se sitúa en el asa $\alpha 6$ - $\beta 6$. De nuevo, un cambio podría alterar de manera indirecta la colocación del succinil-CoA. La estructura de HemA de *R. capsulatus* predice la existencia de puentes de hidrógeno D137-O-N6, S136-O-N1 y S138-O-N6, donde N1 y N6 pertenecen a la porción adenina del CoA (Astner *et al.*, 2005). De manera interesante, la evidencia estructural también indica una interacción directa entre D137 y N161S.

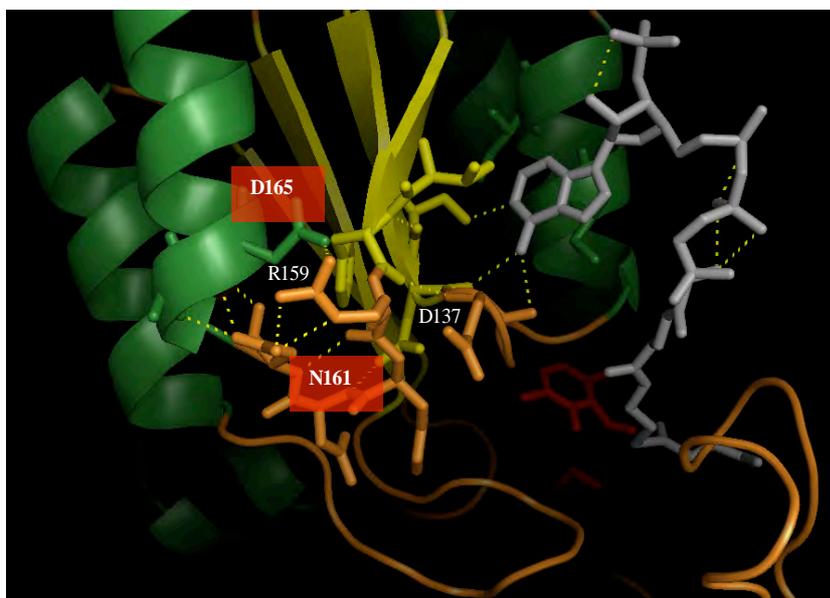


Figura 18. Posible efecto de la mutación de los residuos D165 y N161. El asa $\alpha 6$ - $\beta 6$ muestra interacciones cuya eliminación podría asociarse a un cambio en la libertad conformacional de la lamina $\beta 6$, la cual exhibe contribuye al posicionamiento del sustrato sCoA.

L241H es el único cambio en la región intermedia de la estructura. El cambio L241H no afecta de manera directa a la interfase. Sin embargo, se sitúa en la misma lámina que el residuo T245 ($\beta 9$). En la estructura de *R. capsulatus*, T245 establece uno de 6 puentes de hidrógeno que estabilizan al grupo fosfato del PLP (Astner *et al.*, 2005). Los residuos están separados y muestran orientaciones opuestas. Sin embargo, el cambio por un residuo His podría tener implicaciones para el empaquetamiento de la lámina.

Cambios de superficie distantes respecto del lugar de la catálisis. Los cambios A168V y K171E se situán en la hélice α_6 , al igual que D165V. Los cambios D326G y G381S están tan alejados del lugar de la catálisis como la posición T288. El cambio G381S es interesante, pues se encontró individualmente en tres ocasiones. Más aún, este cambio es suficiente para producir una variante capaz de revertir la auxotrofia de la biotina. El ensayo de complementación genética da evidencia de la importancia de los cambios P62L y G381S: Ambos se asocian por separado al fenotipo de crecimiento. El cambio V22M parece importante por su situación respecto del residuo R21, pero también por su ocurrencia en las variantes HemA^{mut11} y HemA^{mut14}. En estas variantes, las posiciones 165 y 288 mostraron cambios que podrían ser conservativos por razones de físicoquímica y de su presencia entre los ortólogos de HemA. Curiosamente, el cambio V22M podría tener un papel análogo al de una mutación caracterizada en el ortólogo de humanos. La mutación del residuo F165, equivalente a F23 en la enzima de *B. japonicum*, se ha asociado previamente a la anemia sideroblástica. Según la información cristalográfica disponible, la mutación de la posición F23 desestabiliza la unión del succinil-CoA, así como la interacción entre los dominios N- y C-terminales (Astner *et al.*, 2005). Por otro lado, el cambio A30V se encontró en el mismo contexto de secuencia que V22M. La mutación de la posición equivalente en el ortólogo humano (A172) también se asocia a la enfermedad. En este caso, se supone que la mutación afecta indirectamente la unión del PLP, pero también la estabilidad estructural.

La literatura sugiere que los cambios en la interfase alteran negativamente la estabilidad y la afinidad por el substrato succinil-CoA. Esta información no parece extrapolable a un papel en la evolución de una actividad novedosa. Pero este papel podría ser contextual en alguno de dos sentidos. En el primero, las mutaciones tienen efectos comparables entre los ortólogos; sin embargo, el efecto a nivel metabólico difiere en los sistemas biológicos de referencia. En humanos, la mutación tendría consecuencias negativas por la desregulación de la producción del grupo hemo. En *E. coli*, la síntesis de los tetrapirroles quedaría esencialmente inalterada, dado que HemA representa una actividad redundante. En el segundo sentido, las mutaciones también desestabilizan a la estructura y la afinidad por el substrato, pero su efecto global está condicionado por la historia evolutiva de cada variante enzimática. En las variantes HemA^{mut11} y HemA^{mut14}, el efecto de las mutaciones podría ser de tipo sinérgico. Es decir, algunos de los residuos en 165 y 288 no producirían un fenotipo seleccionable en una situación experimental. Los cambios V22M y A30V podrían contribuir a que el crecimiento fuera sensible, aunque por separado podrían ser deletéreos.

Es predecible que las mutaciones en la interfase o en la superficie tengan un efecto levemente negativo en la estructura. Su correlación con la elevación de una actividad promiscua es explicable si se considera que la actividad enzimática es resultado de un balance entre la estabilidad y la flexibilidad de la estructura. Se ha observado que las fluctuaciones estructurales hacen contribuciones directas a la catálisis, la unión de substratos y la regulación alostérica. Las formas en que la actividad se correlaciona con

fluctuaciones conformacionales sugiere que la alteración de estas últimas podrían generar incrementos mínimos, pero biológicamente relevantes, de las actividades promiscuas.

Las estructuras proteínicas fluctúan alrededor de una posición de equilibrio. Los movimientos estructurales incluyen la rotación de enlaces y la vibración de cadenas laterales, pero también desplazamientos concertados de elementos secundarios y de dominios. La dinámica estructural de las enzimas se ha detectado con técnicas espectrométricas de Resonancia Magnética Nuclear (NMR) y de Transferencia de Energía por Resonancia de Fluorescencia (FRET). Por mucho tiempo se ha creído que los movimientos estructurales de mayor escala tienen importancia para la catálisis. En la actualidad hay evidencia de movimientos inducidos por la unión de los sustratos. En el sitio activo de la fosfoglicerato cinasa, por ejemplo, el movimiento de la región de bisagra acerca elementos secundarios para interactuar con el sustrato (White *et al.*, 2008). Otras evidencias sugieren la coordinación entre los movimientos estructurales y la trayectoria de la reacción. En la ciclofilina y la adenilato cinasa, la dinámica conformacional resulta indistinguible respecto del recambio catalítico (Eisenmesser *et al.*, 2002; Eisenmesser *et al.*, 2005; Henzler-Wildman *et al.*, 2007a). Estos trabajos sugieren que los movimientos observados durante la catálisis ya están presentes en la enzima libre. Asimismo, indican que movimientos de menor escala, como las vibraciones de cadenas laterales únicas o distantes del sitio activo, también coinciden con la catálisis. En resumen, fluctuaciones estructurales intrínsecas, con diferentes amplitudes y escalas temporales, se acoplan con la actividad enzimática.

Los elementos de estructura en la interfase de HemA tienen una distribución espacial compleja. Por razones evolutivas, esta distribución podría estar optimizada para la unión y el recambio de los sustratos principales. Las mutaciones en esta región generan defectos de empaquetamiento, o perturbaciones en la libertad conformacional de segmentos estructurales. Algunos segmentos albergan cadenas laterales determinantes de la especificidad de sustrato. Los cambios en la movilidad y orientación relativas de las determinantes podrían provocar un aumento de la actividad promiscua. Esto podría ser por la disminución de las restricciones espaciales para el acceso de sustratos alternativos. También podría deberse a la disminución de la afinidad por el sustrato principal. En cualquier caso ocurriría una relajación de la especificidad enzimática: Sustratos geoméricamente parecidos tendrían oportunidad de participar en el recambio catalítico, aunque con un posicionamiento no óptimo. Asimismo, un aumento en la libertad conformacional podría significar la capacidad de las determinantes de especificidad para establecer una interacción productiva con los sustratos alternativos. En este caso, el aumento de la actividad promiscua sería explicable como una ampliación de la especificidad enzimática.

Los cambios en la superficie podrían tener un efecto más dinámico en el proceso catalítico. Este efecto correspondería a la transmisión vectorial de movimientos vibratorios hasta el sitio activo. Se ha observado que ciertas mutaciones de superficie mejoran la catálisis de la fosforilasa de nucleósidos purínicos

(Saen-Oon *et al.*, 2008). Los análisis cristalográficos de las enzimas silvestre y mutante no muestran diferencias conformacionales significativas. No obstante, simulaciones *in silico* indican que en la mutante tiene un mayor acoplamiento entre la promoción de vibraciones y la coordenada de reacción.

Parece posible que las mutaciones de superficie de las variantes HemA^{mut} favorezcan fluctuaciones estructurales importantes para el recambio catalítico. De algún modo, estas mutaciones de superficie implican la contribución de una fluctuación local y rápida a un desplazamiento estructural amplio e infrecuente (Henzler-Wildman *et al.*, 2007b). Esta jerarquía espacio-temporal de fluctuaciones, que facilitan movimientos conformacionales en la escala del recambio catalítico, se ha conjeturado con base en análisis de la fosfoglicerato cinasa de *E. coli* y *Aquifex aeolicus* (Henzler-Wildman *et al.*, 2007b). La evolución de las variantes HemA^{mut} puede explicarse por alteraciones de la estabilidad estructural. Lo anterior sugiere la ampliación o pérdida de la especificidad enzimática como fundamento del incremento de una actividad promiscua seleccionable. Como resultado de esta desespecialización, las variantes HemA^{mut} podrían exhibir un repertorio de actividades secundarias capaces de adquirir relevancia fisiológica de manera contingente.

HemA^{mut1} revierte la auxotrofía asociada a la delección de un segundo gen de la biosíntesis de biotina

Se supuso que HemA^{mut1} podría exhibir más de una actividad promiscua relevante para la biosíntesis de biotina. La validez de esta suposición se determinó con cepas mutantes $\Delta bioF$, $\Delta bioA$ y $\Delta bioA\Delta bioF$ (Fig. 19). Se sembraron diluciones seriales en medio con biotina en exceso ó en trazas. Los casaminoácidos fueron excluidos del medio de crecimiento. En la condición control, números similares de UFCs se observaron en puntos de crecimiento equivalentes. No obstante, las células mostraron diferencias fenotípicas que sugieren la disminución de la viabilidad por la ausencia de casaminoácidos.

Ante la limitación de vitamina, *hemA^{mut1}* confirió capacidad de crecimiento a las células $\Delta bioA$. Este fenotipo fue similar al mostrado por las células $\Delta bioF$. De manera inesperada, la presencia de *hemA^{mut1}* también se correlacionó con el crecimiento de la cepa $\Delta bioA\Delta bioF$. Aunque comparativamente menor, el crecimiento fue evidente hasta en los puntos correspondientes a las diluciones de menor concentración celular. Las actividades promiscuas evolucionadas suelen ser ineficientes bajo un criterio bioquímico. Sin embargo, favorecen la supervivencia bajo una presión de selección. La evolución *in vitro* de HemA^{mut1} resultó en la generación de actividades promiscuas capaces de sostener la viabilidad celular.

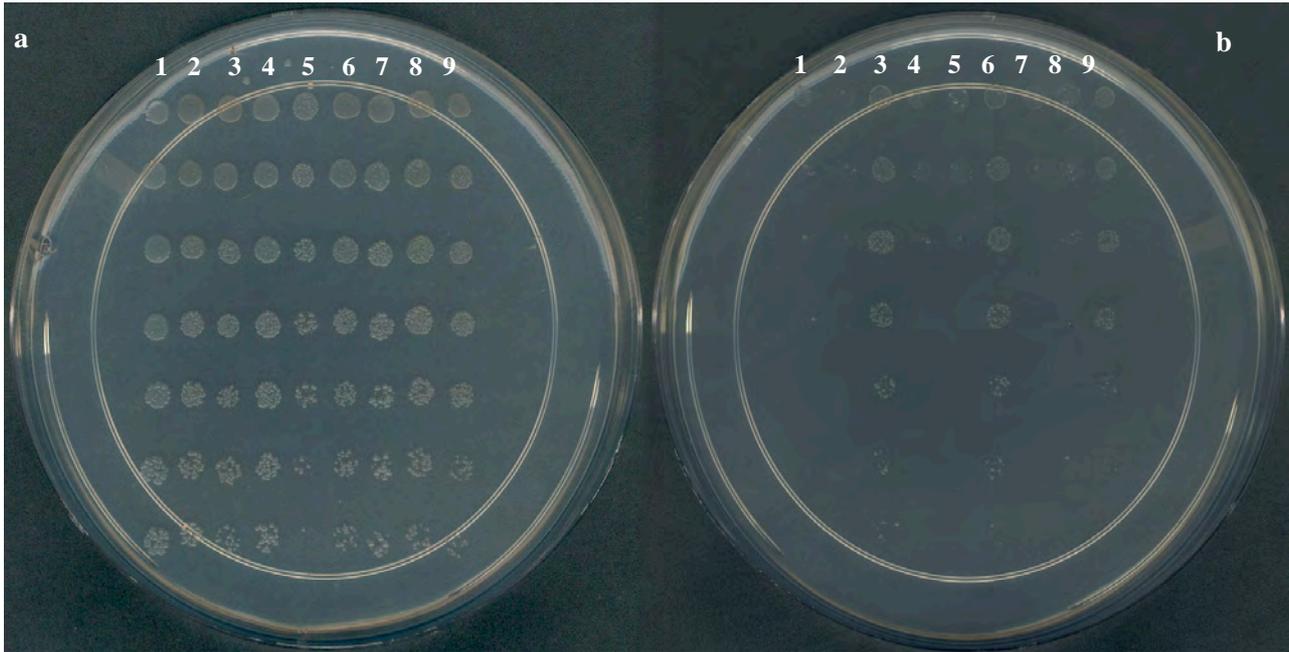


Figura 19. Complementación por el alelo *hemA*^{mut1} en los fondos de delección $\Delta bioF$, $\Delta bioA$ y $\Delta bioA\Delta bioF$. Células $\Delta bioF$ (carriles 1-3), $\Delta bioA$ (carriles 4-6) y $\Delta bioA\Delta bioF$ (carriles 7-9) transformadas con pUC18 (carriles 1, 4 y 7), pUC18/*hemA* (carriles 2, 5 y 8) o pUC18/*hemA*^{mut1} (carriles 3, 6 y 9) fueron lavadas y sembradas en medio complementado con biotina en exceso (panel a) ó en trazas (panel b). En (b), la incubación fue de 12 días. (Resultados representativos de tres experimentos independientes.)

En apariencia, la ausencia del suministro de oligopéptidos y aminoácidos libres comprometen la capacidad de crecimiento. Este compromiso puede observarse inclusive en presencia de un exceso de biotina. Lo anterior sugiere un estrés metabólico. Si hay tal estrés, es significativo que la presencia de *hemA*^{mut1} pueda restablecer el crecimiento de una fracción poblacional. En un ambiente estricto, definido por la ausencia de un conjunto de intermediarios clave como un cofactor y los aminoácidos, la presencia de actividades promiscuas levemente incrementadas es una ventaja para la supervivencia. Esta ventaja estaría representada por su capacidad para restablecer la biosíntesis de un compuesto esencial, pero también por su influencia indirecta en la respuesta metabólica a la ausencia de otros componentes. El hecho de que se trata de reacciones acopladas es interesante por razones evolutivas. Significativamente, estas actividades contribuyen en un mismo proceso metabólico. Una conclusión importante es que actividades promiscuas relacionadas pueden constituir eslabones de una cadena de reacciones que conducen a un compuesto esencial. Esta conjetura tiene implicaciones importantes para el modelo de evolución por reclutamiento de Jensen.

Otras variantes exhiben un repertorio de actividades promiscuas

Se reclonaron cinco alelos mutantes para realizar experimentos de complementación en los tres fondos de mutación (Figs. 20 a 22). Los criterios de elección de los alelos fueron los siguientes: La segregación de las mutaciones en las posiciones 165 y 288 ($hema^{Mut4}$ y $hema^{Mut16}$); la mutación de posiciones importantes por su conservación y su posible papel en la especificidad ($hema^{Mut11}$ y $hema^{Mut14}$); la presencia de mutaciones en un contexto en que no se alteraron las posiciones 165 y 288 ($hema^{Mut3}$). La figura 20 confirma el fenotipo de complementación en el fondo $\Delta bioF$. Después de $hema^{mut1}$, los alelos $hema^{Mut3}$ y $hema^{Mut11}$ confirieron los fenotipos más robustos (carriles 3, 8 y 4, respectivamente). Un crecimiento menos significativo se asoció a $hema^{Mut14}$, $hema^{Mut16}$ y $hema^{Mut4}$ (carriles 7, 6 y 5).

Un comportamiento similar fue observado en el fondo genético $\Delta bioA$ (Fig. 21). No obstante, el crecimiento atribuible $hema^{Mut4}$ fue mayor. El fenotipo de complementación también fue observado en el fondo $\Delta bioF\Delta bioA$ (Fig. 22). En este caso, el fenotipo más robusto se correlacionó con la presencia de los alelos $hema^{Mut3}$ y $hema^{Mut4}$. Los alelos $hema^{Mut16}$ y $hema^{Mut14}$ confirieron la menor capacidad de crecimiento.

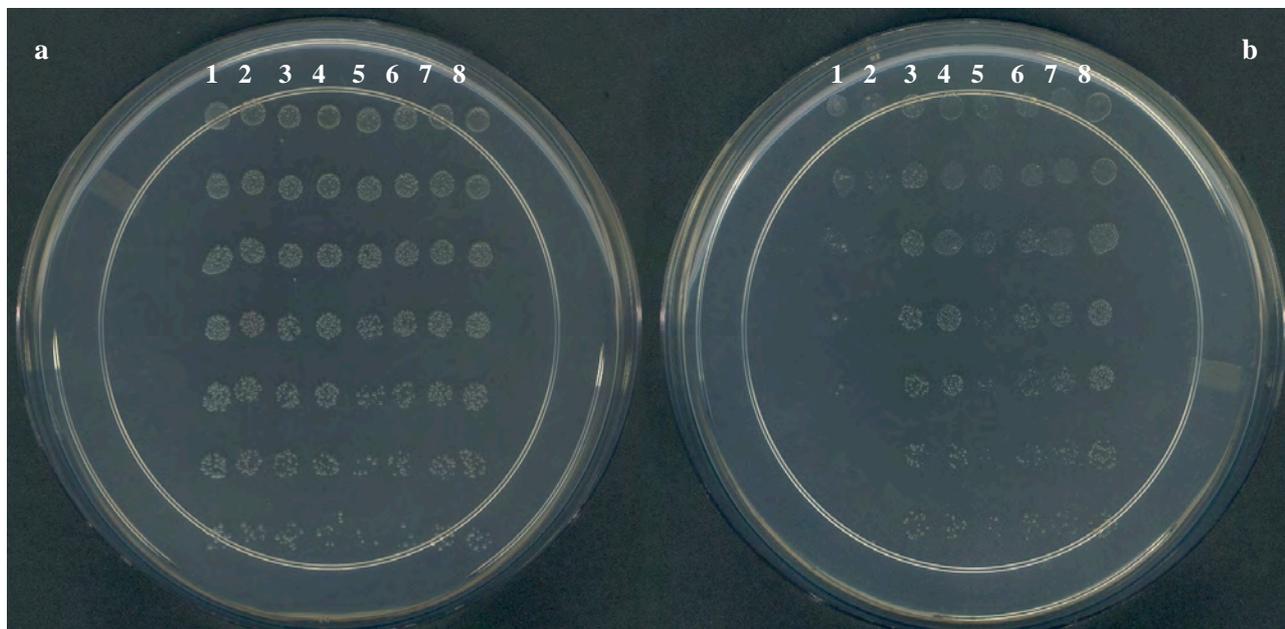


Figura 20. Complementación por los alelos $hema^{Mut}$ en el fondo de delección $\Delta bioF$. El fenotipo de crecimiento conferido por pUC18 (carril 1), $hema$ (carril 2) y $hema^{mut1}$ (carril 3) fue comparado con el asociable a $hema^{mut11}$ (carril 4), $hema^{mut4}$ (carril 5), $hema^{mut16}$ (carril 6), $hema^{mut14}$ (carril 7) y $hema^{mut3}$ (carril 8). Panel a, medio complementado con un exceso de biotina; panel b, medio con trazas de biotina (se observa el crecimiento tras 12 días). (Resultados representativos de tres experimentos independientes.)

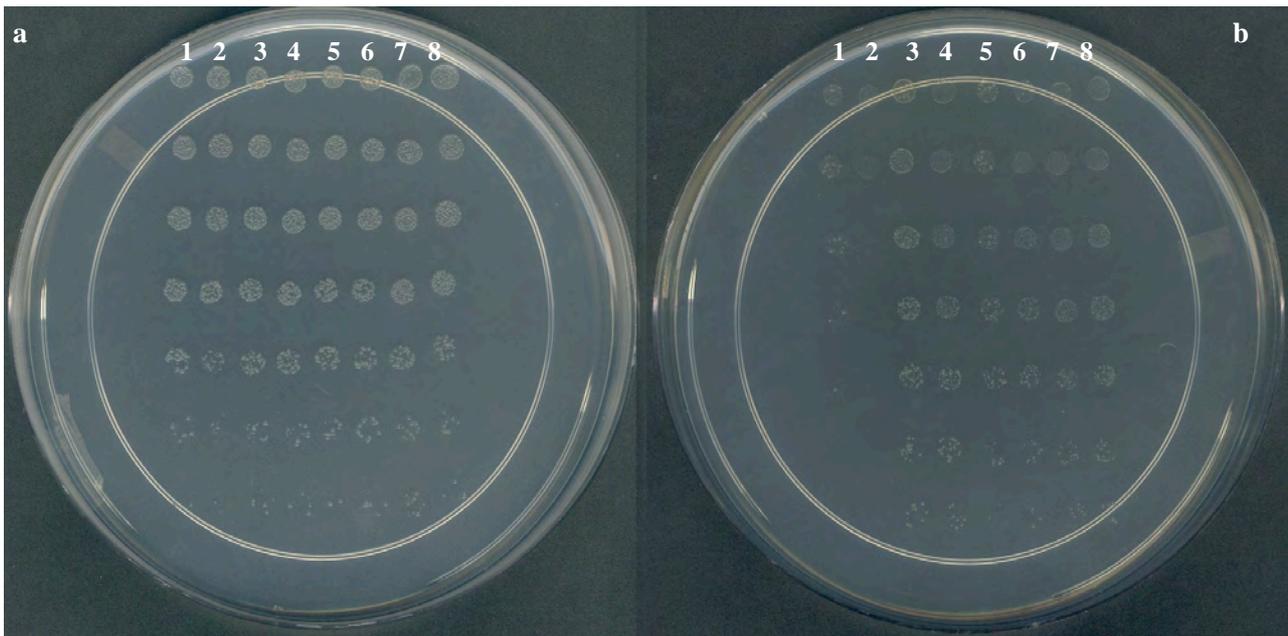


Figura 21. Complementación por los alelos *hemA*^{Mut} en el fondo de delección $\Delta bioA$. El fenotipo de crecimiento conferido por pUC18 (carril 1), pUC18/*hemA* (carril 2) y pUC18/*hemA*^{mut1} (carril 3) fue comparado con el asociable a las construcciones pUC18/*hemA*^{mut11} (carril 4), pUC18/*hemA*^{mut4} (carril 5), pUC18/*hemA*^{mut16} (carril 6), pUC18/*hemA*^{mut14} (carril 7) y pUC18/*hemA*^{mut3} (carril 8). Panel a, medio complementado exceso de biotina; panel b, medio con trazas de biotina (se presenta el crecimiento observado tras 12 días). (Resultados representativos de tres experimentos independientes.)

Una conclusión general atañe a la hipótesis de plasticidad de las posiciones 165 y 288. La evidencia genética sugiere que distintos residuos en cualquiera de estas posiciones puede tener un efecto en el nivel de las actividades promiscuas estudiadas. Asimismo, se tiene evidencia de otras posiciones que podrían actuar de manera plástica. Aunque las variantes revelan nuevas posiciones de plasticidad, éstas podrían afectar de manera diferencial el repertorio de actividades promiscuas.

La promiscuidad catalítica como una componente de la adaptabilidad evolutiva de las enzimas

Las enzimas han alcanzado un *status* de adaptación después de numerosos pasos de mutación y selección. Este proceso evolutivo parece ser complicado, dado que está condicionado por factores diversos. Tales factores se interrelacionan íntimamente, e incluyen las necesidades de mantener la topología y estabilidad del plegamiento, la eficiencia y especificidad de la catálisis, el acomodamiento de la enzima en un contexto metabólico y las interacciones funcionales.



Figura 22. Complementación por los alelos $hemA^{Mut}$ en el fondo de delección $\Delta bioA \Delta bioF$. El fenotipo de crecimiento conferido por pUC18 (carril 1), pUC18/ $hemA$ (carril 2) y pUC18/ $hemA^{mut1}$ (carril 3) fue comparado con el asociado a las construcciones pUC18/ $hemA^{mut11}$ (carril 4), pUC18/ $hemA^{mut4}$ (carril 5), pUC18/ $hemA^{mut16}$ (carril 6), pUC18/ $hemA^{mut14}$ (carril 7) y pUC18/ $hemA^{mut3}$ (carril 8). Panel a, medio complementado exceso de biotina; panel b, medio con trazas de biotina (se presenta el crecimiento observado tras 12 días). (Resultados representativos de tres experimentos independientes.)

Este escenario sugiere que las mutaciones provocan fácilmente alteraciones estructurales y catalíticas, y quizá el colapso de las enzimas. Trueques negativos son esperados, por ejemplo, porque la eficiencia del sitio activo depende de contactos múltiples y precisos entre los residuos de especificidad y los sustratos. Cambios significativos en el sitio activo podrían ser necesarios para generar una nueva actividad. Lo anterior haría poco probable la ocurrencia de novedades evolutivas tras la adaptación. De manera similar, constreñimientos evolutivos más amplios podrían evitar a las enzimas exhibir actividades nuevas o múltiples. Altos niveles de especificidad y eficiencia catalíticas son necesarios para el direccionamiento de intermediarios dentro y entre procesos metabólicos. Entonces, la selección negativa podría operar para disminuir la promiscuidad catalítica. En este sentido, un análisis de redes ha mostrado que las enzimas altamente conectadas, y aquellas asociadas a flujos metabólicos significativos, muestran mayores constreñimientos (Vitkup *et al.*, 2006).

Por otro lado, hay evidencia de que en la naturaleza se evitan algunos constreñimientos evolutivos. Esto podría ocurrir por la selección de estructuras con adaptabilidad evolutiva. La adaptabilidad evolutiva parece una cualidad intrínseca con dos componentes no excluyentes entre sí (Aharoni *et al.*, 2005; O'Loughlin *et al.*, 2006). Por un lado está la robustez, o la capacidad de las estructuras para reducir el efecto

negativo de las mutaciones. Por el otro está la plasticidad, por la que mutaciones discretas producen nuevas actividades.

La robustez se infiere por la observación de la diversidad de secuencias que adoptan un plegamiento dado. Dicha diversidad es más amplia de lo esperado (Shakhnovich *et al.*, 2003). Enzimas a las que se atribuye relaciones de homología por criterios de estructura y catálisis pueden mostrar similitudes de secuencia insignificantes. Esto se correlaciona con estudios de modelaje *in silico*, donde las secuencias termodinámicamente compatibles con un plegamiento muestran identidades tan bajas como 14 % (Koehl & Levitt, 2002). También se ha observado la tolerancia a una mutagénesis extensiva sin alteraciones de estructura o actividad significantes. De acuerdo con lo anterior, la probabilidad de desactivación de una enzima por una mutación puntual aleatoria es ~33% (Guo *et al.*, 2004).

La existencia de la plasticidad se infiere en estudios de mutagénesis al azar y de selección de actividades novedosas. La revisión de estos estudios indica que la evolución *in vitro* suele generar variantes con un aumento de promiscuidad catalítica (Khersonky *et al.*, 2006). Actividades promiscuas con niveles cercanos al umbral de detección pueden mejorarse con unas cuantas mutaciones. Las mutaciones suelen situarse en regiones relativamente distantes del sitio de unión de sustratos y catálisis. De manera relevante, el aumento del nivel de una actividad promiscua no cancela el mantenimiento de la actividad nativa: En general, las actividades promiscuas se incrementan 10 a >1000 veces, mientras que las actividades principales muestran disminuciones menores a 50 veces (Aharoni *et al.*, 2005).

Algunos estudios sugieren la importancia de la promiscuidad catalítica para el establecimiento de rutas metabólicas ancestrales. Un enfoque interesante es el de la evolución de enzimas que catalizan reacciones análogas en rutas metabólicas independientes. Un ejemplo es la evolución *in vitro* de una variante bifuncional de HisA, que muestra actividad TrpF además de su actividad principal (Jürgens *et al.*, 2000). HisA y TrpF participan en las rutas de biosíntesis de la histidina y el triptofano, respectivamente. Los autores sugieren que estas enzimas podrían haber evolucionado a partir de una enzima promiscua como la obtenida experimentalmente. De manera interesante, *Streptomyces coelicolor* y *Mycobacterium tuberculosis* no muestran un ortólogo de TrpF. Sin embargo, tienen un gen bifuncional que contrarresta la doble auxotrofia de mutantes *hisA trpF* de *E. coli* (Barona-Gómez & Hodgson, 2003). El modelo de Jensen sobre la evolución las rutas metabólicas adquiere relevancia con este tipo de observaciones. Sin embargo, todavía es escasa la evidencia de las contribuciones potenciales de la promiscuidad catalítica a la adecuación celular y, en consecuencia, a la innovación evolutiva.

En este trabajo se intentó ampliar la comprensión del fenómeno de la adaptabilidad evolutiva. Se probó el potencial de la promiscuidad catalítica para proveer reacciones dentro y entre rutas metabólicas. Los

resultados coinciden con las observaciones mencionadas en el párrafo anterior y en la sección de antecedentes. Mediante mutagénesis al azar se generó la variante Hema^{mut1}, la cual mostró una actividad promiscua incrementada y correspondiente a la actividad CAPS. La secuencia de esta variante muestra tan solo dos sustituciones. Una sustitución está en una estructura que rodea a la pared del sitio activo, mientras que la otra se sitúa en la superficie. De acuerdo con un ensayo *in vitro* preliminar, la actividad promiscua evolucionada podría estar apenas por arriba del umbral de detección. No obstante, la relevancia biológica de esta actividad queda demostrada por su habilidad para permitir la supervivencia celular bajo la limitación estricta de un compuesto esencial.

La actividad principal de la variante Hema^{mut1} se mostró casi inafectada, aunque no hubo selección experimental para mantenerla. Se ha observado tal comportamiento en otro trabajo de evolución y detección *in vitro* de actividades promiscuas (Aharoni *et al.*, 2005). Sin embargo, el sistema de selección *in vivo* permite extrapolaciones en el terreno de la evolución biológica. En *E. coli*, el ácido α -aminolevulínico se genera a partir de L-glutamato por dos enzimas consecutivas, *i.e.*, la glutamil-tRNA reductasa y glutamato-1-semialdehído 2,1-aminomutasa (Panek & O'Brian, 2002). Esta ruta hace la producción del ácido α -aminolevulínico por Hema redundante, y quizá desventajosa, en el contexto metabólico de *E. coli*. Lo anterior excluye factores de selección asociados a la necesidad metabólica por una actividad o un compuesto. Al considerar esto, puede concluirse que la actividad principal de la variante Hema^{mut1} se mantuvo casi inafectada gracias a la relación entre la estructura y la actividad de la enzima.

La evolución de un generalista tiene cierta direccionalidad cuando una de sus actividades promiscuas es capturada por la selección. Sin embargo, las características estructurales responsables de la adaptabilidad evolutiva podrían ayudar a mantener niveles relativamente altos de la actividad principal, aún si las presiones selectivas que actúan sobre ésta se relajan. Esto podría evitar desbalances metabólicos durante el tiempo evolutivo que sigue inmediatamente a la duplicación génica: Se supone que el aumento del nivel de expresión asociado a la duplicación exacta de una secuencia codificante podría tener poco efecto en la adecuación. Sin embargo, también se espera la relajación de los constreñimientos que operan normalmente sobre la variación de la secuencia. Esto podría deberse a que la presión de selección está distribuida entre las nuevas secuencias.

Las mutaciones podrían acumularse en ambos duplicados. Asimismo, las mutaciones podrían ser toleradas mientras la suma de los niveles de actividad de las enzimas codificadas sean satisfactorias para el metabolismo. Sin embargo, si las mutaciones fueran fuertemente negativas para la actividad principal, no se permitiría a los duplicados lograr un 'equilibrio de adecuación'. Simplemente, funciones esenciales se perderían junto con las novedosas. Entonces, los residuos de plasticidad están distribuidos a lo largo de la secuencia y no implican necesariamente posiciones determinantes para la actividad principal. La presión de selección para la adaptabilidad evolutiva es mayor cuando el cambio evolutivo muestra una alta tasa. Desde

esta perspectiva, la promiscuidad provee una ventaja selectiva inmediata a las células sujetas a condiciones estresantes al introducir actividades potencialmente útiles (O'Brien & Herschlag, 1999). En consecuencia, la adaptabilidad evolutiva podría permitir a las enzimas un comportamiento exploratorio mientras se desempeñan de manera adecuada en el contexto metabólico.

El alelo *hemA*^{mut1} mostró una capacidad contingente para complementar la ausencia de la actividad de BioA. Lo anterior sugiere que la variante HemA^{mut1} exhibe al menos otra actividad promiscua incrementada. Además de mantener su actividad principal, la variante HemA^{mut1} parece capaz de catalizar dos reacciones consecutivas en la ruta de biosíntesis de la biotina. Como lo sugiere la evidencia genética, estas reacciones promiscuas son ventajosas para la célula independientemente de su bajo nivel, por separado o en combinación.

Análisis de estructura y secuencia sugieren que BioF y BioA podrían compartir un ancestro común (Schneider & Lindqvist, 2001; Otsuka *et al.*, 1988). Estas proteínas actúan de manera consecutiva en una ruta y están codificadas por genes contiguos en un operón. Por lo anterior, resulta coherente la hipótesis de que estas enzimas representan un caso de retroevolución. Los resultados muestran que los pasos consecutivos catalizados por BioF y BioA podrían haberse ensamblado por el reclutamiento de un generalista como HemA^{mut1}. Este escenario evolutivo no necesita una enzima ancestral asociada previamente a la biosíntesis de la biotina. El generalista podría haber sido proficiente en cualquier reacción y después ser reclutado desde una ruta preexistente. De manera importante, una transaminación dependiente de un L-aminoácido como donador de grupos amino podría haber participado en la reacción correspondiente a BioA. (HemA^{mut1} probablemente cataliza las transaminaciones con pimeloil CoA y luego con 7-ceto-aminopelargonato empleando L-glicina u otro L-aminoácido.) Eventos evolutivos posteriores podrían haber incluido el reclutamiento de una enzima capaz de usar S-adenosilmetionina como donador de grupos amino. El uso de S-adenosilmetionina como donador de grupos amino se conoce en los ortólogos BioA de *E. coli* y *Bacillus sphaericus*. No obstante, el ortólogo de *B. subtilis* emplea L-lisina (Van Arsdell *et al.*, 2005), lo cual hace posible el escenario evolutivo antes propuesto.

Un aspecto fundamental de las ideas de Jensen es que, tempranamente en la evolución, unas cuantas enzimas ineficientes podrían haber ensamblado largas cadenas de reacciones. Los resultados ilustran que una enzima promiscua podría representar eslabones consecutivos en una cadena de reacciones por la que se obtiene un compuesto esencial. Tal posibilidad no excluye que la enzima participe de manera proficiente en alguna otra ruta. Las observaciones anteriores tienen dos implicaciones relacionadas con la evolución de la complejidad del metabolismo. La probabilidad del aumento de la promiscuidad enzimática por mutaciones discretas podría facilitar la evolución de nuevas actividades tras la duplicación de una secuencia codificante. En este sentido, la complejidad metabólica aumenta por la adición de enzimas parálogas, *i. e.*, por evolución

expansiva. Sin embargo, la complejidad puede mantenerse, e inclusive incrementarse, en un escenario de evolución reductiva si las reacciones son consideradas de manera independiente del número de enzimas que las catalizan. La promiscuidad catalítica podría proveer las actividades paradójicamente ausentes en ciertos organismos.

Parece posible que versiones simplificadas de las rutas existan por el reclutamiento de enzimas promiscuas. De esta manera podrían sintetizarse compuestos requeridos en pequeñas cantidades o de forma intermitente. Lo anterior podría ser relevante para especies bacterianas que se encuentran en el camino evolutivo hacia un estilo de vida parasítico o endosimbiótico. Especies parasíticas como *Mycobacterium leprae*, *Shigella flexneri* y *Salmonella enterica* muestran reducciones del tamaño del genoma (Cole *et al.*, 2001; Wei *et al.*, 2003; Parkhill *et al.*, 2001). Se supone que la pérdida de genes puede asociarse a la relajación de las presiones de selección (Parkhill *et al.*, 2003). Esta pérdida puede ocurrir en escalas de tiempo cortas, como se ha observado experimentalmente en *S. enterica* (Nilsson *et al.*, 2005). Análisis de redes sugieren que las enzimas parálogas que exhiben actividades similares tienden a estar agrupadas en ‘vecindarios metabólicos’. Entonces, si una enzima se pierde, el paso metabólico catalizado por esta enzima podría ser rescatado parcialmente por parálogos promiscuos que exhiben un umbral de la actividad faltante. Se ha sospechado el papel potencial de las enzimas promiscuas en la catálisis de reacciones faltantes. En varias especies del género *Mycoplasma* no hay ortólogos de la enzima con actividad nucleósido bisfosfato cinasa. Sin embargo, la actividad podría ser compensada por enzimas no relacionadas, como cinasas de acetato, fosfoglicerato y piruvato (Pollack *et al.*, 2002).

Se supone que las especies adaptadas a la endosimbiosis encuentran condiciones estables en su hospedero. La asociación intracelular parece llevar a la pérdida de genes de procesos metabólicos centrales (Moran, 2003). Por ejemplo, el género *Buchnera* exhibe sólo algunos genes para el ciclo del ácido tricarboxílico. Los miembros de este género tienen por hospedero áfidos, que les proporcionan aminoácidos no esenciales, CoA y quizá polisacáridos (Zientz *et al.*, 2004). Pero en el género *Buchnera* se retiene el potencial genético para sintetizar aminoácidos esenciales (Moran & Baumann, 2000). Esto refleja una complementariedad entre los repertorios génicos y las capacidades metabólicas del hospedero y el endosimbionte. Es decir, ha habido selección desde el ambiente (el hospedero) hacia funciones génicas del endosimbionte y *vice versa*.

Por otro lado, ningún ambiente se prolonga indefinidamente. Ciertas presiones de selección podrían emerger del ambiente por la heterogeneidad de las fuentes nutrimentales encontradas por el hospedero. Considérese el escenario siguiente. *Blochmannia* es un endosimbionte de hormigas del género *Camponotus*. Estas hormigas consumen partes vegetales y excremento de aves, pero también se dedican a la depredación y carroñeo de insectos. Asimismo, la dieta de *Camponotus* varía entre regiones geográficas y de manera estacional. (Zientz *et al.*, 2004; Wernegreen *et al.*, 2002). Parece posible que las poblaciones de *Blochmannia*

enfrenten fluctuaciones en el aprovisionamiento de algunos nutrientes. En un escenario como éste, el endosimbionte podría necesitar ciertas capacidades metabólicas de manera transitoria.

Una predicción es que en las bacterias endosimbiontes está muy representada la promiscuidad catalítica. Estas bacterias muestran una inusual aceleración de las tasas evolutivas. En *Buchnera spp.*, las tasas de sustituciones nucleotídicas por generación son hasta cuatro veces más elevadas que en *E. coli* y *S. enterica* (Wernegreen, 2002). La relajación de la selección explicaría la acumulación de mutaciones en genes específicos. Pero las mutaciones se fijan en todos los *loci*, inclusive en los que están sujetos a selección (Moran & Baumann, 2000). Las bacterias endosimbiontes forman poblaciones pequeñas y se transmiten maternalmente por cuellos muy estrechos (Fares *et al.*, 2002). Esto supone que mutaciones se fijan por deriva aleatoria. La mayoría de las mutaciones son al menos ligeramente negativas y deben tener un costo para la estabilidad y la función de las proteínas.

A pesar de la acumulación de mutaciones, los endosimbiontes no ven afectada su adecuación. Esto parece deberse a la expresión constitutiva de las chaperoninas GroEL/GroES (Fares *et al.*, 2004; Rutherford 2003). Dicho papel ha sido estudiado en una población hipermutagénica de *E. coli*: La adecuación de esta población disminuye tras 3,000 generaciones de cultivo, pero se ve restablecida al inducirse artificialmente la transcripción de *groE* (Fares *et al.*, 2002). Se infiere que GroEL/GroES mitiga la inestabilidad proteínica en especies donde la deriva aleatoria permite altas tasas evolutivas. Por complementación genética en una cepa *trpE* de *E. coli*, se determinó que la antranilato sintasa de *Buchnera aphidicola* pierde actividad por arriba de los 26 °C. Sin embargo, la coexpresión de *trpE* de *B. aphidicola* con los genes de las chaperoninas aumenta la función enzimática (Huang *et al.*, 2008). Asimismo, la sobreproducción de GroEL/GroES permite el plegamiento de variantes enzimáticas con mutaciones en el centro hidrofóbico y con valores desestabilizadores mayores que 3.5 kcal mol⁻¹ (Tokuriki & Tawfik, 2009). Puede suponerse la disminución de los constreñimientos evolutivos y una concomitante elevación de los niveles de promiscuidad catalítica en las enzimas de las bacterias endosimbiontes.

La restauración de actividades faltantes podría ser un evento frecuente en la evolución, debido a la redundancia asociable a la promiscuidad catalítica de las enzimas homólogas. En *E. coli*, por lo menos cinco miembros de la familia ROK poseen suficiente actividad glucocinasa para permitir la supervivencia de una cepa *pgk*⁻ (Miller & Raines, 2005). Resulta razonable pensar en la promiscuidad catalítica como un catálogo de actividades de respaldo. Desde luego, en este catálogo podrían incluirse actividades llevadas a cabo tanto por proteínas homólogas como por proteínas sin relación evolutiva: A partir de una librería de DNA genómico se seleccionó *in vivo* una proteína, YjbQ, cuya expresión extracromosómica contrarresta la auxotrofia de una mutante *thiE* de *E. coli*. ThiE es la tiamina fosfato sintasa, una enzima con plegamiento de barril (β/α)₈. YjbQ tiene ortólogos en los tres dominios de la vida, pero su plegamiento α/β es diferente. La

función biológica de YjbQ se desconoce, y su actividad tiamina sintasa debe ser de tipo secundario (Morett *et al.*, 2008).

Las enzimas dependientes de PLP son excelentes candidatos para el papel de restauración de actividades faltantes. Esto es sugerido por su papel versátil en el metabolismo y porque se observa la promiscuidad entre miembros de las distintas familias que emplean PLP como cofactor. En apoyo de esta suposición está la noción de que el uso promiscuo de sustratos por las aminotransferasas dependientes de PLP es importante para comunicar rutas metabólicas:

La aspartato aminotransferasa (AAT) resulta un excelente ejemplo de la versatilidad y promiscuidad de las enzimas dependientes de PLP. Esta enzima cataliza la transaminación reversible entre aminoácidos dicarboxílicos y cetoácidos. Su función biológica incluye la síntesis de L-aspartato, L-alanina, aminoácidos aromáticos y la remoción de grupos amino a partir de estos aminoácidos durante el catabolismo. En términos químicos, transfiere el grupo amino del L-aspartato (o L-glutamato) al 2-oxoglutarato (u oxaloacetato). Sin embargo, los ortólogos de la AAT pueden usar aminoácidos aromáticos (Hayashi *et al.*, 1993). Esta enzima también es capaz de catalizar reacciones de racemización α , descarboxilación β y eliminación β (Kochlar & Christen, 1992; Birolo *et al.*, 1995, Graber *et al.*, 1995; Vacca *et al.*, 1997). Estas actividades se catalizan en niveles bajos y quizá son irrelevantes en términos fisiológicos. Sin embargo, estudios de mutagénesis indican que las especificidades de sustrato y de reacción de la AAT pueden alterarse con unas cuantas mutaciones. La relevancia evolutiva de estos experimentos puede verse en dos sentidos. En el primero, la promiscuidad enzimática podría haber facilitado la diversificación funcional de la familia α . En el segundo, un potencial evolutivo es albergado por las enzimas dependientes de PLP de la familia α . El hallazgo de variantes de HemA^{mut} con actividades promiscuas incrementadas apunta a que dicho potencial podría explotarse tanto en el laboratorio como en la naturaleza.

Las fuentes de respaldo también podrían provenir de otras familias de enzimas dependientes de PLP. Por ejemplo, la subunidad β de la triptofano sintasa cataliza la condensación fisiológica de L-serina e indol, pero también muestra actividad L-serina deshidratasa. Asimismo, la treonina sintasa de *Bacillus subtilis* cataliza en bajos niveles las reacciones de L-treonina y L-serina deshidratasa (Skarstedt & Greer, 1973; Laber *et al.*, 1994). Un caso interesante es el de la aminotransferasa de aminoácidos aromáticos de *E. coli*. Esta enzima exhibe un comportamiento recíproco al de la AAT (Hayashi *et al.*, 1993), pues utiliza aromáticos y aspartato con igual eficiencia en ensayos *in vitro*. El solapamiento de especificidades sugiere otra vez la retención evolutiva de actividades de respaldo. La aminotransferasa de aminoácidos aromáticos es sensible a la represión por L-tirosina. *In vivo*, esta enzima compensa la ausencia de AAT cuando la represión es eliminada fisiológicamente ó por mutación (Jensen & Gu, 1996).

La revelación del potencial evolutivo de las enzimas dependientes de PLP puede depender de cambios evolutivos discretos. Se sugiere lo primero por las reacciones promiscuas que ya tienen niveles muy elevados. Por ejemplo, la racemización y la desaminación de L-serina a piruvato son catalizadas en tasas similares por la serina racemasa de cerebro de ratón (Strisovsky *et al.*, 2003). Por otro lado, múltiples enzimas podrían coincidir en una actividad. Un caso interesante es el de la cisteína-S-conjugado β -liasa (E. C. 4.4.1.13). Se desconoce la secuencia codificante de la enzima especializada en esta actividad. Sin embargo, al menos nueve enzimas dependientes de PLP, incluyendo aminotransferasas, catalizan esta reacción en mamíferos (Cooper *et al.*, 2002). La ocurrencia de la promiscuidad implica que un organismo puede tener más actividades asociadas al PLP que genes de enzimas dependientes de PLP. Además, un número de actividades relacionadas con el PLP no se han caracterizado a nivel de secuencia, y tal vez son casos de promiscuidad catalítica.

Los resultados de este trabajo pueden extenderse en un escenario hipotético. *E. coli* tiene un requerimiento absoluto de biotina. El incremento sutil de la promiscuidad catalítica de las variantes HemA^{mut} podría permitir la pérdida definitiva de BioF y BioA. De manera concomitante, las enzimas responsables de la síntesis del ácido α -aminolevulínico podrían removerse, quizá sin comprometer significativamente el suministro del grupo hemo. Resulta interesante la posibilidad de que nuevas variantes HemA^{mut} sean capaces de restaurar la actividad de Kbl *in vivo*. Como resultado, la red metabólica de *E. coli* podría modificarse en diversos pasos. Estos pasos podrían ser remplazados por un intermediario generalista. Los modelos *in silico* de las variantes HemA^{mut} sugieren que parte del potencial evolutivo de la promiscuidad está constreñido por la pérdida de la estabilidad estructural. Por otro lado, los constreñimientos parecen reducirse en la naturaleza a través de la función de las chaperonas. Resulta atractiva la opción de incorporar la sobreproducción de GroEL/GroES en el modelo de estudio de este laboratorio. Así podría probarse el alcance máximo de las variantes HemA^{mut} como proveedoras de actividades novedosas o de respaldo. En un sentido más amplio, podría relacionarse la evolución de las actividades enzimáticas y las redes metabólicas con un mecanismo por el que el impacto de la variación genética se almacena y permanece latente.

Conclusiones

El surgimiento de actividades enzimáticas biológicamente relevantes puede ocurrir por presiones de selección impuestas sobre una actividad promiscua preexistente.

Las actividades promiscuas pueden adquirir relevancia biológica por un aumento marginal. El grado de variación de secuencia que supone este cambio puede ser de unas cuantas sustituciones.

La estabilidad estructural podría verse comprometida por las sustituciones que dan lugar a un incremento de promiscuidad suficiente para la actuación de la selección.

Los compromisos estructurales asociados al aumento de una actividad promiscua seleccionable podrían suponer el aumento de nivel de otras actividades. Estas actividades promiscuas incrementadas podrían seleccionarse en los ambientes apropiados. Esto da origen a un intermediario generalista potencial.

El aumento de las actividades del repertorio de promiscuidad catalítica puede darse a través de diferentes caminos evolutivos. Esto es que diferentes variantes convergen en su capacidad para sostener actividades demandadas por la célula, como resultado de la imposición de presiones de selección (inherentemente impredecibles).

Diferentes variantes enzimáticas provenientes de un mismo precursor (ancestro) pueden actuar como intermediarios generalistas en el establecimiento de los mismos pasos de una ruta metabólica. Esto ocurre gracias a la distribución de posiciones de plasticidad a lo largo de la estructura primaria de las enzimas.

Perspectivas

La confirmación bioquímica de la actividad CAPS de las variantes enzimáticas HemA^{mut}.

El aumento de la actividad CAPS de la variante HemA^{mut1} mediante rondas adicionales de mutagénesis al azar y selección *in vivo*.

La evaluación del papel del residuo D165 de HemA en el establecimiento de interacciones débiles capaces de modular la promiscuidad catalítica.

La identificación del sustrato donador de grupos amino que interviene en la actividad por la que las variantes HemA^{mut} compensan la eliminación de BioA.

La recombinación y segregación de las mutaciones observadas. Así podría determinarse el papel de plasticidad de posiciones diversas. Asimismo, podría evaluarse el papel de la epistasis en el mejoramiento de la actividad CAPS.

La evaluación del potencial de las variantes HemA^{mut} para catalizar otras reacciones de transaminación. El esfuerzo más inmediato podría enfocarse en la actividad de la enzima Kbl de *E. coli*.

La determinación de la estabilidad de las variantes HemA^{mut} para conocer el costo de estabilidad estructural que conlleva la exploración evolutiva de las actividades promiscuas.

El estudio del efecto de la sobreproducción de la chaperona GroEl/GroES en la capacidad de las variantes HemA^{mut} para restablecer la biosíntesis de la biotina, y consecuentemente el crecimiento celular.

El remplazamiento de las reacciones que originan al ácido 5-aminolevulínico en *E. coli* por la reacción dependiente de las variantes HemA^{mut}. En un fondo de remoción de genes para la biosíntesis de la biotina, las variantes HemA^{mut} podrían remplazar cuatro pasos en dos rutas metabólicas.

Referencias

- Afriat, L., C. Roodveldt, G. Manco, & D. S. Tawfik.** 2006. The latent promiscuity of newly identified microbial lactonases is linked to a recently diverged phosphotriesterase. *Biochemistry* **45**:13677-86.
- Aharoni, A., L. Gaidukov, O. Khersonsky, Q. G. S. Mc, C. Roodveldt, & D. S. Tawfik.** 2005. The 'evolvability' of promiscuous protein functions. *Nat Genet* **37**:73-6.
- Albert, R., H. Jeong, & A. L. Barabasi.** 2000. Error and attack tolerance of complex networks. *Nature* **406**:378-82.
- Alexander, F. W., E. Sandmeier, P. K. Mehta, & P. Christen.** 1994. Evolutionary relationships among pyridoxal-5'-phosphate-dependent enzymes. Regio-specific alpha, beta & gamma families. *Eur J Biochem* **219**:953-60.
- Alexeev D., M. Alexeeva, R. L. Baxter, D. J. Campopiano, S. P. Webster, L. Sawyer L.** 1998. The crystal structure of 8-amino-7-oxononanoate synthase: a bacterial PLP-dependent, acyl-CoA-condensing enzyme. *J Mol Biol* **284**:401-19.
- Alves, R., R. A. Chaleil, & M. J. Sternberg.** 2002. Evolution of enzymes in metabolism: a network perspective. *J Mol Biol* **320**:751-70.
- Anfinsen, C. B.** 1973. Principles that govern the folding of protein chains. *Science* **181**:223-30.
- Aroul-Selvam, R., T. Hubbard, & R. Sasidharan.** 2004. Domain insertions in protein structures. *J Mol Biol* **338**:633-41.
- Asano, Y., I. Kira, & K. Yokozeki.** 2005. Alteration of substrate specificity of aspartase by directed evolution. *Biomol Eng* **22**:95-101.
- Astner I., J. O. Schulze, J. van den Heuvel, D. Jahn, W. D. Schubert, D. W. Heinz.** 2005. Crystal structure of 5-aminolevulinic synthase, the first enzyme of heme biosynthesis, and its link to XLSA in humans. *EMBO J* **24** (18):3166-77.
- Barabasi, A. L., & R. Albert.** 1999. Emergence of scaling in random networks. *Science* **286**:509-12.
- Barabasi, A. L., & Z. N. Oltvai.** 2004. Network biology: understanding the cell's functional organization. *Nat Rev Genet* **5**:101-13.
- Barona-Gomez, F., & D. A. Hodgson.** 2003. Occurrence of a putative ancient-like isomerase involved in histidine and tryptophan biosynthesis. *EMBO Rep* **4**:296-300.
- Bartlett, G. J., N. Borkakoti, & J. M. Thornton.** 2003. Catalysing new reactions during evolution: economy of residues and mechanism. *J Mol Biol* **331**:829-60.
- Beadle, B. M., & B. K. Shoichet.** 2002. Structural bases of stability-function tradeoffs in enzymes. *J Mol Biol* **321**:285-96.
- Benkovic, S. J., & S. Hammes-Schiffer.** 2003. A perspective on enzyme catalysis. *Science* **301**:1196-202.
- Berkovitch F., Y. Nicolet, J. T. Wan, J. T. Jarrett, & C. L. Drennan.** 2004. Crystal structure of biotin synthase, an S-adenosylmethionine-dependent radical enzyme. *Science* **303**:76-9.
- Besenmatter, W., P. Kast, & D. Hilvert.** 2007. Relative tolerance of mesostable and thermostable protein homologs to extensive mutation. *Proteins* **66**:500-6.
- Bhor V.M., S. Dev, G. R. Vasanthakumar, P. Kumar, S. Sinha, & A. Surolia.** 2006. Broad substrate stereospecificity of the Mycobacterium tuberculosis 7-keto-8-aminopelargonic acid synthase: Spectroscopic and kinetic studies. *J Biol Chem* **281**:25076-88.

- Birolò L., E. Sandmeier, P. Christen, & R. A. John.** 1995. The roles of Tyr70 and Tyr225 in aspartate aminotransferase assessed by analysing the effects of mutations on the multiple reactions of the substrate analogue serine o-sulphate. *Eur J Biochem* **232**: 859-64.
- Bloom, J. D., S. T. Labthavikul, C. R. Otey, & F. H. Arnold.** 2006. Protein stability promotes evolvability. *Proc Natl Acad Sci U S A* **103**:5869-74.
- Bornscheuer, U. T., & R. J. Kazlauskas.** 2004. Catalytic promiscuity in biocatalysis: using old enzymes to form new bonds and follow new pathways. *Angew Chem Int Ed Engl* **43**:6032-40.
- Camps, M., A. Herman, E. Loh, & L. A. Loeb.** 2007. Genetic constraints on protein evolution. *Crit Rev Biochem Mol Biol* **42**:313-26.
- Cole, S. T., K. Eiglmeier, J. Parkhill, K. D. James, N. R. Thomson, P. R. Wheeler, N. Honore, T. Garnier, C. Churcher, D. Harris, K. Mungall, D. Basham, D. Brown, T. Chillingworth, R. Connor, R. M. Davies, K. Devlin, S. Duthoy, T. Feltwell, A. Fraser, N. Hamlin, S. Holroyd, T. Hornsby, K. Jagels, C. Lacroix, J. Maclean, S. Moule, L. Murphy, K. Oliver, M. A. Quail, M. A. Rajandream, K. M. Rutherford, S. Rutter, K. Seeger, S. Simon, M. Simmonds, J. Skelton, R. Squares, S. Squares, K. Stevens, K. Taylor, S. Whitehead, J. R. Woodward, & B. G. Barrell.** 2001. Massive gene decay in the leprosy bacillus. *Nature* **409**:1007-11.
- Collins, M. O., & J. S. Choudhary.** 2008. Mapping multiprotein complexes by affinity purification and mass spectrometry. *Curr Opin Biotechnol* **19**:324-30.
- Cooper A.** 1999. Thermodynamics of protein folding and stability. *Protein: A Comprehensive Treatise* 2: 217-270. Series Editor: Geoffrey Allen. Publisher: JAI Press Inc.
- Cooper A.J., S. A. Bruschi, & M. W. Anders.** 2002. Toxic, halogenated cysteine S-conjugates and targeting of mitochondrial enzymes of energy metabolism. *Biochem Pharmacol* **64**:553-64.
- Copley, R. R., & P. Bork.** 2000. Homology among (betaalpha)(8) barrels: implications for the evolution of metabolic pathways. *J Mol Biol* **303**:627-41.
- Copley S. D.** 2009. Evolution of efficient pathways for degradation of anthropogenic chemicals. *Nat Chem Biol* **5**: 559-566.
- Creighton, T. E.** 1990. Protein folding. *Biochem J* **270**:1-16.
- Chothia, C.** 1992. *Proteins*. One thousand families for the molecular biologist. *Nature* **357**:543-4.
- Christen, P., P. Kasper, H. Gehring, & M. Sterk.** 1996. Stereochemical constraint in the evolution of pyridoxal-5'-phosphate-dependent enzymes. A hypothesis. *FEBS Lett* **389**:12-4.
- Dryden D.T., A. R. Thomson, & J. H. White.** 2008. How much of protein sequence space has been explored by life on Earth? *J R Soc Interface* **5**: 953-6.
- DePristo, M. A., D. M. Weinreich, & D. L. Hartl.** 2005. Missense meanderings in sequence space: a biophysical view of protein evolution. *Nat Rev Genet* **6**:678-87.
- Diaz-Mejia, J. J., E. Perez-Rueda, & L. Segovia.** 2007. A network perspective on the evolution of metabolism by gene duplication. *Genome Biol* **8**:R26.
- Dill, K. A., S. Bromberg, K. Yue, K. M. Fiebig, D. P. Yee, P. D. Thomas, & H. S. Chan.** 1995. Principles of protein folding--a perspective from simple exact models. *Protein Sci* **4**:561-602.
- Dill, K. A., & H. S. Chan.** 1997. From Levinthal to pathways to funnels. *Protein Sci* **4**:10-9.
- Dobson C.M., & M. Karplus.** 1999. The fundamentals of protein folding: bringing together theory and experiment. *Curr Opin Struct Biol* **9**: 92-101.
- Drummond, D. A., J. D. Bloom, C. Adami, C. O. Wilke, & F. H. Arnold.** 2005. Why highly expressed *Proteins* evolve slowly. *Proc Natl Acad Sci U S A* **102**:14338-43.

- Eisenmesser E.Z., D. A. Bosco, M. Akke, & D. Kern. 2002. Enzyme dynamics during catalysis. *Science* 295: 1520-3.
- Eisenmesser E.Z., O. Millet, W. Labeikovsky, D. M. Korzhnev, M. Wolf-Watz, D. A. Bosco, J. J. Skalicky, L. E. Kay LE, & D. Kern. 2005. Intrinsic dynamics of an enzyme underlies catalysis. *Nature* 438: 117-21.
- Eliot, A. C., & J. F. Kirsch. 2003. Avoiding the road less traveled: how the topology of enzyme-substrate complexes can dictate product selection. *Acc Chem Res* 36:757-65.
- Eliot, A. C., & J. F. Kirsch. 2004. Pyridoxal phosphate enzymes: mechanistic, structural, and evolutionary considerations. *Annu Rev Biochem* 73:383-415.
- Ericsson, M., D. Hanstorp, P. Hagberg, J. Enger, & T. Nyström. 2000. Sorting out viability with optical tweezers. *J Bacteriol* 182: 5551-55.
- Fares M.A., M. X., Ruiz-González, A. Moya, S. F. Elena, E. Barrio. 2002. Endosymbiotic bacteria: groEL buffers against deleterious mutations. *Nature* 417: 398.
- Fares, M. A., A. Moya, & E. Barrio. 2004. GroEL and the maintenance of bacterial endosymbiosis. *Trends Genet* 20:413-6 .
- Fenton, A. W. 2008. Allostery: an illustrated definition for the 'second secret of life'. *Trends Biochem Sci* 33:420-5.
- Fersht, A. R. 1997. Nucleation mechanisms in protein folding. *Curr Opin Struct Biol* 7:3-9.
- Force, A., M. Lynch, F. B. Pickett, A. Amores, Y. L. Yan, & J. Postlethwait. 1999. Preservation of duplicate genes by complementary, degenerative mutations. *Genetics* 151:1531-45.
- Fraser, H. B. 2005. Modularity and evolutionary constraint on proteins. *Nat Genet* 37:351-2.
- Fraser, H. B., A. E. Hirsh, L. M. Steinmetz, C. Scharfe, & M. W. Feldman. 2002. Evolutionary rate in the protein interaction network. *Science* 296:750-2.
- Gavin, A. C., P. Aloy, P. Grandi, R. Krause, M. Boesche, M. Marzioch, C. Rau, L. J. Jensen, S. Bastuck, B. Dumpelfeld, A. Edelmann, M. A. Heurtier, V. Hoffman, C. Hoefert, K. Klein, M. Hudak, A. M. Michon, M. Schelder, M. Schirle, M. Remor, T. Rudi, S. Hooper, A. Bauer, T. Bouwmeester, G. Casari, G. Drewes, G. Neubauer, J. M. Rick, B. Kuster, P. Bork, R. B. Russell, & G. Superti-Furga. 2006. Proteome survey reveals modularity of the yeast cell machinery. *Nature* 440:631-6.
- Gerlt, J. A., & P. C. Babbitt. 2001. Divergent evolution of enzymatic function: mechanistically diverse superfamilies and functionally distinct suprafamilies. *Annu Rev Biochem* 70:209-46.
- Glasner, M. E., J. A. Gerlt, & P. C. Babbitt. 2006. Evolution of enzyme superfamilies. *Curr Opin Chem Biol* 10:492-7.
- Glasner, M. E., J. A. Gerlt, & P. C. Babbitt. 2007. Mechanisms of protein evolution and their application to protein engineering. *Adv Enzymol Relat Areas Mol Biol* 75:193-239, xii-xiii.
- Gogarten, J. P., & L. Olendzenski. 1999. Orthologs, paralogs and genome comparisons. *Curr Opin Genet Dev* 9:630-6.
- Goll, J., & P. Uetz. 2006. The elusive yeast interactome. *Genome Biol* 7:223.
- Gould, S. J., & E. A. Lloyd. 1999. Individuality and adaptation across levels of selection: how shall we name and generalize the unit of Darwinism? *Proc Natl Acad Sci U S A* 96:11904-9.
- Graber, R., P. Kasper, V. N. Malashkevich, E. Sandmeier, P. Berger, H. Gehring, J. N. Jansonius, & P. Christen. 1995. Changing the reaction specificity of a pyridoxal-5'-phosphate-dependent enzyme. *Eur J Biochem* 232:686-90.

- Guo, H. H., J. Choe, & L. A. Loeb.** 2004. Protein tolerance to random amino acid change. *Proc Natl Acad Sci U S A* **101**:9205-10.
- Gutteridge A., & J. M. Thornton.** 2005. Understanding nature's catalytic toolkit. *Trends Biochem Sci* **30**: 622-9.
- Hayashi, H., K. Inoue, T. Nagata, S. Kuramitsu, & H. Kagamiyama.** 1993. *Escherichia coli* aromatic amino acid aminotransferase: characterization and comparison with aspartate aminotransferase. *Biochemistry* **32**:12229-39.
- Hedstrom L.** 2002. Serine protease mechanism and specificity. *Chem Rev* **102**: 4501-24.
- Henzler-Wildman K.A., V. Thai, M. Lei, M. Ott, M. Wolf-Watz, T. Fenn, E. Pozharski, M. A. Wilson, G. A. Petsko, M. Karplus, C. G. Hübner, & D. Kern.** 2007a. Intrinsic motions along an enzymatic reaction trajectory. *Nature* **450**: 838-44.
- Henzler-Wildman K.A., M. Lei, V. Thai, S. J. Kerns, M. Karplus, & D. Kern.** 2007b. A hierarchy of timescales in protein dynamics is linked to enzyme catalysis. *Nature* **450**: 913-6.
- folde. *FEBS Lett* **510**:133-5.
- Hoelzer, G. A., E. Smith, & J. W. Pepper.** 2006. On the logical relationship between natural selection and self-organization. *J Evol Biol* **19**:1785-94.
- Holland, P. W.** 1999. Gene duplication: past, present and future. *Semin Cell Dev Biol* **10**:541-7.
- Honig, B.** 1999. Protein folding: from the Levinthal paradox to structure prediction. *J Mol Biol* **293**:283-93.
- Horowitz, N. H.** 1945. On the Evolution of Biochemical Syntheses. *Proc Natl Acad Sci U S A* **31**:153-7.
- Huang C.Y., C. Y. Lee, H. C. Wu, M. H. Kuo, & C. Y. Lai.** 2008. Interactions of chaperonin with a weakly active anthranilate synthase from the aphid endosymbiont *Buchnera aphidicola*. *Microb Ecol.* **56**:696-703.
- Hughes, A. L.** 2005. Gene duplication and the origin of novel Proteins. *Proc Natl Acad Sci U S A* **102**:8791-2.
- Ifuku O., H. Miyaoka, N. Koga, J. Kishimoto, S. Haze, Y. Wachi, & M. Kajiwara.** 1994. Origin of carbon atoms of biotin. ¹³C-NMR studies on biotin biosynthesis in *Escherichia coli*. *Eur J Biochem* **220**: 585-91.
- Jansonius, J. N.** 1998. Structure, evolution and action of vitamin B6-dependent enzymes. *Curr Opin Struct Biol* **8**:759-69.
- Jensen, R. A.** 1976. Enzyme recruitment in evolution of new function. *Annu Rev Microbiol* **30**:409-25.
- Jensen, R. A., & W. Gu.** 1996. Evolutionary recruitment of biochemically specialized subdivisions of Family I within the protein superfamily of aminotransferases. *J Bacteriol* **178**:2161-71.
- Jeong, H., B. Tombor, R. Albert, Z. N. Oltvai, & A. L. Barabasi.** 2000. The large-scale organization of metabolic networks. *Nature* **407**:651-4.
- Jordan, I. K., I. B. Rogozin, Y. I. Wolf, & E. V. Koonin.** 2002. Essential genes are more evolutionarily conserved than are nonessential genes in bacteria. *Genome Res* **12**:962-8.
- Jürgens C., A. Strom, D. Wegener, S. Hettwer, M. Wilmanns, & R. Sterner.** 2000. Directed evolution of a (beta alpha)₈-barrel enzyme to catalyze related reactions in two different metabolic pathways. *Proc Natl Acad Sci U S A.* **97**:9925-30.
- Kellis, M., B. W. Birren, & E. S. Lander.** 2004. Proof and evolutionary analysis of ancient genome duplication in the yeast *Saccharomyces cerevisiae*. *Nature* **428**:617-24.

- Kerner M.J., D.J., Naylor , Y. Ishihama, T. Maier, H. C. Chang, A. P. Stines, C. Georgopoulos, D. Frishman, M. Hayer-Hartl, M. Mann, F. U. Hartl.** 2005. Proteome-wide analysis of chaperonin-dependent protein folding in *Escherichia coli*. *Cell* **122**:209-20.
- Khersonsky, O., C. Roodveldt, & D. S. Tawfik.** 2006. Enzyme promiscuity: evolutionary and mechanistic aspects. *Curr Opin Chem Biol* **10**:498-508.
- Kirschner, M., & J. Gerhart.** 1998. Evolvability. *Proc Natl Acad Sci U S A* **95**:8420-7.
- Kochhar, S., & P. Christen.** 1992. Mechanism of racemization of amino acids by aspartate aminotransferase. *Eur J Biochem* **203**:563-9.
- Koehl, P., & M. Levitt.** 2002. Protein topology and stability define the space of allowed sequences. *Proc Natl Acad Sci U S A* **99**:1280-5.
- Kondrashov, A. S., S. Sunyaev, & F. A. Kondrashov.** 2002. Dobzhansky-Muller incompatibilities in protein evolution. *Proc Natl Acad Sci U S A* **99**:14878-83.
- Krogan, N. J., G. Cagney, H. Yu, G. Zhong, X. Guo, A. Ignatchenko, J. Li, S. Pu, N. Datta, A. P. Tikuisis, T. Punna, J. M. Peregrin-Alvarez, M. Shales, X. Zhang, M. Davey, M. D. Robinson, A. Paccanaro, J. E. Bray, A. Sheung, B. Beattie, D. P. Richards, V. Canadien, A. Lalev, F. Mena, P. Wong, A. Starostine, M. M. Canete, J. Vlasblom, S. Wu, C. Orsi, S. R. Collins, S. Chandran, R. Haw, J. J. Rilstone, K. Gandi, N. J. Thompson, G. Musso, P. St Onge, S. Ghanny, M. H. Lam, G. Butland, A. M. Altaf-Ul, S. Kanaya, A. Shilatifard, E. O'Shea, J. S. Weissman, C. J. Ingles, T. R. Hughes, J. Parkinson, M. Gerstein, S. J. Wodak, A. Emili, & J. F. Greenblatt.** 2006. Global landscape of protein complexes in the yeast *Saccharomyces cerevisiae*. *Nature* **440**:637-43.
- Labeledan, B., & M. Riley.** 1995. Gene products of *Escherichia coli*: sequence comparisons and common ancestries. *Mol Biol Evol* **12**:980-7.
- Laber, B., K. P. Gerbling, C. Harde, K. H. Neff, E. Nordhoff, & H. D. Pohlenz.** 1994. Mechanisms of interaction of *Escherichia coli* threonine synthase with substrates and inhibitors. *Biochemistry* **33**:3413-23.
- Lamble H. J., C. C. Milburn, G. L. Taylor, D W. Hough, & M. J. Danson.** 2004. Gluconate dehydratase from the promiscuous Entner-Doudoroff pathway in *Sulfolobus solfataricus*. *FEBS Lett.* 2004 Oct 8;576(1-2):133-6.
- Lazcano, A., & S. L. Miller.** 1999. On the origin of metabolic pathways. *J Mol Evol* **49**:424-31.
- Lee D., A. Grant, R. L. Marsden, & C. Orengo.** 2005. Identification and distribution of protein families in 120 completed genomes using Gene3D. *Proteins* **59**: 603-15.
- Lee, D., A. Grant, D. Buchan, & C. Orengo.** 2003. A structural perspective on genome evolution. *Curr Opin Struct Biol* **13**:359-69.
- Leung D.W., E. Chen, & D. V Goeddel.** 1989. A method for random mutagenesis of a defined DNA segment using a modified polymerase chain reaction. *Technique* **1**: 11-15.
- Levinthal C.** Are there pathways for protein folding? 1968. *Extrait du Journal de Chimie Physique* **65**: 44.
- Light, S., & P. Kraulis.** 2004. Network analysis of metabolic enzyme evolution in *Escherichia coli*. *BMC Bioinformatics* **5**:15.
- Light, S., P. Kraulis, & A. Elofsson.** 2005. Preferential attachment in the evolution of metabolic networks. *BMC Genomics* **6**:159.

- Liu, X., K. Fan, & W. Wang.** 2004. The number of protein folds and their distribution over families in *Nature. Proteins* **54**:491-9.
- Loh, E., J. Choe, & L. A. Loeb.** 2007. Highly tolerated amino acid substitutions increase the fidelity of *Escherichia coli* DNA polymerase I. *J Biol Chem* **282**:12201-9.
- Long, M.** 2001. Evolution of novel genes. *Curr Opin Genet Dev* **11**:673-80.
- Long, M., E. Betran, K. Thornton, & W. Wang.** 2003. The origin of new genes: glimpses from the young and old. *Nat Rev Genet* **4**:865-75.
- Luisi, P. L.** 2003. Contingency and determinism. *Philos Transact A Math Phys Eng Sci* **361**:1141-7.
- Link A. J., D. Phillips, & G. M. Church.** 1997. Methods for generating precise deletions and insertions in the genome of wild-type *Escherichia coli*: application to open reading frame characterization. *J Bacteriol* **179**: 6228-37.
- Makarova, K. S., Y. I. Wolf, S. L. Mekhedov, B. G. Mirkin, & E. V. Koonin.** 2005. Ancestral paralogs and pseudoparalogs and their role in the emergence of the eukaryotic *Cell. Nucleic Acids Res* **33**:4626-38.
- Makinoshima, H. A., A. Nishimura, & A. Ishihama.** 2002. Fractionation of *Escherichia coli* populations at different stages during growth transition to stationary phase. *Mol Microbiol* **43**: 269-79.
- Massingham, T., L. J. Davies, & P. Lio.** 2001. Analysing gene function after duplication. *Bioessays* **23**:873-6.
- Matsumura, I., & A. D. Ellington.** 2001. In vitro evolution of beta-glucuronidase into a beta-galactosidase proceeds through non-specific intermediates. *J Mol Biol* **305**:331-9.
- Mehta, P. K., & P. Christen.** 2000. The molecular evolution of pyridoxal-5'-phosphate-dependent enzymes. *Adv Enzymol Relat Areas Mol Biol* **74**:129-84.
- Miller, B. G., & R. T. Raines.** 2004. Identifying latent enzyme activities: substrate ambiguity within modern bacterial sugar kinases. *Biochemistry* **43**:6387-92.
- Moran, N. A.** 2003. Tracing the evolution of gene loss in obligate bacterial symbionts. *Curr Opin Microbiol* **6**:512-8.
- Moran N. A., & P. Baumann.** 2000. Bacterial endosymbionts in animals. *Curr Opin Microbiol* **3**: 270-75.
- Morett, E., G. Saab-Rincon, L. Olvera, M. Olvera, H. Flores, & R. Grande.** 2008. Sensitive genome-wide screen for low secondary enzymatic activities: the YjbQ family shows thiamin phosphate synthase activity. *J Mol Biol* **376**:839-53.
- Nagano, N., C. A. Orengo, & J. M. Thornton.** 2002. One fold with many functions: the evolutionary relationships between TIM barrel families based on their sequences, structures and functions. *J Mol Biol* **321**:741-65.
- Nei, M.** 2005. Selectionism and neutralism in molecular evolution. *Mol Biol Evol* **22**:2318-42.
- Neitz, M., & J. Neitz.** 1995. Numbers and ratios of visual pigment genes for normal red-green color vision. *Science* **267**:1013-6.
- Nilsson, A. I., S. Koskiniemi, S. Eriksson, E. Kugelberg, J. C. Hinton, & D. I. Andersson.** 2005. Bacterial genome size reduction by experimental evolution. *Proc Natl Acad Sci U S A* **102**:12112-6.
- Nobeli, I., R. V. Spriggs, R. A. George, & J. M. Thornton.** 2005. A ligand-centric analysis of the diversity and evolution of protein-ligand relationships in *E. coli*. *J Mol Biol* **347**:415-36.
- O'Brien, P. J., & D. Herschlag.** 1999. Catalytic promiscuity and the evolution of new enzymatic activities. *Chem Biol* **6**:R91-R105.
- O'Grady, R. T.** 1984. Evolutionary theory and teleology. *J Theor Biol* **107**:563-78.

- O'Loughlin, T. L., W. M. Patrick, & I. Matsumura.** 2006. Natural history as a predictor of protein evolvability. *Protein Eng Des Sel* **19**:439-42.
- Ohta, T.** 1992. The nearly neutral theory of molecular evolution. *Ann Rev Ecol Syst* **23**:263-286.
- Omburo, G. A., J. M. Kuo, L. S. Mullins, & F. M. Raushel.** 1992. Characterization of the zinc binding site of bacterial phosphotriesterase. *J Biol Chem* **267**:13278-83.
- Onuchic, J. N., Z. Luthey-Schulten, & P. G. Wolynes.** 1997. Theory of protein folding: the energy landscape perspective. *Annu Rev Phys Chem* **48**:545-600.
- Onuchic, J. N., & P. G. Wolynes.** 2004. Theory of protein folding. *Curr Opin Struct Biol* **14**:70-5.
- Orengo, C. A., F. M. Pearl, J. E. Bray, A. E. Todd, A. C. Martin, L. Lo Conte, & J. M. Thornton.** 1999. The CATH Database provides insights into protein structure/function relationships. *Nucleic Acids Res* **27**:275-9.
- Orengo, C. A., A. E. Todd, & J. M. Thornton.** 1999. From protein structure to function. *Curr Opin Struct Biol* **9**:374-82.
- Orengo C.A., & J. M. Thornton.** 2005. Protein families and their evolution-a structural perspective. *Annu Rev Biochem* **74**: 867-900
- Orr, H. A.** 2005. The genetic theory of adaptation: a brief history. *Nat Rev Genet* **6**:119-27.
- Otsuka A.J., M. R. Buoncristiani, P. K. Howard, J. Flamm, C Johnson, R. Yamamoto, K. Uchida, C. Cook, J. Ruppert, J. Matsuzaki.** 1988. The *Escherichia coli* biotin biosynthetic enzyme sequences predicted from the nucleotide sequence of the bio operon. *J Biol Chem* **263**: 19577-85.
- Pal, C., B. Papp, & L. D. Hurst.** 2003. Genomic function: Rate of evolution and gene dispensability. *Nature* **421**:496-7; discussion 497-8.
- Pal, C., B. Papp, & L. D. Hurst.** 2001. Highly expressed genes in yeast evolve slowly. *Genetics* **158**:927-31.
- Pal, C., B. Papp, & M. J. Lercher.** 2006. An integrated view of protein evolution. *Nat Rev Genet* **7**:337-48.
- Palmer, D. R., J. B. Garrett, V. Sharma, R. Meganathan, P. C. Babbitt, & J. A. Gerlt.** 1999. Unexpected divergence of enzyme function and sequence: "N-acylamino acid racemase" is o-succinylbenzoate synthase. *Biochemistry* **38**:4252-8.
- Pande, V. S., A. Grosberg, T. Tanaka, & D. S. Rokhsar.** 1998. Pathways for protein folding: is a new view needed? *Curr Opin Struct Biol* **8**:68-79.
- Panek, H., & M. R. O'Brian.** 2002. A whole genome view of prokaryotic haem biosynthesis. *Microbiology* **148**:2273-82.
- Parkhill, J., G. Dougan, K. D. James, N. R. Thomson, D. Pickard, J. Wain, C. Churcher, K. L. Mungall, S. D. Bentley, M. T. Holden, M. Sebahia, S. Baker, D. Basham, K. Brooks, T. Chillingworth, P. Connerton, A. Cronin, P. Davis, R. M. Davies, L. Dowd, N. White, J. Farrar, T. Feltwell, N. Hamlin, A. Haque, T. T. Hien, S. Holroyd, K. Jagels, A. Krogh, T. S. Larsen, S. Leather, S. Moule, P. O'Gaora, C. Parry, M. Quail, K. Rutherford, M. Simmonds, J. Skelton, K. Stevens, S. Whitehead, & B. G. Barrell.** 2001. Complete genome sequence of a multiple drug resistant *Salmonella enterica* serovar Typhi CT18. *Nature* **413**:848-52.
- Parkhill, J., M. Sebahia, A. Preston, L. D. Murphy, N. Thomson, D. E. Harris, M. T. Holden, C. M. Churcher, S. D. Bentley, K. L. Mungall, A. M. Cerdeno-Tarraga, L. Temple, K. James, B. Harris, M. A. Quail, M. Achtman, R. Atkin, S. Baker, D. Basham, N. Bason, I. Cherevach, T. Chillingworth, M. Collins, A. Cronin, P. Davis, J. Doggett, T. Feltwell, A. Goble, N. Hamlin, H. Hauser, S. Holroyd, K. Jagels, S. Leather, S. Moule, H. Norberczak, S. O'Neil, D. Ormond, C. Price, E. Rabinowitsch, S. Rutter, M. Sanders, D. Saunders, K. Seeger, S. Sharp, M. Simmonds, J. Skelton, R. Squares, S. Squares, K. Stevens, L. Unwin, S. Whitehead, B. G. Barrell, & D. J. Maskell.** 2003. Comparative analysis

of the genome sequences of *Bordetella pertussis*, *Bordetella parapertussis* and *Bordetella bronchiseptica*. *Nat Genet* **35**:32-40.

Patthy L. 1999. Protein Evolution. *Wiley Blackwell Science, Oxford, U. K.*

Perುದani, R., & A. Peracchi. 2003. A genomic overview of pyridoxal-phosphate-dependent enzymes. *EMBO Rep* **4**:850-4.

Perona, J. J., & C. S. Craik. 1995. Structural basis of substrate specificity in the serine proteases. *Protein Sci* **4**:337-60.

Ploux O., & A. Marquet. 1996. Mechanistic studies on the 8-amino-7-oxopelargonate synthase, a pyridoxal-5'-phosphate-dependent enzyme involved in biotin biosynthesis. *Eur J Biochem* **236**:301-8.

Pollack, J. D., M. A. Myers, T. Dandekar, & R. Herrmann. 2002. Suspected utility of enzymes with multiple activities in the small genome *Mycoplasma* species: the replacement of the missing "household" nucleoside diphosphate kinase gene and activity by glycolytic kinases. *Omic* **6**:247-58.

Poon, A., & L. Chao. 2005. The rate of compensatory mutation in the DNA bacteriophage phiX174. *Genetics* **170**:989-99.

Poon, A., B. H. Davis, & L. Chao. 2005. The coupon collector and the suppressor mutation: estimating the number of compensatory mutations by maximum likelihood. *Genetics* **170**:1323-32.

Rauschel, F. M., & H. M. Holden. 2000. Phosphotriesterase: an enzyme in search of its natural substrate. *Adv Enzymol Relat Areas Mol Biol* **74**:51-93.

Ravasz, E., A. L. Somera, D. A. Mongru, Z. N. Oltvai, & A. L. Barabasi. 2002. Hierarchical organization of modularity in metabolic networks. *Science* **297**:1551-5.

Riley M., T. Abe, M. B. Arnaud, M. K. Berlyn, F. R. Blattner, R. R. Chaudhuri, J. D. Glasner, T. Horiuchi, I. M. Keseler, T. Kosuge, H. Mori, N. T. Perna, G. Plunkett 3rd, K. E. Rudd, M. H. Serres, G. H. Thomas, N. R. Thomson, D. Wishart, & B. L. Wanner. 2006. Escherichia coli K-12: a cooperatively developed annotation snapshot--2005. *Nucleic Acids Res* **34**:1-9.

Rison, S. C., S. A. Teichmann, & J. M. Thornton. 2002. Homology, pathway distance and chromosomal localization of the small molecule metabolism enzymes in *Escherichia coli*. *J Mol Biol* **318**:911-32.

Rison, S. C., & J. M. Thornton. 2002. Pathway evolution, structurally speaking. *Curr Opin Struct Biol* **12**:374-82.

Rocha, E. P., & A. Danchin. 2004. An analysis of determinants of amino acids substitution rates in bacterial Proteins. *Mol Biol Evol* **21**:108-16.

Roodveldt, C., & D. S. Tawfik. 2005. Shared promiscuous activities and evolutionary features in various members of the amidohydrolase superfamily. *Biochemistry* **44**:12728-36.

Ruiz-Mirazo K., J. Peretó, & A. Moreno. 2004. A universal definition of life: autonomy and open-ended evolution. *Origins of Life and Evolution of the Biosphere* **34**: 323-346.

Rutherford S. L. 2003. Between genotype and phenotype: Protein chaperones and evolvability. *Nat Rev* **4**: 263-74.

Saen-Oon S, Ghanem M, Schramm VL, Schwartz SD. 2008. Remote mutations and active site dynamics correlate with catalytic properties of purine nucleoside phosphorylase. *Biophys J* **94**: 4078-88.

Sankoff, D. 2001. Gene and genome duplication. *Curr Opin Genet Dev* **11**:681-4.

Saqi, M. A., & M. J. Sternberg. 2001. A structural census of metabolic networks for *E. coli*. *J Mol Biol* **313**:1195-206.

- Sambrook J., E. F. Fritsch, & T. Maniatis.** 1989. Molecular cloning: A laboratory manual. Second edition. Cold Spring Harbor Laboratory Press.
- Scanlan, T. S., & R. C. Reid.** 1995. Evolution in action. *Chem Biol* **2**:71-5.
- Schmidt, D. M., E. C. Mundorff, M. Dojka, E. Bermudez, J. E. Ness, S. Govindarajan, P. C. Babbitt, J. Minshull, & J. A. Gerlt.** 2003a. Evolutionary potential of (beta/alpha)₈-barrels: functional promiscuity produced by single substitutions in the enolase superfamily. *Biochemistry* **42**:8387-93.
- Schmidt, S., S. Sunyaev, P. Bork, & T. Dandekar.** 2003b. Metabolites: a helping hand for pathway evolution? *Trends Biochem Sci* **28**:336-41.
- Schneider G., & Y. Lindqvist.** 2001. Structural enzymology of biotin biosynthesis. *FEBS Letters* **495**: 7-11.
- Shakhnovich, B. E., N. V. Dokholyan, C. DeLisi, & E. I. Shakhnovich.** 2003. Functional fingerprints of folds: evidence for correlated structure-function evolution. *J Mol Biol* **326**:1-9.
- Shanahan, T.** 2008. Why don't zebras have machine guns? Adaptation, selection, and constraints in evolutionary theory. *Stud Hist Philos Biol Biomed Sci* **39**:135-46.
- Shoichet, B. K., W. A. Baase, R. Kuroki, & B. W. Matthews.** 1995. A relationship between protein stability and protein function. *Proc Natl Acad Sci U S A* **92**:452-6.
- Shoolingin-Jordan, P. M., S. Al-Daihan, D. Alexeev, R. L. Baxter, S. S. Bottomley, I. D. Kahari, I. Roy, M. Sarwar, L. Sawyer, & S. F. Wang.** 2003. 5-Aminolevulinic acid synthase: mechanism, mutations and medicine. *Biochim Biophys Acta* **1647**:361-6.
- Skarstedt M.T., & S. B. Greer.** 1973. Threonine synthetase of *Bacillus subtilis*. The nature of an associated dehydratase activity. *J Biol Chem* **248**: 1032-44.
- Smith, A. J., R. Muller, M. D. Toscano, P. Kast, H. W. Hellinga, D. Hilvert, & K. N. Houk.** 2008. Structural reorganization and preorganization in enzyme active sites: comparisons of experimental and theoretically ideal active site geometries in the multistep serine esterase reaction cycle. *J Am Chem Soc* **130**:15361-73.
- Sonnhammer, E. L., & E. V. Koonin.** 2002. Orthology, paralogy and proposed classification for paralog subtypes. *Trends Genet* **18**:619-20.
- Sterner R., & W. Liebl.** 2001. Thermophilic adaptation of proteins. *Crit Rev Biochem Mol Biol* **36**: 39-106.
- Streit, W. R., & P. Entcheva.** 2003. Biotin in microbes, the genes involved in its biosynthesis, its biochemical role and perspectives for biotechnological production. *Appl Microbiol Biotechnol* **61**:21-31.
- Stríšovský K., J. Jirásková, C. Barinka, P. Majer, C. Rojas, B. S. Slusher, & J. Konvalinka.** 2003. Mouse brain serine racemase catalyzes specific elimination of L-serine to pyruvate. *FEBS Lett.* **535**: 44-8.
- Takano, T., S. Kusakabe, A. Koga, & T. Mukai.** 1989. Polymorphism for the number of tandemly multiplied glycerol-3-phosphate dehydrogenase genes in *Drosophila melanogaster*. *Proc Natl Acad Sci U S A* **86**:5000-4.
- Taylor, J. S., & J. Raes.** 2004. Duplication and divergence: the evolution of new genes and old ideas. *Annu Rev Genet* **38**:615-43.
- Teichmann, S. A., S. C. Rison, J. M. Thornton, M. Riley, J. Gough, & C. Chothia.** 2001. The evolution and structural anatomy of the small molecule metabolic pathways in *Escherichia coli*. *J Mol Biol* **311**:693-708.
- Thornton, J. M., C. A. Orengo, A. E. Todd, & F. M. Pearl.** 1999. Protein folds, functions and evolution. *J Mol Biol* **293**:333-42.
- Todd, A. E., C. A. Orengo, & J. M. Thornton.** 2001. Evolution of function in protein superfamilies, from a structural perspective. *J Mol Biol* **307**:1113-43.

- Tokuriki, N., F. Stricher, J. Schymkowitz, L. Serrano, & D. S. Tawfik.** 2007. The stability effects of protein mutations appear to be universally distributed. *J Mol Biol* **369**:1318-32.
- Tsoka, S., & C. A. Ouzounis.** 2001. Functional versatility and molecular diversity of the metabolic map of *Escherichia coli*. *Genome Res* **11**:1503-10.
- Tokuriki N., & D. S. Tawfik.** 2009. Chaperonin overexpression promotes genetic variation and enzyme evolution. *Nature* 459: 668-73.
- Tsokolov S. A.** 2009. Why is the definition of life so elusive?: Epistemological considerations. *Astrobiology* **9**: 401-412.
- Tung W.L., & K. C. Chow.** A modified medium for efficient electrotransformation of *E. coli*. *TIG* 1995, (11)4:128-9.
- Vacca R.A., S. Giannattasio, R. Graber, E. Sandmeier, E. Marra, & P. Christen.** 1997. Active-site Arg --> Lys substitutions alter reaction and substrate specificity of aspartate aminotransferase. *J Biol Chem.* 272: 21932-7.
- Van Arsdell S. W., J. B. Perkins, R. R. Yocum, L. Luan, C. L. Howitt, N. P. Chatterjee, & J. G. Pero.** 2005. Removing a bottleneck in the *Bacillus subtilis* biotin pathway: BioA utilizes lysine rather than S-adenosylmethionine as the amino donor in the KAPA-to-DAPA reaction. *Biotechnol Bioeng* **91**:75-83.
- Vitkup D., P. Kharchenko, & A. Wagner.** 2006. Influence of metabolic network structure and function on enzyme evolution. *Genome Biol* 7: R39.
- Wagner, A., & D. A. Fell.** 2001. The small world inside large metabolic networks. *Proc Biol Sci* **268**:1803-10.
- Wall, D. P., A. E. Hirsh, H. B. Fraser, J. Kumm, G. Giaever, M. B. Eisen, & M. W. Feldman.** 2005. Functional genomic analysis of the rates of protein evolution. *Proc Natl Acad Sci U S A* **102**:5483-8.
- Warshel, A., P. K. Sharma, M. Kato, Y. Xiang, H. Liu, & M. H. Olsson.** 2006. Electrostatic basis for enzyme catalysis. *Chem Rev* **106**:3210-35.
- White E. M., A. R. Holland, & G. MacDonald.** 2008. Infrared studies reveal unique vibrations associated with the PGK-ATP-3-PG ternary complex. *Biochemistry* 47: 84-91.
- Wei, J., M. B. Goldberg, V. Burland, M. M. Venkatesan, W. Deng, G. Fournier, G. F. Mayhew, G. Plunkett, 3rd, D. J. Rose, A. Darling, B. Mau, N. T. Perna, S. M. Payne, L. J. Runyen-Janecky, S. Zhou, D. C. Schwartz, & F. R. Blattner.** 2003. Complete genome sequence and comparative genomics of *Shigella flexneri* serotype 2a strain 2457T. *Infect Immun* 71:2775-86.
- Webster S.P., D. Alexeev, D. J. Campopiano, R. M. Watt, M. Alexeeva, L. Sawyer, & R. L. Baxter.** 2000. Mechanism of 8-amino-7-oxononanoate synthase: spectroscopic, kinetic, and crystallographic studies. *Biochemistry* 39: 516-28.
- Wernegreen J.J.** 2002. Genome evolution in bacterial endosymbionts of insects. *Nat Rev Genet* 3: 850-61.
- Winstanley, H. F., S. Abeln, & C. M. Deane.** 2005. How old is your fold? *Bioinformatics* **21 Suppl 1**:i449-58.
- Winther, R. G.** 2008. Systemic darwinism. *Proc Natl Acad Sci U S A* **105**:11833-8.

- Wise, E., W. S. Yew, P. C. Babbitt, J. A. Gerlt, & I. Rayment.** 2002. Homologous (beta/alpha)₈-barrel enzymes that catalyze unrelated reactions: orotidine 5'-monophosphate decarboxylase and 3-keto-L-gulonate 6-phosphate decarboxylase. *Biochemistry* **41**:3861-9.
- Wise, E. L., W. S. Yew, J. Akana, J. A. Gerlt, & I. Rayment.** 2005. Evolution of enzymatic activities in the orotidine 5'-monophosphate decarboxylase suprafamily: structural basis for catalytic promiscuity in wild-type and designed mutants of 3-keto-L-gulonate 6-phosphate decarboxylase. *Biochemistry* **44**:1816-23.
- Witting, L.** 2008. Inevitable evolution: back to The Origin and beyond the 20th Century paradigm of contingent evolution by historical natural selection. *Biol Rev Camb Philos Soc.*
- Witting, L.** 2003. Major life-history transitions by deterministic directional natural selection. *J Theor Biol* **225**:389-406.
- Wolfe, K. H., & D. C. Shields.** 1997. Molecular evidence for an ancient duplication of the entire yeast genome. *Nature* **387**:708-13.
- Wolfenden, R., & M. J. Snider.** 2001. The depth of chemical time and the power of enzymes as catalysts. *Acc Chem Res* **34**:938-45.
- Ycas, M.** 1974. On earlier states of the biochemical system. *J Theor Biol* **44**:145-60.
- Yew, W. S., J. Akana, E. L. Wise, I. Rayment, & J. A. Gerlt.** 2005. Evolution of enzymatic activities in the orotidine 5'-monophosphate decarboxylase suprafamily: enhancing the promiscuous D-arabino-hex-3-ulose 6-phosphate synthase reaction catalyzed by 3-keto-L-gulonate 6-phosphate decarboxylase. *Biochemistry* **44**:1807-15.
- Zientz E., T. Dandekar, & R. Gross.** 2004. Metabolic interdependence of obligate intracellular bacteria and their insect hosts. *Microbiol Mol Biol Rev* **68**:745-70.
- Zwanzig, R.** 1997. Two-state models of protein folding kinetics. *Proc Natl Acad Sci U S A* **94**:148-50.

Protocolos

Extracción de DNA cromosómico. La extracción se basó en el protocolo descrito en Sambrook *et al.* (1989) para la extracción de DNA de cultivos celulares en medio líquido. Se transfirió un cultivo de 3 mL ($DO_{540} \sim 1.0$) a tubos eppendorf en volúmenes de 1 mL. Las células se cosecharon mediante centrifugación a 15,000 rpm, por 1 min, y se resuspendieron en 500 μ L de una mezcla Tris-EDTA (10:1). Se repitió el paso anterior. Después se agregaron las siguientes disoluciones: 10 μ L de Proteinasa K (20 mg mL^{-1}), 10 μ L de RNasa pancreática ($20 \text{ } \mu\text{g mL}^{-1}$) y 60 μ L de SDS (10%). El contenido se mezcló por agitación y se incubó a 37 °C hasta que se tornó transparente; posteriormente se pasó por una jeringa de 1000 U para disminuir su viscosidad.

Se usó fenol y cloroformo para eliminar restos de proteína. Se agregaron 800 μ L de una mezcla de fenol-cloroformo (1:1); el contenido se agitó manualmente hasta la formación de una emulsión. Tras una centrifugación de 3 min (15,000 rpm), la fase superior del contenido se transfirió a un tubo nuevo. Se repitió el procedimiento una vez. Se agregaron 800 μ L de cloroformo y se centrifugó nuevamente. La fase superior se transfirió a un tubo nuevo. El contenido se igualó a 500 μ l con Tris-EDTA. El DNA se precipitó sucesivamente con etanol absoluto y al 70 %. En cada precipitación, el DNA y el etanol se mezclaron por agitación; el etanol se decantó tras una centrifugación de 10 min (15,000 rpm). El DNA se secó con calor (60°C) y vacío, por 15 minutos. El DNA se resuspendió en 100 μ L de agua bidestilada.

Extracción de plásmido. Para separar el plásmido de las células se empleó el protocolo comercial 'High Pure Plasmid Isolation Kit' de Roche. Volúmenes de 2 mL de cultivo ($DO_{540} \sim 1.0$) se concentraron en tubos eppendorf mediante centrifugación a 9,000 rpm, por 30 s. Se decantó el sobrenadante y las células se resuspendieron en 250 μ L de amortiguador 1. El contenido se mezcló por inversión y se le agregaron 250 μ L de amortiguador de lisis. El contenido se mezcló por inversión y se incubó a temperatura ambiental por 5 min. Se añadieron 350 μ L de amortiguador de unión enfriado previamente en hielo. El contenido se mezcló por inversión y se incubó en hielo por 5 min. Después se centrifugó (15,000 rpm) por 10 min. La fase líquida se transfirió a la columna-filtro provista por el fabricante. Siguió una centrifugación de 1 min y la eliminación del líquido. Después se agregaron 500 μ L de amortiguador de lavado 1. El amortiguador se eliminó después por centrifugación durante 5 s. Se agregaron 700 μ L de amortiguador de lavado 2. Se hicieron dos centrifugaciones de 1 minuto para eliminar todo el líquido. Las columnas fueron secadas por calor y vacío, durante 5 minutos. Finalmente, el filtro se ajustó a un tubo nuevo y se agregaron 100 μ L de amortiguador de elución pre-calentado. El tubo se centrifugó para recuperar el plásmido.

Purificación de DNA a partir de reacciones de PCR y de geles de electroforesis. Se usó el protocolo comercial de Roche 'High Pure PCR Isolation Kit'. Para purificar a partir de geles de electroforesis, se usaron 300 μL de amortiguador de unión por cada 100 mg de agarosa; el fragmento se disolvió mediante incubación a 65 °C con agitación continua a 1500 rpm, en un Termomixer (Eppendorf). La mezcla se transfirió a la columna y se centrifugó (15,000 rpm). Se repitió el paso anterior con el fin de promover la unión de más DNA a la columna. Después se agregaron 500 μL de amortiguador de lavado 1. Tras centrifugar por 1 min, se agregaron 700 μL de amortiguador de lavado 2. Después de centrifugar, se agregaron 100 μL de amortiguador de elución pre-calentado. El DNA se recuperó en un tubo nuevo.

Digestión y Ligación de DNA. Las reacciones de digestión se prepararon con las proporciones siguientes: 1- 5 μg de DNA; 3- 20 u de enzima de restricción; 5 μL de amortiguador de reacción 10x; agua hasta 50 μL . Las digestiones preparativas se incubaron por 8 h, y las analíticas por 2 h. Después, las enzimas se desactivaron mediante incubación a 65 °C, por 20 minutos. Las reacciones de ligación se prepararon en volúmenes de 20 μL : 1 μL de ligasa de bacteriófago T4; inserto y vector en una proporción molar 3:1 (1-10 $\mu\text{g mL}^{-1}$ total); 2 μL de amortiguador de ligación 10x [Tris-HCl (500 mM), MgCl_2 (100 mM), ATP (10 mM), Ditioneitol (100 mM), BSA (250 $\mu\text{g mL}^{-1}$)]. Las reacciones se incubaron a 15 °C, por 16 horas. La ligasa se desactivó incubándola a 65 °C, por 10 min. Se agregó 1 mL de 1-butanol para precipitar el DNA. El contenido se mezcló por agitación; tras centrifugar a 15,000 rpm, por 10 min, el 1-butanol se decantó y la pastilla se secó por vacío y calor durante 10 min. El DNA se resuspendió en 10 μL de agua bidestilada. La eficiencia de ligación se evaluó mediante electroporación y cuantificación de transformantes. En cada transformación se usaron 2 μL de reacción de ligación. Para evaluar la eficiencia de las digestiones, se realizaron ligaciones en las que no se incluyó inserto; se estimó el porcentaje de vector re-ligado.

Preparación de células electrocompetentes. Las células se lavaron extensivamente para eliminar sales. Se inocularon 50 mL de un pre-cultivo en 950 mL de medio LB-estreptomina pre-calentado a 37 °C. Este cultivo se incubó hasta una $\text{D.O.}_{600} \approx 1.0$; después se mantuvo en hielo por 15 minutos. Desde este punto, la manipulación de las células se realizó en frío (centrífuga, recipientes, agua y disoluciones a 4 °C). El cultivo se distribuyó en 4 tubos. Las células se cosecharon mediante centrifugación (4,000 rpm) por 10 minutos. Se decantó el sobrenadante y las células se resuspendieron en volúmenes de 500 mL de agua bidestilada y esterilizada por autoclave. Las células se cosecharon por centrifugación y se concentraron en 2 volúmenes de 250 mL de glicerol esterilizado (10%). El paso anterior se repitió una vez. Se realizó una tercer centrifugación, tras la cual las células se resuspendieron en 2 mL de una mezcla 1:1 de medio GYT y glicerol (10%). Se prepararon alícuotas de 60 μL en tubos eppendorf. Los tubos se congelaron en hielo seco y se

almacenaron a $-75\text{ }^{\circ}\text{C}$ para su uso ulterior. La competencia de las células se determinó por electroporación en presencia de 1 ng de pUC18. Como control negativo se hicieron electroporaciones sin plásmido. En general, se observaron entre 10^7 a 10^9 transformantes por ng de DNA. Se usó una ultracentrífuga Beckman (modelo J2-21) con rotor JA14. Composición del medio GYT: glicerol (10%), extracto de levadura (0.125 %), tripona (0.25 %).

Transformación. Las transformaciones se lograron mediante electroporación, de acuerdo con las indicaciones de Tung & Chow (1995). Los medios se prepararon de acuerdo con Hanahan (1982) y Sambrook *et al.* (1989). Las células electrocompetentes y el plásmido se descongelaron en hielo. En cada transformación se usaron 0.5-1.0 ng de plásmido (1-2.5 μL de muestra). El plásmido y las células se mezclaron en celdas enfriadas previamente. Se aplicaron pulsos eléctricos de 2.5 kV cm^{-1} (25 μF de capacitancia y $200\ \Omega$ de resistencia). Tras cada pulso, las células se mezclaron con 1 mL de medio SOC [Super optimal broth with catabolite repression; triptona (2%), extracto de levadura (0.5%), NaCl (10 mM), KCl (2.5 mM), MgSO_4 (10 mM), MgCl_2 (10 mM), glucosa (10 mM);]. Tras una incubación a $37\text{ }^{\circ}\text{C}$, por 1 h, se sembraron alícuotas en medio sólido LB-estreptomicina-ampicilina para seleccionar transformantes. Se usaron celdas desechables (0.1 cm de separación de electrodo) de BioRad.

Electroforesis de DNA. Todos los corrimientos se realizaron a 60 V (corriente de $\sim 122\text{ mA}$). Se usaron geles de agarosa (1 %) y amortiguador TBE [Tris (89 mM), H_3BO_3 (89 mM), EDTA (2 mM)]. Las muestras de DNA se mezclaron con agua y amortiguador de carga 6X [azul de bromofenol (0.3 %), xilen-cianol (0.03 %), glicerol (10 %), EDTA (100 mM; pH 7.6, ajustado con Tris)]. Se usaron los marcadores comerciales 1 Kb Plus Ladder (Invitrogen) y Gene RulerTM DNA Ladder Mix (Fermentas Life Sciences). Para la visualización del DNA, el gel se incubó por 5 min en bromuro de etidio ($25\ \mu\text{g mL}^{-1}$), se lavó en el amortiguador de corrimiento y se expuso en una lámpara de luz ultravioleta (Hoefler, Macro Vue Uv25). Los corrimientos se fotografiaron con un equipo Eagle EyeTM de Stratagene.

Electroforesis de Proteínas

La separación electroforética se logró con geles de Tris/glicina SDS-poliacrilamida, de acuerdo con Bollag *et al.* (1996). Se hicieron geles de 0.75 mm de espesor. Para la parte inferior se mezclaron: 2.5 mL de disolución A (acrilamida 30%, *N,N'*-metileno- bis-acrilamida 0.8%); 1.25 mL Disolución B 4x [Tris-HCl (1.5 M, pH 8.8); SDS (0.4 %)]; 50 μL de $(\text{NH}_4)_2\text{S}_2\text{O}_8$ (10%); 5 μL de TEMED; agua hasta un volumen de 5 mL. Para la parte superior se usaron: 670 μL de disolución A; 1 mL de disolución C 4x [Tris-HCl (0.5 M, pH 6.8); SDS (0.4%)]; 30 μL de Persulfato de amonio (10%); 5 μL de TEMED (*N, N, N', N'*-tetrametil-etano-1, 2-diamina), y; agua hasta 4 mL. Las muestras de proteína se mezclaron con amortiguador de carga 5X [Tris-

HCl (60 mM), Glicerol (25 %), SDS (2 %), 2-mercaptoetanol (14.4 mM), azul de bromofenol (0.1 %)]. Antes de cargarse, las muestras se incubaron en agua en ebullición, por 10', y se centrifugaron brevemente. El corrimiento se hizo a voltaje constante (200 V) con corriente de 100-60 mA. Amortiguador de electroforesis: [Tris (25 mM), Glicina (192 mM, pH 8.8), SDS (0.1 %); pH 8.3].