

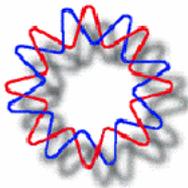


UNIVERSIDAD NACIONAL AUTONOMA DE MEXICO

INSTITUTO DE BIOTECNOLOGÍA

Ampliación de la especificidad de la enzima ácido 7,8-diaminopelargónico sintasa de *E. coli*, para demostrar la posible existencia de intermediarios no específicos

T E S I S
QUE PARA OBTENER EL GRADO DE:
MAESTRO EN CIENCIAS BIOQUIMICAS
P R E S E N T A
QFB. CHRISTIAN TORRES SOSA



TUTOR
DR. J. ENRIQUE MORETT SÁNCHEZ.

Cuernavaca, Mor.

2006



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.

Dedicada a mis tres madres: Olga, Charo y Loli.
A mi papá, a Richi y a Ximena.

Agradecimientos

A mis amigos del laboratorio:

Enrique, agradezco toda la paciencia que invertiste en mí; Humberto, te agradezco todas las asesorías que me diste. Richard, Lety y Mary, porque tuvieron toda la paciencia para enseñarme cuando llegué al laboratorio; Ángel y Alfredo, porque son excelentes compañeros; Gabriel, por aquellas salidas en bicicleta; Doña Juanita y Pancho, que siempre me facilitaron la vida al tener listo el material; Nelly, por todos los trámites que realizaste.

Algunos amigos que me ayudaron a realizar el presente trabajo:

Javier, tu ayuda para poder construir las redes fue crucial para completar este trabajo y especialmente por haberme brindado tu amistad.

Mónica Ivonne, me ayudaste a salir adelante. Nunca lo he de olvidar.

A mis amigos del laboratorio del Dr. Soberón:

Manuel, Fili, Adrián, Anne-Laure, Juanita, Adriana, Etienne, Azucena, Biviana, Lianet, Joelito, Giovanni, Joel, Paco y Gloria.

A mis amigos de síntesis:

Paúl, Santiago y Eugenio.

A mis amigos no relacionados a este ámbito (sin orden de importancia):

Gaby, porque ya estás en Cuernavaca; Za-Nichté, por esas infinitas charlas; Martha, porque no puedo explicar seis años de mi vida sin ti; Rocío, aunque le vayas al Pachuca; Laura, porque fuiste muy buena anfitriona cuando me encontraba en Chapingo; Violeta y Karina, porque son todo un caso; Mónica, por ser una gran flautista; Carmen, por aquél día en la bodeguita; Chucho, Sergio e Hilda, por apoyarme en aquél percance automovilístico; Ana María, porque te quise; Alma, por tu nobleza; Ángeles, por ser una gran fotógrafa; Sandra, por aquellos momentos en la facultad; Esmeralda, porque fuiste una excelente compañera durante tu estancia en el laboratorio.

A mis tíos y primos, cuya lista es extraordinariamente extensa.

ÍNDICE

Resumen	1
Introducción	2
Duplicación génica	3
<i>El destino de los segmentos duplicados</i>	4
<i>Sobre los dominios proteicos</i>	6
Homólogos, ortólogos y parálogos	7
Los modelos de evolución de las vías metabólicas	8
La naturaleza no suele reinventar, recicla las soluciones	12
<i>La especificidad hacia el sustrato es la estrategia dominante</i>	13
<i>El mecanismo químico es la estrategia dominante</i>	14
<i>La arquitectura del sitio activo es la estrategia dominante</i>	14
Las enzimas dependientes de piridoxal fosfato (PLP) catalizan reacciones muy diversas	15
Las enzimas dependientes de PLP fueron reinventadas al menos 5 veces durante la evolución	16
<i>Plegamiento de tipo I</i>	17
<i>Plegamiento de tipo II</i>	18
<i>Plegamiento de tipo III (β/α)₈</i>	19
<i>Plegamiento de tipo IV</i>	19
<i>Plegamiento de tipo V</i>	20
Síntesis de biotina en <i>E. coli</i>	20
<i>Enzima 8-amino-7-oxononanoato sintasa (7-KAPs) codificada por el gen <i>bioF</i></i>	22
<i>Enzima ácido 7,8-diaminopelargónico sintasa (7-DAPA sintasa) codificada por el gen <i>bioA</i></i>	22
<i>Enzima Detiobiotina sintetasa (DTBS) codificada por el gen <i>bioD</i></i>	23
<i>Enzima biotina sintasa codificada por el gen <i>bioB</i></i>	23
Evolución dirigida de proteínas	24
<i>Generación de diversidad</i>	24
<i>PCR mutagénica (epPCR)</i>	25
<i>PCR sexual (DNA shuffling)</i>	26
<i>Mutagénesis de sitio dirigida</i>	27
Planteamiento del problema	30
Objetivos	32
Objetivos particulares	32
Hipótesis	32
Metodología	33
<i>Análisis global de la evolución de las enzimas que unen PLP como cofactor</i>	33
<i>Generación de cepas mutantes de <i>E. coli</i> MC1061Δ<i>bioA</i>/Δ<i>bioF</i> y MC1061Δ<i>bioA</i></i>	34
<i>Evaluación de las cepas mutantes por PCR</i>	36
<i>Evaluación de las cepas mutantes por complementación fenotípica</i>	37
<i>Generación de diversidad</i>	37
<i>Sistema de selección</i>	37
<i>Estrategia para la generación de la librería de variantes de <i>BioA</i> y selección de variantes</i>	38

Resultados y discusión	40
<i>Obtención de las mutantes E. coli MC1061ΔbioA/ΔbioF y MC1061ΔbioA</i>	40
<i>Análisis de las remociones por PCR</i>	41
<i>Análisis de las remociones por complementación</i>	42
<i>Análisis de los genes bioA y bioF en la cepa de E. coli MC1061 wt</i>	43
<i>Análisis de la librería de variantes, primera ronda mutagénica</i>	44
<i>Selección de variantes</i>	50
<i>Análisis de variantes de la primera ronda mutagénica</i>	51
<i>Segunda ronda mutagénica</i>	55
<i>Análisis de la librería de variantes, segunda ronda mutagénica</i>	55
<i>Selección de variantes</i>	57
<i>Análisis global de la evolución de enzimas que unen PLP como cofactor</i>	64
<i>Plegamiento de tipo I</i>	64
<i>Plegamiento de tipo II</i>	79
<i>Plegamiento de tipo III (β/α)₈</i>	81
<i>Plegamiento de tipo IV y V</i>	83
Comentarios adicionales	85
Glosario	87
Referencias	88

RESUMEN

La especificidad es el resultado de la especialización de una enzima determinada en el reconocimiento de algún sustrato(s) y por lo tanto, desempeña un papel particular en el metabolismo celular, logrado a través de millones de años de evolución. En el laboratorio buscamos ampliar la especificidad a una enzima que interviene en síntesis de biotina para obtener un resultado opuesto a este escenario evolutivo. Para lograr esto, establecimos un sistema de selección basado en la auxotrofia a biotina, cuya biosíntesis en *E. coli* participan dos aminotransferasas denotadas como BioA y BioF. Ambas enzimas intervienen en dos pasos consecutivos y dependen de piridoxal fosfato (PLP) para poder realizar sus respectivas actividades. Al remover BioA y BioF de *E. coli*, generamos una cepa auxótrofa de biotina. Al transformar esta cepa con una colección de variantes de BioA, seleccionamos por complementación positiva una variante con una supuesta propiedad bifuncional.

Tal propiedad bifuncional implicaría haber obtenido una variante candidata para haber adquirido la actividad BioF sin perder su propia actividad.

Por otro lado, a partir de la base de datos SWISS PROT, obtuvimos secuencias de enzimas que dependen de PLP. Con ellas construimos una red de identidades cuyos nodos son enzimas y los vértices son el valor de la identidad dentro del intervalo [25%-70%]. Posteriormente, asociamos cada nodo de la red al tipo de plegamiento al que pertenecen, según la clasificación propuesta por Grishin *et al.*, (1995). Con base al gráfico de la red generada, encontramos diversos conjuntos de nodos con altos coeficientes de agrupamiento que en términos generales, representan conjuntos de enzimas con actividades relacionadas. De estos conjuntos de enzimas construimos árboles filogenéticos a partir del principio de máxima parsimonia.

INTRODUCCIÓN

La naturaleza ha generado una enorme diversidad de organismos en la tierra que habitan prácticamente todos los ambientes que hay en ella: Desde las regiones más calientes de la tierra, a las regiones más frías; desde las partes más profundas del océano, hasta la atmósfera. Todo ello debido a la enorme capacidad de adaptar y de modificar todo el repertorio metabólico, puesto que el papel fundamental en el metabolismo lo desempeñan las proteínas. Del éxito de ellas para innovar soluciones y poder afrontar los retos que depara la vida, subyace el éxito de los organismos para colonizar nuevos nichos.

Este trabajo pretende emular uno de los tantos logros de la naturaleza: El cambio de especificidad enzimática. Puesto que suponemos que las enzimas han adquirido nuevas funciones dentro de la célula que las alberga mediante un proceso llamado duplicación y posterior modificación en la especificidad con el que reconoce sustratos, incrementando de este modo el repertorio funcional y posibilita a los organismos subsistir en su medio ambiente.

Se acepta comúnmente que las primeras formas de vida se derivaron de compuestos orgánicos a partir del medio ambiente primitivo. Una vez generadas las primeras formas celulares, éstas seguían dependiendo de diversos compuestos orgánicos que se podían tomar del medio ambiente. Con el surgimiento de las primeras enzimas, esta dependencia se redujo. Aún así, es razonable pensar que las enzimas de aquellas formas de vida no tenían el grado de eficiencia y especialización mostrado por las enzimas modernas.

Evidentemente, las principales presiones de selección que los organismos sobrellevaron están basadas en satisfacer ciertos requerimientos estructurales (síntesis de membranas, DNA, RNA y proteínas) y energéticos (componentes que participan en la degradación y síntesis de catabólicos y la producción de los componentes celulares necesarios para la asimilación de fuentes externas de materia y energía). En los organismos modernos, todos estos requisitos los satisfacen enzimas altamente especializadas cuyas características cinéticas son adecuadas de acuerdo al organismo que las alberga.

Imaginamos un escenario en el que las primeras formas de vida que ya contaban con genomas, contenían codificadas un limitado repertorio enzimático. Durante millones de años los genomas aumentaron de tamaño por medio de duplicaciones de fragmentos o genomas completos. Es razonable pensar que la evolución de los primeros duplicados en etapas tempranas de la vida fueron favorecidas por carecer de mecanismos de reparación de DNA (Koch 1972, en Lazcano 1997).

La universalidad reflejada en el mapa de las vías metabólicas es una expresión del

consenso y origen común del metabolismo intermediario. Además, refleja un proceso evolutivo sujeto a presión adaptativa desde el periodo pre-enzimático hasta las formas de vida contemporáneas (Morowitz *et al.*, 2000).

La diferencia metabólica más importante entre los organismos es la forma específica en que obtienen energía para llevar a cabo los procesos vitales. El metabolismo pues, es la suma de todas las transformaciones químicas que se llevan a cabo en una célula a través de una serie de reacciones catalizadas enzimáticamente y que constituyen las rutas metabólicas. Cada uno de los pasos consecutivos en una ruta metabólica genera un cambio específico y sutil, que puede consistir en la eliminación, transferencia o adición de un átomo o un grupo funcional en particular. El término *metabolismo intermediario* es aplicado con frecuencia al conjunto de las actividades de todas las rutas metabólicas que interconvierten precursores, metabolitos y productos de peso molecular relativamente bajo, generalmente por debajo de 1000 Daltones. Las variantes en el metabolismo intermediario dependen de las condiciones ambientales (Morowitz *et al.*, 2000) y el dominio biológico al que se realice (Arquea, bacteria y Eucaria), al estadio del organismo en su ciclo de vida, y en el caso de los organismos pluricelulares, la función específica de la célula, el linaje celular u órgano, e inclusive el organelo en el que se realice el proceso.

Duplicación génica

La duplicación de genes individuales, de segmentos cromosomales y de genomas completos son la fuente de material genético para novedades evolutivas; incluyendo nuevas funciones génicas o diferentes patrones de expresión que podrían incidir en la especiación. Las duplicaciones génicas son accidentes debido a los distintos procesos celulares como la replicación de DNA, la recombinación y la expresión génica que generan copias de DNA dentro de un genoma. Existen principalmente tres mecanismos para duplicar regiones de DNA: Recombinación no homóloga, deslizamiento en la replicación (replication slippage) y la retro transposición. La recombinación no homóloga consiste en que dos segmentos no homólogos de DNA de doble hélice sufren recombinación. El fenómeno que hace posible esta situación es la existencia de secuencias repetidas. Esto permite que una copia de una repetición en un cromosoma impida la correcta alineación de la recombinación con una copia diferente de la repetición en el cromosoma homólogo, en vez de con la copia que le corresponde. Cuando se produce la recombinación, aparece una remoción en uno de los cromosomas recombinantes y una inserción correspondiente en el otro. Otro mecanismo para generar duplicaciones es el llamado “replication slippage”, el cual ocurre en zonas contiguas a regiones de DNA con secuencias repetidas cortas. Durante la replicación

puede ocurrir un mal apareamiento entre unidades repetidas vecinas produciendo un “deslizamiento”, resultando en la remoción o duplicación de un segmento de DNA, dependiendo si éste se produjo en la dirección $5' \rightarrow 3'$ o en la dirección opuesta. La transposición requiere que un gen sea transcrito en RNA, a partir de éste, una transcriptasa inversa podría producir una copia de DNA de doble cadena, el cual puede integrarse en el genoma en cualquier colocación arbitraria (Wen-Hsiung, 1995). Los genes duplicados de esta manera son llamados “retrogenes”.

El destino de los duplicados

Un duplicado surge en el genoma y debe de sortear diversas dificultades para poder ser observado en comparaciones evolutivas. Naturalmente, uno se pregunta ¿cómo un gen puede adquirir nuevas funciones sin comprometer la función original?, ¿con qué frecuencia un duplicado puede adquirir mutaciones benéficas?

La primera hipótesis sobre el origen de nuevas funciones fue propuesta por Ohno (1970), el cual, asume que después de una duplicación, una de las copias es redundante y por lo tanto, libre de toda restricción para divergir y convertirse en un nuevo gen. Ahora es claro que el proceso no es tan simple como se había pensado. Cuando el grado de divergencia entre las copias no es tan grande, una de las copias puede ser eliminada por recombinación no homóloga o convertida a la copia conservada por conversión génica (Wen-Hsiun, 1995). Más recientemente, se han considerado tres posibles destinos para los duplicados (Lynch *et al.*, 2000), (i) una de las copias puede ser silenciada por mutaciones degenerativas (no funcionalización); (ii) una copia puede adquirir una nueva función bajo mutaciones evidentemente benéficas; de esta forma, se preserva bajo el influjo de la selección natural, con la otra copia manteniendo la función original (neo funcionalización); (iii) ambas copias pueden sufrir mutaciones que reduzcan su capacidad para realizar su función original al nivel de una copia sencilla (sub funcionalización).

Para dilucidar el proceso que prevalece en la evolución de los duplicados, las secuencias de DNA nos ofrecen mucha mayor información que las secuencias proteicas. Uno de los indicadores es la fracción de sustituciones nucleotídicas sinónimas que ocurrieron por sitio sinónimo de DNA desde la duplicación (K_s). El segundo indicador es la fracción de sustitución en sitios no sinónimos (cambio de aminoácidos) que ocurrieron por sitio no sinónimo de DNA desde la duplicación (K_a). K_s nos ofrece una medición relativa del tiempo desde la duplicación; puesto que las sustituciones sinónimas no están sometidas a la misma presión de selección que las sustituciones no sinónimas que cambian un aminoácido en una proteína, razón por el cual, las mutaciones se van acumulando de manera aleatoria y proporcional al tiempo de divergencia. Si la relación

$K_a/K_s \approx 1$, significa que la tasa de reemplazo de los aminoácidos es igual a las sustituciones sinónimas, por lo que se infiere que la copia no está sometida a ninguna restricción selectiva. En aquellas relaciones en el cual $K_a/K_s < 1$, se infiere que la copia está sometida a una fuerte restricción “selección purificante”. En el caso opuesto, en el que $K_a/K_s > 1$, indica que la tasa de reemplazo de sitios no sinónimos es mayor a lo esperado por azar; esto indica que han ocurrido mutaciones benéficas en la evolución de los duplicados. Podemos generar categorías de duplicados de acuerdo a edades con base al valor de sus K_s . Si los duplicados se preservaran de manera indefinida, uno podría suponer que hallaríamos la misma cantidad de genes duplicados dentro de cada categoría de edades, por el contrario, si los duplicados se eliminaran después del evento de duplicación, este número disminuiría al aumentar el valor de K_s . Bajo todos estos supuestos, Lynch *et al.*, (2000) encontraron que el 95% de los duplicados se eliminan dentro de los primeros 50 millones de años. La tasa de duplicación la estimaron dentro de un intervalo de 0.002 y 0.02 por gen, por millón de años. Dado este rango, el 50% de todos los genes en un genoma sufren al menos una duplicación en el rango de 35-350 millones de años. Dada la baja tasa de fijación de los duplicados, ¿cómo es posible que los duplicados existentes hayan logrado preservarse a lo largo de millones de años? En primera instancia habrían dos posibilidades para poder preservarse, la neofuncionalización y la subfuncionalización. Para que los duplicados adquieran nuevas funciones, tendrían que acumular una serie de mutaciones benéficas, pero tales mutaciones suelen ser extremadamente raras. Por otro lado, la subfuncionalización no requiere de selección positiva; en este modelo las dos copias sufren mutaciones degenerativas, pero aún pueden complementar la función del gen parental. Si el gen parental efectuaba dos funciones relacionadas, al duplicarse, cada copia puede especializarse en una de las funciones del gen parental (figura 1). De aquí en adelante, las copias entran en un proceso de evolución acelerada (Lynch *et al.*, 2000)

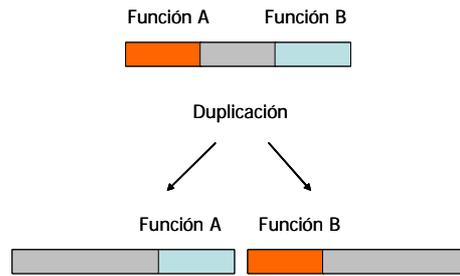


Figura 1. Duplicación y subfuncionalización de un gen parental que realizaba dos funciones. Los genes “hijos” se han especializado en una de las dos funciones. (Modificado de Hughes, 2005).

Sobre los dominios proteicos

Durante el transcurso de la evolución, los genes pueden producir proteínas más complejas al fusionarse, o menos complejas si se fisioan (Kummerfeld *et al.*, 2005), en consecuencia, los dominios son la unidad evolutiva de las proteínas y todas las proteínas consisten en uno o más dominios (Apic *et al.*, 2001). Un dominio se define como una unidad estructural que puede plegarse y ser funcionalmente activo en forma aislada (Ponting *et al.*, 2002). Existen relativamente pocas familias de dominios, sin embargo, ellas pueden albergar muchas más funciones. Los dominios de una misma familia, además de combinarse entre si, también pueden asociarse con otros miembros de diferentes familias. La mayoría de las proteínas, dos tercios en organismos unicelulares y más del 80% en metazoarios, son proteínas multidominios, creadas a partir de distintas combinaciones de dominios. Se ha reportado con base en secuencias proteicas de 40 genomas, 783 familias de dominios, de las que se encuentran combinadas en 1307 combinaciones diferentes (Apic *et al.*, 2001). Una pequeña fracción de los dominios se combina con más de un dominio de una familia distinta. Aproximadamente, de un tercio a la mitad de las familias no se combina con alguna otra familia. Por lo que muy pocas familias son sumamente versátiles en sus patrones de combinación, sin embargo, éstas son las más abundantes en los genomas. Este patrón combinatorio, se ajusta a una función de ley de potencias y puede ser descrito bajo una red libre de escala, en el cual, la mayor parte de los nodos de una red (en este caso un nodo es un dominio) tiene muy pocas conexiones (muy pocas combinaciones con dominios de otra familia), y sin embargo el resto (conocidos como “hubs”) tienden a estar altamente conectados con otros miembros (Jeong *et al.*, 2000).

Puesto que los dominios pueden combinarse para formar proteínas multidominio, no es claro cuánto de las combinaciones de dominios, así como su versatilidad, se debe a la selección o a un proceso aleatorio. Es de suponerse que en un escenario completamente

aleatorio, la abundancia relativa de dominios determina el nivel de combinación con otros dominios. Aquellos dominios que prevalecen en los genomas deberían estar más combinados que aquellos dominios menos representados. Vogel *et al.*, 2005, Compararon la abundancia versus la versatilidad de cada superfamilia en varios genomas. Clasificaron los dominios en diferentes categorías de acuerdo a su clase estructural, plegamiento, filiación filogenética, función molecular y proceso biológico. Todas estas categorías se compararon versus la abundancia. Los datos se graficaron y se ajustaron a una función. Para cada grupo se encontró que (i) la ley de potencias es la función que mejor describe los datos, donde $V \sim A^\theta$ (ii) la pendiente $\theta_{\text{obs}} = 0.42 \pm 0.04$, para todos los genomas.

Dada la observación que el número de combinaciones diferentes es mucho menor a todas las posibles combinaciones, sugiere que inclusive, aún en un escenario aleatorio como el descrito anteriormente, la versatilidad de los dominios está fuertemente limitada por la presión de selección. Para contestar esto, Vogel *et al.*, 2005), simularon un proceso aleatorio en el cual se barajearon los dominios, donde la abundancia, el número y tamaño de los dominios fueron constantes. Tales datos también se ajustaron a una función de ley de potencias; el valor de $\theta_{\text{rand}} = 0.72 \pm 0.01$, fue sensiblemente más alto al observado en los genomas ($\theta_{\text{obs}} = 0.42 \pm 0.04$). Un análisis más detallado de los datos, reveló que los datos observados en los genomas, a diferencia de la simulación, mostraron que una pequeña proporción de combinaciones de dominios son más frecuentes que lo esperado al azar, mientras que una gran proporción de combinaciones están subrepresentadas en la simulación. Cuando una combinación se duplica, la abundancia de cada dominio aumenta, pero la versatilidad permanece constante. La discrepancia entre la simulación y la realidad se debe a que este modelo no toma en cuenta la duplicación de combinación de dominios ya establecidas por barajeo de dominios. Al considerar en la simulación, las duplicaciones de combinaciones de dominios, pudieron reproducir mejor los datos observados en los genomas estudiados. De esta forma, se pudo concluir que la duplicación y retención de combinaciones de dominios debido a la selección, es un factor dentro de la evolución del repertorio proteico.

Homólogos, ortólogos y parálogos

El termino “Homología” fue introducido por Richard Owen en 1843 quien la definió como “El mismo órgano en diferentes animales bajo cada variedad de formas”, además, Owen distinguió claramente la homología de la analogía, el cuál definió como “parte de un órgano de un animal, el cual tiene la misma función en parte o órganos en diferentes animales” (Fitch, 2000; Koonin, 2005). Años más tarde, después del “Origen de las

especies” de Darwin, Huxley reinterpretó la homología en términos evolutivos (Koonin, 2005). Actualmente, homología se define como la relación de dos caracteres que han descendido, usualmente con divergencia, a partir de un carácter ancestral (Fitch, 2000). Los caracteres son cualquier atributo, ya sean morfológicos, moleculares, etológicos, ecológicos, biogeográficos, que podamos observar en un organismo, cuyas diferentes manifestaciones se denominan estados (Morrone, 2000).

El concepto de ortología y paralogía lo introdujo por primera vez Walter Fitch en 1970 (Fitch, 1970). Frecuentemente se comenta que genes (proteínas) ortólogos son aquellas que tienen la misma función, pero se encuentran en diferentes organismos, mientras que genes parálogos son aquellos que tienen diferente función, pero se encuentran en el mismo organismo. La definición original propuesta por Fitch, es diferente, pues él definió como genes ortólogos a aquellos que se encuentran en dos especies diferentes y que derivaron de un ancestro común, a partir de un gen en el último ancestro común, es decir, la divergencia de estos se explica por un proceso de especiación. Una consecuencia implícita en la definición de ortología, es que no hay un requerimiento para que la ortología sea una relación uno a uno, por lo que es válido asignar ortología a grupos enteros. Parálogo se definió originalmente como aquellos genes que derivaron a partir de una duplicación de un solo gen; bajo esta definición, paralogía no implica que dos genes tengan que estar en el mismo genoma. Más recientemente se ha distinguido diferentes tipos de parálogos, de acuerdo al momento de la duplicación (Sonnhammer et al., 2002). Para poder distinguir las distintas clases de parálogos, Sonnhammer ha definido in-parálogos como aquellos parálogos de un linaje determinado que han evolucionado por una duplicación génica que ocurrió posterior a la radiación (especiación). Out-parálogos son aquellos parálogos que dado un linaje han evolucionado por una duplicación génica que ocurrió antes del evento de especiación. Una consecuencia generada a partir de la definición de paralogía, es la relación con la ortología, pues se dice que dos genes son co-ortólogos cuando se ha producido una duplicación de ortólogos posterior al evento de especiación

Los modelos de evolución de vías metabólicas

En un trabajo innovador a su tiempo, Horowitz (1945), supone que el metabolismo ha evolucionado en un orden inverso al de la biosíntesis actual usando intermediarios presentes en el ambiente prebiótico, entonces, la última enzima de alguna ruta biosintética sería la más antigua y la primera enzima la más reciente (figura 2). Este modelo se conoce también como la hipótesis retrógrada. Con el descubrimiento de los operones, Horowitz *et al.*, (1965), modificaron la teoría retrógrada para argumentar que esta agrupación de genes que participan en una misma vía metabólica podría ser

explicada por antiguas duplicaciones en tandem; sin embargo, esta teoría tiene limitaciones (Lazcano *et al.*, 1999) entre las cuales destacan:

- a) La mayoría de los intermediarios metabólicos son químicamente inestables y es difícil explicar su síntesis y acumulación en los ambientes prebióticos.
- b) La hipótesis retrograda no explica el origen de los mecanismos regulatorios de las vías catabólicas y el desarrollo de rutas biosintéticas que involucran reacciones diferentes.

Se conocen muy pocos ejemplos que se ajusten al modelo de Horowitz: 1) Los productos de los genes *hisA* y *hisF*, de *E. coli*. Los productos de los genes *HIS6* y *HIS7* de *Saccharomyces cerevisiae*, los cuales participan en el anabolismo de la histidina (Fani *et al.*, 1997)

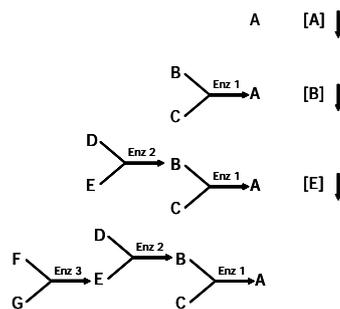


Figura 2. Un organismo heterotrófico para un metabolito esencial “A” suministrado por el medio ambiente. El reclutamiento fortuito de una enzima (Enz 1) capaz de sintetizar “A” a partir de “B” y “C” confiere una ventaja selectiva al organismo. Una disminución en la concentración de “B” y “E” es compensado por aquellos organismos que hayan reclutado las enzimas “Enz 2” y “Enz3”, respectivamente. (Modificado de Rison *et al.*, 2002)

Granick, (1957), propuso que la biosíntesis de algunos productos relativamente complicados podía ser explicada a partir de compuestos más sencillos por una evolución “hacia adelante”. Este modelo asume que los compuestos bioquímicos simples son más antiguos que los compuestos finales complejos y por lo tanto, que las enzimas que catalizan los primeros pasos en una ruta biosintética son más antiguas que las enzimas que catalizan los últimos pasos. De esa manera, los pasos de la ruta biosintética reflejan su historia evolutiva. Waley, (1969), propuso que la evolución de las rutas metabólicas podría ser debida a cambios mínimos en la copia de un gen duplicado teniendo como resultado una nueva enzima capaz de catalizar una reacción similar. A partir de esta idea, y de forma independiente, Ycas, (1974) y Jensen, (1976), propusieron una teoría

de evolución de las vías metabólicas, la cual consiste en que éstas evolucionaron reclutando enzimas poco eficientes de amplia especificidad a partir de otras vías; un modelo que se conoce como evolución por “patchwork” (figura 3). En este modelo, las enzimas pueden tener sustratos y mecanismos catalíticos preferenciales. Posteriormente, la duplicación génica, seguida de mutaciones y especialización definen el papel central que tome la enzima en el metabolismo. A diferencia del modelo de Horowitz, el modelo de evolución por “patchwork” operaría únicamente después del surgimiento de otras vías metabólicas. Actualmente se ha reportado en la literatura una vasta cantidad de ejemplos de evolución por “patchwork”. Teichmann et al., 2001, han estimado que de las 282 familias de proteínas en *E. coli*, 144 familias tienen representantes en más de una vía, por lo que se subraya la importancia del reclutamiento de enzimas bajo el modelo de evolución por “patchwork” en *E. coli*.

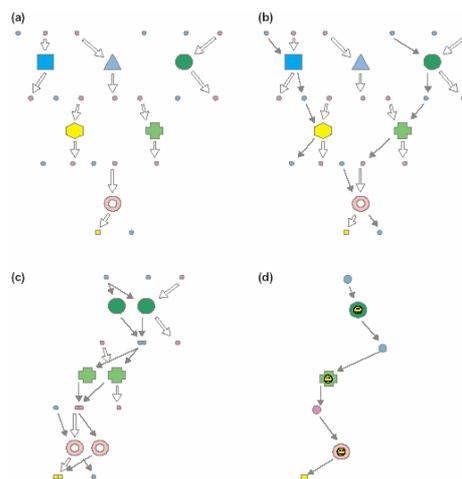


Figura 3. Modelo de evolución por “patchwork”. En este modelo las enzimas tienen un sustrato y mecanismo catalítico favorito (a), pero pueden exhibir amplias especificidades y son capaces de catalizar otras reacciones (b). Por lo tanto, se pueden crear algunas cadenas metabólicas. (c) La duplicación del gen podría ser ventajoso, puesto que se sintetizan más metabolitos. (d) En consecuencia, este modelo debería explicar la existencia de muchas vías metabólicas. (Tomado de Rison *et al.*, 2002).

Desde hace siglos y especialmente a partir de la revolución industrial, el humano ha estado perturbando el medio ambiente y ha sometido a todas las especies a una severa presión de selección, particularmente con la introducción de compuestos xenobióticos. De esta forma, los microorganismos tienen nuevas oportunidades para usar estos compuestos como nuevas fuentes de carbono, característica que suele ser aprovechada en biotecnología para eliminar compuestos tóxicos. Aunque la mayoría de las vías metabólicas evolucionaron hace varios millones de años, aún podemos ser testigos de la

evolución de una nueva vía metabólica, (Copley, 2000), reporta la vía de degradación de pentaclorofenol en *Sphingomonas chlorophenolica*, un pesticida liberado al medio ambiente en 1936 y diferente a cualquier compuesto de origen biótico conocido. El autor propuso que esta vía evolucionó en pocas décadas reclutando enzimas de dos vías catabólicas. La vía consiste de 4 pasos. La primera y tercera enzima de la vía son originarias de una vía para la degradación de compuestos naturales fenólicos clorados. La segunda enzima, una halogenasa reductora, pudo haber surgido a partir de una maleilacetoacetato isomerasa, enzima asociada a la degradación de tirosina. Como podría esperarse de una vía recién formada, algunas enzimas de la vía son poco eficientes y muestran una amplia especificidad por sustratos relacionados.

Debido a las pocas evidencias encontradas que apoyen el modelo de Horowitz, Lazcano *et al.*, (1999), propusieron otro modelo que complementa la hipótesis de Horowitz, este modelo llamado “Origen semienzimático de las vías metabólicas” el cual asume un paso metabólico que ocurrió u ocurre de manera espontánea por ser termodinámicamente favorable y que por ende no requiere catálisis enzimática (Figura 4), pero que al originarse una enzima poco específica que lo catalice, “la adopta” y en lo sucesivo el paso metabólico puede ser o no catalizado por la enzima. Este modelo es especialmente útil para conectar la química prebiótica con el desarrollo de vías metabólicas. Las premisas en el cual está basado este modelo son:

- a) Existió un conjunto de compuestos prebióticos lo suficientemente estables en el océano primitivo.
- b) Compuestos secundarios producidos por algunas vías también son disponibles. Se propone la presencia de intermediarios metabólicos, debido a que los valores de K_m son finitos.
- c) Se considera la existencia de enzimas poco específicas que podrían estar disponibles para futuros eventos de duplicación génica.

El ejemplo propuesto por Lazcano *et al.*, (1999) está basado en el gen de la glutamina amidotransferasa (*hisH*), el cual forma parte de la biosíntesis de histidina y cataliza la reacción mostrada en la figura 3. La reacción adiciona NH_3 en altas concentraciones de éste y en ausencia de la proteína HisH. Proponen que la reacción tomó lugar con NH_3 , seguido con el desarrollo de HisH y el reemplazo de NH_3 por glutamina cuando se agotó el NH_3 en el ambiente.

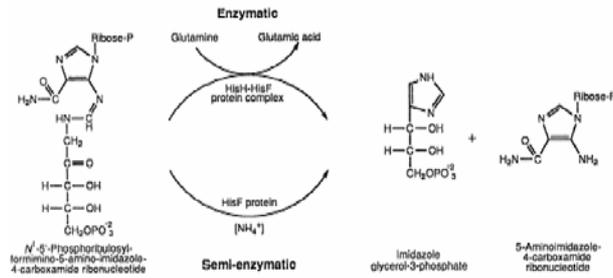


Figura 4. Adición enzimática y no enzimática de nitrógeno a N-5'-fosforibosilformiamino-5-aminoimidazol-4-carboxamida ribotido, para producir imidazol glicerol fosfato (IGP). (Tomado de Lazcano *et al.*, 1999).

Los modelos de Horowitz y Granick son mutuamente excluyentes, aunque en ambos la presión de selección dominante es la conservación de la especificidad por el sustrato, mientras que en el modelo de Waley, Ycas y Jensen (evolución por “patchwork”), y el modelo Lazcano y Miller, la presión imperante es la conservación del mecanismo químico, de hecho, la baja especificidad por el sustrato es lo que permite la incorporación de la enzima preexistente a nuevas rutas metabólicas, estos últimos modelos pueden ser complementarios entre sí y con el de Horowitz o el de Granick.

La naturaleza no suele reinventar, recicla las soluciones

Un corolario importante generado a partir de todos los modelos de formación y evolución de los genes y de las vías metabólicas es el reciclaje de soluciones. Hemos visto que todos los modelos evolutivos requieren la existencia de vías anteriormente establecidas en los organismos. A consecuencia de esto, las proteínas, pese a tener un ilimitado número de secuencias posibles, el número de plegamientos proteicos es muy reducido. Se estima que el número de plegamientos gira alrededor de los 10 000. Si asumimos que el tamaño promedio de una proteína es de aproximadamente de 200 aminoácidos, entonces tenemos $20^{200} \sim 10^{260}$ proteínas posibles, un número muy superior al número de átomos en el universo observable. Pese a que nuestro conocimiento sobre el plegamiento de proteínas es bastante limitado, suponemos razonablemente que solo una pequeña fracción de esta inmensa cantidad de secuencias pueden plegarse. Y si además asumimos que existen 10 millones de especies en la tierra, y que cada especie tiene un promedio de 5000 proteínas, entonces existen 5×10^{10} proteínas diferentes (Koonin *et al.*, 2002).

Un plegamiento puede codificar para múltiples funciones, o inversamente, una función tiene más de una solución estructural. Como paradigma podemos asegurar que existen más funciones que plegamientos. (Todd *et al.*, 2001; Orengo *et al.*, 1994). Con base a este paradigma, asumimos que la función, a diferencia del plegamiento, se ha inventado independientemente varias veces durante toda la historia evolutiva de las especies.

Cada vez parece más claro el origen monofilético de proteínas con el mismo plegamiento. El caso de los barriles TIM ejemplifica esto. Debido a la enorme divergencia en identidad mostrada por proteínas con esta estructura, se ha debatido profundamente si esta estructura es el resultado de un proceso de evolución convergente o si es el resultado de un proceso de evolución divergente. Copley *et al.*, (2000); Nagano *et al.*, (2002), reportaron 21 superfamilias con evidencias de tener un origen monofilético; en particular, la familia de FBPA IA se identificó como un grupo divergente, puesto que no muestran identidad significativa con el resto de las demás familias de los barriles TIM, (Lorentzen *et al.*, 2003). Los miembros de esta familia son claramente homólogos, aunque la identidad entre los miembros de esta familia sea por debajo del 20%. Lorentzen *et al.*, (2003), estos autores arguyen con elementos estructurales que la familia de las FBPA IA son homólogas al resto de las otras familias de barriles TIM. Como consecuencia, el rango de funciones catalíticas que puede albergar una sola clase de plegamiento es considerablemente amplio, por lo tanto la diversidad de funciones que ofrece una determinada superfamilia representa una solución económica en el número de plegamientos requeridos para sostener la vida. (Bartlett *et al.*, 2003; Gerlt *et al.*, 2001). En nuestro estado actual de conocimientos, podemos decir que las proteínas con la misma estructura son de origen monofilético.

La función como se mencionó anteriormente puede explicarse por convergencia o por divergencia, Gerlt *et al.*, 2001); sugieren al menos tres diferentes estrategias para evolución divergente:

La especificidad hacia el sustrato es la estrategia dominante

La especificidad hacia el sustrato es la estrategia dominante en la evolución divergente. Ésta consiste en que la evolución de una proteína recién duplicada está limitada a la unión de un sustrato en particular. Los mecanismos de reacción, catalizadas por la enzima progenitora y la enzima nueva no tienen que estar necesariamente relacionadas, únicamente lo que se conservaría en este modelo es la especificidad hacia el sustrato. Como se mencionó anteriormente, pocos son los ejemplos de vías que se ajusten al menos de manera parcial. El caso mejor documentado es en dos proteínas de la vía biosintética de histidina, (HisA y HisF) (Fani *et al.*, 1997; Gerlt *et al.*, (2001).

El mecanismo químico es la estrategia dominante

El mecanismo químico es posiblemente la estrategia más importante y una consecuencia de la evolución por “patchwork”. La enzima progenitora y la enzima nueva pueden tener la misma función, pero difieren en la especificidad hacia el sustrato.

Alternativamente, la progenitora y la nueva enzima pueden catalizar diferentes reacciones, reteniendo una estructura común, catalizando algunos pasos limitantes y/o estabilizando un estado de transición intermediario común entre la progenitora y la nueva enzima. Gerlt *et al.*, (2001), han documentado varios ejemplos como el caso de las superfamilias Enolasa, Amidohidrolasa/Fosfotriesterasa, etc.

La arquitectura del sitio activo es la estrategia dominante

La estrategia basada en la conservación de la arquitectura del sitio activo consiste en que la enzima progenitora y la nueva enzima no retienen la especificidad hacia el sustrato, tampoco comparten algún paso en la química de la catálisis. En lugar de esto, el sitio activo alberga reacciones alternas usando una parte de los grupos funcionales en diferentes contextos mecanísticos y metabólicos. Un ejemplo de esta estrategia de divergencia, lo suministran dos enzimas homólogas: La orotidina 5-fosfato descarboxilasa y la D-arabino-3-hexulosa-6-fosfato sintasa (Gerlt *et al.*, 2000).

Una vez comentado algunos de los “algoritmos” conocidos de evolución de la vasta diversidad proteica, estimamos que al seleccionar un sistema que nos permita simular en el laboratorio la evolución de la especificidad enzimática, tenemos que considerar una superfamilia enzimática que muestre diversidad funcional y flexibilidad mecanística, para que de esta forma tengamos mayores oportunidades de éxito. Puesto que no es posible igualar lo que la naturaleza ha logrado durante muchos millones de años. Por esta razón, una buena elección será aquella que considere las enzimas dependientes de piridoxal fosfato (PLP), por la razón de ser el conjunto de enzimas conocido más versátil, debido a su flexibilidad mecanística. En consecuencia, esta inherente capacidad de innovación explica en parte, que haya conquistado diversos nichos en el metabolismo de todos los organismos conocidos.

Las enzimas dependientes de piridoxal fosfato (PLP) catalizan reacciones muy diversas

Las enzimas dependientes de PLP son quizás unos de los biocatalizadores más versátiles, puesto que intervienen en un amplio rango de reacciones metabólicas (Alexeev *et al.*, 1998). Para estas enzimas se han reportado cinco de las seis principales actividades enzimáticas de la Enzyme Comision (EC): Oxidoreductasas, transferasas, hidrolasas, liasas e isomerasas. Con la excepción de las funciones asociadas a las actividades de ligasa, las demás actividades se encuentran representadas en esta clase de

enzimas (figura 5). La actividad mejor representada es la de transferasas, seguida de las liasas (de un total de más de 113 EC diferentes). Además de su versatilidad enzimática, las enzimas dependientes de PLP participan en variados procesos celulares, generalmente involucrados en la biosíntesis de aminoácidos y metabolitos derivados de éstos, y en la síntesis de aminoácidos glicosilados (He, 2002). Como en los casos reportados por Gerlt *et al.*, (2001); las enzimas dependientes de PLP comparten una porción de su mecanismo catalítico como estrategia de evolución divergente: El PLP se une covalentemente a las enzimas mediante una unión por base de Schiff (imina) formada por la condensación de su grupo aldehído con el grupo ϵ -amino de un resto de lisina de la enzima. Esta base, que está conjugada al anillo piridínico de la coenzima, es el foco de actividad de la coenzima. El mecanismo catalítico común en estas enzimas es el siguiente:

a) Transferencia del grupo PLP del residuo de lisina del centro activo de la proteína al grupo amino del sustrato para formar una aldimina “externa” (dentro del centro activo, pero no unida covalentemente a la enzima).

b) Escisión de un enlace C^α del sustrato adyacente al enlace C-N imina, para generar un reactivo quininoide formado del complejo PLP. El destino del intermediario quininoide, y por lo tanto, el curso de la reacción dependen de las características individuales de la enzima.

Como todo cofactor, el PLP desempeña un papel central en la catálisis de este tipo de enzimas, puesto que las enzimas que dependen de él muestran múltiples orígenes (ver mas adelante) y el papel único del PLP hacen suponer que éste apareció en la escena evolutiva antes del surgimiento de las apoenzimas (Mehta *et al*, 1998), esta suposición es consistente con el modelo del origen semienzimático propuesto por Lazcano *et al.*, (1999).

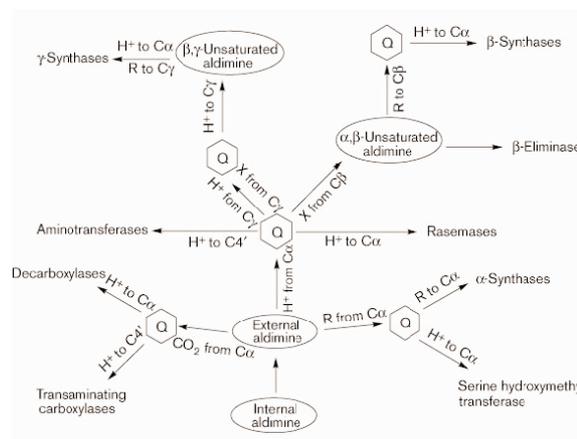


Figura 5. Se muestra la diversidad de reacciones catalizadas por enzimas dependientes de piridoxal fosfato. La “Q” indica el intermediario quininoide. (Tomado de Schneider *et al.*, 2000).

Las enzimas dependientes de piridoxal fosfato (PLP) fueron reinventadas al menos cinco veces durante la evolución

De acuerdo a la clasificación propuesta por Grishin *et al.*, (1995), las enzimas dependientes de PLP pueden agruparse en cinco diferentes tipos de plegamientos (tipo I, II, III, IV y V) y en subclases con base a la comparación de su secuencia primaria. Partiendo de la premisa: “mismo plegamiento, mismo origen” comentado anteriormente; entonces podremos asegurar que las enzimas dependientes de PLP se han reinventado al menos cinco veces en el transcurso de la evolución (Mehta *et al.*, 1998). Pese al origen independiente de esta clase de enzimas, el dominio de unión al PLP es bastante similar entre los 5 plegamientos. De acuerdo con Denessiouk *et al.*, (1999), hay 7-9 segmentos que pueden ser sobrepuestos con base a los carbonos α de las enzimas Aspartato aminotransferasa de *E. coli* (AspAT, plegamiento I), Alanina racemasa de *Bacillus stearothermophilus* (AlaR, plegamiento III), en los carbonos α de los dominios carboxilos terminal de la subunidad β de la enzima Triptofano sintasa de *Salmonella typhimurium* (TrpS β , plegamiento II), de la enzima Ácido D-amino Transferasa de *Bacillus sphaericus* (D-AAT, plegamiento IV) y de la enzima Glucógeno fosforilasa de músculo esquelético de conejo (GP, plegamiento V). Cada una representante de los 5 tipos de plegamientos que se reportan para enzimas que usan PLP como cofactor. Estos segmentos no comparten los mismos aminoácidos, pero estructuralmente son similares e incluyen una lámina β adyacente, una hélice- α cercana al amino terminal en el cual se “anclan” los grupos fosfatos del cofactor PLP. Esta unidad estructural, desempeña un papel fundamental en el reconocimiento del PLP en los 5 tipos de plegamientos. Sin embargo, el PLP se orienta de manera distinta en los 5 plegamientos a pesar de la similitud de estos motivos estructurales.

La uniformidad catalítica de estas enzimas, contrasta con su múltiple origen evolutivo. Por lo que se supone que esta uniformidad se origina por restricciones inherentes a la química del PLP, más que a sucesos aleatorios.

Plegamiento de tipo I

La familia de enzimas con el plegamiento de tipo I es la mejor estudiada, y se le ha dividido en ocho subfamilias (figura 6). Estructuralmente consiste en un gran dominio cuya característica central es una lámina β de siete hebras y un dominio pequeño incluyendo la parte C-terminal de la cadena.

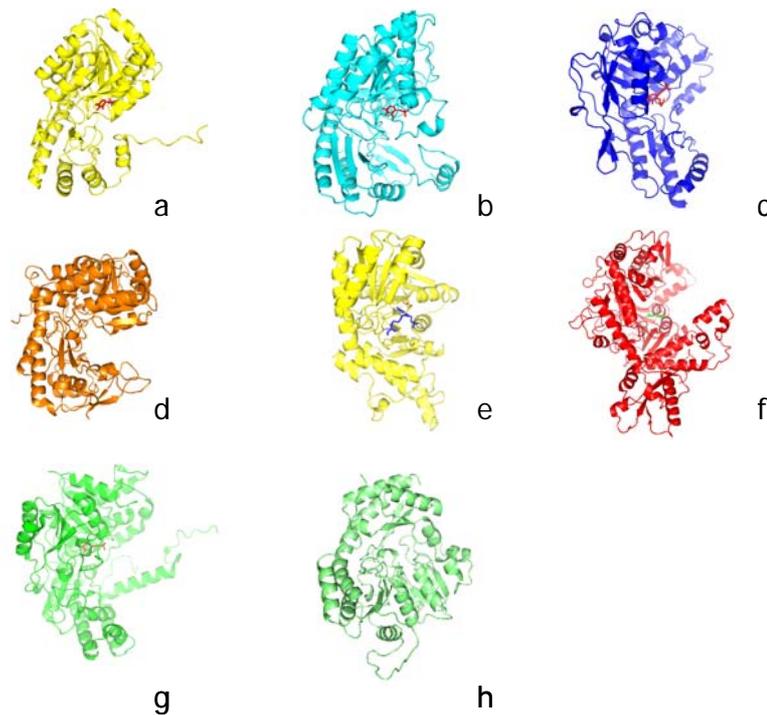


Figura 6. Dominios que pertenecen al plegamiento de tipo I. (a) Aspartato aminotransferasa PDB 1ahe. Plegamiento tipo I, subclase I aminotransferasa. (b) 7,8-diamino pelargónico sintasa de *E coli*. PDB 1qj5. Plegamiento tipo I, subclase II aminotransferasa. (c) Fosfoserina aminotransferasa de *E coli*. PDB 1bjn. Plegamiento tipo I, subclase Fosfoserina aminotransferasa. (d) Tirosina fenol liasa de *Citrobacter freundii*. PDB 2tpl. Plegamiento tipo I, subclase tirosina fenol liasa. (e) Cistationina β -liasa de *E. coli*. PDB 1cl1. Plegamiento tipo I, subclase Cistationina β -liasa. (f) Ornitina descarboxilasa de *Lactobacillus sp.* PDB 1ord. Plegamiento tipo I, subclase Ornitina descarboxilasa. (g) Serina hidroximetiltransferasa de *Oryctolagus cuniculus* PDB 1ls3. Plegamiento tipo I, subclase Serina hidroximetiltransferasa. (h) Ácido 3-Amino-5-hidroxibenzoico sintasa de *Amycolatopsis mediterranei*. PDB 1b9h. Plegamiento de tipo I, subclase Ácido 3-Amino-5-hidroxibenzoico.

Este dominio se pliega dentro de una lámina β de tres o cuatro hebras cubiertas por una hélice a un lado, pese a esto, únicamente cinco elementos estructurales correspondientes a láminas β : β 4, β 7, β 8, β 9 y β 10 solapan estrictamente a otros miembros de esta familia (Kack *et al.*, 1999).

Esta familia, además de ser la más grande, es la más diversa en terminos de funciones enzimáticas. Son invariablemente homodiméricas y sus sitios activos se encuentran

ubicados en la interfaz de ambos monómeros, en el cual, cada monómero contribuye con residuos esenciales para los dos sitios activos. Cada monómero está formado por un dominio pequeño y un grande (dominio en terminología bioquímica se refiere a regiones proteicas con función asignada, indistintamente de su estructura tridimensional (Ponting *et al.*, 2002), en ciertos casos, cada monómero se mueve en asociación con el sustrato para crear una conformación cerrada que podría contribuir a la especificidad y al tipo de reacción (Eliot *et al.*, 2004).

Plegamiento de tipo II

Al igual que las enzimas que pertenecen al plegamiento de tipo I, los miembros con plegamiento de tipo II, se definió únicamente bajo criterios estructurales. A diferencia de las enzimas que pertenecen al plegamiento de tipo I, los sitios activos de las enzimas del plegamiento de tipo II están compuestos de residuos de uno de los monómeros (figura 7). También difieren en que contienen dominios adicionales para regulación alostérica (Eliot *et al.*, 2004), además los miembros de esta familia tienen la característica principal de catalizar reacciones en las que los carbonos C^α y C^β participan en los cambios de covalencia. (Mehta *et al.*, 1998).

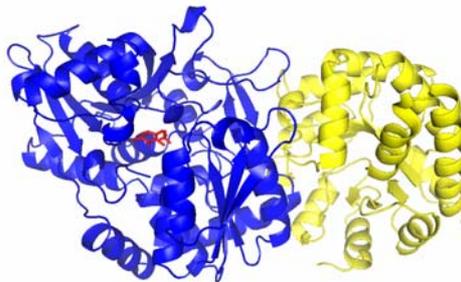


Figura 7. Unidad biológica de la enzima Triptofano sintasa β de *Salmonella typhimurium*. PDB 11s3. Plegamiento tipo II. En color azul se muestra el dominio catalítico, en amarillo se muestra el dominio regulador.

Plegamiento de tipo III

Los miembros con plegamiento de tipo III pertenecen a una familia pequeña, las enzimas que pertenecen a esta familia y son las únicas dependientes de PLP que están relacionadas estructuralmente a otras enzimas que no dependen de PLP, puesto que muestran un plegamiento tipo TIM barril (figura 8). Al igual que otros barriles TIM, el sitio activo está acomodado en el carboxilo terminal de las láminas β (Schneider *et al.*, 2000). Sorprendentemente el PLP se une covalentemente a un residuo de lisina con la enzima, al igual que otras enzimas que pertenecen a otros plegamientos con el grupo

fosfato unido al amino terminal de una hélice α . Estas enzimas son dímeros obligados, puesto que cada monómero contribuye con residuos para ambos sitios activos (Eliot *et al.*, 2004).

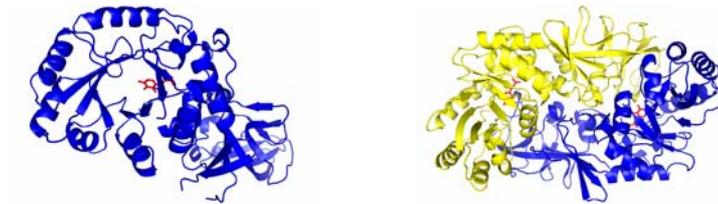


Figura 8. Dominio catalítico (izquierda) y unidad biológica (derecha) de la enzima Alanina racemasa de *Bacillus stearothermophilus*, PDB 1sft. Plegamiento tipo III

Plegamiento de tipo IV

Los miembros de la familia con plegamiento de tipo IV son similares a las enzimas de plegamiento I y II, puesto que son homodímeros con el sitio activo colocado en la interfaz entre ambos monómeros (figura 9). El cofactor se encuentra enlazado en un sitio que es aproximadamente la imagen especular de los sitios de las enzimas de plegamiento I y II (Eliot *et al.*, 2004).

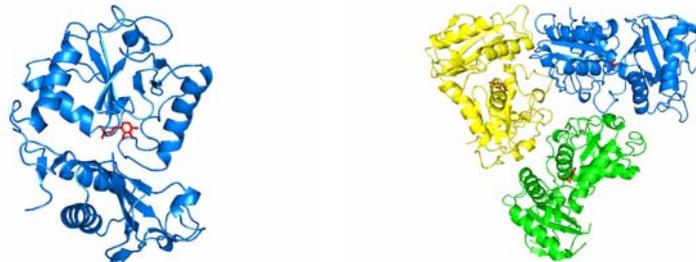


Figura 9. Dominio catalítico (izquierda) y unidad biológica (derecha) de la enzima D-Aminoácido aminotransferasa de *Bacillus sp.* PDB 1aog. Plegamiento tipo IV.

Plegamiento de tipo V

En los miembros con plegamiento de tipo V (figura 10), el PLP no actúa como catalizador electrofílico, en lugar de esto, el grupo fosfato es el encargado de realizar la transferencia de protones.

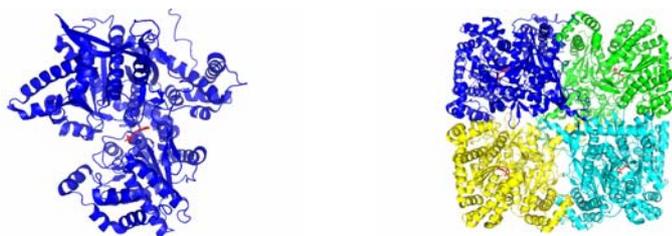


Figura 10. Monómero y unidad biológica de la enzima Maltodextrina fosforilasa de *Oryctolagus cuniculus*. PDB 1abb. Plegamiento tipo V.

Se ha demostrado que diferentes estructuras proteicas pueden usar al PLP como cofactor, en ningún caso el plegamiento determina el tipo de reacción, cada tipo de plegamiento contiene múltiples tipos de reacciones enzimáticas y todos los tipos de reacciones se hallan en al menos dos plegamientos (plegamientos I y II).

Hasta ahora hemos visto que las enzimas dependientes de PLP se encuentran principalmente involucradas en algunos pasos de la síntesis de aminoácidos, y debido a su versatilidad catalítica, han conquistado nichos en diferentes vías metabólicas. La síntesis de biotina en *E. coli* no es la excepción, puesto que en esta vía podemos encontrar dos enzimas que dependen de PLP. Creemos que esta vía es un interesante modelo de estudio porque nos permitiría establecer un sistema de selección al generar auxotofría hacia biotina al remover uno o los dos genes que codifican para proteínas que dependen de PLP (ver más adelante).

Síntesis de Biotina en *E. coli*

La biotina, (vitamina H) es un cofactor esencial para una clase importante de enzimas metabólicas: La Biotina carboxilasas y las descarboxilasas. En *E. coli*, solo la enzima acetil CoA carboxilasa usa biotina, por lo que se requiere en una ínfima cantidad tal como ~100-200 moléculas por célula (Choi-Rhee *et al.*, 2005). La ruta sintética de biotina se distribuye ampliamente entre los microorganismos (figura 11). En algunos microorganismos como *Bacillus subtilis* y *Bacillus sphaericus* difieren en el primer paso de la biosíntesis respecto a *E. coli*, puesto que el pimeloil-CoA se produce a partir

de ácido pimélico y en *E. coli*, se sintetiza a partir de L-alanina y/o acetato vía acetil-CoA (Eisenberg, 1985; Rodionov *et al.*, 2002).

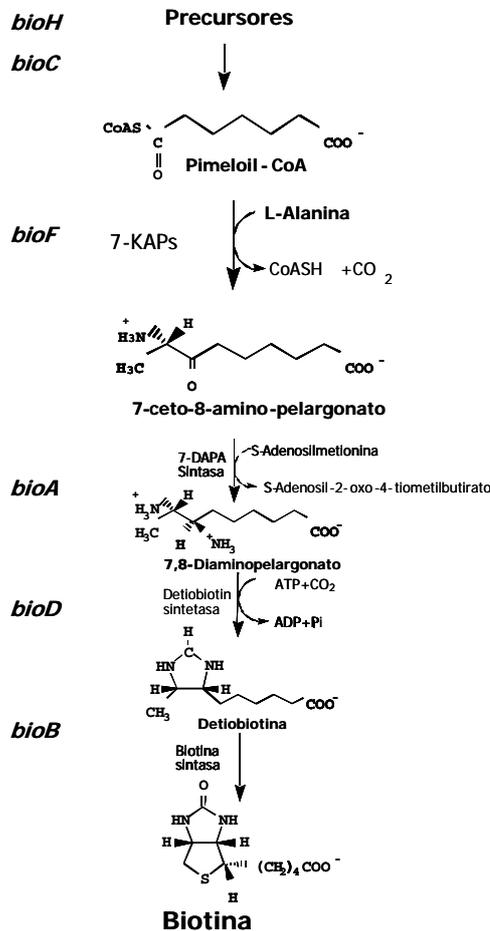


Figura 11. Esquema de la síntesis de biotina en *E. coli*. A la izquierda, se muestran los genes que codifican para la enzima que realiza la catálisis en cada paso.

Los genes biosintéticos de biotina en *E. coli* se localizan en dos loci: Uno con el operón *bioBFCD* y el gen *bioA* divergente respecto a estos cuatro (figura 12). Los genes, *bioC* y *bioH*, podrían estar involucrados en etapas tempranas de esta biosíntesis y tienen una localización genómica distinta.

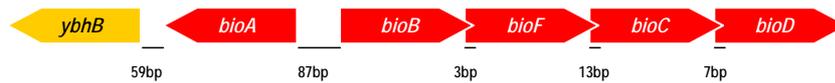


Figura 12. Estructura de unidades transcripcionales que contienen genes que codifican para proteínas involucradas en los últimos pasos de biotina en *E. coli*. Las líneas inferiores indican el espacio intergénico o el espacio en que dos genes vecinos se superlapan. El sentido de la transcripción está indicado como flechas de color rojo que a su vez, son genes que pertenecen a una unidad transcripcional.

El control de la síntesis de biotina en *E. coli* lo media BirA, una proteína bifuncional, que actúa como una ligasa biotina-proteína y también como represor transcripcional del operón de biotina. El operón *bioBFCD* contiene dos promotores no superlapanos en *E. coli*, dentro del espacio intergénico *bioA-bioB* (Shiuan *et al.*, 1994).

Enzima 8-amino-7-oxononanoato sintasa (7-KAPs) codificada por el gen bioF

El producto de *bioF* participa en la primera reacción de la vía y depende de piridoxal fosfato (PLP) para poder realizar su actividad. Esta enzima cataliza la conversión del ácido pimeloil-CoA hacia el ácido 7-ceto-8-aminopelargónico (EC 2.3.1.47). Es simétricamente homodimérica, con masa molecular de 42kD y exhibe un plegamiento clasificado dentro del plegamiento de tipo I, subclase II (Alexeev *et al.*, 1998).

Enzima ácido 7,8-diaminopelargónico sintasa (7-DAPA sintasa) codificada por el gen bioA

El producto de *bioA* participa en el antepenúltimo paso de la vía y al igual que el producto de *bioF*, depende de PLP. Es una enzima homodimérica, con masa molecular de 94kD y contiene 429 residuos por monómero. Cataliza la conversión del ácido 7-ceto-8-aminopelargónico hacia el ácido 7,8-diaminopelargónico (EC 2.3.1.62). La enzima usa S-adenosilmetionina como donador del grupo amino, la única aminotransferasa que usa un agente metilante para este fin. Estudios estructurales demuestran que esta enzima tiene una estructura clasificada dentro del plegamiento de tipo I, subclase II (Kack, Sandmark *et al.*, 1999). La estructura general de 7-DAPA es muy similar a la estructura de 7-KAPS, aunque la identidad entre ambas es muy baja (16%). Lo que sugiere una relación evolutiva entre ellas.

Enzima Detiobiotina sintetasa (DTBS) codificada por el gen bioD

El producto de bioD cataliza el penúltimo paso de la vía: La formación de detiobiotina a partir del ácido 7,8-diaminopelargónico y CO₂. DTBS es un homodímero, con una masa molecular de 24 kD por dominio (Huang, *et al.*, 1994).

Enzima biotina sintasa, codificada por el gen bioB

La Biotin sintasa es la responsable de la síntesis de biotina a partir de detiobiotina y azufre. cataliza el último paso de la vía. Se ha reportado que la Biotin sintasa produce menos de una molécula de biotina por molécula de proteína. tiene características muy particulares, puesto que mediante estudios *in vitro*, se ha reportado que Lo que impone un auténtico cuello de botella en la vía (Choi-Rhee *et al.*, 2005). La fuente del átomo de azufre para formar biotina, proviene del complejo [2Fe-2S], produciendo la inactivación y posterior degradación proteolítica de la enzima.

Por las razones que se han venido comentando, hemos seleccionado la dependencia de PLP como modelo de estudio, también hemos considerado la síntesis de biotina en *E. coli* porque podemos establecer un sistema de selección. Ahora bien, para modificar ciertas enzimas en algunos de sus parámetros cinéticos, debemos introducir cambios en las secuencias proteicas. Como desconocemos *a priori* el papel que desempeña cada aminoácido, tenemos la opción de usar herramientas que nos permitan introducir mutaciones aleatorias dentro de cada secuencia y seleccionar una de estas bajo el sistema de selección propuesto.

Evolución dirigida de proteínas

Durante millones de años la evolución natural ha introducido mutaciones, insertado, eliminado, recombinado y seleccionado variantes que le conferían una ventaja evolutiva a los organismos. Sin embargo, esta escala de tiempo no es practica para el trabajo en el laboratorio, por esta razón, a principio de los 90s los investigadores comenzaron a hablar de evolución dirigida como un conjunto de metodologías que permitían inducir mutaciones aleatorias en las partes codificantes de un gen o en segmentos específicos de ellos y seleccionar las enzimas con las propiedades deseadas en semanas o en meses.

La motivación principal para el desarrollo de estas técnicas fue el mejoramiento de enzimas de interés biotecnológico y el de comprender los mecanismos evolutivos. La mutagénesis sitio-dirigida solo puede ser aplicada en casos donde se conoce el papel de algunos aminoácidos importantes en las características catalíticas, selectividad y/o estabilidad de las enzimas. En la evolución dirigida no se requiere de este conocimiento para el mejoramiento de las enzimas. Se parte de la premisa de que no se tiene los suficientes conocimientos para diseñar “racionalmente” proteínas con las propiedades deseadas.

La evolución dirigida comienza con la generación de una población heterogénea de macromoléculas (ya sea de DNA o de RNA) y se realizan repetidas rondas de amplificación selectiva basada en propiedades físicas o bioquímicas de estas macromoléculas. Uno de los aspectos críticos y también una de las más grandes limitaciones en evolución dirigida es la selección de variantes proteicas con alguna propiedad deseada, especialmente si esta propiedad expresa una señal muy baja. Los requerimientos para una selección estricta (distinguir por ejemplo, una variante proteica con alguna propiedad entre los millones que conforman una población) y de alta sensibilidad (la intensidad de la señal buscada) generalmente se oponen. De allí es el reto impuesto al experimentalista para encontrar la combinación adecuada de presión de selección para obtener el resultado deseado.

Generación de diversidad

La diversidad genética se introduce al comienzo de los experimentos de evolución dirigida para crear una librería mutagénica. En primer lugar, se desea que todas las posibles mutaciones en estas librerías ocurrieran con la misma frecuencia, aunque en la actualidad no se cuenta con metodologías para ello. Por otro lado, la naturaleza además de mutaciones puntuales también genera inserciones y remociones, las cuales, con las tecnologías desarrolladas hasta el momento, no podemos igualar. Aún con todas estas limitaciones tecnológicas, se han modificado enzimas en sus parámetros cinéticos, estabilidad a temperaturas, modificación de pHs óptimos, etc. A continuación se enumeran algunas de las técnicas de mutagénesis más usadas.

PCR mutagénica (epPCR)

Con mucha frecuencia se inician los experimentos de evolución dirigida al crear diversidad a partir de PCR mutagénica. Existen dos metodologías generales para realizar una PCR mutagénica. En la más común se usa una amplia variedad de concentraciones de Mg^{2+} , $MnCl_2$ y diferentes proporciones de dNTPs para mutagenizar fragmentos de DNA de interés (Leung *et al.*, 1989). La PCR estándar genera una tasa

intrínseca de mutación muy pequeña para generar variantes potenciales; de acuerdo a las condiciones de reacción, de 0.1×10^{-4} a 2.0×10^{-4} por proceso de extensión (aproximadamente 1 nucleótido por cada 1000, en una PCR con 30 ciclos de extensión; Cadwell, 1992).

En la literatura podemos encontrar protocolos para PCR mutagénica que han reportado frecuencias de sustitución en el intervalo de 0.01 a 10%, de acuerdo con la metodología empleada (Vanhercke *et al.*, 2005). En la segunda metodología se usan nucleósidos análogos mutagénicos, los cuales se parecen estéricamente a los cuatro nucleósidos que conforman al DNA (A,T,G y C), con la particularidad de que adoptan frecuentemente una forma tautomérica que le permite aparearse con otro nucleótido diferente al complementario que sustituyó. Usando estos nucleósidos la frecuencia de mutación asciende de 9.7 a 32.2 sustituciones por cada 1000 nucleótidos (Zaccolo *et al.*, 1996; Penning *et al.*, 2001). En caso de obtenerse sustituciones favorables en distintos genes parentales, la PCR mutagénica no permite la combinación de mutaciones benéficas ni la aleatorización de secuencias > 3000 pb, lo cual restringe el número de variantes potenciales para explorar. Por otro lado, la PCR mutagénica presenta diversas fuentes de sesgos. La incorporación preferencial de las transiciones sobre las transversiones. Aún dentro de las transiciones, la transición AT→GC se favorece; lo mismo sucede con las transversiones, el cuál también está favorecida la transversión AT→TA. Existe un inherente sesgo debido a la naturaleza del código genético; únicamente se puede explorar un promedio de 5.6 aminoácidos por mutaciones sencillas por codón (Miyazaki, 1999). También podemos considerar el sesgo introducido en el proceso de amplificación de PCR. Debido a la naturaleza exponencial del proceso de amplificación, las moléculas que se amplifican en los primeros ciclos de extensión estarán sobre representadas al final del proceso. Esta fuente de sesgo sería difícil de detectar al secuenciar algunas clonas, en particular si hablamos de librerías grandes.

PCR sexual (DNA shuffling)

Esta técnica es un método para recombinación homóloga *in vitro* de genes mutantes (o de genes homólogos) por medio de una fragmentación por DNasa I y reensamblados en genes de tamaño completo por repetidos ciclos de apareamiento y extensión (PCR). Los fragmentos obtenidos después de la digestión funcionan como primers y se aparean con regiones con mayor identidad de otros genes, recombinando de esta forma mutaciones puntuales ya existentes, si es que las secuencias (genes) ya habían sido obtenidas a través de un proceso mutagénico, y al mismo tiempo, introduciendo nuevas mutaciones puntuales (Stemmer, 1994), (figura 13). El PCR sexual es un método combinatorio (recombinación), el cual permite explorar un número más grande de variantes potenciales en comparación al método anteriormente descrito.

A diferencia de la PCR mutagénica, esta técnica permite la aleatorización de secuencias > 3000 pb, además tiene una tasa de mutagénesis similar a la PCR mutagénica (0.77%). Se ha reportado que la DNasa I corta preferentemente en sitios adyacentes a nucleótidos de pirimidina, (Sambrock *et al.*,1989), lo cual reduce el número de variantes potenciales. La ventaja de esta metodología es que puede ser combinada con la PCR mutagénica o algún otro método que induzca mutaciones puntuales.

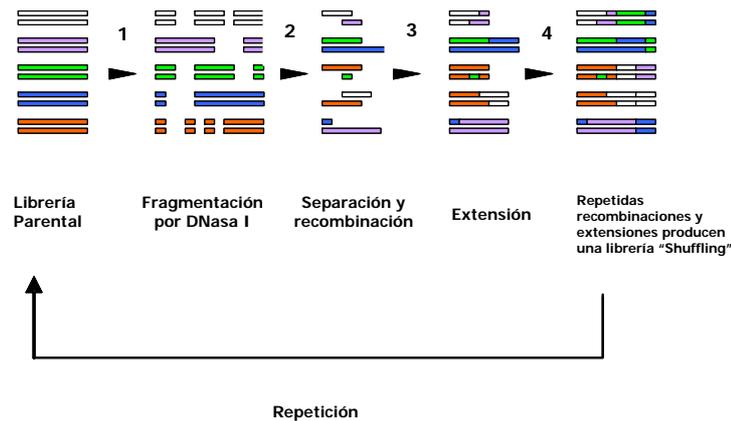


Figura 13. (1) Una librería de genes parentales (las muestras pueden provenir de ciclos de PCR mutagénica o algún otro método que induzcan mutaciones puntuales) es fragmentada por DNasa I. (2) Estos fragmentos son desnaturalizados y enfriados para permitir la recombinación. (3) Esto genera nuevas recombinaciones. Las recombinaciones se extienden. (4) Se repite el proceso de re-alineamiento, recombinación y extensión. Así se produce una librería de nuevos genes de tamaño completo que contienen varias recombinaciones (Modificado de Penning *et al.*, 2001).

Mutagénesis sitio dirigida

El concepto básico de la mutagénesis sitio dirigida es reemplazar una porción de un gen clonado en un plásmido por oligonucleótidos que contengan las mutaciones. El diseño de los oligonucleótidos debe incluir unas secuencias complementarias a la hebra patrón (del gen) flanqueado por ambos lados del sitio(s) de mutación. Los sitios de mutación pueden tener una secuencia preestablecida, de acuerdo con el tipo de aminoácido (s) que se quiera incorporar, o bien, pueden ser de secuencia aleatoria al usar oligonucleótidos cuya zona de mutación contenga las cuatro bases (o menos si se prefiere). El grado de aleatorización, y consecuentemente el número promedio de sustituciones de aminoácidos es un importante parámetro que puede ser controlado con esta metodología (Skandalis *et al.*, 1997).

Los métodos de mutagénesis sitio dirigida pueden clasificarse dentro de dos categorías: Aquellos que usan PCR y los que no usan PCR. Casi todos los métodos que no usan

PCR fueron los primeros en diseñarse (Hutchison *et al.*, 1978) y anteceden a la invención de la PCR. Por la sencillez metodológica, los métodos más usados son los que usan PCR que a continuación se describen.

QuikChange™ es un método disponible comercialmente (figura 14), la técnica para inducir mutaciones en sitios determinados consiste en principio, clonar un gen de la proteína de interés en un plásmido y transformar éste en una cepa *dam*⁺ de *E. coli* para que se metile. Una vez metilado, el plásmido se usa como templado en una reacción de PCR, con un par de oligos complementarios que contengan una o más mutaciones deseadas. Más adelante, el producto de PCR será una mezcla de plásmidos con muescas que contengan cadenas de DNA metiladas y no metiladas. A continuación, los grupos metilos se eliminan al tratar el producto de PCR con la endonucleasa DpnI (G^{me} A↓ TC). Las muescas se reparan al transformar el plásmido en la cepa *XLI Blue* de *E. coli*. (Weiner *et al.*, 1994).

GeneTailor™ es otro método para mutagénesis disponible comercialmente. La técnica consiste en incubar el plásmido (con el gen clonado en él) con una DNA metilasa antes de la reacción de mutagénesis. La metilasa metila residuos de citosina en determinados sitios específicos de dsDNA. La amplificación, vía PCR, se efectúa con un par de oligos que sobrelapan parcialmente, uno de los cuales, lleva las mutaciones deseadas. Al principio del proceso de la PCR, solo una de las hebras tendrá la mutación deseada, pero al avanzar en los ciclos de PCR, ambas hebras tendrán las mutaciones. Para eliminar el templado (wt) metilado, el producto de PCR es transformado en la cepa electrocompetente DH5α™-TI^R de *E. coli*; que contiene la endonucleasa *McrBr* que suele digerir DNA metilado de doble cadena, dejando de este modo, el plásmido no metilado y con las mutaciones deseadas (Kelleher, 1991; Waite-Rees *et al.*, 1991).

El método “overlap extension” (Reikofski *et al.*, 1992) usa cuatro oligonucleótidos, dos de los cuales tienen la mutación deseada y son parcialmente complementarios, al menos en la zona que se quiere mutar (figura 15). Los otros dos oligonucleótidos flanquean el gen. El método consiste en realizar dos PCR, una de las cuales usarán los oligonucleótidos A y C (figura 14) y la otra reacción usará los oligonucleótidos B y D. Los oligonucleótidos B y C son los que llevan la mutación deseada. Los productos de PCR son usados para una tercera reacción de PCR para hibridar los dos fragmentos. El producto resultante, se extiende en otra PCR usando los oligos A y D.

Una variante de la metodología “Overlap extension” es el llamado método “Megaprimer extension” (Kammann, *et al.*, 1989), el cual consiste en efectuar una PCR inicial con dos oligonucleótidos, uno de estos oligonucleótidos flanquea al gen en cuestión, y otro oligonucleótido hibrida internamente, en el sitio donde se desea inducir algún cambio. Éste mismo oligonucleótido llevará las mutaciones deseadas. El producto resultante, será usado como oligonucleótido grande (Megaprimer) para una segunda reacción en

combinación con otro oligonucleótido que flanquea al gen en el lado opuesto del oligonucleótido flanqueante de la primera reacción. En los métodos “Overlap extension” y “Megaprimer extension”, los oligonucleótidos flanqueantes llevan incorporados sitios de restricción para poder clonar el gen en un plásmido de expresión.

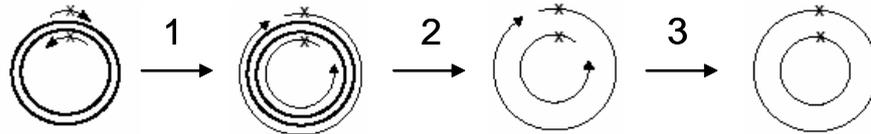


Figura 14. 1.- Método QuikChange™. Dos oligonucleótidos que contienen la mutación(s) deseada, X, son usados en una reacción de PCR, generando un plásmido con muescas. 2.- La enzima de restricción DpnI digiere las cadenas metiladas provenientes del templado. 3.- El plásmido con muescas es transformado en células electrocompetentes para cubrir las muescas (Modificado de Antikainen *et al.*, 2005).

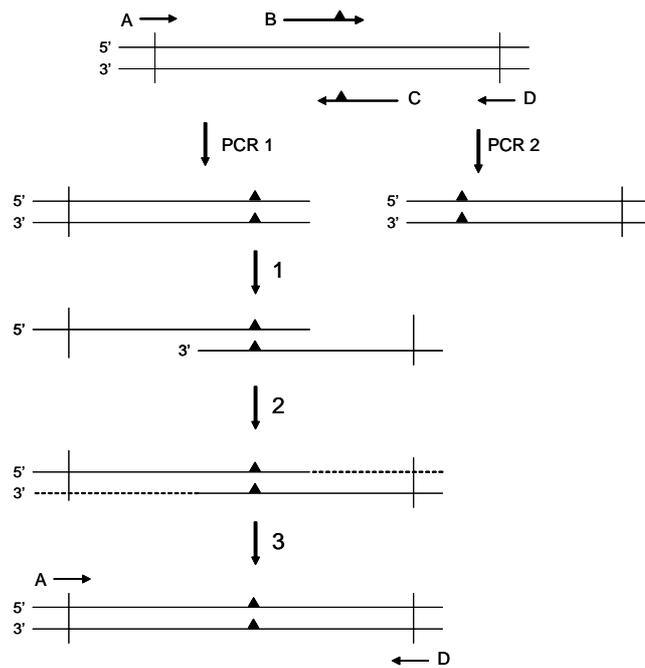


Figura 15. Método “Overlap extension”. Se realizan dos PCR paralelas, una con oligonucleótidos A y C, y otra con los primers B y D. Los oligonucleótidos que llevan la mutación (s) son B y C. 1.- Se realiza una PCR usando el producto de la PCR 1 y 2, sin oligonucleótidos de por medio. 2.- La reacción de PCR polimeriza las zonas faltantes. 3.- Como la reacción es poco eficiente, una PCR con oligonucleótidos externos (A y D) amplifica el gen que incluye las mutaciones deseadas (Modificado de Antikainen *et al.*, 2005).

PLANTEAMIENTO DEL PROBLEMA

Hemos visto que la naturaleza favorece el reciclaje de soluciones, en vez de inventar de “novo” cada solución. La idea fundamental de este trabajo es ampliar la especificidad por el sustrato de la enzima ácido 7,8 diaminopelargónico sintasa, (a la nos referiremos como BioA). Para que una variante de esta enzima pueda reconocer sustratos diferentes a los suyos (particularmente de la enzima 8-amino-7-oxononanoato sintasa, que será referida como BioF). Para seleccionar una variante con estas propiedades, es necesario un sistema de selección que nos permita discriminar cuál o cuales variantes han adquirido la propiedad de reconocer sustratos ajenos. El sistema de selección que diseñamos es la auxotrofía a biotina. Dicho sistema de selección tiene las siguientes características:

- 1.- La biotina es un cofactor indispensable para la célula. La remoción de uno o más genes que participen en la vía biosintética de biotina generan cepas no viables en un medio sin biotina.
- 2.- En la vía de síntesis de biotina participan dos enzimas parálogas (BioA y BioF), ambas dependientes de otro cofactor esencial, el piridoxal fosfato.
- 3.- Estas dos enzimas son aminotransferasas, que además de compartir la misma estructura, básicamente tienen el mismo mecanismo catalítico.
- 4.- La remoción de al menos uno de los genes que codifican para estas proteínas, generan una cepa auxótrofa a biotina.
- 5.- Si removemos *bioF* del genoma de *E. coli*, nos permite seleccionar variantes de BioA que pudieran complementar la función de BioF.
- 6.- Si removemos *bioA* del genoma de *E. coli*, nos permite seleccionar variantes de BioF que pudieran complementar la función de BioA.
- 7.- Si removemos *bioA* y *bioF* del genoma de *E. coli*, nos permitiría seleccionar variantes de BioA o de BioF que pudieran complementar la función del otro gen sin perder parte de su propia función.
- 8.- Debido a las ínfimas cantidades de biotina requeridas por los microorganismos para subsistir, es de suponerse que toda enzima que haya sido evolucionada hacia la función de alguna de las enzimas que participan en la vía, tiene oportunidades de ser seleccionada, aún cuando su actividad adquirida por evolución dirigida sea muy pequeña.

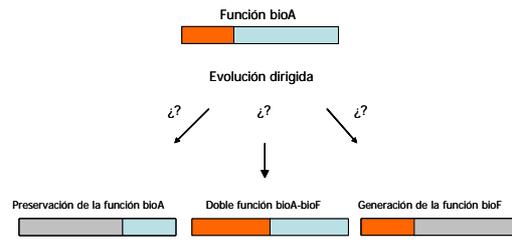


Figura 16. Posibles destinos de la función de BioA al ser sujeto de mutaciones aleatorias a lo largo del gene. De izquierda a derecha, es posible que BioA no pueda reconocer sustratos ajenos a los suyos; en un escenario más optimista, BioA podría adquirir la capacidad de reconocer los sustratos de BioF, sin perder su función original (ampliación); otra posibilidad es que las actividades BioA y BioF no puedan ser realizadas por una misma enzima, en este caso, BioA (migración).

Objetivos

Determinar si en la evolución de la especificidad de BioA pudieron existir ancestros comunes con características de intermediarios no específicos, como modelo para comprender la evolución de enzimas parálogas α -oxoaminas sintetas dependientes de PLP.

Objetivos particulares

- Análisis global de enzimas que usan PLP como cofactor, mediante la generación de una red de identidades.
- Obtención y selección fenotípica de variantes de BioA capaces de complementar la función de BioF, sin perder su propia actividad.
- Análisis *in vitro* de estas variantes, para demostrar la ampliación de la especificidad hacia la función de BioF sin perder la actividad como producto de BioA.
- Ensayos enzimáticos usando sustratos de otras enzimas (por definir) dependientes de PLP.

Hipótesis

Debido a la semejanza en el mecanismo catalítico y en el plegamiento (tipo I, subclase II) entre BioA y BioF de *E. coli*, pensamos que es factible la ampliación de la especificidad de BioA, puesto que con los métodos empleados en evolución dirigida y un sistema de selección eficiente, pensamos que tendremos mayores posibilidades de encontrar variantes de BioA que reconozcan los mismos sustratos de BioF, sin perder la capacidad de seguir reconociendo a los suyos, aunque haya un detrimento en sus parámetros cinéticos.

Metodología

Análisis global de la evolución de enzimas que unen PLP como cofactor

Realizamos una búsqueda de todas las enzimas que unen PLP como cofactor, en la base de datos SWISS-PROT (Bairoch *et al.*, 1991); para ello, SWISS-PROT da la opción de búsqueda de secuencias proteicas a partir de cofactores. A partir de este universo enzimático, efectuamos un corte al 70% de identidad para eliminar la redundancia, mediante el uso de CD-HIT (Li *et al.*, 2006). De aquí en adelante, identificamos los dominios a los que pertenecían cada secuencia con el software HMMER versión 2.3.1 (Gough *et al.*, 2001), que asocia cada secuencia a familias definidas en la base de datos SUPERFAMILY versión 1.65 (SCOP). La esperanza que se consideró para la identificación de dominios es $e \leq 0.001$. Una vez que identificamos los dominios en las secuencias proteicas, éstos son escindidos.

La relación de identidad entre todos los dominios proteicos se realizó con PSI-BLAST (Altschul *et al.*, 1997) hasta 5 iteraciones con una esperanza de $e \leq 0.001$. Con los datos obtenidos en la 5ª iteración de PSI-BLAST, efectuamos un corte al eliminar aquellos pares de enzimas cuya identidad es menor al 25%. Generamos una red de identidades usando CYTOSCAPE (Shannon *et al.*, 2003), considerando que los atributos de los nodos son dominios y las conexiones son identidades.

Por lo tanto, generamos un gráfico de una perspectiva global de enzimas dependientes de PLP, y asociamos al tipo de plegamiento al cual pertenecen, de acuerdo a la clasificación de Grishin *et al.*, 1995). La red obtenida es de 922 nodos (dominios estructurales) con 21 288 conexiones entre ellas.

Con los resultados obtenidos con PSI-BLAST, y graficados con CYTOSCAPE, obtuvimos varios grupos de secuencias. Cada grupo de secuencias se alinearon con CLUSTAL W (Thompson *et al.*, 1994). Verificamos cada alineamiento para limpiar las secuencias de gaps, eliminar fragmentos que no se alinearon con el resto de las secuencias y verificar si en éstas habían conjuntos de secuencias que tuvieran fusionados otros dominios. Cuando se encontró otros dominios fusionados (que modifican el valor de la identidad), se separaron y se desecharon.

A partir de cada grupo de alineamientos, construimos árboles filogenéticos sin raíz usando el principio de parsimonia (Felsenstein, 1993).

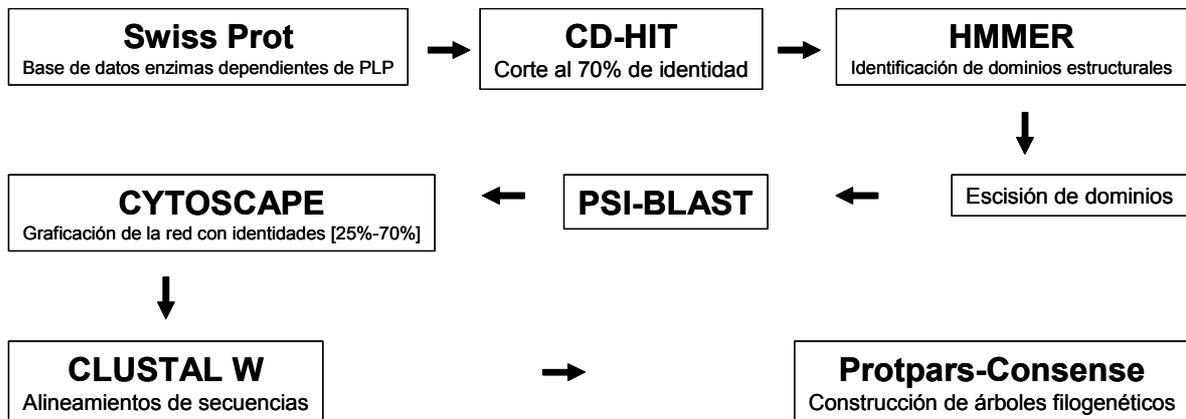


Figura 17. Diagrama de flujo para obtener la red de identidades de enzimas que usan PLP como cofactor.

Generación de cepas mutantes de *E. coli* MC1061 Δ bioA/ Δ bioF y MC1061 Δ bioA

El primer paso para lograr los objetivos de este proyecto es la obtención de mutantes de *E. coli* auxótrofas a biotina para seleccionar las variantes de *bioA* y *bioF* de interés. La cepa seleccionada para este trabajo es la MC1061. En el laboratorio ya se contaba con la cepa MC1061 Δ bioF, por lo que se procedió a remover el gen *bioA* en las cepas MC1061wt y MC1061 Δ bioF, para producir mutantes con remoción sencilla y doble, respectivamente. Para la remoción de los genes ubicados en determinados operones, se tuvo que tener cuidado de no afectar los genes vecinos, así como sus secuencias promotoras. De acuerdo a la estructura del operón de biotina en *E. coli* (figura 18), la remoción de *bioA* podría afectar las secuencias promotoras del ORF *ybhB* y de los genes *bioB*, *F*, *C*, *D*.

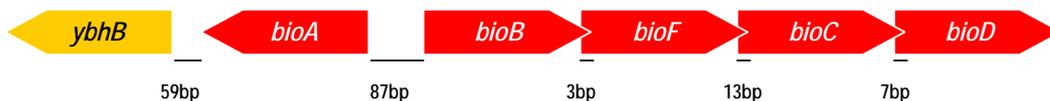


Figura 18. Estructura del operón de biotina en *E. coli*. En la parte inferior se indica la distancia entre cada gen. Para los genes *bioB*, *F*, *C* y *D* se indica los pares de bases que se solapan.

Por lo tanto, se consideró no remover completamente *bioA* al eliminar la parte central de éste y dejando solo los extremos (figura 19).

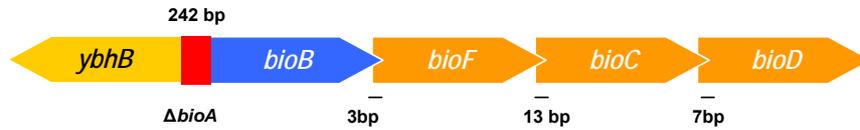


Figura 19. Estructura del operón de biotina en *E. coli* después de la remoción de *bioA*.

La remoción se llevó a cabo recombinando al cromosoma de *E. coli* un fragmento homólogo que excluya la región que se quiere remover y que incluya secuencias adyacentes a esta región (figura 20).

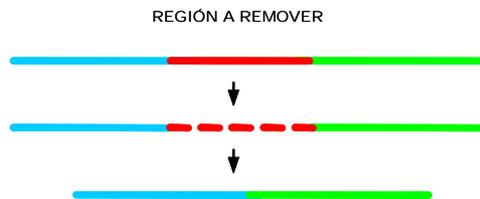


Figura 20. La región que se desea remover está ilustrada en color rojo. Las regiones adyacentes a ella se marcan en azul y verde.

Para producir el fragmento homólogo al cromosoma se comienza con la amplificación de las regiones 5' (fragmento *ybhB*, 1094 bp) y 3' (fragmento *bioB*, 1150 bp) de *bioA* y posterior ensamble. El ensamble es posible mediante el diseño de un oligo de dirección "forward" (*bioBfw*) que es complementario al extremo 3' de la región 5' adyacente a *bioA* (figura 21).

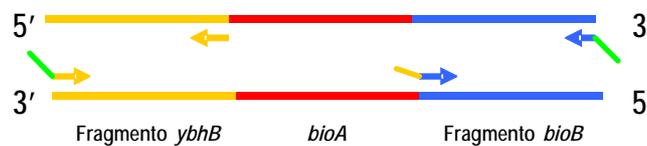


Figura 21. Se esquematiza la obtención de los dos fragmentos adyacentes a *bioA*. A la izquierda, en amarillo se muestra el fragmento *ybhB* y a la derecha, en azul el fragmento *bioB*. Al purificar ambos fragmentos, son ensamblados posteriormente mediante PCR. El ensamble es posible porque el extremo 5' del oligo *bioBfw* es complementario al fragmento *ybhB*.

Una vez que ya se cuenta con este fragmento, se clona en el vector pK03 para inducir remociones (Link *et al.*, 1997). Debido a que múltiples intentos por clonar este

fragmento fallaron, se decidió clonarlo mediante el uso de TOPO, para posteriormente clonarlo en el vector pK03.

El procedimiento para inducir remociones mediante el uso del vector pK03 (que confiere resistencia a cloranfenicol) comienza con la transformación de las cepas MC1061wt y MC1061 Δ *bioF* (para la mutante sencilla y doble, respectivamente) con la construcción

pK03-inserto homólogo. A continuación, las transformantes se extienden sobre cajas de LB con cloranfenicol y se incuban a 43°C, toda la noche. Solo crecen colonias cuyo plásmido ha sido integrado al cromosoma, puesto que este plásmido no puede replicarse autónomamente a 43°C. Se pican unas 1-3 colonias y se cultivan en tubos de LB con estreptomycin (la cepa es resistente a este antibiótico) a 30°C durante una hora para después extenderlas sobre cajas de LB con sacarosa al 10%, sin cloranfenicol. Todas aquellas células que porten al plásmido mueren en presencia de sacarosa, puesto que el vector contiene el gen *SacB* que codifica para la producción de levansacarasa a partir de sacarosa. De aquí se evalúa cuales colonias tienen el fenotipo deseado mediante réplica en caja. La réplica se efectúa con cajas con LB sin cloranfenicol, LB con cloranfenicol, y medio mínimo sin biotina. Las colonias candidatas son aquellas que se desarrollen en la caja de LB sin cloranfenicol, que no se desarrollen en la caja de LB con cloranfenicol ni en la caja de medio mínimo sin biotina.

Caso especial es la obtención de la cepa MC1061 Δ *bioA*/ Δ *bioF*, puesto que la cepa de partida MC1061 Δ *bioF*, es por si misma auxótrofa a biotina y en este caso no podríamos seleccionar candidatas por auxotrofia a biotina. Para evitar la evaluación de la remoción de *bioA* por PCR en cada colonia candidata (screening), complementamos la cepa MC1061 Δ *bioF* en “trans” con la construcción pUC18/*bioF*.

Evaluación de las cepas mutantes por PCR

Diseñamos oligos para verificar las remociones en las cepas MC1061 Δ *bioA*/ Δ *bioF* y MC1061 Δ *bioA*.

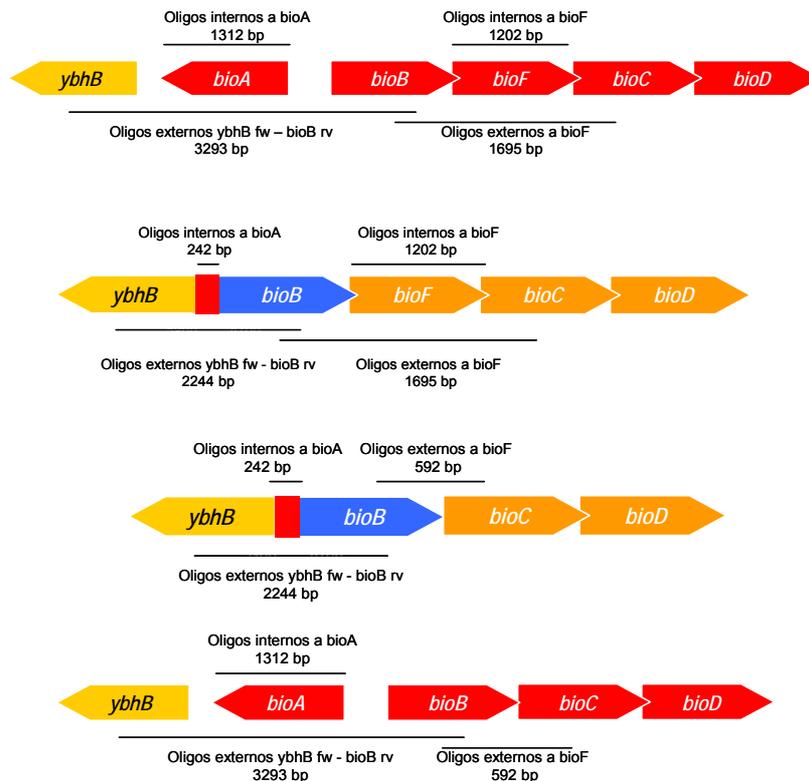


Figura 22. Estructura del operón de las distintas cepas mutantes. También se indica el tamaño de los fragmentos obtenidos al realizar PCR con los oligos indicados.

Evaluación de las cepas mutantes por complementación fenotípica

Para poder evaluar las mutantes derivadas de la cepa MC1061, hicimos las construcciones pUC18/*bioF*, pUC18/*bioA* con sitios de restricción EcoRI-SacI y SacI-HindIII, respectivamente, para ensamblarlos y generar una construcción doble pUC18*bioF/bioA*- con un solo promotor ubicado corriente arriba de *bioF*. Esta construcción se usó para evaluar la cepa MC1061 Δ *bioA*/ Δ *bioF*. Asimismo evaluamos las cepas mutantes MC1061 Δ *bioF* y MC1061 Δ *bioA* con las construcciones sencillas pUC18/*bioF* y pUC18/*bioA*.

Generación de diversidad

Usamos PCR mutagénica para generar diversidad sobre la secuencia de *bioA* y ligadas al vector pUC18.

Sistema de selección

La auxotrofia hacia la biotina es un método de selección extremadamente sensible para encontrar variantes que complementen esta auxotrofia, puesto que esta vitamina se requiere en cantidades extremadamente pequeñas (alrededor de 200 moléculas por

célula, Choi-Rhee *et al.*, 2005). Para ello necesitamos materiales y reactivos libres de biotina. Todos los reactivos utilizados fueron de calidad analítica y el material se lavó previamente con mezcla crómica para eliminar trazas de biotina. Otro punto importante a considerar fue el número máximo de colonias a sembrar en cajas de petri con medio mínimo. Debido a la experiencia adquirida durante este proyecto, el crecimiento de fondo de este sistema de selección es bastante alto, dificultando enormemente la labor de hallar clonas cuya variante tenga la actividad catalítica deseada; razón suficiente para que de esta manera las siembras por cajas no superaran finalmente las 50 000 ufc (ver apéndice para los detalles en la preparación del medio de selección).

Estrategia para la generación de librería de variantes de BioA y selección de variantes de éstas que pudieran haber ampliado su especificidad de sustrato

A partir de haber generado diversidad, ligamos en pUC18 y transformamos en la cepa de *E. coli* MC1061 wt. Es muy importante contar con alícuotas de células electrocompetentes de alta eficiencia de transformación. Las librerías se generaron transformando células electrocompetentes con un alícuota de 2 μ l de la ligación. Posteriormente se sembraron en cajas de LB suplementadas con ampicilina, 5, 10, y 20 μ l de células transformadas, para poder contar el número de UFC por cada 2 μ l de ligación. Evidentemente se desconoce *a priori* el tamaño de la librería, así que la estimación podrá realizarse ya sea al efectuar diluciones o sembrando un alícuota mayor al rango típico de [5,20] μ l, de acuerdo si el número de ufc es contable o no. Una vez que se tiene la librería de mutantes, ésta se transforma en células electrocompetentes de una de las cepas mutantes (MC1061 Δ bioA/ Δ bioF o MC1061 Δ bioF). La selección se lleva a cabo sembrando en cajas con medio mínimo, explorando 10 veces más el número de clonas que el tamaño estimado de la librería. Las cajas se incuban a 37° C y se mantienen así hasta la aparición de colonias candidatas (comúnmente entre 3 y 8 días), a partir de este tiempo el crecimiento de fondo aumenta gradualmente hasta dificultar la búsqueda de más colonias candidatas. Es importante comentar que frecuentemente el crecimiento de fondo aparece al cabo de tres días.

En los casos donde se obtienen colonias candidatas, éstas son cultivadas en LB para poder extraer el plásmido asociado a ellas y en consecuencia volver a transformar en células electrocompetentes de la misma cepa mutante para confirmar el fenotipo. El procedimiento general de todo el proceso se muestra en la figura 23.

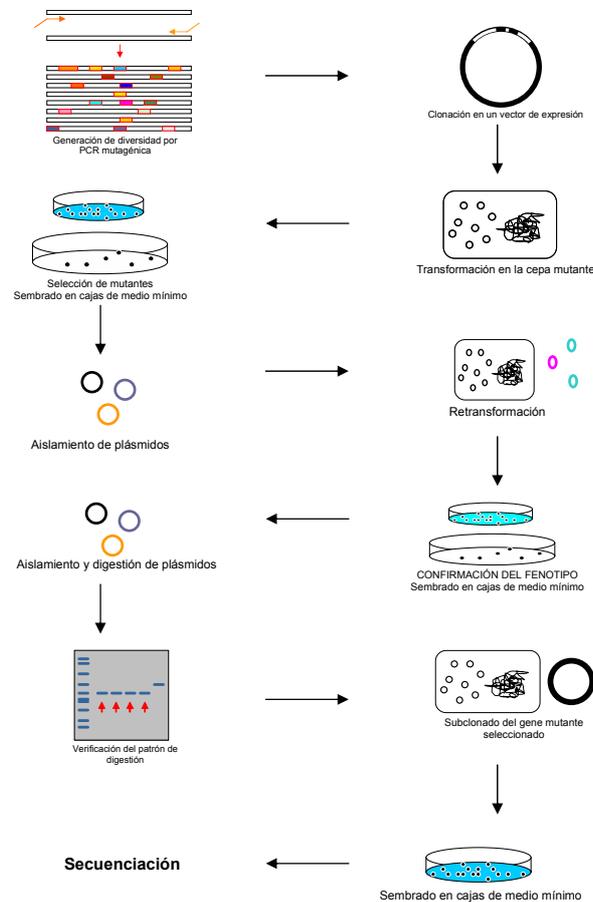


Figura 23. Se parte de la extensión vía PCR mutagénica de un gen en particular para generar diversidad. Esta diversidad se clona en un vector de expresión (pUC18) y se transforma por electroporación en una cepa que adquiera DNA exógeno en alta eficiencia. Mientras más grande sea el número de transformantes, se deduce que mayor será el número de variantes distintas para su posterior exploración. Una vez que se ha obtenido un determinado número de transformantes, a estas se les extrae sus respectivos plásmidos y se vuelve a electroporar en una cepa de interés. La selección positiva se realiza al extender estas transformantes en medio mínimo adicionado con el antibiótico que funge como marcador (para pUC18 es ampicilina). Es importante poner énfasis en el lavado de células, puesto que la biotina residual puede generar intensos falsos positivos. Todas las colonias que logren desarrollarse serán consideradas candidatas y se recogen para cultivarse en medio rico (LB). Se vuelve a lavar cuidadosamente y se retransforma en la misma cepa para poder continuar la selección positiva (confirmación del fenotipo). En caso de que estas clonas mantengan el fenotipo, es necesario subclonar esta variante, para realizar de nueva cuenta la selección positiva y verificar si aún al ser subclonado mantiene el fenotipo. En caso afirmativo, se secuencia esta variante.

Resultados y discusión

OBTENCIÓN DE LAS MUTANTES E. coli MC1061ΔbioA/bioF Y MC1061ΔbioA

Mutante E. coli MC1061ΔbioA (mutante sencilla)

Para la obtención de la mutante sencilla se tuvieron que evaluar alrededor de 360 colonias para encontrar dos colonias con el fenotipo deseado (auxotrofia a biotina, figura 24). Las colonias que mostraron el fenotipo deseado se recuperaron de la caja con LB sin cloranfenicol

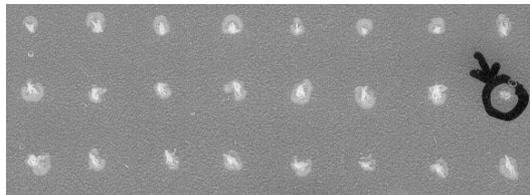


Figura 24. Obtención de la cepa MC1061ΔbioA. Caja con medio mínimo sin biotina. La colonia marcada en círculo es una de las dos colonias que mostraron auxotrofia a biotina.

Mutante E. coli MC1061ΔbioA/ΔbioF (mutante doble)

Para obtener la doble mutante, la cepa parental MC1061ΔbioF se complementó con bioF en pUC18, puesto que esta cepa es de por sí auxótrofa a biotina. En total, para la obtención de la doble mutante se tuvieron que picar alrededor de 320 colonias para encontrar 15 colonias candidatas con el fenotipo deseado (figura 25). Las colonias que mostraron el fenotipo deseado fueron recuperadas de la caja con LB sin cloranfenicol.

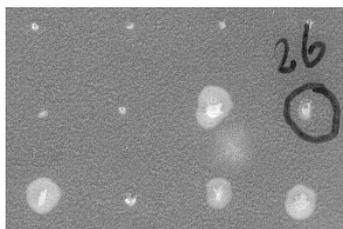


Figura 25. Obtención de la cepa MC1061ΔbioA/ΔbioF. Caja con medio mínimo con ampicilina y sin biotina. La colonia marcada en círculo es una de las 15 colonias que mostraron auxotrofia a biotina.

El siguiente paso para la obtención de la doble mutante fue “curarlo” de bioF en pUC18. Esto se efectuó picando una colonia de esta mutante y cultivándola en 5 ml de LB sin antibiótico, 37°C durante 12 horas. Se tomó un alícuota de 300μl y se inoculó en un matraz de 300ml sin antibiótico.

El proceso de curación se monitoreó tomando un alícuota para extenderlo en cajas con LB ampicilina y LB sin ampicilina. En este paso no se observó crecimiento en la caja con LB ampicilina. Las colonias que crecieron en las cajas con LB sin ampicilina se picaron mediante réplica en otra caja con LB ampicilina. Ninguna colonia volvió a crecer en esta última caja. Estas colonias “curadas” se evaluaron por PCR para verificar la pérdida de *bioF* en pUC18. Para esto se usaron oligos internos a *bioF*.

Análisis de las remociones por PCR

Para el análisis por PCR se usó DNA genómico de las cepas MC1061 wt, MC1061 Δ *bioA* (mutante sencilla), MC1061 Δ *bioA*/ Δ *bioF* (mutante doble) y MC1061 Δ *bioF* (mutante sencilla). En la tabla 1 se muestran los tamaños esperados de los fragmentos y en la figura 26 el gel con los productos de PCR.

Templado	Pares de Oligos	Tamaño del producto de PCR (bp)
MC1061 wt	Oligos <i>bioA</i> internos	1312
	Oligos externos (<i>ybhB</i> fw – <i>bioB</i> rv)	3293
	Oligos <i>bioF</i> internos	1202
	Oligos <i>bioF</i> externos	1695
MC1061Δ<i>bioA</i>	Oligos <i>bioA</i> internos	242
	Oligos externos (<i>ybhB</i> fw – <i>bioB</i> rv)	2244
	Oligos <i>bioF</i> internos	1202
	Oligos <i>bioF</i> externos	1695
MC1061Δ<i>bioF</i>-<i>bioA</i>	Oligos <i>bioA</i> internos	242
	Oligos externos (<i>ybhB</i> fw – <i>bioB</i> rv)	2244
	Oligos <i>bioF</i> internos	-
	Oligos <i>bioF</i> externos	592
MC1061Δ<i>bioF</i>	Oligos <i>bioA</i> internos	1312
	Oligos externos (<i>ybhB</i> fw – <i>bioB</i> rv)	3293
	Oligos <i>bioF</i> internos	-
	Oligos <i>bioF</i> externos	592

Tabla 1. Cepas mutantes y tamaños de los fragmentos al extender DNA genómico de éstas con los oligos indicados.

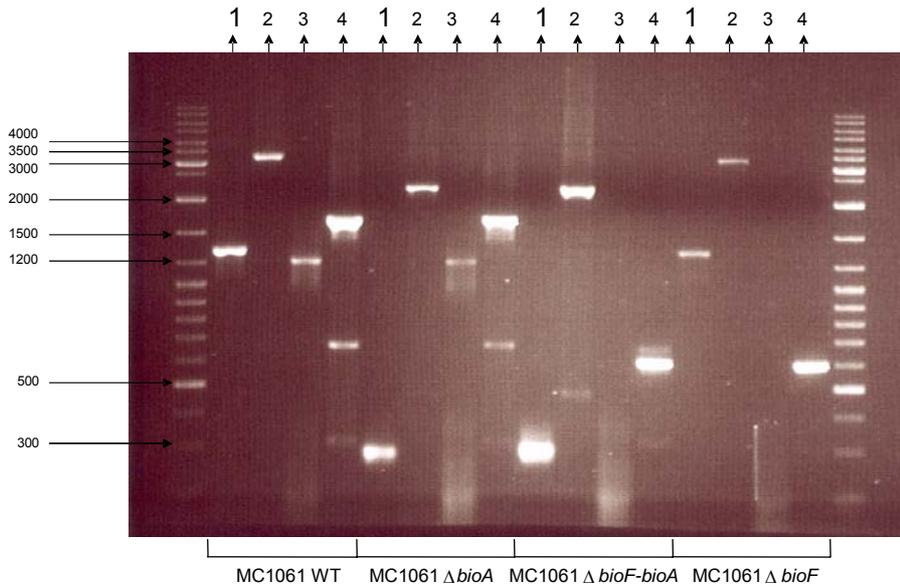


Figura 26. Productos de PCR para demostrar la remoción en las cepas de *E. coli* MC1061 Δ *bioA* MC1061 Δ *bioA*/ Δ *bioF*. Cada cepa fue evaluada con cuatro oligos. Los oligos fueron enumerados de la siguiente forma: 1.-Oligos *bioA* internos; 2.- Oligos externos (*ybhB* fw – *bioB* rv) 3.- Oligos *bioF* internos 4.- Oligos *bioF* externos. Las cepas son indicadas en las cajas mostradas en la parte inferior. El gel se hizo con agarosa al 1.2%.

Análisis de las remociones por complementación

Una vez establecida las remociones en ambas cepas, se procedió a complementar la auxotrofia para evaluar si durante el proceso de remoción se indujeron mutaciones polares en las cepas MC1061 Δ *bioA* y MC1061 Δ *bioA*/ Δ *bioF*.

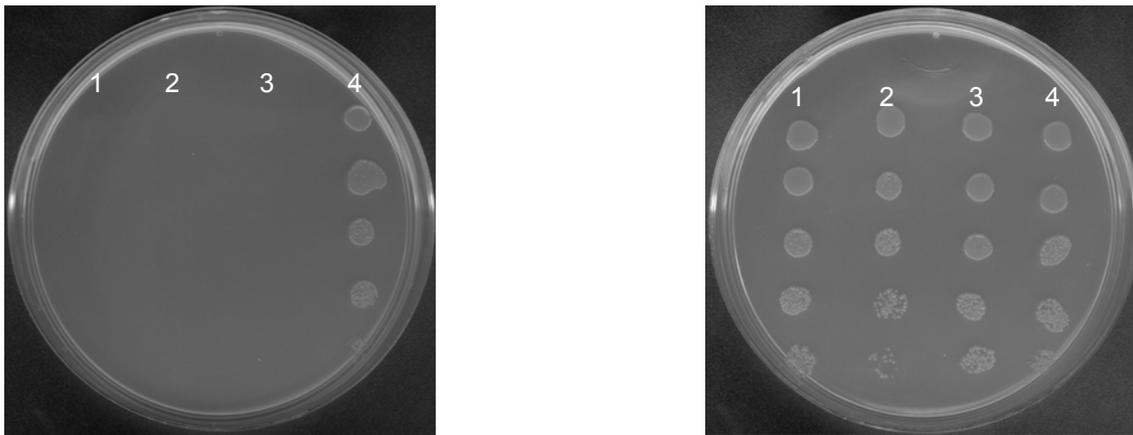


Figura 27. Complementación por gota de la cepa MC1061 Δ *bioA*/ Δ *bioF*. Izquierda, caja con medio mínimo y ampicilina. 1- pUC18, 2.- pUC18/*bioF*, 3.- pUC18/*bioA*, 4.- pUC18*bioF*/*bioA*. Derecha, caja con medio mínimo adicionada con biotina.

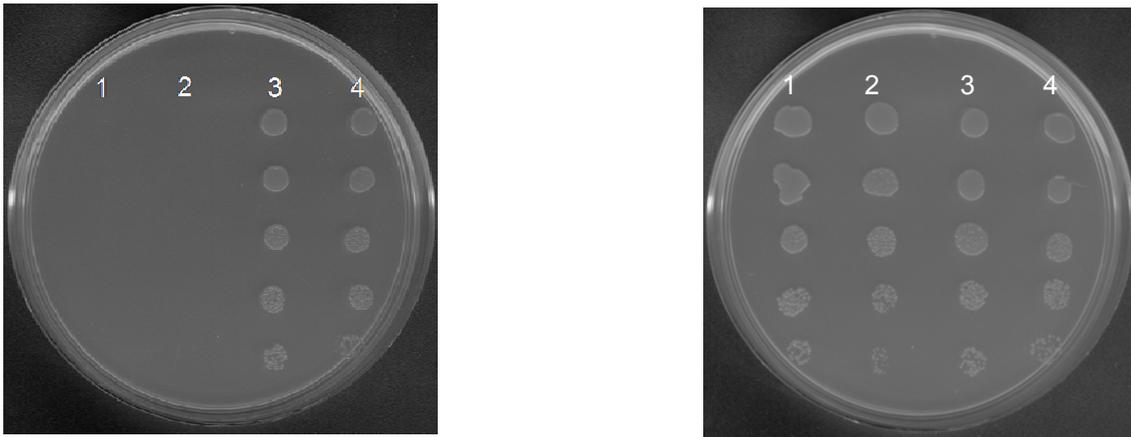


Figura 28. Complementación por gota de la cepa MC1061 Δ *bioA*. Izquierda, caja con medio mínimo y ampicilina. 1.- pUC18, 2.- pUC18/*bioF*, 3.- pUC18/*bioA*, 4.- pUC18*bioF/bioA*. Derecha, caja con medio mínimo adicionada con biotina.

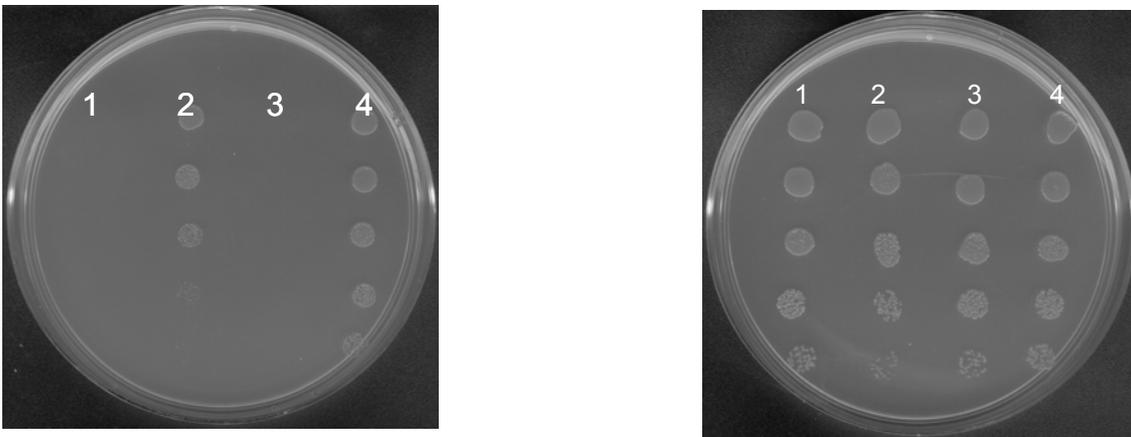


Figura 29. Complementación por gota de la cepa MC1061 Δ *bioF*. Izquierda, caja con medio mínimo y ampicilina. 1.- pUC18, 2.- pUC18/*bioF*, 3.- pUC18/*bioA*, 4.- pUC18*bioF/bioA*. Derecha, caja con medio mínimo y ampicilina adicionada con biotina.

La cepa MC1061 Δ *bioA*/ Δ *bioF* puede crecer en medio mínimo sin biotina cuando es transformada con la construcción pUC18 *bioF/bioA* (figura 27). La cepa MC1061 Δ *bioA* puede crecer en medio mínimo sin biotina cuando es transformada con las construcciones pUC18 *bioA* y pUC18 *bioF/bioA* (figura 28). La cepa MC1061 Δ *bioF* puede crecer en medio mínimo sin biotina cuando es transformada con las construcciones pUC18 *bioF* y pUC18 *bioF/bioA* (figura 29). Por lo que se demuestra que no se introdujeron mutaciones que afecten la expresión de los genes del resto del operón.

Análisis de los genes bioA y bioF en la cepa de E. coli MC1061 wt

Antes de comenzar con la producción de las librerías de variantes secuenciamos los genes *bioA* y *bioF* de la cepa MC1061 de *E. coli*. Las amplificaciones se realizaron con la polimerasa Vent™ con actividad exonucleasa 3'→5'. Una vez amplificados estos genes se clonaron en pUC18 para su posterior secuenciación. Se compararon nueve secuencias de *bioA*, tres obtenidas muy anteriormente y secuenciadas parcialmente.

Las otras seis se secuenciaron completamente de los cuales, cuatro de estas clonas presentan exactamente la misma secuencia respecto a las tres anteriores y solo dos tienen cambios de aminoácidos, lo cuál sugiere que los cambios de aminoácidos en estas dos secuencias es producto de mutaciones generadas durante el proceso de amplificación por PCR. Al comparar todas las secuencias de BioA de *E. coli* reportadas en el NCBI encontramos que presenta muchos cambios aminoacídicos a lo largo de su secuencia. Ya se disponía de una construcción pUC18/*bioF* y se comparó con todas las secuencias reportadas en el NCBI de *bioF*. Para el caso de *bioF* se encontró que el gen en la construcción pUC18/*bioF* tiene la misma secuencia que la reportada en el NCBI.

Análisis de la librería de variantes, primera ronda mutagénica

Generamos diversidad en el gen *bioA* de *E. coli* usando PCR mutagénica. Posteriormente se efectuó la ligación con pUC18 como vector. A continuación transformamos la ligación en la cepa de *E. coli* MC1061 wt y sembramos en cajas de LB adicionadas con ampicilina como marcador. Después de 12 horas de incubación, se picaron 40 colonias y a éstas les fue extraído el plásmido para verificar el porcentaje de ligación. Con base al patrón de restricción (EcoRI-HindIII) todas las colonias analizadas mostraron la presencia del gen *bioA*. De aquí se tomaron al azar 15 muestras y se enviaron a secuenciar, de las cuales 13 secuencias fueron aptas para su análisis

Número de clonas analizadas	13
Número total de cambios	33
Espacio explorado	16 770 nucleótidos
Tasa mutagénica	0.197% (≈2.54 cambios de nucleótidos por secuencia)
Transversiones	8 (24%)
Transiciones	25 (76%)

Mutantes	Número de cambios					
	1	2	3	4	5	6
bioA 1	wt					
bioA 2	C 304 G*					
bioA 3	T 231 C	G 587 C*	A 822 G			
bioA 4	A 91 T*	A 311 G	T 317 C	A 899 G		
bioA 5	T 167 C	G 189 C*	C 465 T			
bioA 6	T 561 C	C 866 T				
bioA 7	T 116 A*	T 120 G	A 205 G	A 230 G	A 628 T*	T 819 C

bioA 8	wt					
bioA 9	A 538 G	T 928 G	T 980 C			
bioA 10	T 120 C	A 718 G	T 1136 C			
bioA 11	T 118 A*	A 664 G	A 805 G	A 857 T*	T1074 A*	
bioA 12	G 291 A	A 879 C				
bioA 13	T 103 C					

Tabla 2. Se muestran todos los cambios generados mediante PCR mutagénica en las 13 clonas que se secuenciaron. En negritas se muestran todas las transversiones. Los restantes son transiciones.

A partir de estos datos evaluamos la diversidad de la librería mutagénica, con base a la estadística reportada por Patrick *et al.*, 2003. Asumiendo una distribución de Poisson, generan un programa llamado PEDEL (Program for Estimating Diversity in Error-prone PCR Libraries). Este programa estima el número de variantes de secuencia única, C, así como la composición de sublibrerías con X mutaciones, entre otras. Los parámetros de entrada para PEDEL es el largo de la secuencia N en pares de bases, el tamaño de la librería L y el número medio de mutaciones puntuales λ .

X	Px	Lx	Vx	Cx	Cx/Vx (%)
0	0.079	7.9×10^4	1.0×10^0	1.0×10^0	100
1	0.20	2.0×10^5	3.9×10^3	3.9×10^3	100
2	0.25	2.5×10^5	7.5×10^6	2.5×10^5	3.4
3	0.22	2.2×10^5	9.6×10^9	2.2×10^5	2.2×10^{-3}
4	0.14	1.4×10^5	9.3×10^{12}	1.4×10^5	1.5×10^{-6}
5	0.069	6.9×10^4	7.2×10^{15}	6.9×10^4	9.7×10^{-10}
6	0.029	2.9×10^4	4.6×10^{18}	2.9×10^4	6.4×10^{-13}
7	0.011	1.1×10^4	2.5×10^{21}	1.1×10^4	4.2×10^{-16}
8	0.0034	3.4×10^3	1.2×10^{24}	3.4×10^3	2.8×10^{-19}
9	0.00096	9.6×10^2	5.2×10^{26}	9.6×10^2	1.8×10^{-22}
10	0.00024	2.4×10^2	2.0×10^{29}	2.4×10^2	1.2×10^{-25}
11	0.000056	5.6×10^1	7.0×10^{31}	5.6×10^1	8.0×10^{-29}
12	0.000012	1.2×10^1	2.2×10^{34}	1.2×10^1	5.3×10^{-32}
13	0.0000023	2.3×10^0	6.6×10^{36}	2.3×10^0	3.5×10^{-35}
14	0.00000042	4.2×10^{-1}	1.8×10^{39}	4.2×10^{-1}	2.3×10^{-38}
15	0.000000071	7.1×10^{-2}	4.6×10^{41}	7.1×10^{-2}	1.5×10^{-41}
16	0.000000011	1.1×10^{-2}	1.1×10^{44}	1.1×10^{-2}	1.0×10^{-44}
17	0.0000000017	1.7×10^{-3}	2.5×10^{46}	1.7×10^{-3}	6.8×10^{-48}
18	0.00000000024	2.4×10^{-4}	5.3×10^{48}	2.4×10^{-4}	4.5×10^{-51}
19	0.000000000032	3.2×10^{-5}	1.1×10^{51}	3.2×10^{-5}	3.0×10^{-54}

20	0.00000000000040	4.0×10^{-6}	2.0×10^{53}	4.0×10^{-6}	2.0×10^{-57}
Número esperado de secuencias distintas: 0.72×10^6 . Para una librería de $1 \times 10^6 = 720\,000$					

x	Número exacto de mutaciones por secuencia
Px	Probabilidad de Poisson de X mutaciones, dada una tasa mutagénica
Lx	Número esperado de secuencias en la librería con exactamente X mutaciones
Vx	Número de posibles variantes con exactamente X mutaciones
Cx	Número esperado (aprox) de distintas variantes en la librería con exactamente X mutaciones

Tabla 3. Datos generados en PEDEL.

Una vez teniendo la distribución de variantes y su respectiva división en sublibrerías con X cambios cada una, nos preguntamos acerca del espacio proteico explorado de acuerdo con esta distribución de variantes a nivel de DNA, por ejemplo, de acuerdo con la tabla 3, en la librería existen 69 000 variantes únicas con cinco mutaciones, ¿Cuántos cambios netos de aminoácidos podemos explorar con variantes con 5 cambios?, ¿con X cambios? Para saber esto, tenemos que considerar que muchas de las mutaciones son neutrales, la naturaleza del código genético que solo permite explorar en promedio 5.6 aminoácidos cuando se muta una posición por codón (Miyazaki *et al.*, 1999), además la tasa de transiciones y transversiones producida por la PCR mutagénica generan secuencias sesgadas. En consecuencia, calculamos la probabilidad de generar cambios netos de aminoácidos cuando han ocurrido cierto número de mutaciones en el gen:

Sean **n** mutaciones en el gen, **m** cambios netos de aminoácidos, **n-m** no cambios de aminoácidos:

$$\binom{n}{m} P_c^m (1 - P_c)^{n-m}$$

Probabilidad de tener **m** cambios de aminoácidos en la proteína, dado que hubiera **n** mutaciones en la secuencia de DNA.

Para calcular esto, primero tenemos que hallar P_c , que es la probabilidad de cambio de un aminoácido cuando ocurre un cambio al azar en cualquiera de los tres nucleótidos del codón.

$$P_c = \frac{P_{c1} + P_{c2} + P_{c3}}{3}$$

Donde P_c es la probabilidad de cambio de un aminoácido a otro cuando sucede un cambio al azar en un codón en cualquier nucleótido de éste. P_{c1} , P_{c2} y P_{c3} son las probabilidades de cambio de aminoácido cuando ha ocurrido una mutación en la 1era, 2da y 3era posición, respectivamente.

Calculando P_{c1} , P_{c2} y P_{c3}

Para el cálculo de P_{c1} , tenemos que hallar la suma del producto de la frecuencia de los codones por la probabilidad de cambio de aminoácido en algún codón de la secuencia de interés, es decir:

$$P_{c1} = \sum_{i=1}^{61} f_{\text{codón } i} (\theta_{\text{codón } i})$$

Donde $\theta_{\text{codón } i}$ es la probabilidad de cambio de un aminoácido, cuando ocurre una mutación en el primer nucleótido del codón. La frecuencia $f_{\text{codón } i}$ es el cociente de las veces que aparece el codón i entre el número de codones de la secuencia. Para este análisis, los tres codones de paro no son tomados en cuenta.

Para calcular $\theta_{\text{codón } i}$ se toma en cuenta la tasa de transiciones y transversiones. A partir de los datos generados al secuenciar cierto número de secuencias con mutaciones, obtuvimos que la probabilidad de transiciones $P_t = 0.76$ y la probabilidad de transversiones $P_\tau = 0.24$

Para el caso de los siguientes tripletes:

1era posición	Aminoácido codificado	Probabilidad de cambio de aa $P_{c_{xyz}}$
TTA	L	0.24
CTA	L	0.24
ATA	I	1
GTA	V	1

Tabla 4.

Los codones TTA y CTA codifican el mismo aminoácido. En caso de producirse una transición T→C no se produciría cambio alguno de aminoácido. Solo de ocurrir una transversión se producirá un cambio de aminoácido. Entonces, calculamos la probabilidad de cambio de aminoácido del codón TTA ($\theta_{\text{codón TTA}}$)

$$\theta_{\text{codón TTA}} = P_t \cdot \theta_{\text{codón TTA}}(t) + P_\tau \cdot \theta_{\text{codón TTA}}(\tau)$$

Donde $\theta_{\text{codón TTA}}(t)$ es la probabilidad de cambio de aminoácido del codón TTA, cuando ha ocurrido una transición. $\theta_{\text{codón TTA}}(\tau)$ es la probabilidad de cambio de aminoácido cuando ha ocurrido una transversión, recordando que $P_t = 0.76$ y $P_\tau = 0.24$

En este caso en particular tenemos:

$$\theta_{\text{codón TTA}} = (0.76)(0) + (0.24)(1) = 0.24$$

Así sucesivamente calculamos la probabilidad de cambio de aminoácido de cada codón presente en la secuencia. La probabilidad de cambio de aminoácido en la 1era posición finalmente se calcula de la siguiente forma:

$$P_{c1} = \sum_{i=1}^{61} f_{\text{codón } i} (\theta_{\text{codón } i}) = 0.945$$

De la misma manera se calcula P_{c3} , puesto que P_{c2} siempre es 1 (todo cambio de

nucleótido en la 2da posición de cualquier codón produce un cambio de aminoácido)

$$P_{c3} = \sum_{i=1}^{61} f_{\text{codón } i} (\theta_{\text{codón } i}) = 0.395$$

La probabilidad de cambio de aminoácido al modificar al azar cualquiera de los tres nucleótidos del codón es:

$$P_c = \frac{(0.945 + 1 + 0.395)}{3} = 0.781$$

Una vez que establecimos el valor de P_c , pudimos encontrar la distribución de aminoácidos cambiados en función del número de mutaciones.

$$\binom{n}{m} P_c^m (1 - P_c)^{n-m}$$

Donde “n” es el número de mutaciones a nivel de DNA y “m” es el número de aminoácidos cambiados al haber “n” mutaciones. Evidentemente el rango de “m” es [0,n], puesto que el número de aminoácidos cambiados no puede ser mayor que el número de mutaciones

X	Número de aminoácidos cambiados										
	0	1	2	3	4	5	6	7	8	9	10
1	21.9	78.1									
2	4.8	34.2	61								
3	1.1	11.2	40.1	47.7							
4	0.23	3.3	17.5	41.7	37.2						
5	0.05	0.9	6.4	22.8	40.7	29.1					
6	0.01	0.24	2.1	10	26.8	38.2	22.7				
7	0.002	0.060	0.64	3.8	13.7	29.3	34.8	17.8			
8	0.0005	0.02	0.19	1.3	6	17.1	30.5	31.1	13.9		
9	0.0001	0.004	0.05	0.44	2.4	8.4	20	30.6	27.3	10.8	
10	0.00003	0.0009	0.0144	0.1375	0.859	3.68	10.9	22.3	29.9	23.7	8.5

Tabla 5. Se muestra la distribución de variantes proteicas en función del número de mutaciones a nivel DNA. Por ejemplo, con una población de variantes a nivel DNA con 3 mutaciones en cada molécula de esta población ($X=3$), tenemos que el 1.1% de ésta no genera ningún cambio de aminoácido. El 11.2% de las variantes generan un cambio de aminoácido. Con 40.1% se observan dos cambios aminoacídicos y el 47.7% generan tres cambios de aminoácidos. Los números en rojo indican el número de cambio de aminoácidos más abundantes a cada distancia mutacional X.

Selección de variantes

Una vez que obtuvimos la librería estimada en 1.0×10^6 comenzamos la selección de variantes. Independientemente exploramos la librería en las cepas mutantes de *E. coli* MC1061 Δ bioA/ Δ bioF y MC1061 Δ bioF. Cada librería se exploró 10 veces más que el tamaño estimado ($\approx 10\,000\,000$ variantes). En consecuencia, obtuvimos un número indeterminado de variantes candidatas que sucesivamente fuimos eliminando hasta seleccionar 3 clonas que parecían mantener el fenotipo en la cepa MC1061 Δ bioA/ Δ bioF y una sola clona en la cepa MC1061 Δ bioF. A estas tres clonas les fueron extraídos sus respectivos plásmidos para reintroducirlos en la cepa MC1061 Δ bioA/ Δ bioF.

Análisis de variantes de la primera ronda mutagénica

A continuación realizamos el ensayo de complementación por gota para poder detectar pequeñas diferencias en la intensidad de los fenotipos mostrados por las tres variantes candidatas (figura 30).

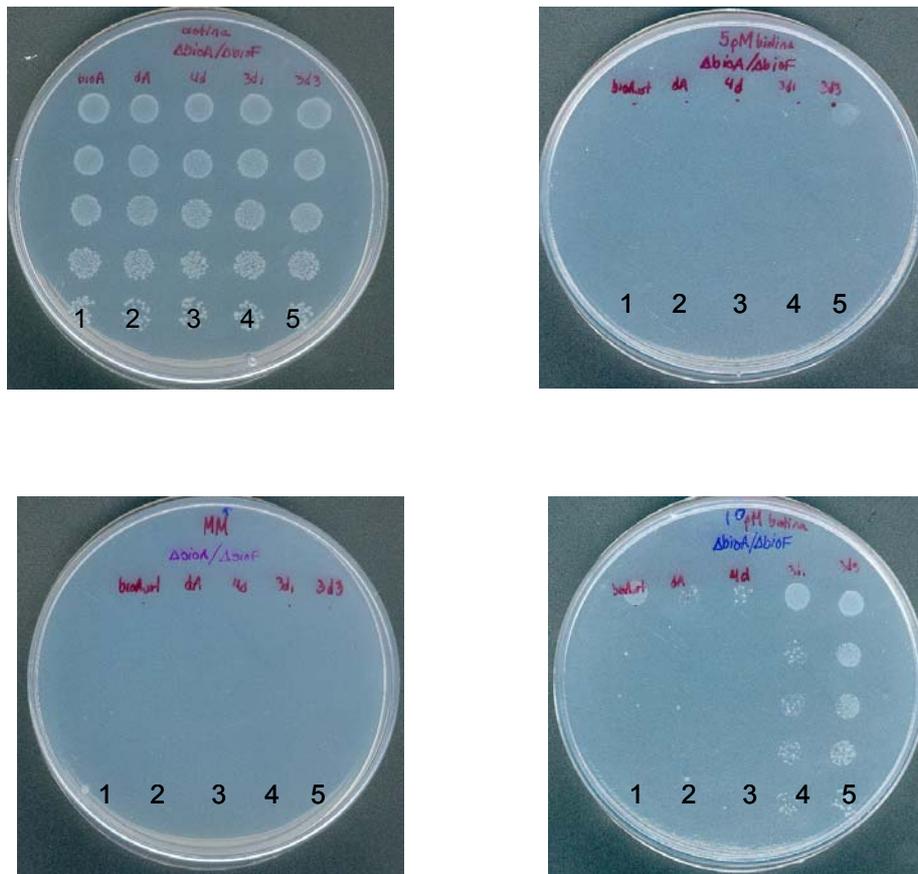


Figura 30. Prueba de complementación de las variantes dA, 4d, 3d1 y 3d3 (de izquierda a derecha, carriles 2, 3, 4 y 5, respectivamente), como control negativo se usó bioA wt (carril 1). Todas estas fueron transformadas en la cepa *E. coli* MC1061 Δ bioA/ Δ bioF. Al secuenciar estas variantes se demostró que las variantes 3d1 y 3d3 tienen la misma secuencia. La variante 3d1 fue aislada a partir de la selección en la cepa MC1061 Δ bioF. Las demás variantes fueron aisladas en la cepa MC1061 Δ bioA/ Δ bioF. La prueba fue preparada con cajas con medio mínimo + biotina 10 mM, (arriba, izquierda), con cajas adicionadas con 5pM (arriba, derecha), 10pM (abajo, derecha) y 0pM (abajo, izquierda) de biotina. Todas las cajas tenían ampicilina como marcador.

Las variantes 3d1 y 3d3 fueron las únicas que complementaron a la cepa de *E. coli* MC1061 Δ bioA/ Δ bioF. Al secuenciar estas dos variantes, se encontró que tenían la misma secuencia de DNA, y en consecuencia, son la misma variante (figura 31). Estas dos variantes se seleccionaron de manera independiente en dos cepas distintas: *E. coli* MC1061 Δ bioF y *E. coli* MC1061 Δ bioA/ Δ bioF. Las demás variantes no pudieron

complementar a la cepa *E. coli* MC1061 Δ *bioA*/ Δ *bioF* mediante la prueba de complementación por gota, por lo que fueron descartadas para futuros análisis.

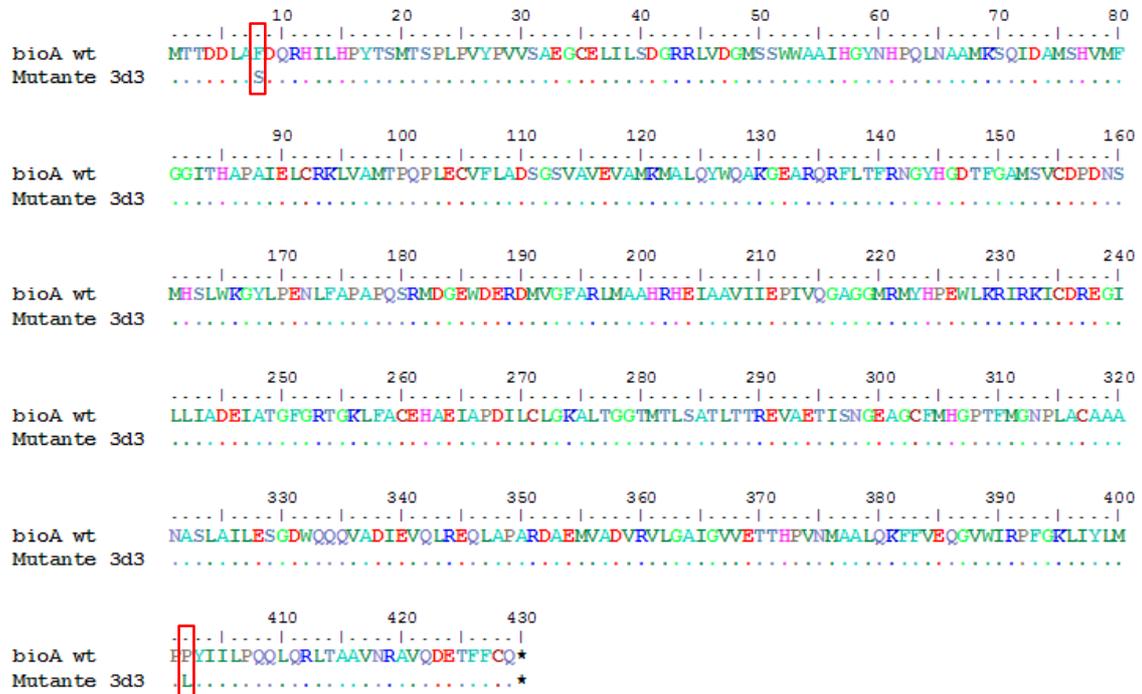


Figura 31. Secuencia de la mutante 3d3 y de BioA wt. Los cambios observados en la mutante 3d3 respecto a BioA wt están enmarcados en rojo.

No es claro cómo las mutaciones que se encuentran alejadas del sitio activo puedan afectar el reconocimiento de los sustratos. En particular la mutación P402S está ubicada en una hélice alfa que, mediante estudios cristalográficos se ha determinado que interacciona con la otra cadena (figuras 32-35). Esto podría ser importante, puesto que el sitio activo se ubica en la interfaz de los monómeros. De tener una actividad real BioF-BioA, pensamos que se debe a la modificación de esta interfaz. Tal actividad debe ser probada mediante ensayos enzimáticos *in vitro*, por la razón de que esta variante presenta problemas de reproducibilidad *in vivo*.

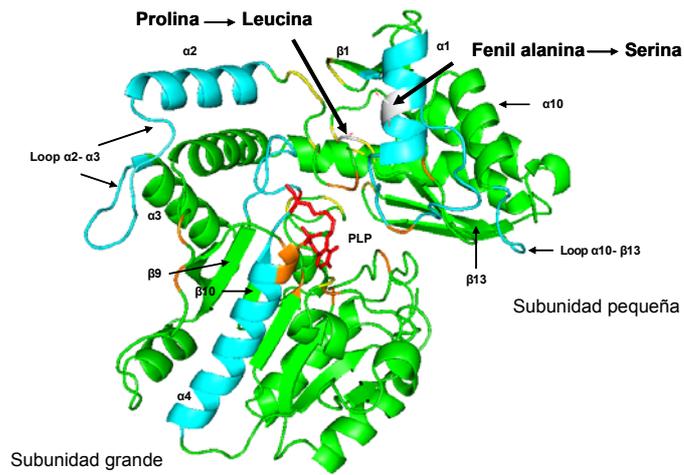


Figura 34. Estructura tridimensional del monómero BioA. En color cian se muestran las partes del monómero que interaccionan con el otro monómero para formar el dímero. En color Naranja se muestran los residuos involucrados en el reconocimiento de los sustratos. En amarillo están marcados aquellos residuos que interaccionan entre ambas subunidades.

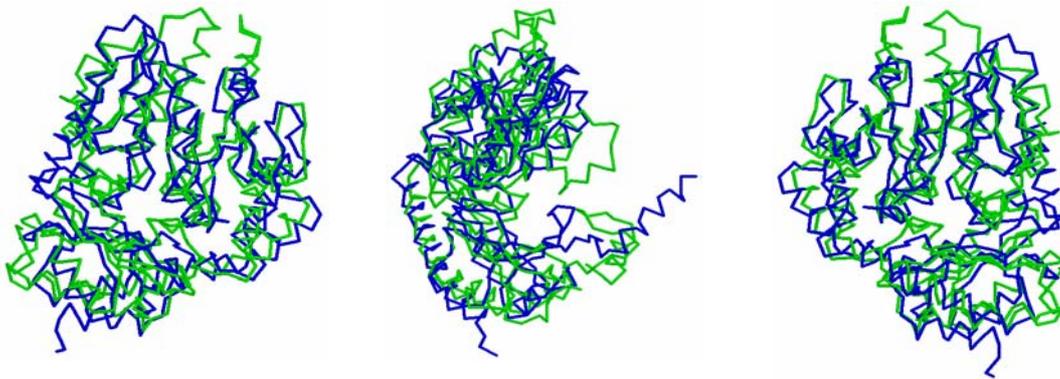


Figura 35. Solapamiento estructural entre los monómeros de las enzimas. Verde: Enzima 7,8-diamino pelargónico sintasa (BioA). Azul: Enzima oxanonanoato sintasa (BioF).

Segunda ronda mutagénica

Continuamos con la generación de diversidad usando como templado la variante ganadora 3d3 obtenida en la primera ronda mutagénica. Aquí se construyó una librería mutagénica estimada en 2×10^6 de variantes

Análisis de la librería de variantes, segunda ronda mutagénica

Número de clonas analizadas	9
Número total de cambios	70
Espacio explorado	11167 nucleótidos
Tasa mutagénica	0.62685% (≈ 8.1 cambios de nucleótidos por secuencia)
Transversiones	21 (30%)
Transiciones	49 (70%)

Mutantes	Número de mutaciones										
	1	2	3	4	5	6	7	8	9	10	11
3d3 1	T118C	C465T	T478C	C535A	G618T	T639A*	A687G	G796A			
3d3 2	A170T*	T175C	A302T*	T410C	A685G	G867A	T875C	G954A	A960T*	G1203T	A1253G
3d3 3	C535A	T811C									
3d3 4	A91G	A803T*	A879G	A937G	A961G	C1104T	G1158A	T1177C	T1187A*	T1215A*	
3d3 5	A170G	T266C	G511C	C535G	A745T*	A879T*	A967G	T1068A*	T1136C	A1183T*	
3d3 6	T86A*	T110A*	A202G	G349A	G662A	C732T	T778C	T809A*	A879T*	A976T*	G1057A
3d3 7	A351G	A881G	T978C	A1034G	T1127A*	A1168G					
3d3 8	A58T*	A205G	T240C	G306A	T341C	T502A*	A503C	T703C	T1151C		
3d3 9	C458T	A1023T*	T1053A*								

Tabla 6. Se muestran todos los cambios generados mediante PCR mutagénica en las 9 clonas que se secuenciaron. En negritas se muestran todas las transversiones. Los restantes son transiciones.

X	Px	Lx	Vx	Cx	Cx/Vx (%)
0	0.00030	6.1×10^2	1.0×10^0	1.0×10^0	100
1	0.0025	4.9×10^3	3.9×10^3	2.8×10^3	72
2	0.010	2.0×10^4	7.5×10^6	2.0×10^4	2.7×10^{-1}
3	0.027	5.4×10^4	9.6×10^9	5.4×10^4	5.6×10^{-4}
4	0.054	1.1×10^5	9.3×10^{12}	1.1×10^5	1.2×10^{-6}
5	0.088	1.8×10^5	7.2×10^{15}	1.8×10^5	2.5×10^{-9}
6	0.12	2.4×10^5	4.6×10^{18}	2.4×10^5	5.2×10^{-12}
7	0.14	2.8×10^5	2.5×10^{21}	2.8×10^5	1.1×10^{-14}
8	0.14	2.8×10^5	1.2×10^{24}	2.8×10^5	2.3×10^{-17}
9	0.13	2.5×10^5	5.2×10^{27}	2.5×10^5	4.8×10^{-20}
10	0.10	2.0×10^5	2.0×10^{29}	2.0×10^5	1.0×10^{-22}
11	0.075	1.5×10^5	7.0×10^{31}	1.5×10^5	2.1×10^{-25}
12	0.051	1.0×10^5	2.2×10^{34}	1.0×10^5	4.5×10^{-28}
13	0.031	6.3×10^4	6.6×10^{36}	6.3×10^4	9.5×10^{-31}
14	0.018	3.6×10^4	1.8×10^{39}	3.6×10^4	2.0×10^{-33}
15	0.0098	2.0×10^4	4.6×10^{41}	2.0×10^4	4.3×10^{-36}
16	0.0050	1.0×10^4	1.1×10^{44}	1.0×10^4	9.0×10^{-39}
17	0.0024	4.7×10^3	2.5×10^{47}	4.7×10^3	1.9×10^{-41}
18	0.0011	2.1×10^3	5.3×10^{48}	2.1×10^3	4.1×10^{-44}
19	0.00046	9.1×10^2	1.1×10^{51}	9.1×10^2	8.6×10^{-47}
20	0.00018	3.7×10^2	2.0×10^{53}	3.7×10^2	1.8×10^{-49}

Número esperado de secuencias distintas: 2.0×10^6 . Para una librería de 2×10^6

x	Número exacto de mutaciones por secuencia
Px	Probabilidad de Poisson de X mutaciones, dada una tasa mutagénica
Lx	Número esperado de secuencias en la librería con exactamente X mutaciones
Vx	Número de posibles variantes con exactamente X mutaciones
Cx	Número esperado (aprox) de distintas variantes en la librería con exactamente X mutaciones

Tabla 7. Datos generados a partir de PEDEL durante la segunda ronda mutagénica.

Selección de variantes

Una vez que obtuvimos la 2ª librería estimada en 2×10^6 comenzamos la selección de variantes con la metodología anteriormente descrita. Debido al tamaño de esta librería, el número de variantes exploradas fue de alrededor de 12 millones (por la dificultad de

explorar 10 veces el tamaño estimado de la librería) y explorada en la cepa mutante de *E. coli* MC1061 Δ *bioA*/ Δ *bioF*. Es importante considerar que el crecimiento de fondo de esta segunda ronda mutagénica fue mucho más intenso en comparación con la primera ronda, dificultando de esta manera la selección de variantes con las propiedades catalíticas deseadas. En consecuencia, se obtuvo una enorme cantidad de clonas candidatas que tuvieron que ser verificados sus fenotipos de modo iterativo al picar las colonias candidatas para ser cultivadas en medio LB, posteriormente su lavado y siembra en medio mínimo. Todo esto con el fin de seleccionar variantes que sostuvieran el fenotipo deseado. Se encontraron varias clonas que mantenían un fenotipo bastante robusto después de varias rondas de selección, pero al purificar sus respectivos plásmidos y digerir éstos con las enzimas de restricción EcoRI-HindIII, presentaron un patrón de restricción extraño (figura 36). Es importante considerar que el banco se amplificó en la cepa de *E. coli* MC1061/pREP4 (con resistencia a Kanamicina), pero esta cepa ya había sido usada anteriormente sin ningún problema en el laboratorio para amplificar bancos mutagénicos de otros proyectos.

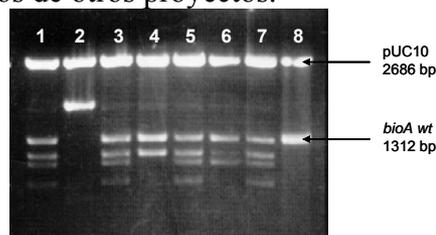


Figura 36. Corrimiento en gel de agarosa al 1.2%. Fragmentos de tamaño conocido: pREP 3740 bp, pUC18 2686 bp, bioA 1312 bp. Los demás fragmentos son de identidad desconocida.

Puesto que las clonas seleccionadas podían ser cultivadas en medio mínimo, pensamos que pese a que estas clonas mostraban un patrón de restricción no esperado cuando estas fueron digeridas con enzimas de restricción, el fenotipo era producido por una variante de BioA. Por lo tanto, se hicieron muchos intentos por aislar la supuesta variante proteica. Usamos PCR con oligos específicos para BioA para amplificar variantes de esta proteína, pero no hubo extensión en muchas de las clonas. En los casos en que se pudo lograr la amplificación, los plásmidos se purificaron, se digirieron, se clonaron y se transformaron en la cepa *E. coli* MC1061 Δ *bioA*/ Δ *bioF* para confirmar el fenotipo observado, sin embargo, estas transformantes no pudieron sostener el fenotipo. También se intentaron aislar variantes proteicas al efectuar una digestión con las enzimas de restricción EcoRI-HindIII y posterior purificación a partir de gel para clonar y transformar en la misma cepa. Aún así tampoco pudimos reproducir el fenotipo deseado.

Debido a que la exploración del banco en medio sólido es una limitante cuando éstos son grandes y en especial el sistema de selección basado en la auxotrofia a biotina que particularmente presenta un ruido de fondo considerablemente alto (por lo que no se

puede sembrar mas de 50 000 colonias por caja). Por estas razones, comenzamos a explorar la librería en medio líquido.

El ensayó partió con matraces con 20, 15, 10, 5 y 0 pM de biotina, respectivamente. El medio mínimo líquido es el mismo que se usó en el mismo medio en sólido, únicamente la diferencia es la sustitución de agar por agua desionizada.

Se inoculó una muestra de 20µl del banco en matraces erlenmeyer con 50 ml de medio mínimo, los cuales fueron previamente tratados con mezcla crómica para eliminar cualquier traza de biotina. El cultivo se incubó a 37°C, 200 rpm, hasta observar crecimiento.

A continuación realizamos el ensayo al efectuar transferencias con 10µl del cultivo de acuerdo a la tabla 8.

Concentración inicial de biotina (pM)	Transferencias				
	1era	2da	3era	4a	5a
20	Crecimiento	Crecimiento	Contaminado	Contaminado	Contaminado
15	Crecimiento	Crecimiento	Contaminado	Contaminado	Contaminado
10	Crecimiento	Crecimiento	Crecimiento	Crecimiento	Crecimiento
5	Crecimiento	Crecimiento	Crecimiento	Crecimiento	Crecimiento
0	No se observó crecimiento				

Tabla 8. Selección positiva en medio líquido.

De modo general, se observó crecimiento a las 24 horas en el primer ensayo, y se incrementó el tiempo de incubación en función de las transferencias efectuadas, hasta incubarse durante una semana la última transferencia. Después de 5 transferencias se tomó un alícuota del cultivo y se sembró en cajas de LB amp y medio mínimo + biotina (figuras 37 y 38).



Figura 37. Izquierda: Caja de medio LB amp con 100µl del cultivo de la 5ª transferencia con 10pM de biotina. Derecha: Caja con medio mínimo + 10 pM de biotina con 100µl del cultivo de la 5ª transferencia con 10pM de biotina.



Figura 38. Izquierda: Caja de medio LB amp con 100 μ l del cultivo de la 5^a transferencia que contenía 5pM de biotina. Derecha: Caja con medio mínimo + 5 pM de biotina con 100 μ l del cultivo de la 5^a transferencia que contenía 5pM de biotina.

Una vez que obtuvimos colonias candidatas, verificamos el fenotipo de estas colonias con respecto a un control negativo, mediante la prueba de complementación por gota para poder detectar pequeñas diferencias en la intensidad de sus respectivos fenotipos (figura 39).

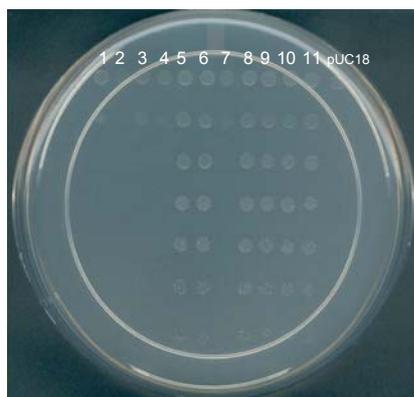


Figura 39. Prueba de complementación por gota en medio mínimo sin biotina y con ampicilina. Carriles [1,4] colonias obtenidas del cultivo con 10pM de biotina en la 5^a transferencia. Etiquetadas de la siguiente forma: 10₁, 10₂, 10₃, 10₄. Carriles [5,11] colonias obtenidas del cultivo con 5pM de biotina en la 5^a transferencia. Etiquetadas de la siguiente forma: 5₁, 5₂, 5₃, 5₄, 5₅, 5₆, 5₇. Como control negativo: pUC18.

Con esta prueba de complementación por gota demostramos que estas clonas podían crecer independientemente de biotina, por lo que se verificó los patrones de restricción de éstas.

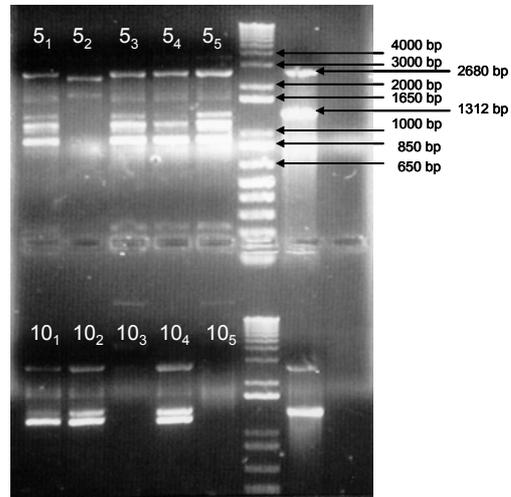


Figura 40. Corrimiento en gel de agarosa al 1.2%. Carriles superiores: Patrón de restricción (EcoRI-HindIII) de clonas obtenidas en cajas de medio mínimo + 5pM de biotina. Carriles inferiores: Patrón de restricción (EcoRI-HindIII) de clonas obtenidas en cajas de medio mínimo + 10pM de biotina. En los carriles superiores e inferiores los fragmentos de tamaño conocido son: pREP 3740 bp, pUC18 2680 bp, *bioA* 1312 bp. Los demás fragmentos son de identidad desconocida.

Con base en este gel (figura 40), encontramos que el patrón de restricción de plásmidos aislados de las clonas candidatas mostraban un patrón similar al mostrado en la figura 36.

Quisimos discernir el fragmento responsable de la complementación de las actividades BioF-BioA, para ello, transformamos la cepa *E. coli* MC1061 Δ *bioA*/ Δ *bioF* con el plásmido purificado de estas clonas. Transformamos entonces y seleccionamos en medio LB con ampicilina, kanamicina y ampicilina + kanamicina (figura 41).



Figura 41. Al transformar las mutantes 5₁, 5₂, 5₃, 5₄, 5₅, 5₆, 5₇ con los y seleccionando en cajas de LB con antibióticos. Izquierda: Medio LB + Ampicilina. Centro: Medio LB + Kanamicina. Derecha: Medio LB + Ampicilina/Kanamicina. Todas las cajas fueron sembradas con 20 μ l de células recuperadas en medio SOC.

A continuación observamos los fenotipos de cada clona después de la segregación con la prueba de complementación por gota (figuras 42-44).

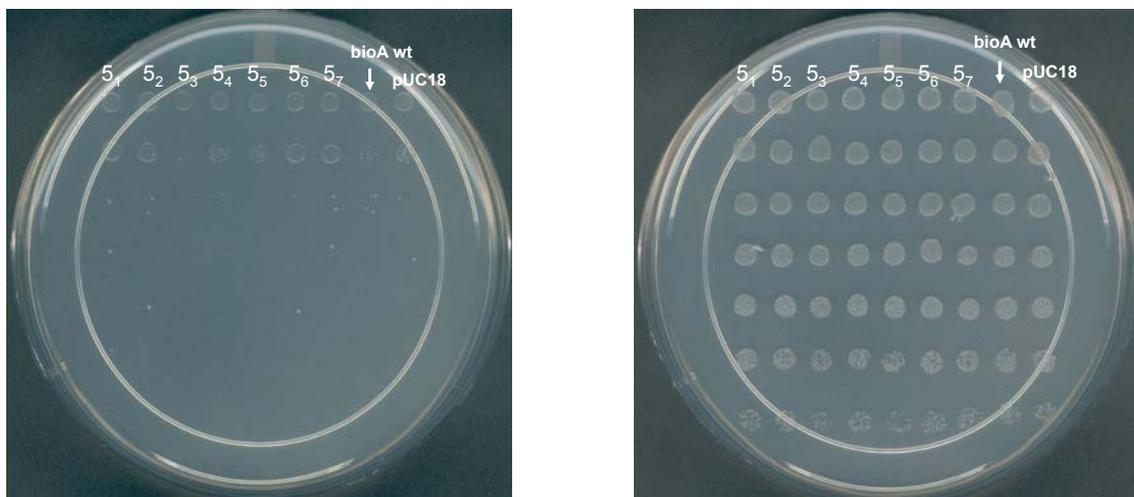


Figura 42. Prueba por complementación por gota. Izquierda: Complementación en medio mínimo con Ampicilina. Como controles negativos se usó BioA wt y pUC18. Derecha: Caja con medio mínimo Ampicilina + biotina.

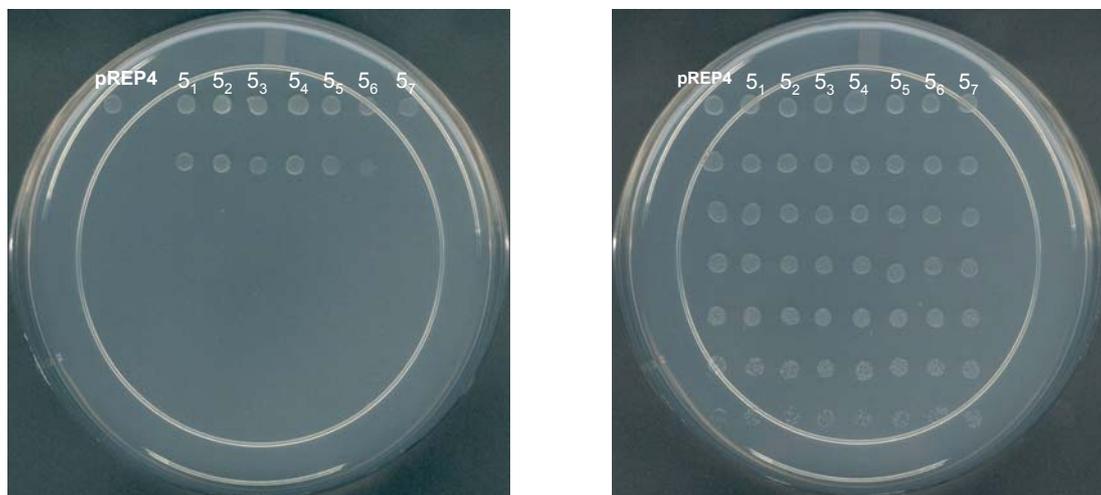


Figura 43. Prueba por complementación por gota. Izquierda: Complementación en medio mínimo con Kanamicina. Como control negativo se usó pREP4. Derecha: Caja con medio mínimo Kanamicina + biotina.

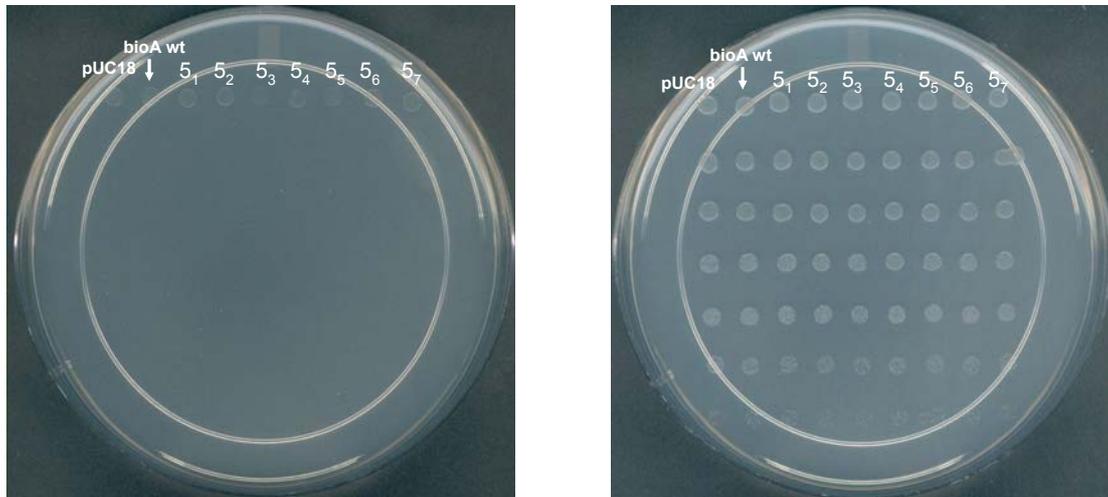


Figura 44. Prueba por complementación por gota. Izquierda: Complementación en medio mínimo con Ampicilina/Kanamicina. Como control negativo se usó pREP4. Derecha: Caja con medio mínimo Ampicilina/Kanamicina + biotina.

Como se puede observar en estas figuras, no se pudo segregar el fragmento responsable de la complementación de la actividad BioF-BioA.

Es posible que el fenotipo exhibido por esta cepa no esté asociado a una variante de BioA, puesto que el patrón de restricción sugiere que de algún modo, el plásmido se recombinó con el cromosoma y que posiblemente haya tomado un gen que al expresarse en alto número de copias, pueda complementar la actividad BioF-BioA. Con base a la experiencia en el grupo de investigación, sabemos que esto es posible. Anteriormente, en el laboratorio se había usado la auxtrofia a tiamina como método de alta sensibilidad para seleccionar eficientemente variantes con actividades muy pequeñas. Para este fin, se había obtenido una librería de variantes de una proteína con estructura $\beta_8\alpha_8$ con actividad isomerasa (Triosa fosfato isomerasa), para seleccionar al menos una variante que adquiriera de novo la actividad sintasa. Tal actividad sintasa la desempeña la enzima ThiE (tiamina sintasa, estructura $\beta_8\alpha_8$), pero al ser removido del cromosoma de *E.coli*, permitió establecer un buen sistema de selección. De modo casual, un gen hipotético se clonó dentro de la colección de variantes de la enzima con actividad sintasa. Este gen hipotético fue capaz de complementar la función de la enzima ThiE en tres días de incubación a 37°C. El gen hipotético se identificó como yjbQ y no tiene relación en secuencia con cualquier otra proteína en *E. coli*.

Evidentemente no es fácil explicar cómo una proteína de función desconocida, pueda

contener una actividad, aunque pequeña, idéntica a otra que muestra una actividad compleja. Si bien la naturaleza ha especializado a las enzimas para una función en particular, parece que muchas de estas aún conservan vestigios de otras actividades relacionadas. En caso de expresarse en alto número de copias, tal vez podrían complementar funciones enzimáticas relacionadas con ellas. Esto bien podría estar ocurriendo en el caso de las aminotransferasas en *E. coli* como BioF y BioA.

Por otro lado, una constante en el sistema de selección basado hacia la auxotrofia a biotina, es la inconsistencia fenotípica de las variantes seleccionadas. Con base en la experiencia obtenida en otros proyectos que se realizaron simultáneamente a éste, hemos establecido la falta de reproducibilidad bajo este sistema. El fenotipo de la mutante 3d3 ha sido elusivo en intentos posteriores de complementación por gota, aunque fue mucho más reproducible cuando la cepa que contenía la variante 3d3 fue estriada con asa (dato no mostrado). Aunque esta forma para evaluar la complementación respecto a los controles negativos no es la más indicada, por la razón de que no se puede controlar el número de UFC que se está estriando.

Análisis global de la evolución de enzimas que unen PLP como cofactor

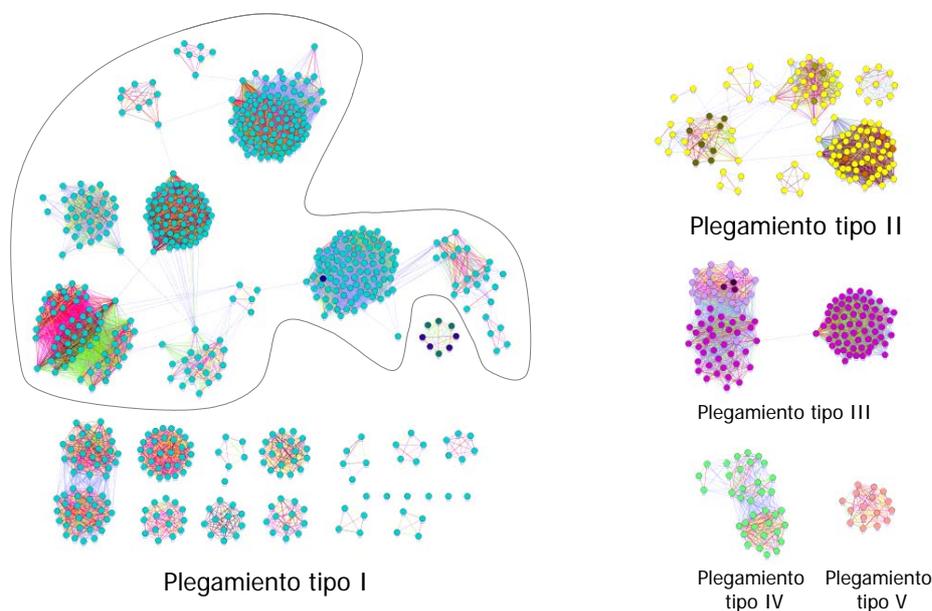


Figura 45. Perspectiva global de dominios que usan PLP como cofactor. El conjunto de enzimas que está encuadrado dentro de la línea, son aquellas que están conectadas por al menos un elemento de la red.

Mostramos una perspectiva global de las enzimas que unen PLP, cuyos nodos son enzimas dependientes de PLP y las conexiones contienen como atributo el valor de la identidad dentro del intervalo [25%-70%]. Podemos observar que las enzimas tienden a agruparse de tal modo que miembros de enzimas dentro de cada grupo están altamente conectadas entre ellas, y son pocas las conexiones que exhiben hacia el exterior. Cada grupo fue definido y nombrado de modo arbitrario, respetando la agrupación que se puede ver en la figura 41.

Plegamiento tipo I

En la figura 45 podemos ver que el plegamiento más abundante es el del tipo I. Es posible que esta superfamilia sea la primera en surgir durante la evolución, puesto que además de ser abundante, se encuentra en los tres dominios y es funcionalmente más diverso (Christen *et al.*, 2001).

En la figura 46 se muestran los grupos 2b, 2e y 19. El grupo 2b es pequeño y homogéneo. La única actividad contenida es la L-quinurenina hidrolasa (enzimas anotadas como KYNU, EC 3.7.1.3). Por otro lado, el grupo 2e contiene 10 actividades enzimáticas distintas, en el que la actividad acetilornitina aminotransferasa (ARGD, EC

2.6.1.11) es por mucho la más abundante, seguida de la actividad ácido-7,8-diaminopelargónico sintasa (enzimas anotadas como BIOA, EC 2.6.1.62). Las otras actividades son relacionadas: LAT, EC 2.6.1.36; AGT2, EC 2.6.1.44; AGT2, EC 2.6.1.40; RHBA, DAT y ECTB EC 2.6.1.76; OAT, 2.6.1.13; TPA, EC 2.6.1.77; OAPT, EC 2.6.1.18. La única actividad completamente distinta es la 2,2-dialquilglicina descarboxilasa, representada por una sola enzima (DGDA_BURCE, EC 4.1.1.64).

Es posible que debido a la abundancia en este grupo, la actividad acetilornitina aminotransferasa, haya sido la actividad que dio origen al resto de las actividades de este conjunto de proteínas.

Otro grupo representado en esta figura es el 19, el cual consiste de 3 actividades enzimáticas: TYDC3, EC 4.1.1.25; DDC, EC 4.1.1.28; y DCHS, EC 4.1.1.22. El grupo 2b se conecta con el grupo 2e mediante la conexión de KYNU_YEAST (*Saccharomyces cerevisiae*) y ARGD_PYRFU (*Pyrococcus furiosus*), El grupo 19 establece una conexión con dos grupos mediante DDC_BOVIN (*Bos taurus*) - ARGD_LISN con el grupo 2e y DDC_BOVIN - GLYA_CHMUL (*Chlamydia muridarum*) del grupo 2d.

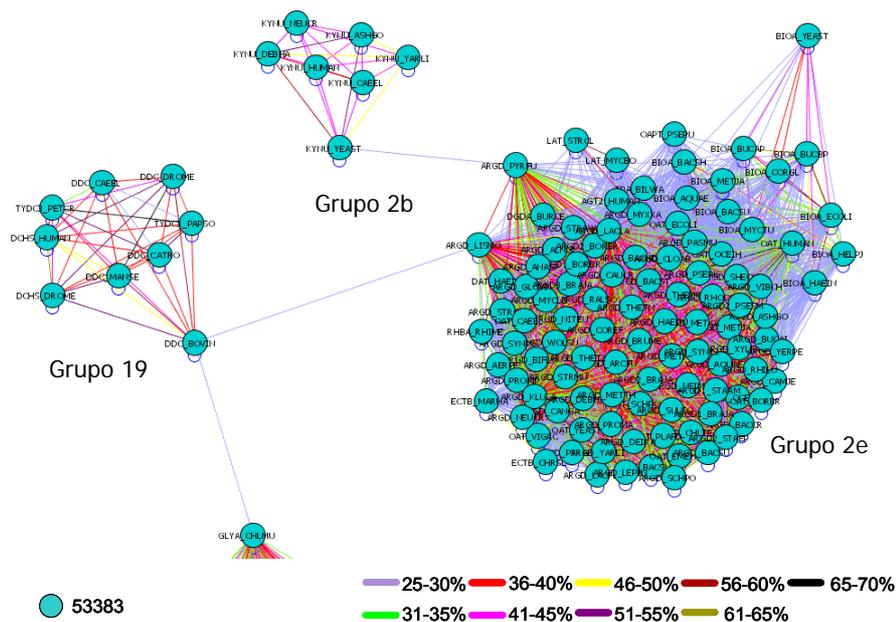


Figura 46. Para todas las enzimas clasificadas dentro del plegamiento de tipo I, el dominio estructural está numerado como 53383 (PLP-dependente transferasa, según SCOP). Las identidades entre cada dominio fueron marcadas como conexiones y la intensidad en la escala de grises representa el grado de ésta.

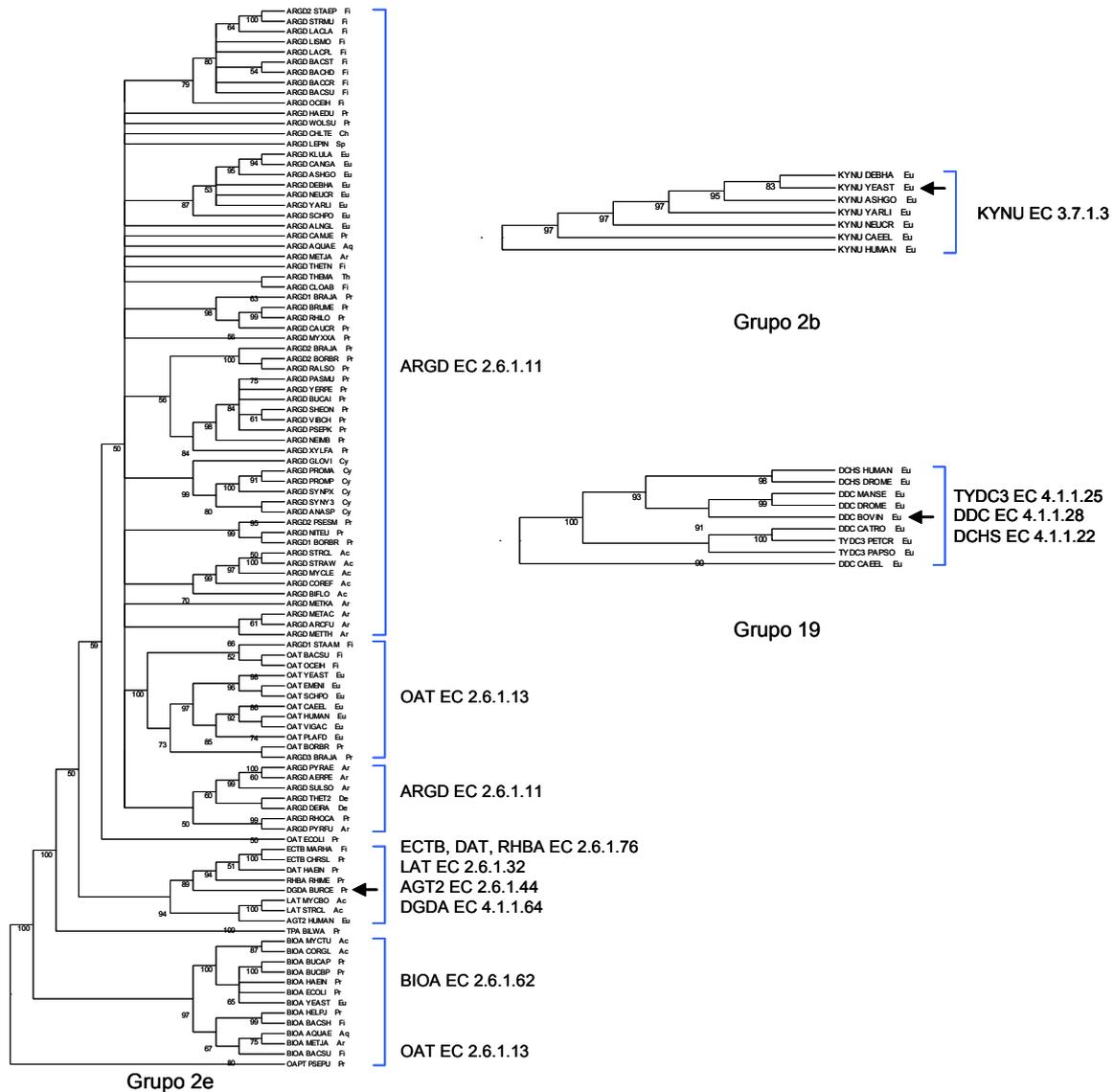


Figura 47. Árboles filogenéticos sin raíz, generados a partir del principio de máxima parsimonia. Pl, Planctomycetes; Ch, Chlorobi; Pr, Proteobacteria; De, Deinococcus; Cy, Cyanobacteria; Ac, Actinobacteria; Fi, Firmicutes; Ba, Bacteroides; Eu, Eukaryota; Sp, Spirochaetes; Ar, Arquea; Chl, Chloroflexi.

En la figura 47 se muestran tres árboles construidos bajo el principio de máxima parsimonia. El árbol que pertenece al grupo 2e tiende a mostrar, en términos generales, que las enzimas suelen agruparse en ramas que reflejan la misma actividad enzimática. Con la excepción de una de las ramas, en el cuál se tiene representados cuatro actividades enzimáticas distintas, especialmente, aquella denotada como DGDA, puesto que es la única enzima de este grupo que exhibe una actividad no relacionada al resto de

las enzimas con las que se agrupa. Curiosamente, DGDA contiene un semejante papel catalítico respecto a todas las enzimas encontradas para el grupo 19, sin embargo, esto no se refleja en una identidad de secuencia mayor al 25%. Las enzimas en el grupo 2b son heterogéneas en cuanto a función.

El grupo 2d es una familia de enzimas homogéneas en cuanto a su actividad enzimática (figura 48). Las enzimas que pertenecen a esta familia están anotadas como GLYA y tienen un EC 2.1.2.1; 13 de ellas, conectan con METC_RHILV (*Rhizobium leguminosarum*, EC 4.4.1.8) del grupo 2c, como es mostrado en la tabla 9.

Grupo 2d	Organismo	Dominio	Grupo 2c	Organismo	Dominio
GLYA_GLOVI	<i>Gloeobacter violaceus</i>	Bacteria; Cyanobacteria	METC_RHILV	<i>Rhizobium leguminosarum</i>	Bacteria; Proteobacteria
GLYA_HAEIN	<i>Haemophilus influenzae</i>	Bacteria; Proteobacteria			
GLYA_RICCN	<i>Rickettsia conorii</i>	Bacteria; Proteobacteria			
GLYA_SULSO	<i>Sulfolobus solfataricus</i>	Archaea; Crenarchaeota			
GLYA_BORBR	<i>Bordetella bronchiseptica</i>	Bacteria; Proteobacteria			
GLYA2_PSEAE	<i>Pseudomonas aeruginosa</i>	Bacteria; Proteobacteria			
GLYA_HELHP	<i>Helicobacter hepaticus</i>	Bacteria; Proteobacteria			
GLYA_CAUCR	<i>Caulobacter crescentus</i>	Bacteria; Proteobacteria			
GLYA_SALTI	<i>Salmonella typhi</i>	Bacteria; Proteobacteria			
GLYA_AERPE	<i>Aeropyrum pernix</i>	Archaea; Crenarchaeota			
GLYA_PROMM	<i>Prochlorococcus marinus</i>	Bacteria; Cyanobacteria			
GLYA_ENTFA	<i>Enterococcus faecalis</i>	Bacteria; Firmicutes			
GLYA_LEPIN	<i>Leptospira interrogans</i>	Bacteria; Spirochaetes			

Tabla 9. Conexiones mediante pares de proteínas entre los grupos 2d y el grupo 2c. Cada par tiene al menos un 25% de identidad.

Por otro lado, el grupo 2d, también conecta con el grupo 2g mediante la conexión entre dos enzimas (Bacteria-Bacteria) GLYA_GLOVI (*Gloeobacter violaceus*) - GCSP_BACTN (*Bacteroides thetaiotaomicron*, EC 1.4.4.2). En el grupo 2f encontramos 4 actividades enzimáticas, pero las dos más abundantes son las actividades 8-amino-7-oxononanoato sintasa (BIOF, EC 2.3.1.47) y 5-aminolevulinato sintasa (HEM, EC 2.3.1.37). El resto de las actividades enzimáticas son KBL, EC 2.3.1.29 y LCB, EC 2.3.1.50. La única conexión que une al grupo 2f con el resto de la red es por

medio de BIOF_SERMA (*Serratia marcescens*, EC 2.3.1.47), conectando a una enzima del grupo 2c METC_RHILV EC

4.4.1.8, la unión de estos grupos por medio de este par de enzimas que efectúan reacciones tan diferentes dan idea de la plasticidad mostrada por esta superfamilia.

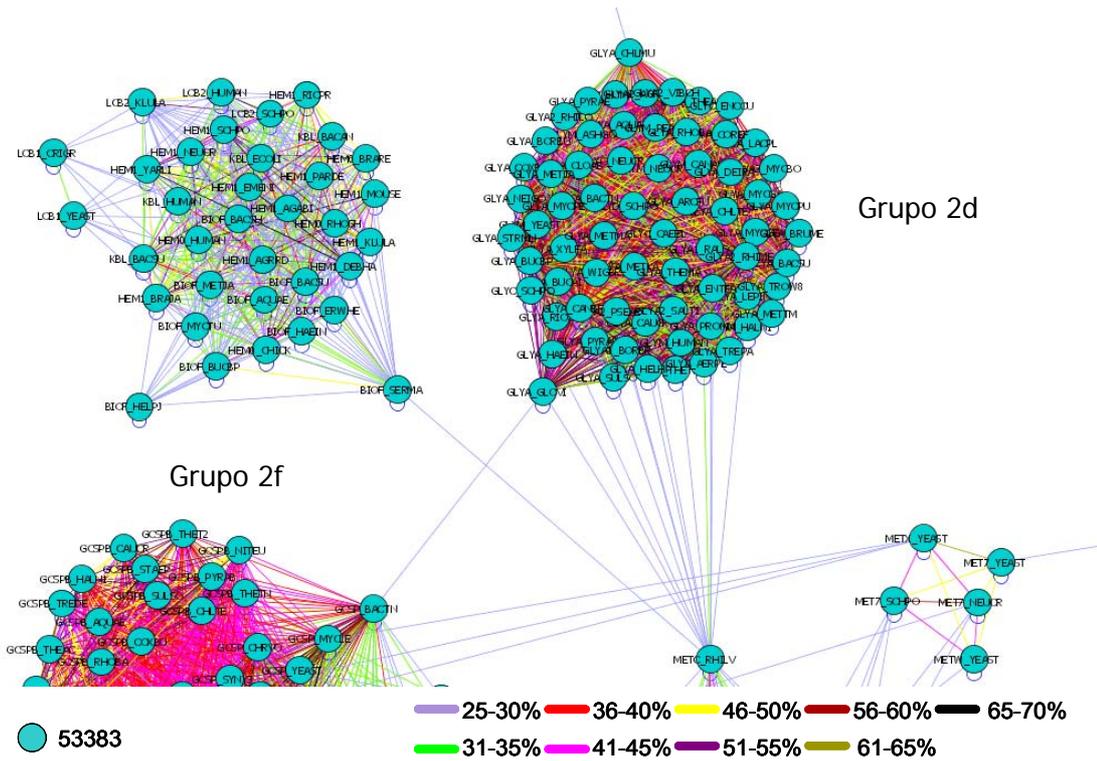


Figura 48. Para todas las enzimas clasificadas dentro del plegamiento de tipo I, el dominio estructural está numerado como 53383 (PLP-depend transferase, según SCOP). Las identidades entre cada dominio fueron marcadas como conexiones y el color indicado representa el grado de ésta.

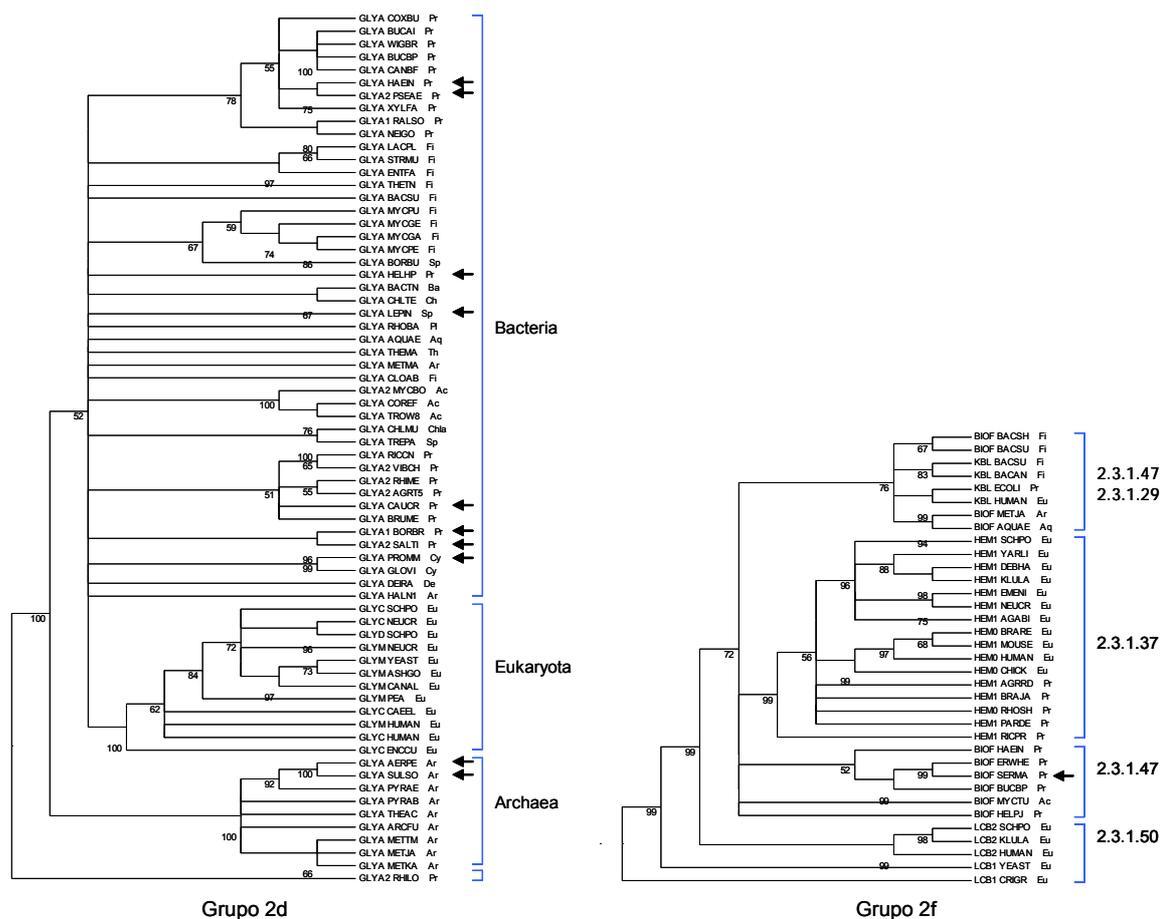


Figura 49. Árboles filogenéticos generados a partir del principio de máxima parsimonia. Pl, Planctomycetes; Ch, Chlorobi; Pr, Proteobacteria; De, Deinococcus; Cy, Cianobacteria; Ac, Actinobacteria; Fi, Firmicutes; Ba, Bacteroides; Eu, Eukaryota; Sp, Spirochaetes; Ar, Arquea; Chl, Chloroflexi. Las enzimas involucradas en la conexión con otros grupos están indicadas con flechas negras.

Los dos árboles producidos (figura 49), muestran ramas que separan enzimas de acuerdo al dominio biológico al que pertenecen (grupo 2d) y a la actividad (grupo 2f). El grupo 2d tiene muchas ramas colapsadas, por la razón de que estas secuencias son poco informativas.

El grupo 2g contiene exclusivamente la actividad glicina descarboxilasa (EC 1.4.4.2). Este es el único caso de asimetría reportado para enzimas que dependen de PLP en el plegamiento de tipo I, puesto que a pesar de ser activa en forma dimérica está constituida por el mismo dominio estructural, pero cada dominio es divergente. En algunos microorganismos la enzima glicina descarboxilasa se encuentra fusionada, y en otros, las subunidades amino y carboxilo terminal se encuentran como genes aparte. La subunidad amino terminal es anotada de forma genérica como GCSPA, la parte carboxilo terminal es anotada como GCSPB; cuando ambos genes se encuentran fusionados, se denota como GCSP.

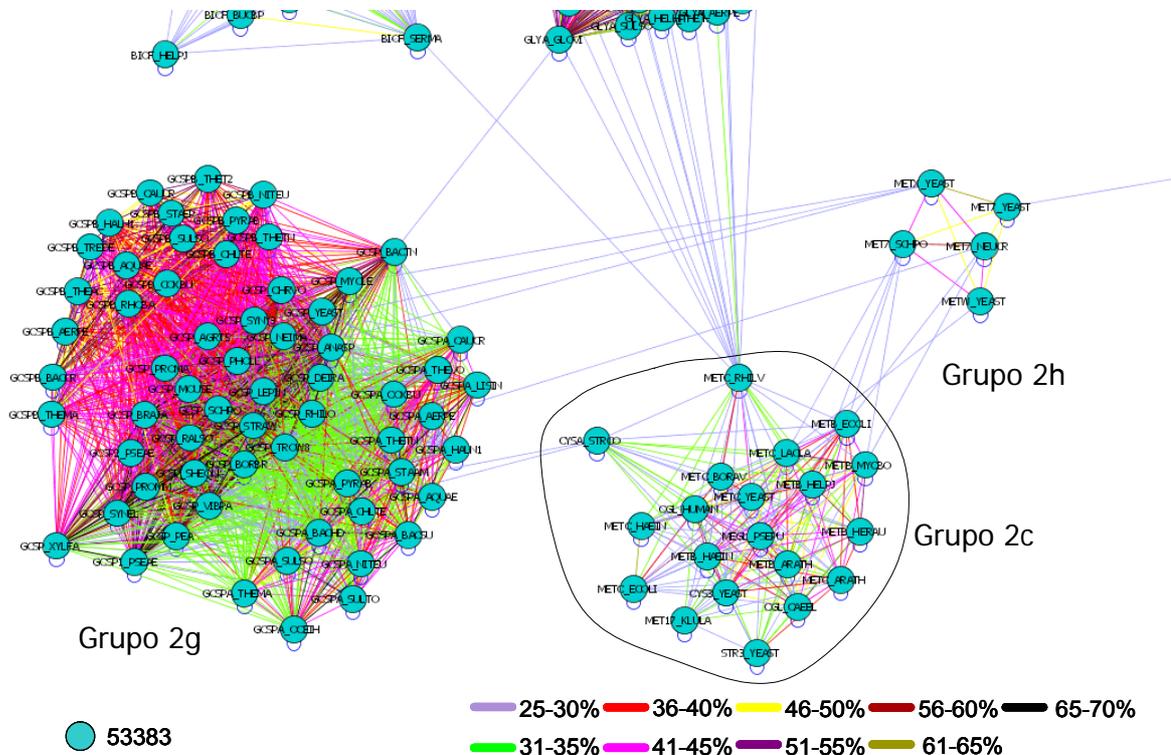


Figura 50. Para todas las enzimas clasificadas dentro del plegamiento de tipo I, el dominio estructural está numerado como 53383 (PLP-dependente transferase, según SCOP). Las identidades entre cada dominio fueron marcadas como conexiones y el color indicado representa el grado de ésta.

La identidad entre las GCSPB-GCSP (la subunidad “B” de éstas) es mayor que entre la identidad mostrada en GCSPA-GCSP (la subunidad “A”).

El grupo 2g presenta conexiones con tres grupos: Grupo 2d, grupo 2h y el grupo 2c. La conexión con el grupo 2d es mediante GCSP_BACTN (*Bacteroides thetaiotaomicron*) – GLYA GLOVI (*Gloeobacter violaceus*). Las conexiones con los grupos 2h y 2c, se resumen en las tablas 10 y 11.

Grupo 2g	Organismo	Dominio	Grupo 2h	Organismo	Dominio
GCSPB_AERPE	<i>Aeropyrum pernix</i>	Archaea; Crenarchaeota	METX_YEAST	<i>Saccharomyces cerevisiae</i>	Eukaryota; Fungi
GCSPB_BACCR	<i>Bacillus cereus</i>	Bacteria; Firmicutes	METX_YEAST	<i>Saccharomyces cerevisiae</i>	Eukaryota; Fungi
GCSPB_THEMA	<i>Thermotoga maritima</i>	Bacteria; Thermotogae	METX_YEAST	<i>Saccharomyces cerevisiae</i>	Eukaryota; Fungi
GCSP_SYNEL	<i>Synechococcus elongatus</i>	Bacteria; Cyanobacteria	MET7_NEUCR	<i>Neurospora crassa</i>	Eukaryota; Fungi

Tabla 10. Conexiones mediante pares de proteínas entre los grupos 2g y el grupo 2h. Cada par tiene al menos un 25% de identidad.

Grupo 2g	Organismo	Dominio	Grupo 2c	Organismo	Dominio
GCSP_XYLFA	<i>Xylella fastidiosa</i>	Bacteria; Proteobacteria	CYSA_STRCO	<i>Streptomyces coelicolor</i>	Bacteria; Actinobacteria
GCSP1_PSEAE	<i>Pseudomonas aeruginosa</i>	Bacteria; Proteobacteria	CYSA_STRCO	<i>Streptomyces coelicolor</i>	Bacteria; Actinobacteria

Tabla 11. Conexiones mediante pares de proteínas entre los grupos 2g y el grupo 2c. Cada par tiene al menos un 25% de identidad.

Los grupos 2c y 2h están formados por pocos elementos, pero el grupo 2c es sorprendentemente diverso, porque contiene 6 actividades enzimáticas distintas, incluyendo una enzima bifuncional, denotada como MET17_KLULA (*Kluyveromyces lactis*) que exhibe las actividades EC 2.5.1.47 y EC 2.5.1.49. Las demás actividades son: CGL y CYSA, EC 4.4.1.1; METB, EC 2.5.1.48; METC, STR3 y CBL, EC 4.4.1.8; MEGL EC 4.4.1.11. Por otro lado, el grupo 2h contiene la actividad EC 2.5.1.48, que como se puede ver, es la misma actividad que muestran algunas enzimas del grupo 2c. Además, el grupo 2h conecta con el grupo 2a con el par MET7_YEAST - HIS8 PESEST (*Pseudomonas stutzeri*)

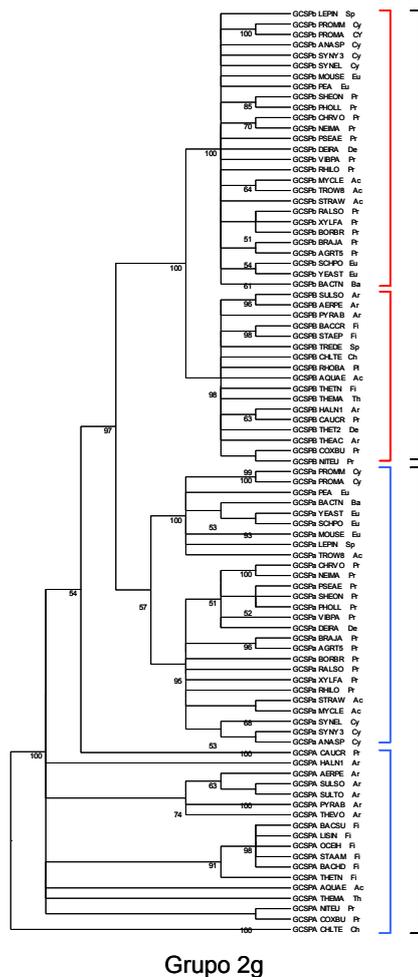
Las conexiones entre los grupos 2h y 2c, se resumen en la tabla 12.

Grupo 2h	Organismo	Dominio	Grupo 2c	Organismo	Dominio
MET7_SCHPO	<i>Schizosaccharo myces pombe</i>	Eukaryota; Fungi	METC_LACLA	<i>Lactococcus lactis</i>	Bacteria; Firmicutes
MET7_SCHPO	<i>Schizosaccharo myces pombe</i>	Eukaryota; Fungi	METB_HELPJ	<i>Helicobacter pylori</i>	Bacteria; Proteobacteria
MET7_SCHPO	<i>Schizosaccharo myces pombe</i>	Eukaryota; Fungi	METB_ECOLI	<i>Escherichia coli</i>	Bacteria; Proteobacteria
MET7_SCHPO	<i>Schizosaccharo myces pombe</i>	Eukaryota; Fungi	METB_MYCBO	<i>Mycobacterium bovis</i>	Bacteria; Actinobacteria
MET7_NEUCR	<i>Neurospora crassa</i>	Eukaryota; Fungi	METB_ECOLI	<i>Escherichia coli</i>	Bacteria; Proteobacteria
MET7_NEUCR	<i>Neurospora crassa</i>	Eukaryota; Fungi	METB_MYCBO	<i>Mycobacterium bovis</i>	Bacteria; Actinobacteria
METW_YEAST	<i>Neurospora crassa</i>	Eukaryota; Fungi	METC_LACLA	<i>Lactococcus lactis</i>	Bacteria; Firmicutes

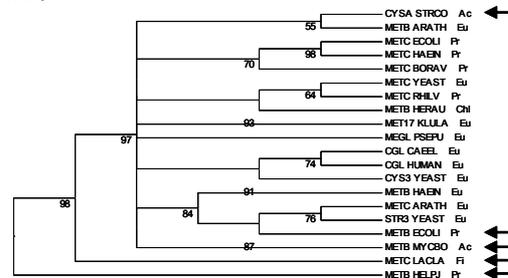
Tabla 12. Conexiones mediante pares de proteínas entre los grupos 2h y el grupo 2c. Cada par tiene al menos un 25% de identidad.

El árbol construido para el grupo 2g (figura 51), muestra la misma regularidad que la red para este mismo grupo (figura 50), en la que la subunidad carboxilo terminal (GCSPB), se encuentra más relacionada con la subunidad carboxilo terminal de GCSP, en organismos que mantienen ambas subunidades como genes aparte (GCSPA y GCSPB), en comparación con la subunidad amino terminal (GCSPA) versus subunidad amino terminal de GCSP.

Por último, los árboles de los grupos 2c y 2h no aportan suficiente información, puesto que varias ramas en el grupo 2c tienen valores bajos de bootstrap y el grupo 2h es pequeño.

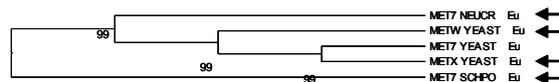


GCSPB
Subunidad carboxilo terminal



Grupo 2c

GCSPA
Subunidad amino terminal



Grupo 2h

Figura 51. Árboles filogenéticos generados a partir del principio de parsimonia. En el grupo 2g, la fusión GCSP fue dividida en subunidades y denotadas como GCSPa y GCSPb, para las subunidades amino y carboxilo terminal, respectivamente. Pl, Planctomycetes; Ch, Chlorobi; Pr, Proteobacteria; De, Deinococcus; Cy, Cyanobacteria; Ac, Actinobacteria; Fi, Firmicutes; Ba, Bacteroides; Eu, Eukaryota; Sp, Spirochaetes; Ar, Arquea; Chl, Chloroflexi.

Por otro lado, el grupo 2a (figura 52) es básicamente homogéneo, puesto que todo el grupo presenta la actividad histidinol fosfato aminotransferasa (HIS8, EC 2.6.1.9), exceptuando tres elementos de este conjunto que además de presentar la actividad Treonina fosfato descarboxilasa (COBC_PSEDE, COBD_SALTI y COBD THETN, los tres con un EC 4.1.1.81), es el elemento más aislado de todo este conjunto de enzimas. El grupo 6 muestra dos actividades enzimáticas, Aspartato aminotransferasa (AAT, EC 2.6.1.1) y Alanina aminotransferasa (ALAT, EC 2.6.1.2). La enzima HIS8_PHOLL (*Photorhabdus luminescens*) del grupo 2a, es la única que presenta una fusión. Las conexiones entre los grupos 2a y 6 son mostradas en la tabla 13.

Grupo 2a	Organismo	Dominio	Grupo 6	Organismo	Dominio
HIS82_RALSO	<i>Pseudomonas solanacearum</i>	Bacteria; Proteobacteria	AAT_THET8	<i>Thermus thermophilus</i>	Bacteria; Deinococcus- Thermus
HIS81_ARCFU	<i>Archaeoglobus fulgidus</i>	Archaea; Euryarchaeota	AAT1_METJA	<i>Methanococcus jannaschii</i>	Archaea; Euryarchaeota
HIS8_HALVO	<i>Halobacterium volcanii</i>	Archaea; Euryarchaeota	AAT_THET8	<i>Thermus thermophilus</i>	Bacteria; Deinococcus- Thermus
HIS8_METMA	<i>Methanosarcina mazei</i>	Archaea; Euryarchaeota	AAT_PYRHO	<i>Pyrococcus horikoshii</i>	Archaea; Euryarchaeota
HIS8_METJA	<i>Methanococcus jannaschii</i>	Archaea; Euryarchaeota	AAT1_METJA	<i>Methanococcus jannaschii</i>	Archaea; Euryarchaeota
HIS8_SULTO	<i>Sulfolobus tokodaii</i>	Archaea; Crenarchaeota	AAT1_METJA	<i>Methanococcus jannaschii</i>	Archaea; Euryarchaeota
HIS8_PYRFU	<i>Pyrococcus furiosus</i>	Archaea; Euryarchaeota	AAT_PYRKO	<i>Pyrococcus kodakaraensis</i>	Archaea; Euryarchaeota
HIS8_METJA	<i>Methanococcus jannaschii</i>	Archaea; Euryarchaeota	AAT_PYRHO	<i>Pyrococcus horikoshii</i>	Archaea; Euryarchaeota
HIS8_DEIRA	<i>Deinococcus radiodurans</i>	Bacteria; Deinococcus- Thermus	AAT_METEX	<i>Methylobacterium extorquens</i>	Bacteria; Proteobacteria
HIS8_METMA	<i>Methanosarcina mazei</i>	Archaea; Euryarchaeota	AAT_AQUAE	<i>Aquifex aeolicus</i>	Bacteria; Aquificae
HIS8_METJA	<i>Methanococcus jannaschii</i>	Archaea; Euryarchaeota	AAT_PYRAB	<i>Pyrococcus abyssi</i>	Archaea; Euryarchaeota
HIS8_METMP	<i>Methanococcus maripaludis</i>	Archaea; Euryarchaeota	AAT_PYRAB	<i>Pyrococcus abyssi</i>	Archaea; Euryarchaeota
HIS8_METMP	<i>Methanococcus maripaludis</i>	Archaea; Euryarchaeota	AAT_RICCN	<i>Rickettsia conorii</i>	Bacteria; Proteobacteria
HIS8_METMP	<i>Methanococcus maripaludis</i>	Archaea; Euryarchaeota	AAT_PYRHO	<i>Pyrococcus horikoshii</i>	Archaea; Euryarchaeota
HIS8_METMP	<i>Methanococcus maripaludis</i>	Archaea; Euryarchaeota	AAT_METJA	<i>Methanococcus jannaschii</i>	Archaea; Euryarchaeota

Tabla 13. Conexiones mediante pares de proteínas entre los grupos 2a y el grupo 6. Cada par tiene al menos un 25% de identidad.

El grupo 2f1 contiene 8 enzimas que presentan de dos a tres fusiones, cuyas actividades enzimáticas son DCOS, DCOR EC, 4.1.1.17; DCLY, DCLZ EC 4.1.1.18 y ADIA, SPEA, EC 4.1.1.19.

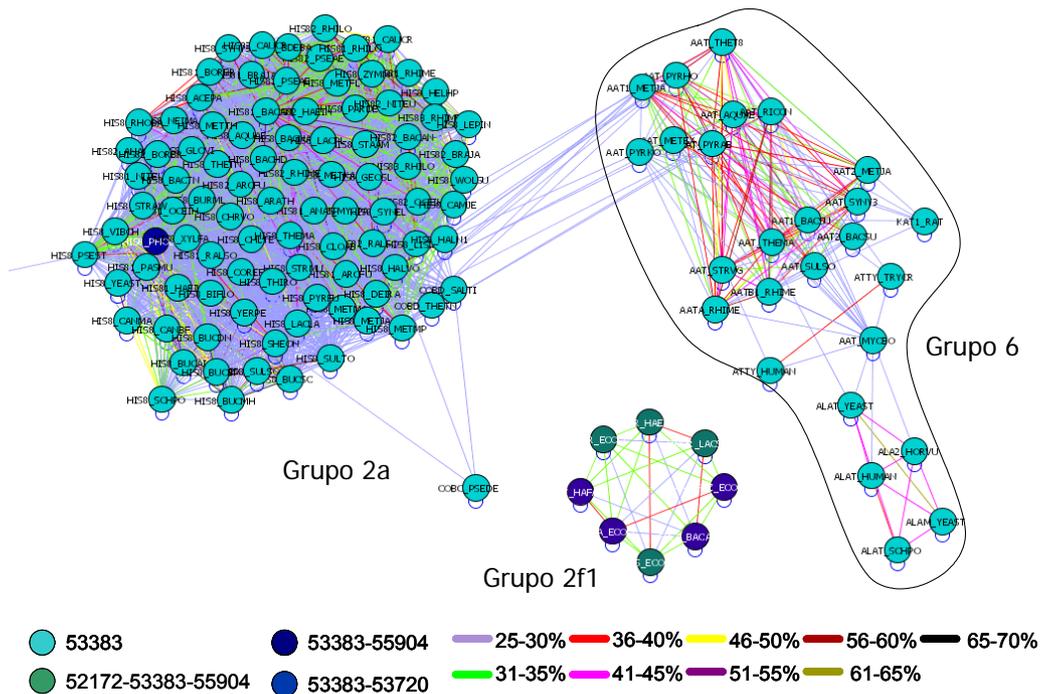


Figura 52. Para todas las enzimas clasificadas dentro del plegamiento de tipo I, el dominio estructural está numerado como 53383 (PLP-dependent transferase, según SCOP). Las identidades entre cada dominio fueron marcadas como conexiones y el color indicado representa el grado de ésta.

Podríamos suponer que los elementos más aislados (con menor número de conexiones), pertenecerían a un grupo monofilético más independiente que el resto, pero tal caso no es así. A partir del árbol generado por el grupo 2a (figura 53), podemos observar que las proteínas codificadas por los genes COBC y COBD están más relacionadas que lo que sugiere la red de identidades. También podemos observar que las proteínas suelen agruparse de acuerdo al tipo de organismo. Aunque algunos clados tienden a estar mezclados. Puesto que las enzimas que son albergadas por organismos que pertenecen al dominio Arquea, son las que básicamente sostienen la conexión entre los grupos 2a y 6, se puede conjeturar que la función HIS8, EC 2.6.1.9 se haya originado dentro de este dominio, y que de allí se haya “dispersado” hacia otros organismos pertenecientes a otros dominios. Incluyendo la diversificación hacia otras actividades enzimáticas, por ejemplo, la diversificación de la actividad HIS8, EC 2.6.1.9 hacia la actividad AAT, EC 2.6.1.1 (enzimas ubicadas en el grupo 6). La dirección de HIS8, EC 2.6.1.9 a AAT, EC 2.6.1.1 la conjeturamos por la abundancia de HIS8, EC 2.6.1.9 respecto a la abundancia de AAT, EC 2.6.1.1.

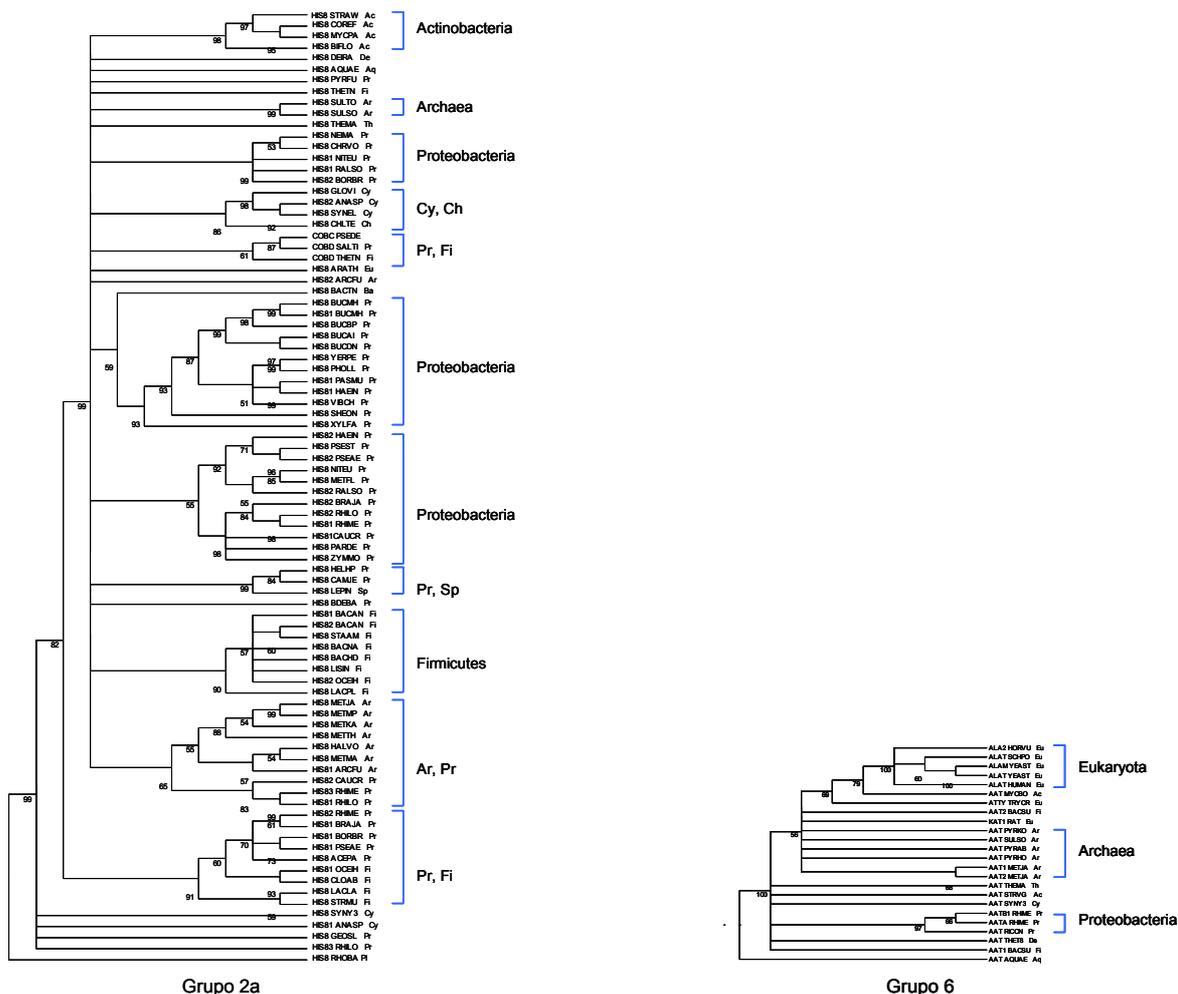


Figura 53. Árboles filogenéticos generados a partir del principio de máxima parsimonia. Pl, Planctomycetes; Ch, Chlorobi; Pr, Proteobacteria; De, Deinococcus; Cy, Cyanobacteria; Ac, Actinobacteria; Fi, Firmicutes; Ba, Bacteroides; Eu, Eukaryota; Sp, Spirochaetes; Ar, Arquea; Chl, Chloroflexi.

Como se muestra en la figura 54, los grupos 13 y 5 son grupos cercanos, puesto que se observa un número significativo de conexiones entre ellos. Los dos grupos contiene una sola actividad enzimática: Cisteína desulfurasa (EC 2.8.1.7). Aunque la identidad entre las enzimas del grupo 13 es mayor y contiene enzimas ubicadas en organismos de los tres dominios.

De esta misma figura, parece que la separación entre los grupos 13 y 5, sería un evento más reciente en comparación con la divergencia (medido con la identidad) de grupos que son conectados por un solo elemento, o en mayor medida, de grupos aislados con la “máxima resolución” (al 25% de identidad) que se usó para generar la red de enzimas

de dependen de PLP.

El grupo 17 muestra tres actividades relacionadas, Aspartato aminotransferasa (EC 2.6.1.1), Ácido amino aromático aminotransferasa (EC 2.6.1.57) y Tirosina aminotransferasa (EC 2.6.1.5) y comprende los dominios de Eucaria y Bacteria.

El grupo 3 contiene enzimas con la actividad Fosfoferina aminotransferasa (EC 2.6.1.52) y al igual que el grupo 17, comprende los dominios Eucaria y Bacteria. El grupo 16 contiene enzimas con dos actividades enzimáticas distintas, Treonina aldolasa (EC 4.1.2.5) y Alanina racemasa (EC 5.1.1.1). Por su parte, el grupo 18 es homogéneo y muestra la actividad 1-aminociclopropano-1-carboxilato sintasa (EC 4.4.1.14).

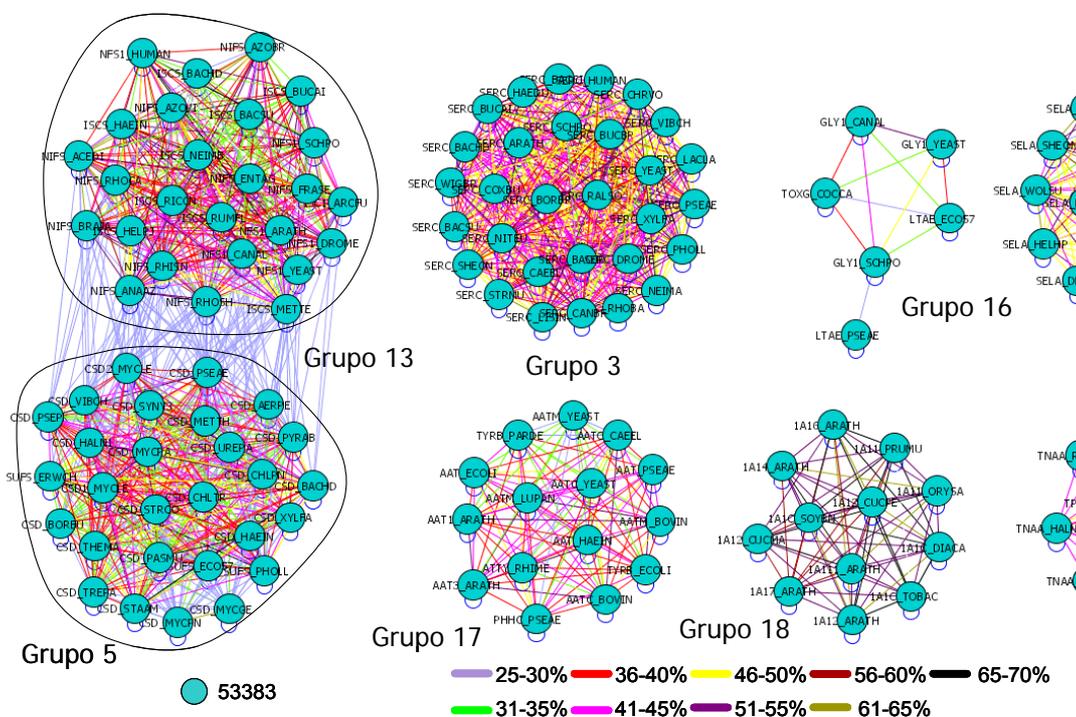


Figura 54. Para todas las enzimas clasificadas dentro del plegamiento de tipo I, el dominio estructural está numerado como 53383 (PLP-dependente transferase, según SCOP). Las identidades entre cada dominio fueron marcadas como conexiones y el color indicado representa el grado de ésta.

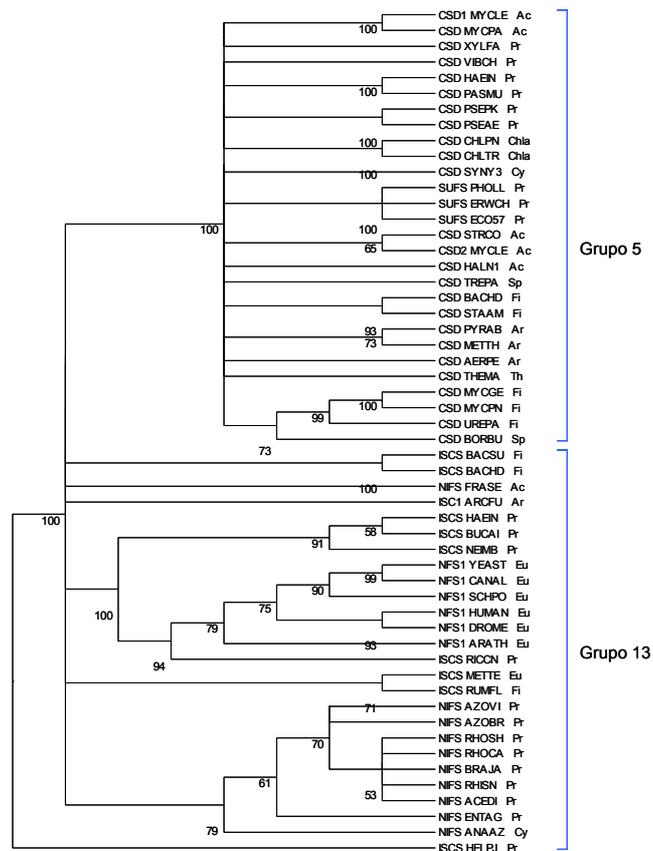


Figura 55. Árboles filogenéticos generados a partir del principio de máxima parsimonia. Pl, Planctomycetes; Ch, Chlorobi; Pr, Proteobacteria; De, Deinococcus; Cy, Cyanobacteria; Ac, Actinobacteria; Fi, Firmicutes; Ba, Bacteroides; Eu, Eukaryota; Sp, Spirochaetes; Ar, Arquea; Chl, Chloroflexi.

Por último, en la figura 56 tenemos 7 grupos pequeños y un conjunto de 6 enzimas que no pudieron conectarse con ningún otro miembro de la red. Las actividades presentes son L-seril-tRNA selenio transferasa (EC 2.9.1.1), Enfingosina-1-fosfato liasa (EC 4.1.2.27), Glutamato descarboxilasa alfa (EC 4.1.1.15), Triptofanasa (EC 4.1.99.1), Glutamato descarboxilasa (EC 4.1.1.15) e Histidina descarboxilasa (EC 4.1.1.22) en los grupos 4, 23, 20, 26 y 28, respectivamente.

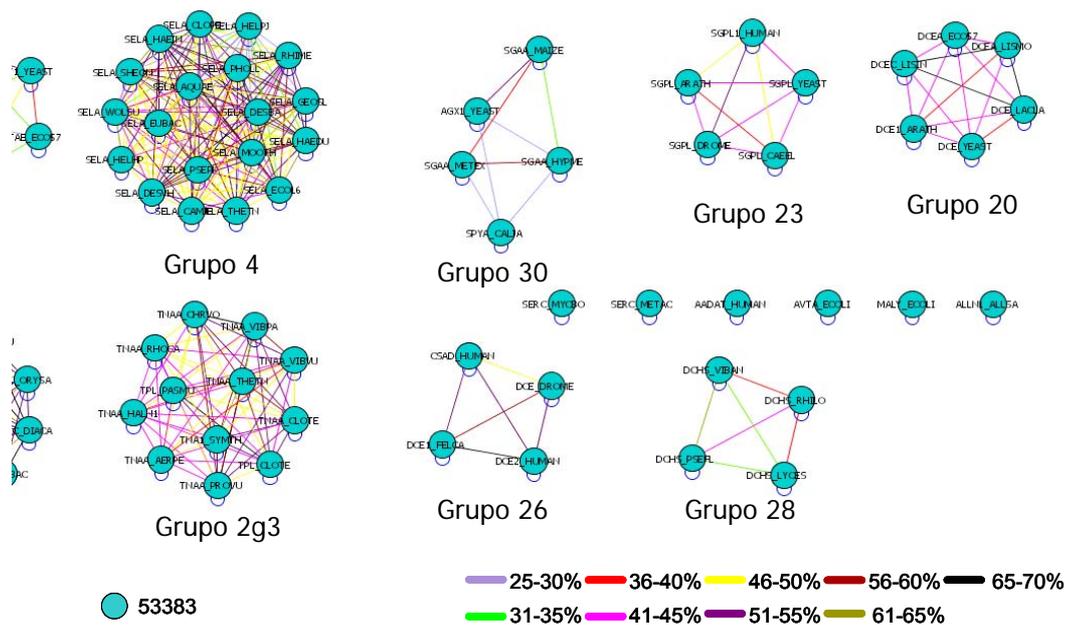


Figura 56. Para todas las enzimas clasificadas dentro del plegamiento de tipo I, el dominio estructural está numerado como 53383 (PLP-dependent transferase, según SCOP). Las identidades entre cada dominio fueron marcadas como conexiones y el color indicado representa el grado de ésta.

Plegamiento tipo II

El plegamiento de tipo II, mostrado en amarillo (dominio 53686, clasificación estructural SCOP), es representado principalmente por el grupo 1b1. Este grupo es capaz de realizar la actividad Triptofano sintasa (gen TRPB, EC 4.2.1.20) cuando se asocia con otro dominio para generar una enzima heterodimérica. El dominio asociado presenta un plegamiento de $(\beta/\alpha)_8$ (gen TRPA). Únicamente en enzimas que pertenecen a *Coprinus cinereus*, *Neurospora crassa* y *Schizosaccharomyces pombe* se presenta naturalmente la fusión. Es interesante que la identidad entre los dominios TRPB sea alta, a diferencia de los dominios TRPA (dato no mostrado).

Los grupos 1a, 1c, 22 y 29 están estrechamente ligados, puesto que comparten diferentes conexiones entre ellos, pero presentan distintas actividades enzimáticas. Las actividades reportadas para estos grupos son: treonina desaminasa (EC 4.3.1.19) para dominios en el grupo 1a; cisteína sintasa (EC 2.5.1.47) y cistationina beta-sintasa (EC 4.2.1.22) para dominios en el grupo 1c; treonina sintasa (EC 4.2.3.1) para dominios en el grupo 22; Para el grupo 29 tenemos las actividades L-serina deshidratasa (EC 4.3.1.17), además de una enzima bifuncional cuyas actividades son L-serina deshidratasa y treonina desaminasa (EC 4.3.1.19). Es posible que esta bifuncionalidad

sea indicio de la existencia de ancestros de amplia especificidad, puesto que la actividad treonina desaminasa, también está presente en dominios dentro del grupo 1a.

El grupo 1d muestra la misma actividad que el grupo 22: treonina sintasa (EC 4.2.3.1), pero no están conectados porque la identidad entre miembros de ambos grupos no supera el 25% de identidad. El grupo 21 presenta una actividad enzimática similar a los demás grupos: D-serina deshidratasa (EC 4.3.1.18), pero no está conectado a ningún grupo. Por último, el grupo 27 presenta una actividad diferente a los demás grupos con el mismo plegamiento, puesto que se reporta la actividad 1-aminociclopropano-1-carboxilato desaminasa (EC 3.5.99.7).

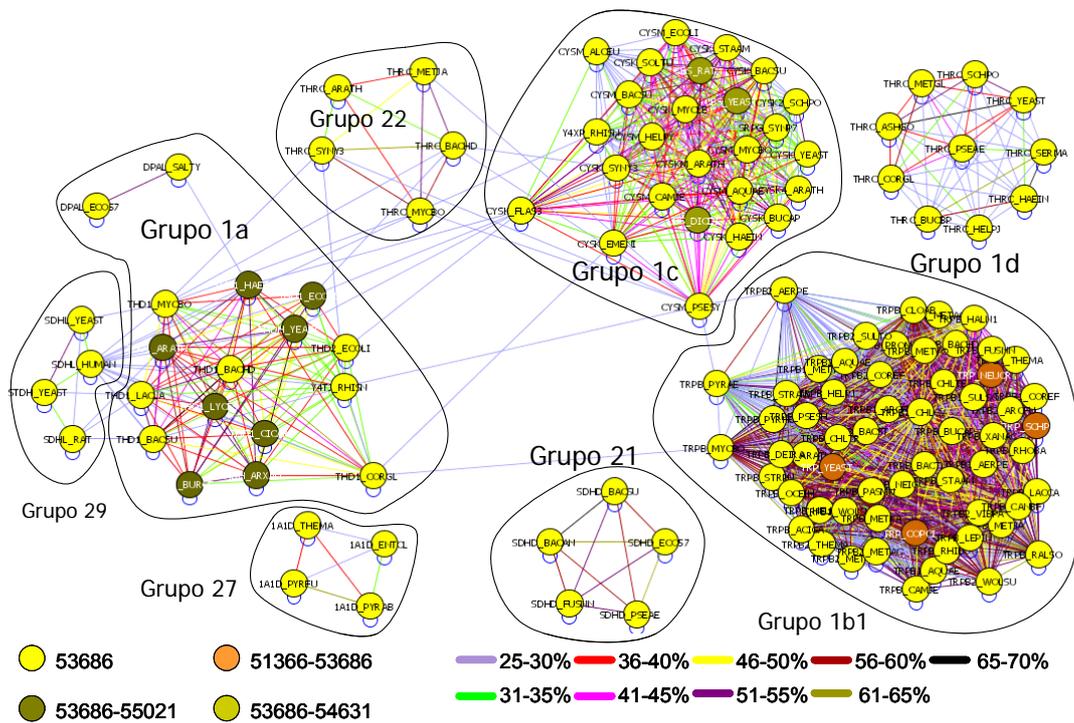


Figura 57. Para todas las enzimas clasificadas dentro del plegamiento de tipo II, el dominio estructural está numerado como 53686 (Tritophan synthase beta subunit like PLP-dependent transferase enzymes, según SCOP). Las identidades entre cada dominio fueron marcadas como conexiones y la intensidad en la escala de grises representa el grado de ésta.

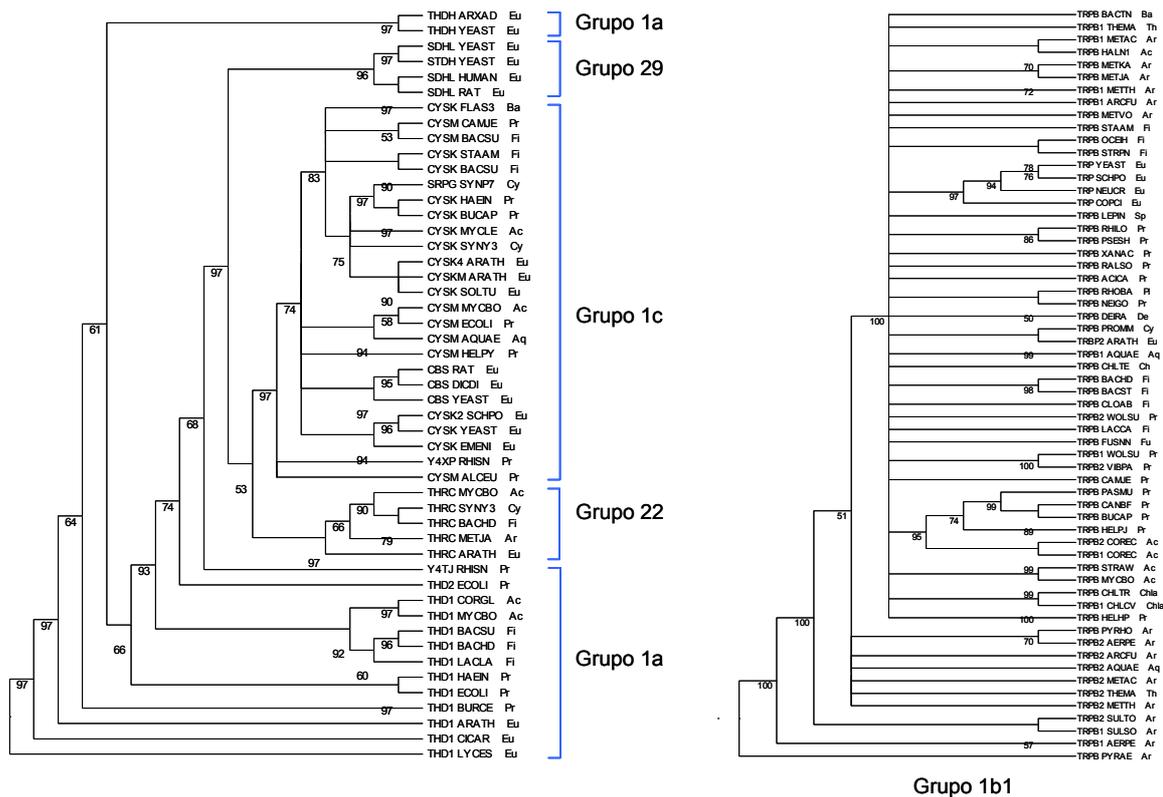


Figura 58. Árboles filogenéticos generados a partir del principio de parsimonia Pl, Planctomycetes; Ch, Chlorobi; Pr, Proteobacteria; De, Deinococcus; Cy, Cyanobacteria; Ac, Actinobacteria; Fi, Firmicutes; Ba, Bacteroides; Eu, Eukaryota; Sp, Spirochaetes; Ar, Arquea; Chl, Chloroflexi; Chla, Chlamydia;

Plegamiento tipo III (β/α)₈

Este dominio es el único dependiente de PLP, cuyo plegamiento está relacionado con otras enzimas que no dependen de PLP. Para este caso, encontramos 3 grupos claramente establecidos. Los grupos 9, 11a y 14 son homogéneos, puesto que presentan las actividades: alanina racemasa (EC 5.1.1.1), diaminopimelato descarboxilasa (EC 4.1.1.20) y arginina descarboxilasa (EC 4.1.1.19). Evidentemente, los grupos más relacionados son los 11a y 14, no solo por identidad, si no también por la similitud de sus actividades (figuras 59-60). Los grupos 9 y 11a presentan fusiones en todos sus miembros, en cambio, el grupo 14 tiene pocos miembros que presenten fusiones. Además, el grupo 9 establece una conexión con el grupo 11a mediante DCDA_HELPJ (*Helicobacter pylori*) – ALR_CARHY (*Carboxydotherrmus hydrogenoformans*).

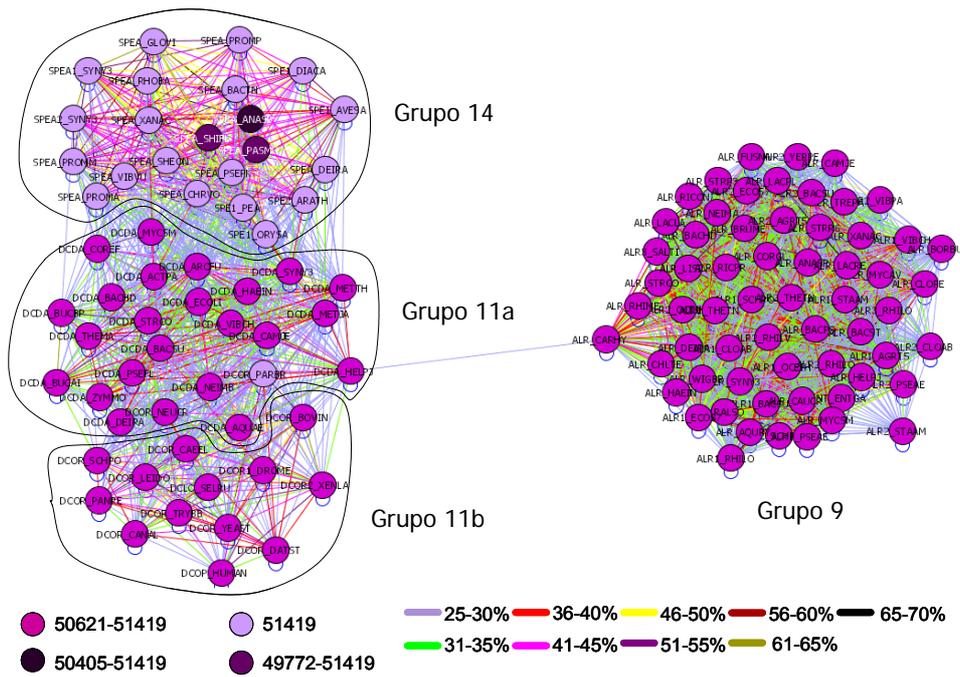


Figura 59. Para todas las enzimas clasificadas dentro del plegamiento de tipo III, el dominio estructural está numerado como 50621 (PLP-binding barrel, según SCOP). Las identidades entre cada dominio fueron marcadas como conexiones y el color indicado representa el grado de ésta.

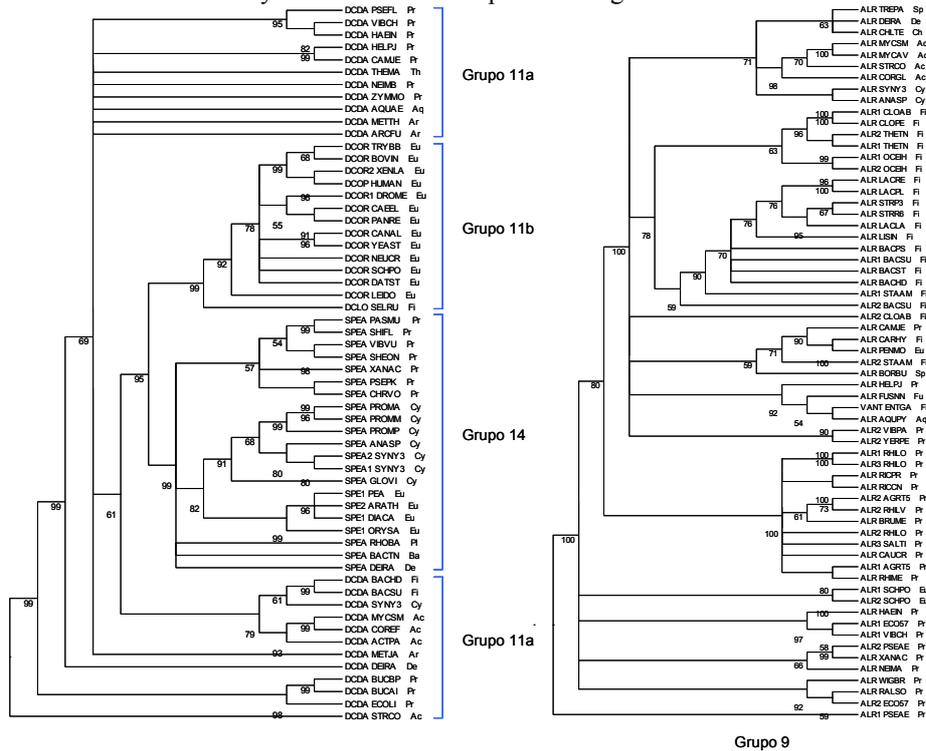


Figura 60. Árboles filogenéticos generados a partir del principio de parsimonia Pl, Planctomycetes; Ch, Chlorobi; Pr, Proteobacteria; De, Deinococcus; Cy, Cyanobacteria; Ac, Actinobacteria; Fi, Firmicutes; Ba, Bacteroides; Eu, Eukaryota; Sp, Spirochaetes; Ar, Arquea; Chl, Chloroflexi; Chla, Chlamydia; Aq, Aquificae.

Plegamiento tipo IV y V

Por último, los plegamientos IV y V son los menos abundantes. En el caso de las enzimas clasificadas dentro del plegamiento IV forma un solo grupo cerrado (grupo 10), aunque tiende a agruparse en tres subconjuntos con base a la identidad (figuras 61-62). En este grupo podemos encontrar 3 actividades enzimáticas distintas. Las enzimas anotadas como PABC muestran un EC 4.1.3.38. Las anotadas como DAAA tienen un EC 2.6.1.21; el resto de las enzimas (ILVE, BCATS, TOXF) muestran un EC 2.6.1.42.

Las enzimas dentro del plegamiento V (grupo 25) forman un grupo muy pequeño y uniforme, pues solo se observa una sola actividad enzimática: EC 2.4.1.1. Esta clase de enzimas son las únicas que usan PLP de manera muy distinta al resto de las enzimas dependientes de PLP.

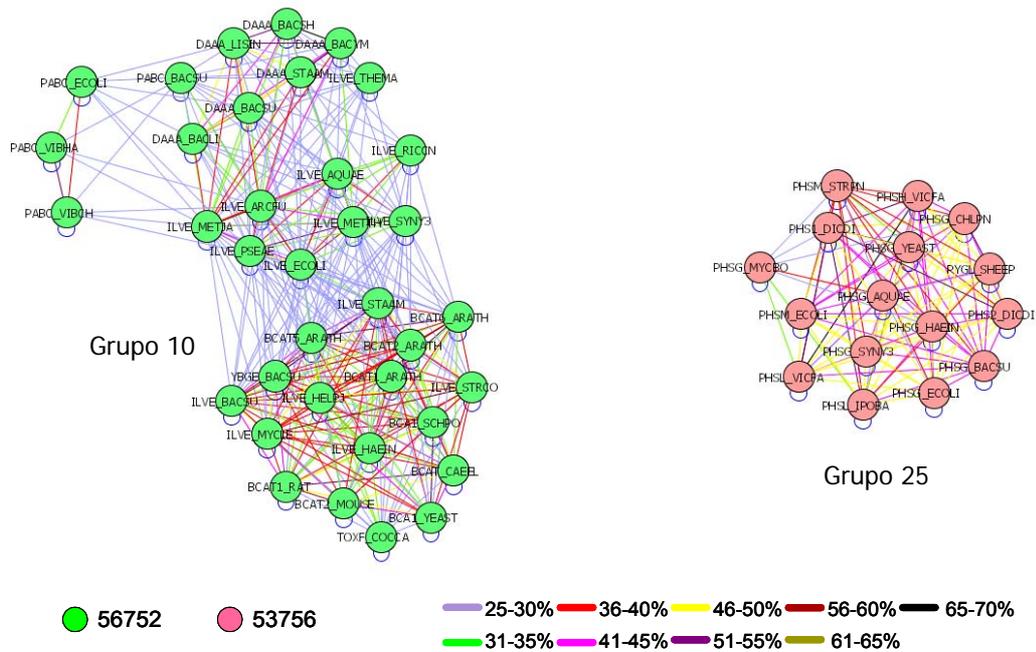


Figura 61. Para todas las enzimas clasificadas dentro del plegamiento de tipo IV, el dominio estructural está numerado como 56752 (D-aminoacid aminotransferase-like PLP-dependent enzymes, según SCOP). Para el caso de las enzimas clasificadas dentro del plegamiento de tipo V, el dominio estructural está numerado como 53756 (UDP_Glycosyltransferase/glycogen phosphorylase, según SCOP). Las identidades entre cada dominio fueron marcadas como conexiones y el color indicado representa el grado de ésta.

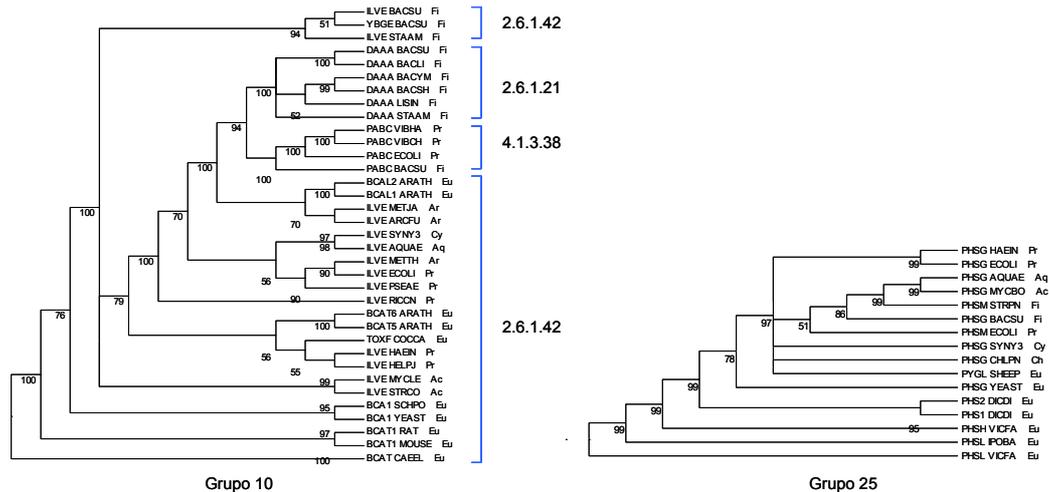


Figura 62. Árboles filogenéticos generados a partir del principio de máxima parsimonia PI, Planctomycetes; Ch, Chlorobi; Pr, Proteobacteria; De, Deinococcus; Cy, Cyanobacteria; Ac, Actinobacteria; Fi, Firmicutes; Ba, Bacteroides; Eu, Eukaryota; Sp, Spirochaetes; Ar, Arquea; Chl, Chloroflexi; Chla, Chlamydia; Aq, Aquificae.

Es evidente que mediante la duplicación las enzimas han conquistado nichos dentro del organismo que los alberga. La enorme flexibilidad mostrada por este cofactor ha permitido que esta clase de enzimas puedan catalizar diversas clases de reacciones. Éste es el único cofactor que participa en la transformación de aminoácidos, su múltiple origen evolutivo, su amplia distribución y por su indispensable papel catalítico; es posible que el PLP haya hecho presencia en la escena evolutiva antes del surgimiento de las apoenzimas (Mehta *et al.*, 1998).

No es claro el hecho que ciertos grupos sean completamente homogéneos en funciones enzimáticas, a diferencia de otros grupos que son funcionalmente diversos, recordando, que las identidades entre enzimas que pertenecen al mismo grupo oscilan en un rango de 25-70%. Además, resulta interesante que dentro de cada grupo heterogéneo (en función), podemos encontrar miembros que suelen estar conectados a otros miembros con EC distintos, en contra parte, la conectividad de otros miembros se limita a miembros con el mismo EC. Para el primer caso, estas enzimas podrían ser blanco de experimentos de evolución dirigida.

Ahora bien, se tiene que cuantificar la diversidad funcional dentro de cada grupo, puesto que no se cuantificó el efecto del primer corte al 70% de identidad, ¿este corte fue proporcional en todos los grupos?, ¿fue proporcional en todos los clados?, ¿fue proporcional en todos los EC? De la misma manera, es posible que la base de datos SWISS PROT contenga sesgos que nos esté impidiendo observar el tamaño relativo de cada grupo, por lo cual, tendría que obtenerse secuencias a partir de todos los genomas secuenciados.

Comentarios adicionales

En el laboratorio hemos empleado la auxotrofia a vitaminas como medio de selección de variantes. Puesto que las vitaminas se requieren en cantidades muy pequeñas, nos planteamos como premisa fundamental que un sistema de selección positiva basado en este tipo de auxotrofia nos daría oportunidad de seleccionar variantes con actividades muy pequeñas. Como experiencia en el laboratorio se usó la auxotrofia a tiamina como método muy sensible para seleccionar variantes con actividades muy pequeñas. Con este sistema de selección se logró la migración catalítica de una actividad isomerasa, hacia una actividad muy reducida de sintasa.

Con base a estos resultados, consideramos usar la síntesis de biotina como sistema de selección. Una diferencia sustancial entre estos sistemas, es que en el caso de la tiamina, son varias las enzimas en *E. coli* que usan esta vitamina como cofactor, a diferencia de la biotina que solo una enzima usa esta vitamina como cofactor. Suponemos que por esta razón el ruido de fondo en el sistema de selección por auxotrofia a biotina es bastante alto, oscureciendo la búsqueda de variantes.

El ruido de fondo es dependiente del número de UFC sembradas en cajas. En nuestro caso difícilmente pudimos sobrepasar las 50 000 UFC por caja. Cuando se comenzó el proyecto sembrábamos 5000 UFC por caja. A esta densidad de UFC el ruido de fondo fue nulo, pero fue inviable explorar un banco de 10^6 variantes. Nosotros interpretamos que la dependencia del ruido de fondo con respecto a la densidad de colonias es debido a la biotina residual producida por la muerte de cierta cantidad de células, de hecho, por experiencia mostramos que el lavado de células debía hacerse con cuidado para evitar la lisis celular, para que esta no fuera fuente de biotina. Además de lavar escrupulosamente el material de vidrio con mezcla crómica para eliminar cualquier traza de vitaminas.

Una situación paradójica a este hecho es la necesidad de agregar cierta cantidad muy pequeña de vitamina en las pruebas de complementación por gota. Con lo aprendido en el sistema de selección de tiamina, se determinó que las variantes que adquirieron la actividad de sintasa, podían complementar esta actividad solo si se agregaban trazas de tiamina, por lo que suponemos que ésta es requerida para activar la maquinaria metabólica sin que sea suficientemente alta para que los controles negativos puedan crecer y dando oportunidad para que las variantes seleccionadas puedan complementar la auxotrofia a tiamina.

Por otro lado, es factible aumentar la tasa mutagénica para tener mejores oportunidades de seleccionar variantes con la propiedad bifuncional deseada. De acuerdo con las tablas 2 y 3, para una tasa mutagénica reducida del 0.197% (aproximadamente 2.54 cambios de nucleótidos por secuencia), la capacidad de explorar el espacio-secuencia cae drásticamente cuando ocurren de uno a dos cambio nucleotídicos. De acuerdo con esto, se exploró el 100% de todas las posibles variantes con un solo cambio (sin considerar que en cada posición puede estar presente cualquiera de los cuatro nucleótidos). Situación diferente sucede cuando han ocurrido más de dos cambios, en la que se observa una disminución en la exploración del espacio-secuencia al 3.4% de todas las variantes posibles. Es posible que al aumentar la tasa mutagénica, mejoren las oportunidades de explorar otras combinaciones que permitan seleccionar aquellas variantes que hayan adquirido una actividad bifuncional BioF-BioA más robusta.

Glosario

Familia: “Grupo de enzimas homólogas (frecuentemente ortólogos) que catalizan la misma reacción (mecanismo y especificidad de sustrato). Frecuentemente, los miembros de las familias comparten una identidad mayor al 30%, sin embargo, las enzimas ortólogas pueden presentar una identidad menor al 30%, por lo tanto, sin alguna información estructural, puede ser difícil distinguir si una enzima pertenece a una familia determinada” (Gerlt *et al.*, 2001).

Superfamilia: “Grupos de enzimas homólogas que catalizan: (a) La misma reacción química con diferente especificidad de sustrato. (b) Diferentes reacciones que comparten algún atributo mecanístico (parte de una reacción, algún intermediario o algún estado de transición)”. “Es común que los miembros de las superfamilias presenten una identidad menor al 50%, inclusive menos del 20%” (Gerlt *et al.*, 2001).

Referencias

- Alexeev D, Alexeeva M, Baxter RL, et al. 1998. The crystal structure of 8-amino-7-oxononanoate synthase: a bacterial PLP-dependent, acyl-CoA-condensing enzyme. *J Mol Biol*, 284:401-19.
- Altschul SF, Madden TL, Schaffer AA, et al. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res*, 25:3389-402.
- Antikainen NM and Martin SF. 2005. Altering protein specificity: techniques and applications. *Bioorg Med Chem*, 13:2701-16.
- Apic G, Gough J and Teichmann SA. 2001. Domain combinations in archaeal, eubacterial and eukaryotic proteomes. *J Mol Biol*, 310:311-25.
- Bairoch A and Boeckmann B. 1991. The SWISS-PROT protein sequence data bank. *Nucleic Acids Res*, 19 Suppl:2247-9.
- Bartlett GJ, Borkakoti N and Thornton JM. 2003. Catalysing new reactions during evolution: economy of residues and mechanism. *J Mol Biol*, 331:829-60.
- Cadwell RC and Joyce GF. 1992. Randomization of genes by PCR mutagenesis. *PCR Methods Appl*, 2:28-33.
- Choi-Rhee E and Cronan JE. 2005. Biotin synthase is catalytic in vivo, but catalysis engenders destruction of the protein. *Chem Biol*, 12:461-8.
- Christen P and Mehta PK. 2001. From cofactor to enzymes. The molecular evolution of pyridoxal-5'-phosphate-dependent enzymes. *Chem Rec*, 1:436-47.
- Copley SD. 2000. Evolution of a metabolic pathway for degradation of a toxic xenobiotic: the patchwork approach. *Trends Biochem Sci*, 25:261-5.
- Denessiouk KA, Denesyuk AI, Lehtonen JV, et al. 1999. Common structural elements in the architecture of the cofactor-binding domains in unrelated families of pyridoxal phosphate-dependent enzymes. *Proteins*, 35:250-61.
- Eisenberg MA. 1985. Regulation of the biotin operon in *E. coli*. *Ann N Y Acad Sci*, 447:335-49.
- Eliot AC and Kirsch JF. 2004. Pyridoxal phosphate enzymes: mechanistic, structural, and evolutionary considerations. *Annu Rev Biochem*, 73:383-415.
- Fani R, Tamburini E, Mori E, et al. 1997. Paralogous histidine biosynthetic genes: evolutionary analysis of the *Saccharomyces cerevisiae* HIS6 and HIS7 genes. *Gene*, 197:9-17.
- Fitch WM. 1970. Distinguishing homologous from analogous proteins. *Syst Zool*, 19:99-113.
- Fitch WM. 2000. Homology a personal view on some of the problems. *Trends Genet*, 16:227-31.
- Gerlt JA and Babbitt PC. 2000. Can sequence determine function? *Genome Biol*, 1:REVIEWS0005.
- Gerlt JA and Babbitt PC. 2001. Divergent evolution of enzymatic function: mechanistically diverse superfamilies and functionally distinct suprafamilies. *Annu Rev Biochem*, 70:209-46.

- Gough J, Karplus K, Hughey R, et al. 2001. Assignment of homology to genome sequences using a library of hidden Markov models that represent all proteins of known structure. *J Mol Biol*, 313:903-19.
- Granick S. 1957. Speculations on the origins and evolution of photosynthesis. *Ann N Y Acad Sci*, 69:292-308.
- Grishin NV, Phillips MA and Goldsmith EJ. 1995. Modeling of the spatial structure of eukaryotic ornithine decarboxylases. *Protein Sci*, 4:1291-304.
- He XM and Liu HW. 2002. Formation of unusual sugars: mechanistic studies and biosynthetic applications. *Annu Rev Biochem*, 71:701-54.
- Horowitz NH and Netzenberg RL. 1965. Biochemical Aspects of Genetics. *Annu Rev Biochem*, 34:527-64.
- Huang W, Lindqvist Y, Schneider G, et al. 1994. Crystal structure of an ATP-dependent carboxylase, dethiobiotin synthetase, at 1.65 Å resolution. *Structure*, 2:407-14.
- Hutchison CA, 3rd, Phillips S, Edgell MH, et al. 1978. Mutagenesis at a specific position in a DNA sequence. *J Biol Chem*, 253:6551-60.
- Jensen RA. 1976. Enzyme recruitment in evolution of new function. *Annu Rev Microbiol*, 30:409-25.
- Jeong H, Tombor B, Albert R, et al. 2000. The large-scale organization of metabolic networks. *Nature*, 407:651-4.
- Kack H, Sandmark J, Gibson K, et al. 1999. Crystal structure of diaminopelargonic acid synthase: evolutionary relationships between pyridoxal-5'-phosphate-dependent enzymes. *J Mol Biol*, 291:857-76.
- Kammann M, Laufs J, Schell J, et al. 1989. Rapid insertional mutagenesis of DNA by polymerase chain reaction (PCR). *Nucleic Acids Res*, 17:5404.
- Kelleher JE and Raleigh EA. 1991. A novel activity in Escherichia coli K-12 that directs restriction of DNA modified at CG dinucleotides. *J Bacteriol*, 173:5220-3.
- Koonin EV. 2005. Orthologs, paralogs, and evolutionary genomics. *Annu Rev Genet*, 39:309-38.
- Koonin EV, Wolf YI and Karev GP. 2002. The structure of the protein universe and genome evolution. *Nature*, 420:218-23.
- Kummerfeld SK and Teichmann SA. 2005. Relative rates of gene fusion and fission in multi-domain proteins. *Trends Genet*, 21:25-30.
- Lazcano A and Miller SL. 1999. On the origin of metabolic pathways. *J Mol Evol*, 49:424-31.
- Li W and Godzik A. 2006. cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics*.
- Link AJ, Phillips D and Church GM. 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.
- Lorentzen E, Pohl E, Zwart P, et al. 2003. Crystal structure of an archaeal class I aldolase and the evolution of (betaalpha)₈ barrel proteins. *J Biol Chem*, 278:47253-60.
- Lynch M and Conery JS. 2000. The evolutionary fate and consequences of duplicate genes. *Science*, 290:1151-5.
- Miyazaki K and Arnold FH. 1999. Exploring nonnatural evolutionary pathways by saturation mutagenesis: rapid improvement of protein function. *J Mol Evol*, 49:716-20.

- Morowitz HJ, Kostelnik JD, Yang J, et al. 2000. The origin of intermediary metabolism. *Proc Natl Acad Sci U S A*, 97:7704-8.
- Nagano N, Orengo CA and Thornton JM. 2002. One fold with many functions: the evolutionary relationships between TIM barrel families based on their sequences, structures and functions. *J Mol Biol*, 321:741-65.
- Orengo CA, Jones DT and Thornton JM. 1994. Protein superfamilies and domain superfolds. *Nature*, 372:631-4.
- Patrick WM, Firth AE and Blackburn JM. 2003. User-friendly algorithms for estimating completeness and diversity in randomized protein-encoding libraries. *Protein Eng*, 16:451-7.
- Penning TM and Jez JM. 2001. Enzyme redesign. *Chem Rev*, 101:3027-46.
- Ponting CP and Russell RR. 2002. The natural history of protein domains. *Annu Rev Biophys Biomol Struct*, 31:45-71.
- Reikofski J and Tao BY. 1992. Polymerase chain reaction (PCR) techniques for site-directed mutagenesis. *Biotechnol Adv*, 10:535-47.
- Rison SC and Thornton JM. 2002. Pathway evolution, structurally speaking. *Curr Opin Struct Biol*, 12:374-82.
- Rodionov DA, Mironov AA and Gelfand MS. 2002. Conservation of the biotin regulon and the BirA regulatory signal in Eubacteria and Archaea. *Genome Res*, 12:1507-16.
- Schneider G, Kack H and Lindqvist Y. 2000. The manifold of vitamin B6 dependent enzymes. *Structure Fold Des*, 8:R1-6.
- Shannon P, Markiel A, Ozier O, et al. 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res*, 13:2498-504.
- Shiuan D, Lin KC and Campbell A. 1994. Transcriptional analysis of the Escherichia coli bio operon. *Gene*, 145:1-7.
- Skandalis A, Encell LP and Loeb LA. 1997. Creating novel enzymes by applied molecular evolution. *Chem Biol*, 4:889-98.
- Sonnhammer EL and Koonin EV. 2002. Orthology, paralogy and proposed classification for paralog subtypes. *Trends Genet*, 18:619-20.
- Stemmer WP. 1994. DNA shuffling by random fragmentation and reassembly: in vitro recombination for molecular evolution. *Proc Natl Acad Sci U S A*, 91:10747-51.
- Teichmann SA, Rison SC, Thornton JM, et al. 2001. The evolution and structural anatomy of the small molecule metabolic pathways in Escherichia coli. *J Mol Biol*, 311:693-708.
- Thompson JD, Higgins DG and Gibson TJ. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res*, 22:4673-80.
- Todd AE, Orengo CA and Thornton JM. 2001. Evolution of function in protein superfamilies, from a structural perspective. *J Mol Biol*, 307:1113-43.
- Vanhercke T, Ampe C, Tirry L, et al. 2005. Reducing mutational bias in random protein libraries. *Anal Biochem*, 339:9-14.
- Vogel C, Teichmann SA and Pereira-Leal J. 2005. The relationship between domain duplication and recombination. *J Mol Biol*, 346:355-65.
- Waite-Rees PA, Keating CJ, Moran LS, et al. 1991. Characterization and expression of the Escherichia coli Mrr restriction system. *J Bacteriol*, 173:5207-19.
- Waley SG. 1969. Some aspects of the evolution of metabolic pathways. *Comp Biochem Physiol*, 30:1-11.

- Weiner MP, Costa GL, Schoettlin W, et al. 1994. Site-directed mutagenesis of double-stranded DNA by the polymerase chain reaction. *Gene*, 151:119-23.
- Ycas M. 1974. On earlier states of the biochemical system. *J Theor Biol*, 44:145-60.
- Zaccolo M, Williams DM, Brown DM, et al. 1996. An approach to random mutagenesis of DNA using mixtures of triphosphate derivatives of nucleoside analogues. *J Mol Biol*, 255:589-603.