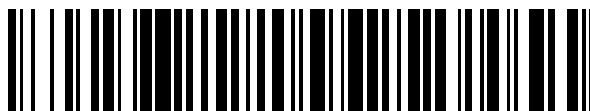


19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 718 844**

51 Int. Cl.:

**C12N 1/00** (2006.01)

**C12N 1/06** (2006.01)

**C12N 15/52** (2006.01)

12

## TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **06.05.2011 PCT/US2011/035639**

87 Fecha y número de publicación internacional: **10.11.2011 WO11140516**

96 Fecha de presentación y número de la solicitud europea: **06.05.2011 E 11721398 (3)**

97 Fecha y número de publicación de la concesión europea: **02.01.2019 EP 2566953**

54 Título: **Métodos para controlar el flujo en rutas metabólicas mediante la reubicación de enzimas**

30 Prioridad:

**07.05.2010 US 332624 P**

45 Fecha de publicación y mención en BOPI de la traducción de la patente:

**04.07.2019**

73 Titular/es:

**GREENLIGHT BIOSCIENCES, INC. (50.0%)  
200 Boston Avenue, Suite 3100**

**Medford, MA 02155, US y  
THE BOARD OF TRUSTEES OF THE LELAND  
STANFORD JUNIOR UNIVERSITY (50.0%)**

72 Inventor/es:

**SWARTZ, JAMES, R.**

74 Agente/Representante:

**LEHMANN NOVO, María Isabel**

**ES 2 718 844 T3**

Aviso: En el plazo de nueve meses a contar desde la fecha de publicación en el Boletín Europeo de Patentes, de la mención de concesión de la patente europea, cualquier persona podrá oponerse ante la Oficina Europea de Patentes a la patente concedida. La oposición deberá formularse por escrito y estar motivada; sólo se considerará como formulada una vez que se haya realizado el pago de la tasa de oposición (art. 99.1 del Convenio sobre Concesión de Patentes Europeas).

## DESCRIPCIÓN

Métodos para controlar el flujo en rutas metabólicas mediante la reubicación de enzimas.

## Antecedentes de la Invención

5 Se ha demostrado la utilidad de la producción de productos químicos mediante rutas enzimáticas sintéticas en huéspedes microbianos incluyendo isoprenoides, poliquétidos, péptidos no ribosómicos, bioplásticos y bloques de construcción química. Debido a la modularidad inherente de la información biológica, la biología sintética mantiene un gran potencial para expandir aún más esta lista de compuestos que se producen microbianamente. La inclusión de una nueva ruta bioquímica en la red metabólica de una célula huésped o de modificar la expresión de enzimas en una ruta bioquímica nativa puede interrumpir los mecanismos reguladores sutiles que la célula evolucionó durante milenios. de hecho, el rendimiento final de un compuesto se limita a menudo por efectos perjudiciales sobre el metabolismo del metabolismo celular de ingeniería genética que son difíciles de predecir debido a la comprensión limitada de las interacciones complejas que se producen dentro de la célula. El consumo no regulado de recursos celulares, la carga metabólica de la producción de proteínas heterólogas y la acumulación de los intermediarios/productos de la ruta que son inhibidores o tóxicos para el huésped son todos los problemas significativos que pueden limitar el rendimiento total.

20 El concepto de ingeniería metabólica, que se puede definir como modificación intencionada de redes metabólicas y celulares mediante el empleo de diversas técnicas experimentales para lograr objetivos deseados, ha surgido para cumplir este propósito. Lo que distingue la ingeniería metabólica de la ingeniería genética y mejora de la cepa, es que considera redes metabólicas y otras redes celulares para identificar objetivos que se van a diseñar. En este sentido, el flujo metabólico es un concepto esencial en la práctica de la ingeniería metabólica. Aunque los niveles de expresión del gen y las concentraciones de proteínas y metabolitos en la célula pueden proporcionar pautas del estado de la red metabólica, tienen limitaciones inherentes para describir completamente el fenotipo celular debido a la falta de información sobre las correlaciones entre estos componentes celulares. Los flujos metabólicos representan las velocidades de reacción en rutas metabólicas y sirven para integrar estos factores a través de una estructura matemática. de esta manera, los flujos metabólicos se pueden considerar como un modo de representar el fenotipo de la célula como resultado de la interacción entre diversos componentes celulares; los perfiles de flujo metabólico que se observaron reflejan las consecuencias de la transcripción, traducción y reacciones enzimáticas interconectadas que incorporan regulaciones complejas.

30 Síntesis libre de células que ofrece ventajas sobre métodos de producción *in vivo*. Sistemas exentos de células, si no todos, de los recursos metabólicos de la célula hacia la producción exclusiva de una ruta. Además, la falta de una pared celular *in vitro* es ventajosa ya que permite el control del entorno de síntesis. El potencial redox, el pH, o la resistencia iónica también se pueden alterar con mayor flexibilidad que *in vivo* ya que no se refiere a crecimiento o viabilidad celular. Además, la recuperación directa de productos se puede lograr fácilmente.

35 El documento WO2008094546 describe células huésped modificadas genéticamente que exhiben niveles de actividad modificados de uno o más productos genéticos de manera que, cuando se produce una enzima citocromo P450 en la célula huésped modificada genéticamente, los niveles de actividad modificados de uno o más productos genéticos proporcionan una producción y/o actividad mejorada de la enzima citocromo P450.

## Compendio de la Invención

40 La invención es como se expone en las reivindicaciones adjuntas. Se proporcionan composiciones y métodos para controlar el flujo de ruta metabólica a través de la manipulación de enzimas objetivo involucradas en una ruta de interés, incluyendo la manipulación para mantener o alterar la concentración celular de enzimas de ruta clave durante una fase de crecimiento celular, seguida por manipulación a (a) aumentar las concentraciones de enzimas de ruta clave y/o (b) disminuir las concentraciones de enzimas competitivas durante una fase de producción en donde se produce el producto de la ruta de interés. La fase de crecimiento celular implica células intactas, mientras que la fase de producción se lleva a cabo, en general, de tales células. En particular, la presente invención proporciona secuencias genéticas modificadas que codifican una o más enzimas clave en una ruta de interés para reubicar la enzima clave a un compartimento celular o extra celular en el que no se localiza naturalmente y donde la enzima clave no participa sustancialmente en el flujo de la ruta de la célula intacta cuando se reubica así, por ejemplo, en el espacio periplásmico.

50 En algunas realizaciones, las secuencias genéticas que codifican una o más enzimas clave en una ruta de interés se modifican para dar como resultado la reubicación de una o más enzimas a un compartimento celular o extracelular diferente del compartimento natural, por ejemplo, a un compartimento extra-citoplasmático diferente o que se secreta fuera de la célula al medio circundante.

55 En realizaciones específicas, las secuencias genéticas que codifican una o más enzimas clave en una ruta de interés se modifican para codificar una secuencia peptídica que proporciona la dirección periplásmica del polipéptido, para reubicar, o secuestrar, la enzima en el espacio periplásmico de la célula. En algunas realizaciones, la enzima de ruta modificada es una enzima de entrada de ruta, como se define en la presente. En otras realizaciones, la enzima de ruta modificada es una enzima limitante de la velocidad.

Para la mayoría de los propósitos, la enzima periplásmicamente dirigida o de otra manera reubicada se sobreexpresa en la célula, con relación al nivel de expresión en una célula nativa, uniendo operablemente la secuencia codificante a un promotor constitutivo o inducible de alto nivel. En ciertas realizaciones de la invención, una copia nativa de la enzima objetivo, o una isozima de la enzima objetivo, se expresa en la célula a niveles fisiológicamente normales, por ejemplo, del promotor nativo. En algunas realizaciones, las enzimas en la ruta distinta de la enzima objetivo se sobreexpresan, es decir, se expresan a niveles mayores que los del nivel fisiológicamente normal.

Durante la fase de crecimiento celular, la enzima reubicada, que puede secuestrarse en el periplasma, por ejemplo, no afecta el flujo de la ruta. Para iniciar la fase de producción, se lisan las células, en cuyo punto la enzima reubicada se une con las enzimas citoplásmicas en la ruta de interés, permitiendo la producción a alto nivel del producto de interés.

En algunas realizaciones, se proporcionan métodos para producir un producto de interés a alta velocidad de flujo, el método comprende: células de crecimiento que se modifican genéticamente para sobreexpresar al menos una enzima reubicada en una ruta de interés a una densidad celular deseada; lisar las células; y producir el producto de la ruta en un sistema libre de células que comprende el lisado. Uno o más sustratos, nutrientes, cofactores, tampones, agentes reductores, y/o sistemas de generación de ATP, pueden unirse al sistema libre de células.

En otro aspecto, se proporciona una célula modificada genéticamente que sobreexpresa al menos una enzima reubicada en una ruta de interés.

En otro aspecto, se proporcionan lisados de tal célula modificada genéticamente, cuyo lisado se puede combinar con uno o más de los sustratos, nutrientes, cofactores, tampones, agentes reductores y/o sistemas generadores de ATP, para generar un sistema libre de células para producir un producto de interés.

Los detalles de una o más realizaciones de la invención se exponen en la presente. Otras características, objetos y ventajas de la invención serán evidentes a partir de la descripción, las figuras, los ejemplos y las reivindicaciones.

#### Breve Descripción de los Dibujos

*La figura 1* ilustra la localización periplásmica de AroG genéticamente modificada.

*La figura 2* representa datos de crecimiento de cultivos celulares de BL21 (DE3) que expresan OmpA-aroG o un control del vector vacío pACYC, lo que indica que la expresión periplásmica de AroG no tiene ningún efecto negativo sobre el crecimiento celular.

*La figura 3* muestra la actividad específica de 3-desoxi- D-arabino-heptulosonato-7-fosfato (DAHP) sintasa expresada periplásmicamente.

*La figura 4* representa la ruta para la biosíntesis de ácido shikímico.

*La figura 5* representa la ruta para la biosíntesis de amorfadieno.

#### Descripción detallada

La presente invención se basa en la idea de que las células manipuladas genéticamente se pueden diseñar para producir un elemento funcional (por ejemplo, una enzima) que tendría un impacto negativo sobre la salud de la célula, pero para la reubicación de ese elemento funcional fuera de la célula o en una ubicación secuestrada dentro de la celda. En una realización, dicha localización secuestrada es el espacio periplásmico de la célula. En ciertas realizaciones, el elemento funcional es una enzima clave que controla el flujo en una ruta de interés.

Por ejemplo, en un aspecto, se proporciona una célula con al menos una enzima que controla el flujo en una ruta de interés, en donde la enzima es modificada genéticamente para reubicar la enzima clave a un compartimento celular o extracelular de origen no natural (es decir, un compartimento celular o extracelular distinto del compartimento en el que se produce naturalmente la enzima), y en donde la enzima no participa en el flujo de ruta de la célula intacta cuando se reubica de esta manera. Rutas ejemplares de interés incluyen, pero no se limitan a, la síntesis de shikimato, diversos isoprenoides y terpenoides, poli-3-hidroxitirato, isobutanol y 1-butanol, como se detalla en la presente memoria descriptiva.

En ciertas realizaciones, la enzima se modifica genéticamente para incluir una secuencia peptídica que proporciona la dirección periplásmica del polipéptido, es decir, donde la enzima es secuestrada en el periplasma de la célula. En ciertas realizaciones, la enzima es una enzima de entrada de ruta. En ciertas realizaciones, la enzima es una enzima limitante de la velocidad. En ciertas realizaciones, la enzima aumenta la velocidad de suministro de precursor a la ruta de interés o suministra cualquier otro sustrato o cofactor requerido. En ciertas realizaciones, una contraparte nativa de la enzima se expresa a niveles citoplásmicos normales. En determinadas formas de realización, la contraparte nativa se lleva a cabo. En ciertas realizaciones, la enzima se sobreexpresa en la célula. En ciertas realizaciones, la enzima está presente en un vector episómico o en un cromosoma. En ciertas realizaciones, al menos dos enzimas (por ejemplo, dos, tres, cuatro, cinco o más enzimas) en la ruta de interés se modifican genéticamente para comprender una secuencia peptídica que proporciona la dirección periplásmica del polipéptido. En ciertas

realizaciones, el medio de crecimiento celular se ha modificado por la adición o mejora de un factor (por ejemplo, un nutriente, co-factor, agente reductor) que aumenta o conserva la actividad de la enzima.

5 En otro aspecto, se proporciona un sistema para producir un producto de una ruta de interés, sistema que comprende una celda de la presente invención; y opcionalmente uno o más sustratos, enzimas, nutrientes, co-factores, tampones, agentes reductores y sistemas generadores de ATP.

En otro aspecto, se proporciona un sistema para producir un producto de una ruta de interés, donde el sistema comprende una célula de la presente invención; y opcionalmente uno o más sustratos, enzimas, nutrientes, co-factores, tampones, agentes reductores y sistemas generadores de ATP.

10 En otro aspecto, se proporciona un sistema para producir un producto de una ruta de interés, donde el sistema comprende un lisado de una célula de la presente invención; y opcionalmente uno o más sustratos, enzimas, nutrientes, cofactores, tampones, agentes reductores y sistemas generadores de ATP. En ciertas realizaciones, el sistema incluye además uno o más lisados celulares adicionales.

15 En otro aspecto más, se proporciona un método para producir un producto de una ruta de interés, el método comprende cultivar una célula de la presente invención a una densidad celular deseada; lisar las células; y combinar el lisado con uno o más sustratos, enzimas, nutrientes, cofactores, tampones, agentes reductores, o sistemas de generación de ATP, en los que las enzimas en la ruta de interés producen la producción del producto deseado. En ciertas realizaciones, el método comprende además combinar el lisado con uno o más lisados celulares adicionales.

En otro aspecto más, se proporciona un vector que codifica una enzima modificada genéticamente para comprender una secuencia peptídica que proporciona la dirección periplásmica del polipéptido.

20 Secuestro periplásmico

En algunas realizaciones de la invención, la reubicación de la enzima es al espacio periplásmico. En tales aspectos, la presente invención proporciona métodos para la generación de células; y sus usos, en los que una o más enzimas clave en una ruta de interés se modifican genéticamente para incorporar una secuencia peptídica que proporciona la dirección periplásmica del polipéptido. Secuencias de péptido de señal de dirección periplásmica (también denominadas señales de direccionamiento o secuencias de señal) normalmente se encuentran en el extremo N-terminal de las proteínas secretoras bacterianas. Varían en longitud de aproximadamente 15 a aproximadamente 25 70 aminoácidos. Las secuencias de aminoácidos primarias de los péptidos de señal también varían, pero generalmente tienen una estructura general común que incluye las siguientes partes: i) la parte N-terminal tiene una longitud variable y generalmente lleva una carga neta positiva ; ii) un núcleo hidrófobo central de aproximadamente 6 a aproximadamente 15 aminoácidos; iii) la parte final incluye cuatro a seis aminoácidos que definen el sitio de escisión para peptidasas de señal.

35 Las secuencias de péptido de señal de dirección periplásmica adecuadas para uso en la presente invención se derivan generalmente de una proteína que se secreta en una bacteria Gram negativa. La proteína secretada puede ser codificada por la bacteria, o por un bacteriófago que infecta a la bacteria. Ejemplos de fuentes bacterianas Gram-negativas adecuadas de proteínas secretadas incluyen, pero no se limitan a, miembros de los géneros *Escherichia*, *Pseudomonas*, *Klebsiella*, *Salmonella*, *Caulobacter*, *Methylomonas*, *Acetobacter*, *Achromobacter*, *Acinetobacter*, *Aeromonas*, *Agrobacterium*, *Alcaligenes*, *Azotobacter*, *Burkholderia*, *Citrobacter*, *Comamonas*, *Enterobacter*, *Erwinia*, *Rhizobium*, *Vibrio*, and *Xanthomonas*.

40 Existen tres rutas para la translocación a través de la membrana citoplásmica: (i) SecB-dependiente, (ii) partícula de reconocimiento de señal (SRP), y (iii) rutas de translocación de arginina gemela (TAT). Las rutas de las partículas de reconocimiento de la señal SecB y dependientes de la señal utilizan tanto el translocado como el translocado SecYEG. La ruta de translocación de arginina gemela utiliza el complejo de TatABCE. La translocación dependiente de SecB se utiliza más comúnmente, pero esta ruta no es capaz de transportar proteínas plegadas. Plegamiento citoplasmático rápido puede necesitar el uso de rutas SRP o TAT. Los ejemplos de proteínas secretadas por bacterias que tienen péptidos de señal de dirección periplásmica incluyen, pero no se limitan a proteínas codificadas por los siguientes genes: *ompA*, *genIII*, *E coli* fosfatasa alcalina, *lamB*, *malE*, *secE*, *secY*, y *prfA-4*. Un entendido en la técnica puede identificar fácilmente la señal de péptido de dirección periplásmica situado en el extremo N-terminal de cada una de estas proteínas, y de otras proteínas secretoras bacterianas. También se conoce por un entendido en la técnica que algunas sustituciones de aminoácidos, adiciones, y/o supresiones se pueden hacer en un péptido de 50 señal de dirección periplásmica mientras que retiene su función de dirección. de esta manera, un péptido de señal de dirección periplásmica funcional de uso en la presente invención puede ser completamente natural o una secuencia modificada.

Las etapas en el proceso de secuestro periplásmico incluyen: i) el análisis de ruta para identificar una enzima de entrada clave (s) para secuestrar al periplasma, ii) la construcción de casetes de expresión para la dirección periplásmica de la enzima (s) incluyendo la selección de péptidos de señal y la optimización de la expresión, iii) la verificación de la enzima diana activa, periplásmica, activa y iv) la demostración de crecimiento celular metabólicamente saludable seguido de un flujo incrementado al producto de interés post-lisis en una reacción activa libre de células.

Las proteínas de fusión de la presente invención comprenden una señal de dirección periplásmica (PerS) y una enzima de ruta, por ejemplo, una enzima de entrada de ruta y/o una enzima limitante de la velocidad. Generalmente, el péptido de señal periplásmica óptimo para cada proteína dirigida al periplasma se determina empíricamente a partir de una selección de tales péptidos. La eficacia de la secreción dependerá de diversos parámetros, por ejemplo, del péptido de señal utilizado, siendo la proteína diana, la cepa hospedadora utilizada y/o el nivel de expresión. Una biblioteca de genes modificados con regiones de 5'variables que codifican para diferentes péptidos de señal periplásmica, se pueden crear utilizando PCR u Otros métodos familiares para los entendidos en la técnica. Esta biblioteca se subclona en un vector que permite la expresión controlada (por ejemplo, un vector que permite la expresión controlada utilizando el sistema de inducción T7 tal como un vector de la serie pET), y se ensaya respecto a la eficacia de la exportación, así como a la actividad de la proteína diana (véase, por ejemplo, Dahl et al., J. Biol. Chem. (1992) 267:4882-4888; Chen et al., J. Biol. Chem. (1992) 267:12375-12379; Publicación No. 2007/0111283; Merulhao et al. J Microbiol. Biotechnol. (2007) 17: 1236-1241; y Mergulhao et al., Biotechnology Advances (2005) 23:177-202). Se incluyen señales de dirección periplásmica ejemplares, sin limitación, en la tabla 1.

Tabla 1. Señales de Dirección Periplásmica

Nombre	Péptido de señal	Ruta	Fuente
MalEss	MKIKTGARILALSALTTMMFSASALA (SEQ ID NO:1)	Sec	<i>E. coli</i>
PhoAss	MKQSTIALALLPLLFTPVTKA (SEQ ID NO:2)	Sec	<i>E. coli</i>
LamBss	MMITLRKPLAVAVAAGVMSAQAMA (SEQ ID NO:3)	Sec	<i>E. coli</i>
MglBss	MNKKVLTLSAVMASMLFGAAHA (SEQ ID NO:4)	Sec	<i>E. coli</i>
PelBss	MKYLLPTAAAGLLLLAAQPAMA (SEQ ID NO:5)	Sec	<i>E. caratovora</i>
DsbAss	MKKIWLALAGLVLAFSASA (SEQ ID NO:6)	SRP	<i>E. coli</i>
SfmCss	MMTKIKLLMLIIFYLIISASAHA (SEQ ID NO:7)	SRP	<i>E. coli</i>
TolBss	MKQALRVAFGLILWASVLHA (SEQ ID NO:8)	SRP	<i>E. coli</i>
TorTss	MRVLLFLLLSLFMLPAFS (SEQ ID NO:9)	SRP	<i>E. coli</i>
TorAss	MANNDLFQASRRRFLAQLGGLTVAGMLGPSLLTPRRA TA (SEQ ID NO:10)	TAT	<i>E. coli</i>

Las tablas 4 y 5 de los Ejemplos proporcionan cebadores ejemplares útiles en la incorporación de una de las secuencias de la Tabla 1 en el extremo N-terminal de una proteína de interés, e incluye secuencias para sitios de restricción útiles en subclonación. Se entiende que un entendido en la técnica sería capaz de codificar una señal de dirección periplásmica, como se ejemplifica en la tabla 1, utilizando secuencias de ácido nucleico diferentes de las ejemplificadas en las Tablas 4 y 5 en base a la naturaleza degenerada del código genético. Mutaciones silenciosas en la secuencia de ácido nucleico (es decir, que no afectan a la secuencia de aminoácidos) no afectarán la actividad de direccionamiento periplásmico. Mutaciones no silenciosas en la secuencia de ácido nucleico (es decir, que afectan a la secuencia de aminoácidos) también son posibles que no afectarían sustancialmente la dirección periplásmica. En ciertas realizaciones, uno, dos, tres, cuatro o cinco mutaciones en una señal de dirección periplásmica de la tabla 1 Logra la dirección al periplasma. En ciertas realizaciones, al menos 90%, 95%, 98%, o 99% de homología en una secuencia de aminoácidos de la Tabla 4 y/o 5 logra la dirección al periplasma de la proteína de interés. En ciertas realizaciones, el uso del codón en la secuencia de ácido nucleico que codifica la señal de dirección periplásmica se optimiza para el organismo huésped.

Un sitio de escisión se localiza opcionalmente entre la señal de dirección periplásmica (PerS) y la enzima para permitir la separación de estos péptidos. El sitio de escisión puede ser cualquier sitio que pueda utilizarse para separar los PerS y la enzima. Puede utilizarse cualquier uso de sitio de escisión de cualquier método para la escisión de proteínas. Los PerS de *E. coli* pueden contener dentro de su secuencia señal un motivo reconocido por peptidasa líder (Lep) para procesamiento y escisión de la secuencia de señales. Otros métodos que pueden encontrar utilidad incluyen métodos de escisión de proteasa, por ejemplo, trombina, proteasa del factor Xa y otras endo-peptidasas, tales como tripsina. Los genes que codifican la proteína de fusión pueden sintetizarse para incluir un sitio de escisión para una de estas proteasas entre el péptido PerS y la secuencia de enzimas. Otro sistema para la fusión y escisión es el sistema de dominio de unión de integrina/quitina que hace uso de las propiedades de autoescisión de proteínas integrantes (véase, por ejemplo, Chong et al. Gene (1997) 192: 271-281).

Las secuencias de ADN que codifican señales de direccionamiento periplásmicas útiles en la invención pueden ser las secuencias de codificación naturales presentes en los genes a partir de los cuales derivan. Adicionalmente, la secuencia de codificación se puede traducir de nuevo utilizando la secuencia de aminoácidos de la señal de dirección periplásmica, opcionalmente utilizando codones optimizados. Un fragmento de ADN que codifica una señal de dirección periplásmica que se usa en una proteína de fusión que codifica un fragmento de ácido nucleico aislado se puede obtener utilizando cualquier método tal como aislamiento de la naturaleza, síntesis química, técnicas recombinantes, o amplificación tal como mediante la utilización de PCR.

Ácidos nucleicos, polipéptidos y células para uso en la presente invención

Los ácidos nucleicos utilizados para poner en práctica esta invención, ya sea ARN, iARN, ácido nucleico antisentido, ADNc, ADN genómico, vectores, cromosomas artificiales, los virus o los híbridos de los mismos pueden aislarse de una variedad de fuentes, genéticamente diseñadas, amplificadas y/o expresadas/generadas de manera recombinante. Una molécula de ácido nucleico o moléculas de ácido nucleico que codifican cualquiera de las enzimas asociadas con la invención se pueden introducir en una célula o células utilizando métodos y técnicas que son estándar en la técnica. Por ejemplo, se pueden introducir moléculas de ácido nucleico mediante protocolos estándar tales como transformación que incluye transformación química y electroporación, transducción, bombardeo de partículas, etc. La expresión de una molécula de ácido nucleico (s) que codifica una enzima también puede lograrse integrando la molécula de ácido nucleico en el genoma. La molécula de ácido nucleico (s) se puede integrar en el ADN genómico de una célula utilizando técnicas estándar bien conocidas en la técnica. Los polipéptidos recombinantes generados a partir de estos ácidos nucleicos pueden ser individualmente aislados o clonados y probados para una actividad deseada. Se puede usar cualquier sistema de expresión recombinante, incluyendo, pero sin limitarse a, bacterias, sistemas de expresión de células de mamíferos, levaduras, insectos o vegetales. Los ácidos nucleicos para uso en la presente invención se pueden sintetizar *in vitro* mediante técnicas de síntesis química bien conocidas, como se describe en, por ejemplo, Adams et al, J Am Chem Soc (1983) 105: 661; Belosov et al. Nucleic Acids Res (1997) 25: 3440-3444; Frenkel et Al. Free Radic. Biol. Med. (1995) 19: 373-380; Blommers et al. Biochemistry (1994) 33: 7886-7896; Narang et al. Meth. Enzymol. (1979) 68: 90; Brown et al. Meth. Enzymol. (1979) 68: 109; Beaucage et al. Tetrahedron Letters (1981) 22: 1859; y la patente de los Estados Unidos 4,458,066.

Las células huésped de interés para ingeniería de ruta incluyen una amplia variedad de microorganismos heterotróficos y autotróficos, incluyendo, pero sin limitarse a, bacterias, hongos y protozoarios. Las células huésped preferidas incluyen aquellas para las cuales se conocen medios por los cuales un polipéptido puede dirigirse a un compartimiento celular o compartimientos extracelulares. La invención abarca cualquier tipo de célula que expresa recombinantemente ácidos nucleicos asociados con la invención, incluyendo células procarióticas y eucarióticas. En algunas realizaciones, la célula es una célula bacteriana, tal como *Escherichia* spp., *Streptomyces* spp., *Zymonas* spp., *Acetobacter* spp., *Citrobacter* spp., *Synechocystis* spp., *Rhizobium* spp., *Clostridium* spp., *Corynebacterium* spp., *Streptococcus* spp., *Xanthomonas* spp., *Lactobacillus* spp., *Lactococcus* spp., *Bacillus* spp., *Alcaligenes* spp., *Pseudomonas* spp., *Aeromonas* spp., *Azotobacter* spp., *Comamonas* spp., *Mycobacterium* spp., *Rhodococcus* spp., *Gluconobacter* spp., *Ralstonia* spp., *Acidithiobacillus* spp., *Micrococcus* spp., *Geobacter* spp., *Geobacillus* spp., *Arthrobacter* spp., *Flavobacterium* spp., *Serratia* spp., *Saccharopolyspora* spp., *Thermus* spp., *Stenotrophomonas* spp., *Chromobacterium* spp., *Sinorhizobium* spp., *Saccharopolyspora* spp., *Agrobacterium* spp. y *Pantoea* spp. La célula bacteriana puede ser una célula Gram-negativa tal como célula *Escherichia coli* (*E. coli*), o una célula Gram-positiva tal como una especie de *Bacillus*. En otras realizaciones, la célula es una célula fúngica tal como células de levadura, por ejemplo, *Saccharomyces* spp., *Schizosaccharomyces* spp., *Pichia* spp., *Paffia* spp., *Kluyveromyces* spp., *Candida* spp., *Talaromyces* spp., *Brettanomyces* spp., *Pachysolen* spp., *Debaryomyces* spp., *Yarrowia* spp. Las cepas de levadura poliploide industrial. Otros ejemplos no limitativos de hongos incluyen *Aspergillus* spp. *Penicillium* spp. *Fusarium* spp. *Rhizopus* spp. *Acremonium* spp., *Neurospora* spp. *Sordaria* spp. *Magnaporina* spp. *Allomyces* spp. *Ustilago* spp. *Botrytis* spp. y *Trichoderma* spp. En otras realizaciones, la célula es una célula de algas, una célula de planta, o una célula de mamífero. Se debe apreciar que algunas células compatibles con la invención pueden expresar una copia endógena de uno o más de los genes asociados con la invención, así como una copia recombinante. Las especies de interés incluyen, sin limitación, especies *S. cerevisiae*, *E. coli*, *Pseudomonas* species, *Klebsiella* y especies de *Synechocystis*. Para evitar la degradación no deseada de la proteína reubicada, la cepa huésped puede ser modificada para remover varias proteasas compartimentales (por ejemplo, proteasas periplásmicas) y/o para aumentar con proteínas tales como chaperonas y maturasas para ayudar con el plegamiento de proteínas; tales modificaciones y aumentos emplean métodos familiares para los entendidos en la técnica; véanse, por ejemplo, las patentes de los Estados Unidos 4.946.783 y 6.921.659, y Chen et al., *Biotechnology and Bioengineering* (2004) 85: 463-474.

En algunas realizaciones uno o más genes asociados con la invención se expresan recombinantemente en una célula bacteriana. Las células bacterianas de acuerdo con la invención se pueden cultivar en medios de cualquier tipo (rico o mínimo) y cualquier composición. En algunas realizaciones, las células son cultivos en medio mínimo. Como se entenderá por un entendido en la técnica, la optimización de rutina permitiría el uso de una variedad de tipos de medios. El medio seleccionado puede complementarse con diversos componentes adicionales. Algunos ejemplos no limitativos de componentes suplementarios incluyen glucosa, antibióticos, IPTG, tetraciclina o anhidro-tetraciclina (aTc) de inducción de gen y Suplemento Mineral de Trazas de ATCC. de manera similar, otros aspectos del medio, y las condiciones de crecimiento de las células de la invención pueden optimizarse a través de la experimentación de rutina. Por ejemplo, el pH y la temperatura son ejemplos no limitativos de factores que pueden optimizarse. En

algunas realizaciones, se puede optimizar la concentración y la cantidad de un componente complementario. En algunas realizaciones, como a menudo el medio se complementa con uno o más componentes suplementarios, y la cantidad de tiempo en que se cultiva el medio.

5 Las técnicas para la manipulación de ácidos nucleicos, por ejemplo, subclonación, sondas marcadoras (por ejemplo, marcaje de cebador aleatorio utilizando polimerasa de Klenow, traducción de muesca, amplificación), la secuenciación, la hibridación y similares se describen bien en la literatura científica y de patentes, véase, por ejemplo, Sambrook, Ed. Molecular Cloning: A Laboratory Manual (2ª Ed.) Volúmenes 1-3, Cold Spring Harbor Laboratory (1989); Ausubel, Ed, Current Protocols in Molecular Biology, John Wiley & Sons, Inc. Nueva York (1997); y Tijssen, Ed. Laboratory Techniques In Biochemistry and Molecular Biology: Hybridization with Nucleic Acid Probes, Parte I. Theory and Nucleic Acid Preparation, Elsevier, N.Y. (1993).

10 Se debe apreciar que los genes que codifican enzimas asociadas con la invención pueden obtenerse de una variedad de fuentes. Como alguien de experiencia ordinaria en la técnica sería consciente, los genes homólogos para estas enzimas existen en muchas especies y pueden ser identificados por búsquedas de homología, por ejemplo, a través de una búsqueda de proteína BLAST, disponible en el sitio de internet NCBI ([www.ncbi.nlm.nih.gov](http://www.ncbi.nlm.nih.gov)). Los Genes que codifican para estas enzimas pueden ser amplificados por PCR a partir de ADN procedente de cualquier fuente que contenga la enzima dada, por ejemplo, utilizando cebadores degenerados, como se entenderá por un entendido en la técnica. En algunas realizaciones, el gen que codifica para una enzima dada puede ser sintético. Cualquier medio para obtener los genes que codifican para las enzimas descritas aquí son compatibles con los aspectos de la presente invención.

15 La presente divulgación también proporciona polipéptidos aislados codificados por los ácidos nucleicos. Tales polipéptidos son útiles, por ejemplo, solos o como proteínas de fusión. Los polipéptidos asociados con la invención pueden aislarse a partir de muestras biológicas que incluyen homogenatos de tejidos o de células, y también se pueden expresar de manera recombinante en una variedad de sistemas de expresión procariótica y eucariótica mediante la construcción de un vector de expresión apropiado para el sistema de expresión, introducir el vector de expresión en el sistema de expresión, y aislar la proteína expresada recombinantemente. Los polipéptidos también se pueden sintetizar químicamente utilizando métodos bien establecidos de síntesis de péptidos.

20 Se pueden utilizar una variedad de metodologías bien conocidas por el entendido en la materia para obtener polipéptidos aislados asociados con la invención. El polipéptido puede ser purificado a partir de células que producen naturalmente el polipéptido por medio cromatográfico o reconocimiento inmunológico. Alternativamente, se puede introducir un vector de expresión en células para provocar la producción del polipéptido. En otro método, los transcritos de ARNm pueden ser microinyectados o de otra manera introducidos en células para provocar la producción del polipéptido codificado. La traducción de ARNm en extractos libres de células tales como el sistema de lisado de reticulocitos también se pueden usar para producir polipéptido. Los entendidos en la técnica también pueden seguir fácilmente métodos conocidos para aislar polipéptidos. Estos incluyen, pero no se limitan a, inmunocromatografía, HPLC, cromatografía de tamaño-exclusión, cromatografía de intercambio iónico y cromatografía de afinidad-afinidad.

25 La expresión de las moléculas de la invención se puede determinar utilizando métodos de rutina conocidos por los entendidos en la técnica. Estos métodos incluyen, pero no se limitan a: amplificación directa de ARN, transcripción inversa de ARN a ADNc, RT-PCR en tiempo real, amplificación de ADNc, hibridación, y métodos de ensayo de base inmunológica, que incluyen, pero no se limitan a inmunohistoquímica, ensayo de captura de emparejado de anticuerpos, ELISA, y Ensayo de inmunospot ligado a enzimas (ensayo de ELISA). Por ejemplo, la determinación de la presencia del nivel de moléculas de ácido nucleico de la invención la invención en un sujeto o tejido puede llevarse a cabo mediante cualquier ensayo de determinación de ácido nucleico estándar, incluyendo la reacción en cadena de la polimerasa, o ensayando con sondas de hibridación marcadas. Tales métodos de hibridación incluyen, pero no se limitan a técnicas de micromatrices.

30 La presente divulgación por lo tanto involucra en un aspecto enzimas, genes que codifican aquellas enzimas, modificaciones funcionales y variantes de lo anterior, así como usos relacionados con las mismas. Los homólogos y alelos de los ácidos nucleicos de la invención pueden ser identificados por técnicas convencionales. También están abarcados por la invención los ácidos nucleicos que se hibridan bajo condiciones rigurosas a los ácidos nucleicos descritos en la presente. El término "condiciones estrictas", como se utiliza en la presente, se refiere a parámetros con los cuales la técnica es familiar. Se pueden encontrar parámetros de hibridación de ácidos nucleicos en referencias que compilan tales métodos, por ejemplo, Molecular Cloning: A Laboratory Manual, J Sambrook et al. eds. Segunda Edición, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 1989, o Current Protocols in Molecular Biology, FM Ausubel, et al. eds. John Wiley & Sons, Inc. New York. Más específicamente, condiciones severas, como se utiliza en la presente, se refiere, por ejemplo, a la hibridación a 65° C en tampón de hibridación (3,5 x SSC, 0,02% de Ficoll, 0,02% de polivinilpirrolidona, 0,02% de Albúmina de Suero Bovino, 2,5 mM de NaH<sub>2</sub>PO<sub>4</sub> (pH 7), SDS al 0,5%, EDTA 2 mM). El SSC es cloruro sódico 0,15 M/citrato sódico 0,015 M, pH 7; SDS es dodecilsulfato sódico; y EDTA es ácido etilendiaminotetraacético. Después de la hibridación, la membrana sobre la cual se transfiere el ADN se lava, por ejemplo, en 2 x SSC a temperatura ambiente y a continuación a 0,1-0,5 x SSC/0,1 x SDS a temperaturas de hasta 68° C.

Existen otras condiciones, reactivos, etc. que pueden utilizarse, lo que da como resultado un grado de rigurosidad similar. El entendido en la materia estará familiarizado con dichas condiciones, y por lo tanto no se dan aquí. Se entenderá, sin embargo, que el entendido en la materia será capaz de manipular las condiciones de una manera que permita al usuario la identificación clara de homólogos y alelos de ácidos nucleicos de la invención (por ejemplo, utilizando condiciones de rigurosidad inferior). El entendido en la técnica también está familiarizado con la metodología para seleccionar células y bibliotecas para la expresión de tales moléculas que luego se aíslan de forma rutinaria, seguido por aislamiento de la molécula de ácido nucleico pertinente y secuenciación.

En general, los homólogos y alelos comparten típicamente al menos 75% de identidad de nucleótidos y/o al menos un 80% de identidad de aminoácidos con las secuencias de ácidos nucleicos y polipéptidos, respectivamente, en algunos casos, comparten al menos 90% de identidad de nucleótidos y/o al menos 90 o 95% de identidad de aminoácidos y en todavía otros casos comparten al menos 95% de identidad de nucleótidos y/o al menos 99% de identidad de aminoácidos. En algunas realizaciones, los homólogos y alelos comparten al menos 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, o 99% de identidad de nucleótidos con las secuencias de ácidos nucleicos y/o 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, o 99% de identidad con las secuencias de los polipéptidos.

La homología se puede calcular utilizando diversas herramientas de software públicamente disponibles desarrolladas por NCBI (Bethesda, Maryland) que se puede obtener a través del sitio de internet de NCBI. Las herramientas ejemplares incluyen el software BLAST, también disponible en el sitio de internet NCBI ([www.ncbi.nlm.nih.gov](http://www.ncbi.nlm.nih.gov)). Se pueden obtener alineamientos Pairwise y ClustalW (matriz BLOSUM30) así como el análisis hidropático de Kyte-Doolittle utilizando el software de análisis de la secuencia del MacVector (Oxford Molecular Group). También se contemplan en la presente los complementos de Watson-Crick de los ácidos nucleicos anteriores.

En la detección e identificación de genes, técnicas conocidas por los entendidos en la técnica tales como transferencias Southern, se pueden aplicar transferencias Northern y protocolos de amplificación tales como la reacción en cadena de la polimerasa utilizando cebadores que hibridan con las secuencias presentadas.

La presente divulgación también incluye ácidos nucleicos degenerados que incluyen codones alternativos a los presentes en los materiales nativos. Por ejemplo, los residuos de serina se codifican por los codones TCA, AGT, TCC, TCG, TCT Y AGC. Cada uno de los seis codones es equivalente a los fines de codificar un residuo de serina. De este modo, será evidente para un entendido en la materia que cualquiera de las que se pueden emplear tripletes de nucleótidos que codifican serina para dirigir el aparato de síntesis de proteínas, *in vitro* o *in vivo*, para incorporar un residuo de serina en un polipéptido alargador. De manera similar, los tripletes de secuencia de nucleótidos que codifican para otros residuos de aminoácidos incluyen, pero no se limitan a: CCA, CCC, CCG y CCT (codones de prolina); CGA, CGC, CGG, CGT, AGA y AGG (codones de arginina); ACA, ACC, ACG y ACT (codones de treonina); AAC y AAT (codones de asparagina); y ATA, ATC y ATT (codones de isoleucina). Otros residuos de aminoácidos pueden ser codificados de manera similar por múltiples secuencias de nucleótidos. de esta manera, la invención abarca ácidos nucleicos degenerados que difieren de ácidos nucleicos biológicamente aislados en secuencia de codones debido a la degeneración del código genético. La invención también abarca la optimización del codón para adaptarse al uso de codones óptimo de una célula huésped.

La presente divulgación también proporciona moléculas de ácido nucleico modificadas que incluyen adiciones, sustituciones y deleciones de uno o más nucleótidos. En realizaciones preferidas, estas moléculas de ácido nucleico modificadas y/o los polipéptidos que codifican para retener al menos una actividad o actividad función de la molécula de ácido nucleico no modificado y/o de los polipéptidos, tales como actividad enzimática. En ciertas realizaciones, las moléculas de ácido nucleico modificadas codifican polipéptidos modificados, de preferencia polipéptidos que tienen sustituciones conservadoras de aminoácidos como se describen en otra parte en la presente. Las moléculas de ácido nucleico modificadas están estructuralmente relacionadas con las moléculas de ácido nucleico no modificadas y en realizaciones preferidas están suficientemente relacionadas estructuralmente con las moléculas de ácido nucleico no modificadas de manera que las moléculas de ácido nucleico modificadas y no modificadas se hibridan bajo condiciones rigurosas conocidas por un entendido en la técnica.

Por ejemplo, se pueden preparar moléculas de ácido nucleico modificadas que codifican polipéptidos que tienen cambios de aminoácidos únicos. Cada una de estas moléculas de ácido nucleico puede tener una, dos o tres sustituciones de nucleótidos exclusivas de cambios de nucleótidos correspondientes a la degeneración del código genético como se describe en la presente memoria. Igualmente, se pueden preparar moléculas de ácido nucleico modificadas que codifican polipéptidos que tienen dos cambios de aminoácidos que tienen, por ejemplo, 2-6 cambios de nucleótidos. Numerosas moléculas de ácido nucleico modificadas como estas serán fácilmente contempladas por un entendido en la técnica, incluyendo, por ejemplo, sustituciones de nucleótidos en codones que codifican los aminoácidos 2 y 3, 2 y 4, 2 y 5, 2 y 6, y así sucesivamente. En el ejemplo anterior, cada combinación de dos aminoácidos se incluye en el conjunto de moléculas de ácido nucleico modificadas, así como todas las sustituciones de nucleótidos que codifican las sustituciones de aminoácidos. Las moléculas de ácido nucleico adicionales que codifican polipéptidos que tienen sustituciones adicionales (es decir, 3 o más), adiciones o deleciones (por ejemplo, mediante la introducción de un codón de parada o un sitio de empalme (s)) también se pueden preparar y se abarcan por la invención como se contempla fácilmente por un entendido en la técnica. Cualquiera de los ácidos nucleicos o



polipéptidos anteriores se puede probar mediante experimentación de rutina para la retención de la relación o actividad estructural a los ácidos nucleicos y/o polipéptidos descritos en la presente.

La presente divulgación abarca variantes de los polipéptidos descritos en la presente. Como se utiliza en la presente, una "variante" de un polipéptido es un polipéptido que contiene una o más modificaciones a la secuencia de aminoácidos primaria del polipéptido. Las modificaciones que crean una variante de enzima pueden hacerse a una enzima, por ejemplo, 1) para alterar la distribución celular de la enzima; 2) para reducir o eliminar una actividad de la enzima; 3) para mejorar una propiedad de una enzima estabilidad de la proteína en un sistema de expresión o la estabilidad de la unión de proteína-proteína; 4) para proporcionar una nueva actividad o propiedad a una enzima, tal como la adición de un epítipo antigénico o adición de una porción detectable; o 5) para proporcionar una unión equivalente o mejor entre una enzima y un sustrato enzimático.

Las modificaciones de un polipéptido se hacen típicamente al ácido nucleico que codifica el polipéptido, y puede incluir supresiones, mutaciones puntuales, truncaciones, sustituciones de aminoácidos y adiciones de aminoácidos o restos no-aminoácidos. Alternativamente, se pueden hacer modificaciones directamente al polipéptido, tal como por escisión, adición de una molécula enlazadora, adición de una porción detectable, tal como biotina, adición de un ácido graso y similares. Las modificaciones también abarcan proteínas de fusión. Un entendido en la técnica estará familiarizado con métodos para predecir el efecto sobre la conformación de proteínas de un cambio en la secuencia de proteínas, y, por lo tanto, "diseñar" un polipéptido variante de acuerdo con métodos conocidos. Un ejemplo de tal método se describe por Dahiyat y Mayo en Science 278: 82-87, 1997, por lo que las proteínas pueden diseñarse *de novo*. El método puede aplicarse a una proteína conocida para variar solamente una porción de la secuencia polipeptídica. Aplicando los métodos de cálculo de Dahiyat y Mayo, las variantes específicas de un polipéptido se pueden proponer y examinar para determinar si la variante retiene una conformación deseada.

En general, las variantes incluyen polipéptidos que se modifican específicamente para alterar una característica del polipéptido no relacionada con su actividad fisiológica deseada. Por ejemplo, los residuos de cisteína pueden ser sustituidos o suprimidos para evitar enlaces disulfuro no deseados. de manera similar, ciertos aminoácidos pueden ser cambiados para mejorar la expresión de un polipéptido eliminando la proteólisis por proteasas en un sistema de expresión (por ejemplo, residuos de aminoácidos dibásicos en sistemas de expresión de levaduras en los que está presente la actividad de la proteasa KEX2).

Las mutaciones de un ácido nucleico que codifican un polipéptido para preservar preferentemente el marco de lectura de aminoácidos de la secuencia codificante, y, preferentemente, no crean regiones en el ácido nucleico que probablemente se hibridan para formar estructuras secundarias, tales como púas o bucles, pueden ser perjudiciales para la expresión del polipéptido variante.

Las mutaciones se pueden realizar mediante la selección de una sustitución de aminoácido, o mediante mutagénesis aleatoria de un sitio seleccionado en un ácido nucleico que codifica el polipéptido. Los polipéptidos variantes se expresan entonces y se analizan para una o más actividades para determinar qué mutación proporciona un polipéptido variante con las propiedades deseadas. Otras mutaciones pueden hacerse a variantes (o a polipéptidos no-variantes) que son silenciosas en cuanto a la secuencia de aminoácidos del polipéptido, pero que proporcionan codones preferidos para la traducción en un huésped particular. Los codones preferidos para la traducción de un ácido nucleico en, por ejemplo, *E. coli*, son bien conocidos por los entendidos en la técnica. Todavía otras mutaciones pueden hacerse a las secuencias no codificadoras de un gen o clon de ADNc para mejorar la expresión del polipéptido. La actividad de variantes de polipéptidos puede ser ensayada mediante la clonación del gen que codifica el polipéptido variante en un vector de expresión bacteriano o mamífero, introducir el vector en una célula huésped apropiada, que expresa el polipéptido variante, y la prueba de una capacidad funcional de los polipéptidos como se describe en la presente.

Los entendidos en la materia también se darán cuenta de que se pueden hacer sustituciones conservadoras de aminoácidos en polipéptidos para proporcionar variantes funcionalmente equivalentes de los polipéptidos anteriores, es decir, las variantes retienen las capacidades funcionales de los polipéptidos. Como se utiliza en la presente, una "sustitución conservativa de aminoácidos" se refiere a una sustitución de aminoácido que no altera la carga relativa del aminoácido o características de tamaño de la proteína en la que se hace la sustitución de aminoácidos. Variantes pueden prepararse de acuerdo con métodos para alterar la secuencia polipeptídica conocida por un entendido en la técnica, tales como las que se encuentran en las referencias que compilan tales métodos, por ejemplo, Molecular Cloning: A Laboratory Manual, J Sambrook et al. eds. Segunda Edición, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 1989, o Current Protocols in Molecular Biology, F.M. Ausubel, et al. eds. John Wiley & Sons, Inc. New York. Variantes funcionalmente equivalentes ejemplares de los polipéptidos incluyen sustituciones conservadoras de aminoácidos en las secuencias de aminoácidos de proteínas descritas en la presente. Sustituciones conservadoras de aminoácidos incluyen sustituciones hechas entre aminoácidos dentro de los siguientes grupos: (a) M, I, L, V; (b) F, Y, W; (c) K, R, H; (d) A, G; (e) S, T; (f) Q, N; y (g) E, D.

En general, se prefiere que menos de todos los aminoácidos se cambien cuando se preparan polipéptidos variantes. En donde se sabe que los residuos de aminoácidos particulares confieren función, dichos aminoácidos no serán reemplazados, o alternativamente, se sustituirán por sustituciones conservadoras de aminoácidos. Preferiblemente, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 residuos pueden cambiarse al preparar polipéptidos

variantes. Se prefiere en general que se realice el número menor de sustituciones. de esta manera, un método para generar polipéptidos variantes es sustituir todos los otros aminoácidos para un aminoácido individual particular, después la actividad de ensayo de la variante, luego repetir el proceso con uno o más de los polipéptidos que tienen la mejor actividad.

- 5 Las sustituciones de aminoácidos conservadoras en la secuencia de aminoácidos de los polipéptidos para producir variantes funcionalmente equivalentes de polipéptidos que se realizan típicamente por alteración de un ácido nucleico que codifica un polipéptido. Tales sustituciones se pueden hacer por una variedad de métodos conocidos por un entendido en la técnica. Por ejemplo, se pueden hacer sustituciones de aminoácidos por mutación dirigida por PCR, mutagénesis dirigida al sitio de acuerdo con el método de Kunkel (Kunkel, Proc. Nat. Acad. Sci. Estados Unidos  
10 (1985) 82: 488-492), o mediante síntesis química de un gen que codifica un polipéptido.

Vectores y construcciones de expresión para uso en la presente invención

- Los vectores útiles para la transformación de un fragmento de ADN aislado que codifica una proteína de fusión de la presente invención en células huésped adecuadas son bien conocidos en la técnica. Como se utiliza en la presente, un "vector" puede ser cualquiera de un número de ácidos nucleicos en los cuales una secuencia o secuencia deseada  
15 las secuencias se pueden insertar por restricción y ligación para el transporte entre diferentes entornos genéticos o para la expresión en una célula huésped. Los vectores se componen típicamente de ADN, aunque también están disponibles vectores de ARN. Vectores que incluyen, pero no se limitan a: plásmidos, fosmidos, fagémidos, genomas de virus y cromosomas artificiales. Típicamente el vector contiene secuencias que dirigen la transcripción y traducción del gen relevante, marcador seleccionable, y secuencias que permiten la replicación autónoma o la integración cromosómica. Los vectores adecuados comprenden una región 5' del gen que alberga los controles de iniciación de la transcripción y una región 3' del fragmento de ADN que controla la terminación de la transcripción. También se  
20 pueden usar vectores que promueven la integración del gen quimérico que codifica una proteína de fusión de la invención en el genoma de la célula huésped. Tales vectores pueden ser para la integración aleatoria, la integración dirigida al sitio, o para la recombinación homóloga. Un vector puede tener características que permiten diferentes tipos de recombinación homóloga o doble-cruzamiento de recombinación homóloga. Se pueden integrar una o  
25 múltiples copias en un genoma de células huésped.

- Un vector de clonación es uno que es capaz de replicarse autónomamente o integrarse en el genoma en una célula huésped, y que además se caracteriza por uno o más sitios de restricción de endonucleasa en los cuales el vector se puede cortar de una manera determinable y en el cual una secuencia de ADN Deseada puede ligarse de manera que  
30 el nuevo vector recombinante retenga su capacidad de replicarse en la célula huésped. En el caso de plásmidos, la replicación de la secuencia deseada puede ocurrir muchas veces al aumentar el plásmido en número de copias dentro de la bacteria huésped o solo un solo tiempo por huésped antes de que el huésped reproduzca por mitosis. En el caso de fagos, la replicación puede ocurrir activamente durante una fase lítica o pasivamente durante una fase lisogénica.

- Un vector de expresión es aquel en el cual se puede insertar una secuencia de ADN deseada por restricción y ligación de manera que se une operablemente a secuencias reguladoras y se puede expresar como un transcrito de ARN. Los vectores pueden contener además una o más secuencias marcadoras adecuadas para utilizarse en la  
35 identificación de células que tienen o no han sido transformadas o transfectadas con el vector. Marcadores incluyen, por ejemplo, genes que codifican proteínas que aumentan o disminuyen la resistencia o sensibilidad a antibióticos u otros compuestos, genes que codifican enzimas cuyas actividades son detectables por ensayos estándar conocidos en la técnica (por ejemplo,  $\beta$ -galactosidasa, luciferasa o fosfatasa alcalina) y genes que afectan visiblemente al fenotipo de células transformadas o transfectadas, huéspedes, colonias o placas (por ejemplo, proteína verde fluorescente). Los vectores preferidos son aquellos capaces de replicación autónoma y expresión de los productos  
40 génicos estructurales presentes en los segmentos de ADN al que están unidos operativamente.

- Como se utiliza en la presente, una secuencia de codificación y secuencias reguladoras se dice que están "operablemente" unidas cuando están covalentemente enlazadas en tal VT a fin de colocar la expresión o transcripción de la secuencia codificante bajo la influencia o control de las secuencias reguladoras. Si se desea que las secuencias codificantes se traduzcan en una proteína funcional, se dice que dos secuencias de ADN se unen operablemente si la inducción de un promotor en las secuencias reguladoras 5' da como resultado la transcripción de la secuencia codificante y si la naturaleza del enlace entre las dos secuencias de ADN no lo hace (1) resultado en la introducción de una mutación de marco-desplazamiento, (2) interfiere con la capacidad de la región promotora para  
45 dirigir la transcripción de las secuencias codificantes o (3) interfiere con la capacidad del transcrito de ARN correspondiente que se va a traducir en una proteína. de esta manera, una región promotora estaría unida operativamente a una secuencia de codificación si la región promotora es capaz de efectuar la transcripción de esa  
50 secuencia de ADN de manera que el transcrito resultante pueda traducirse en la proteína o polipéptido deseado.

- Cuando la molécula de ácido nucleico que codifica cualquiera de las enzimas de la invención reivindicada se expresa en una célula, una variedad de secuencias de control de transcripción (por ejemplo, secuencias promotoras/potenciadoras) se pueden utilizar para dirigir su expresión. El promotor puede ser un promotor nativo, es decir, el promotor del gen en su contexto endógeno, que proporciona una regulación normal de la expresión del gen.  
60 En algunas realizaciones, el promotor puede ser constitutivo, es decir, el promotor no es regulado permitiendo la

transcripción continua de su gen asociado. También se pueden utilizar una variedad de promotores condicionales, tales como promotores controlados por la presencia o ausencia de una molécula.

La naturaleza precisa de las secuencias reguladoras necesarias para la expresión del gen puede variar entre especies o tipos de células, pero, en general, incluirá, según sea necesario, secuencias no transcritas 5' y no traducidas 5' implicadas en la iniciación de la transcripción y traducción, respectivamente, tal como una caja TATA, una secuencia de remate, una secuencia CAAT Y similares. En particular, dichas secuencias reguladoras no transcritas 5' incluirán una región promotora que incluye una secuencia promotora para el control transcripcional del gen unido operativamente. Las secuencias reguladoras también pueden incluir secuencias potenciadoras o secuencias activadoras corriente arriba según se desee. Los vectores de la invención pueden incluir opcionalmente secuencias líderes o de señal 5'. La elección y diseño de un vector apropiado se encuentra dentro de la capacidad y discreción de un entendido en la técnica.

Los vectores de expresión que contienen todos los elementos necesarios para la expresión están disponibles comercialmente y son conocidos por los entendidos en la técnica. Véase, por ejemplo, Sambrook et al. *Molecular Cloning: Laboratory Manual*, Segunda Edición, Cold Spring Harbor Laboratory Press, 1989. Las células se modifican genéticamente mediante la introducción en las células de ADN (o ARN) heterólogo. ADN (o ARN heterólogo) se coloca bajo el control operable de los elementos de transcripción para permitir la expresión del ADN heterólogo en la célula huésped. En algunas realizaciones, dos o más de los ácidos nucleicos de la invención pueden clonarse en el mismo vector de expresión o plásmido.

Los métodos de la invención pueden hacer uso de la expresión constitutiva o regulada de diversas secuencias de codificación. La expresión puede ser regulada por diversas indicaciones, por ejemplo, por inducción por productos químicos, cambio de fase de crecimiento, agotamiento de un nutriente, cambios de temperatura y/o luz. En algunas realizaciones, los promotores inducibles se regulan por la presencia de un agente inductor, por ejemplo, un producto químico tal como lactosa, arabinosa o tetraciclina, como se conoce en la técnica. Típicamente en donde se indica la expresión de " alto nivel ", la concentración de la proteína expresada en la célula es al menos aproximadamente 2 veces por encima de los niveles basales; al menos aproximadamente 10 veces por encima de los niveles basales; al menos aproximadamente 25 veces por encima de los niveles basales; al menos aproximadamente 50 veces por encima de los niveles basales; o más, por ejemplo, entre aproximadamente 2 veces hasta aproximadamente 100 veces por encima de los niveles basales.

Los vectores de expresión y clonación usualmente contienen un promotor que es reconocido por el organismo huésped y está operablemente enlazado a la secuencia codificante de interés. Los promotores son secuencias no traducidas localizadas corriente arriba (5') al codón de inicio de un gen estructural que controla la transcripción de una secuencia de ácido nucleico particular a la que están operablemente enlazados. Tales promotores caen típicamente en dos clases: inducible y constitutivo. Los promotores inducibles son promotores que inician niveles incrementados de transcripción a partir del ADN bajo su control en respuesta a algún cambio en las condiciones de cultivo, por ejemplo, Presencia o ausencia de un nutriente o un cambio de temperatura. En este momento se conocen bien un gran número de promotores reconocidos por una variedad de células huésped potenciales, por ejemplo, para *E. coli* véase, por ejemplo, Hawley y McClure *Nucleic Acids Res* (1983) 11: 2237-55; para *B. subtilis* véase, por ejemplo, Ishii et al. *Nucleic Acids Res* (2001) 29: 278-280; *Saccharomyces cerevisiae* véase, por ejemplo, Chang et al., *Nucleic Acids Res.* (2011) 39:D647-52. Véase también Madigan, Martinko, and Parker, eds., *Brock Biology of Microorganisms*. 9a Ed. Prentice Hall. Upper Saddle River, NJ. Así como se puede usar el promotor nativo, para la mayoría de los fines se prefieren los promotores heterólogos, ya que generalmente permiten una mayor transcripción y mayores rendimientos.

Promotores adecuados para uso con huéspedes procarióticos incluyen los sistemas promotores de la beta-lactamasa y lactosa, fosfatasa alcalina, un triptófano (*trp*) sistema promotor y numerosos promotores híbridos tales como el promotor *lac*. Sin embargo, también son adecuados otros promotores bacterianos o bacteriófagos conocidos, por ejemplo, El promotor *lacI*, el promotor *lacZ*, el promotor T3, el promotor T7, el promotor arabinosa, el promotor *gpt*, el promotor *lambda PR*, el promotor *lambda PL*, promotores de operones que codifican enzimas glicolíticas tales como 3-fosfoglicerato quinasa (PGK) y el promotor de fosfatasa ácida. Se han publicado sus secuencias de nucleótidos, permitiendo de esta manera un entendido en la materia los ligarlos operablemente a una secuencia de interés utilizando enlazadores o adaptadores. Promotores para uso en sistemas bacterianos también contendrán un Shine-Dalgarno (S.D.) secuencia enlazada operativamente a la secuencia codificante (véase, por ejemplo, Shine y Dalgarno, *Nature* (1975) 254: 34-8; Madigan, Martinko y Parker, eds. *Brock Biology of Microorganisms*. 9ª ed. Prentice Hall. Upper Saddle River, NJ). En ciertos casos, también, la célula huésped puede ser modificada genéticamente para ajustar las concentraciones de las proteínas del transportador de metabolito o inductor de manera que todas las células en un cultivo serán inducidas equivalentemente.

También se conocen en la técnica promotores adecuados para células eucarióticas, por ejemplo, células de levadura. Virtualmente todos los genes eucarióticos tienen una región rica en AT localizada aproximadamente de 25 a 30 bases corriente arriba del sitio en donde se inicia la transcripción. Otra secuencia encontrada de 70 a 80 bases corriente arriba del inicio de la transcripción de muchos genes es una región CXCAAT donde X puede ser cualquier nucleótido. En el extremo 3' de la mayoría de los genes eucariotas se encuentra una secuencia AATAAA que puede ser la señal para la adición de la cola poli A al extremo 3' de la secuencia codificante. Todas estas secuencias se insertan

adecuadamente en vectores de expresión eucariotas. Los ejemplos de secuencias promotoras adecuadas para uso con huéspedes de levadura incluyen los promotores para la 3-fosfogliceratasa u otras enzimas glicolíticas, tales como enolasa, gliceraldehído-3-fosfato deshidrogenasa, hexoquinasa, piruvato descarboxilasa, fosfofructoquinasa, glucosa-6-fosfato isomerasa, 3-fosfoglicerato mutasa, piruvato quinasa, triosefosfato isomerasa, fosfoglucosa isomerasa y glucoquinasa.

Otros promotores de levadura, que son promotores inducibles que tienen la ventaja adicional de la transcripción controlada por condiciones de crecimiento, son las regiones promotoras del alcohol deshidrogenasa 2, isocitocromo C, fosfatasa ácida, enzimas degradativas asociadas con el metabolismo de nitrógeno, metalotioneína y enzimas responsables de la utilización de maltosa y galactosa. Los mejoradores de la levadura también se usan ventajosamente con promotores de levadura.

Puede ser deseable ajustar experimentalmente la velocidad de expresión para optimizar la eficiencia de exportación. Una mala translocación puede dar como resultado una capacidad insuficiente de maquinaria de exportación. Los métodos para el ajuste de la velocidad de expresión incluyen, sin limitación, modificación del número de copias del plásmido que lleva el gen que codifica para la proteína a ser exportada al periplasma. Replicones conocidos y utilizados en la técnica incluyen P15A (10 copias/célula), ColA (30 copias/célula), ColEI (40 copias/célula) y RSF1030 (> 100 copias/célula). El sitio de unión al ribosoma en la 5'UTR del gen que codifica para la proteína a ser exportada al periplasma puede ser modificado, en donde se puede crear y probar una biblioteca de sitios de unión a ribosoma con resistencias variables; ver, por ejemplo, Salis et al, Nature Biotechnology (2009) 27: 946-950; y Simmons et al. Nature Biotechnology (1996) 14: 629-634, cada uno de los cuales se incorpora aquí por referencia. La región promotora corriente arriba del gen que codifica la proteína a ser exportada puede ser modificada para ajustar la velocidad de transcripción, en donde se puede crear y probar una biblioteca de regiones promotoras con resistencias variables; ver, por ejemplo, Alper et al. PNAS (2005) 102: 12678-12683; y De Mey et al. BMC Biotechnology (2007) 7: 34.

#### Flujo metabólico

"Flujo" o "flujo metabólico" se refiere a la velocidad de que las moléculas pasan a través de una ruta o reacción de interés. Entre los factores que el flujo de control es la velocidad de catálisis de enzimas en la ruta, la disponibilidad de sustrato, la concentración de enzimas en una célula y/o la proximidad de enzimas en una ruta.

En tanto es deseable una alta velocidad de flujo a través de una ruta de interés, al mismo tiempo puede crear problemas de toxicidad si un producto que no se acumula normalmente a altos niveles en la célula se produce a una alta velocidad con relación a la que se produce en condiciones normales. Se entiende que una alta velocidad de flujo es específica de la ruta, y se refiere a la concentración del producto de ruta a lo largo del tiempo, tal como, por ejemplo, la producción de un producto a una velocidad de aproximadamente 0,1 a aproximadamente 20 gramos de producto/L/h

Una célula sometida a esfuerzo produce un número de proteínas indeseables para mantener biocatálisis activa, tales como nucleasas, proteínas de choque térmico, proteasas y similares.

Los métodos de la invención proporcionan un medio para controlar el flujo a través de una ruta, de tal manera que una célula sana (por ejemplo, con fisiología sustancialmente normal) se puede cultivar a alta densidad, por ejemplo, de aproximadamente 30 a aproximadamente 300 OD<sub>550</sub> durante cuyo período de tiempo se incrementa la concentración de enzimas implicadas en una ruta deseada sin dar como resultado un efecto perjudicial (para la salud de las células) aumento en el flujo de ruta o acumulación tóxica de productos metabólicos. OD<sub>550</sub> se refiere a la densidad óptica a 550 nm, en la que 1 OD<sub>550</sub> es aproximadamente 10<sup>9</sup> células/mL (*E. coli*).

Los métodos para determinar las velocidades de flujo son conocidos y usados en la técnica; véase, por ejemplo, Wiechert et al, Metab. Eng. (2001) 3: 265-283, y Wiechert et al. Metab. Eng. (2001) 3: 195-206; y textos de ingeniería metabólicos tales como Lee y Papoutsakis, Eds., Metabolic Engineering, Marcel Dekker, Inc. New York (1999); Stephanopoulos, Nielsen, y Aristidou, Eds., Metabolic Engineering: Principles and Methodology, Academic Press, New York (1998); Nielsen y Eggeling, Eds., Metabolic Engineering, Springer, London (2001), cada una de las cuales se incorpora en la presente por referencia. El flujo se puede calcular a partir de cantidades medibles utilizando técnicas tales como el análisis flujo metabólico (MFA), por ejemplo, por medición directa de la velocidad de conversión del sustrato isotópicamente marcado.

#### Rutas de Interés

Como se utiliza en la presente, el término "ruta de enzima" o "ruta de interés" se refiere a un sistema celular para convertir un sustrato a un producto de interés, donde el sistema comprende una pluralidad de enzimas y puede comprender adicionalmente sustratos accionados por una o más de las enzimas, productos de las reacciones catalizadas por enzimas, co-factores utilizados por las enzimas y similares. El sistema puede estar presente en una célula intacta, o en un lisado de una célula. Se conocen muchas rutas metabólicas y se han descrito en sistemas microbianos, y son accesibles en bases de datos públicas; véase, por ejemplo, Smolke, Ed., The Metabolic Pathway Engineering Handbook: Tools and Applications, CRC Press, New York (2009); Stephanopoulos, Nielsen, and Aristidou, Eds., Metabolic Engineering: Principles and Methodology, Academic Press, New York (1998); Greenberg,

Metabolic Pathways: Energetics, Tricarboxylic Acid Cycle, and Carbohydrates, Academic Press, New York (1967); la colección de D. M. Greenberg titulada Metabolic pathways, volúmenes 1-7.

Las rutas de interés incluyen, por ejemplo, rutas involucradas en carbohidratos, aminoácido, ácido nucleico, esteroide, ácido graso, y la biosíntesis del producto natural, y abarca la síntesis de diversos compuestos y materiales químicos, que incluyen, pero no se limitan a:

- 5 a) antibióticos; por ejemplo, actinomicina, bleomicina, rifamicina, cloranfenicol, tetraciclina, lincomicina, eritromicina, estreptomicina, ciclohexida, puromicina, cicloserina, bacitracina, penicilina, cefalosporina, vancomicina, polimixina y gramicidina;
- b) biotensioactivos; por ejemplo, ramnolípidos, soforolípidos, glicolípidos y lipopéptidos;
- 10 c) combustibles biológicos; por ejemplo, bioetanol, biodiesel y biobutanol;
- d) aminoácidos; por ejemplo, L-glutamato, L-lisina, L-fenilalanina, ácido L-aspártico, L-isoleucina, L-valina, L-triptófano, L-prolina (hidroxiprolina), L-treonina, L-metionina y D-p-hidroxifenilglicina;
- e) ácidos orgánicos; por ejemplo, ácido cítrico, ácido láctico, ácido glucónico, ácido acético, ácido propiónico, ácido succínico, ácido fumárico y ácido itacónico;
- 15 f) ácidos grasos; por ejemplo, ácido araquidónico, ácido graso poliinsaturado (PUBA) y ácido gamma-linoleico;
- g) alcoholes y polioles, por ejemplo, glicerol, manitol, eritritol, xilitol, poli-3-hidroxibutirato, isobutanol y 1-butanol;
- h) sabores y fragancias; por ejemplo, vainillina, benzaldehído, dihidroxi-acetona, 4-(R)-decanolida y 2-actil-1-pirrolida;
- i) nucleótidos; por ejemplo, ácido 5'-guanílico y ácido 5'-inosínico;
- 20 j) vitaminas; por ejemplo, vitamina C, vitamina F, vitamina B2, provitamina D2, vitamina B12, ácido fólico, nicotinamida, biotina, ácido 2-ceto-L-gulónico y provitamina Q10;
- k) pigmentos; por ejemplo, astaxafina,  $\beta$ -caroteno, leucoponeno, montantubrina, y rubristatina;
- l) azúcares y polisacáridos; por ejemplo, ribosa, sorbosa, xantano, gelán y dextrano; y
- m) biopolímeros y plásticos; por ejemplo, polihidroxialcanoatos (PHA), ácido poli-gamma-glutámico y 1,3-propanodiol.

Otros ejemplos de rutas de interés incluyen la síntesis de diversos metabolitos de E coli. Un metabolito es cualquier sustancia usada o producida durante el metabolismo (es decir, una enzima, sustrato o producto). Para los fines de la presente invención, a menudo se trata de un metabolito, aunque no siempre, el producto de una enzima en la ruta de interés. Los metabolitos de E coli ejemplares incluyen, pero no se limitan a, ácido 2,3-dihidroxibenzóico, 2-cetoglutarato, 3-fosfoglicerato, 4-hidroxibenzoato, 6-fosfogluconato, acetoacetil-CoA, acetil-CoA, acetilfosfato, adenina, adenosina, fosfosulfato de adenosina, ADP, ADP-glucosa, alanina, AMP, antranilato, arginina, asparagina, aspartato, ATP, carbamilaspartato, cis-aconitato, citrato, citrulina, CMP, coenzima A, CTP, AMP cíclico, citidina, citosina, dAMP, dATP, dCTP, desoxiadenosina, desoxiguanosina deoxirribosa-5-P, dGMP, dihidroorotato, dihidroxiacetona-fosfato, dTDP, dTTP, etirrosa-4-fosfato, FAD, flavina-mononucleótido, fructosa-1,6-bisfosfonato, fructosa-6-fosfato, fumarato, GDP, gluconato, gluconolactona, glucosamina-6-fosfato, glucosa-6-fosfato, glucosa-1-fosfato, glutamato, glutamina, glutatión, disulfuro de glutatión, gliceraldehído-3-fosfato, glicerato, 35 glicerol-3-fosfato, GMP, GTP, guanina, guanosina, histidina, histidinol, homocisteína, difosfato de inosina, monofosfato de inosina, trifosfato de inosina, isoleucina, lisina, malato, malonil-CoA, metionina, mio-inositol, N-glucosamina-1P, N-Acetil-ornitina, NAD +, NADH, NADP +, NADPH, ornitina, oxaloacetato, fenilalanina, fenilpiruvato, fosfoenolpiruvato, prolina, propionil-CoA, PRPP, Piruvato, quinolinato, riboflavina, ribosa-5-fosfato, ribulosa-5-fosfato, S-adenosil-L-metionina, serina, ácido shikímico, succinato, succinil-CoA, treonina, triptófano, 40 tirosina, UDP, UDP-glucosa, UDP-glucuronato, UDP-N-acetilglucosamina, uridina, UTP, valina y xilulosa-5-fosfato.

En ciertas realizaciones, la ruta de interés proporciona la síntesis de ácido shikímico y/o shikimato (shikimato es la forma aniónica del ácido shikímico) e intermedios sintéticos A los mismos (por ejemplo, como se proporciona en la Figura 4), un isoprenoide o terpeno (por ejemplo, amorfadieno, farneeno, licopeno, astaxantina, vitamina A, mentol, Beta-caroteno), poli-3-hidroxibutirato, isobutanol y 1-butanol (véase, por ejemplo, los Ejemplos 1-5 y las Figuras 4 y 5, 45 proporcionados en la presente memoria).

Un número de reacciones pueden ser catalizadas por enzimas en una ruta de interés. Clases amplias de enzimas, que pueden ser identificadas por el número de clasificación de enzimas, entre paréntesis, se incluyen, pero no se limitan a:

- 50 (EC 1) oxidorreductasas; por ejemplo, deshidrogenasas, oxidasas, reductasas, oxidorreductasas, sinasas, oxigenasas, monooxigenasas, dioxigenasas, lipoxigenasas, hidrogenasas, transhidrogenasas, peroxidasas, catalasas, epoxidasas, hidroxilasas, desmetilasas, desaturasas, dismutasas, hidroxitransferasas, deshalogenasas y desyodasas;

(EC2) transferasas; por ejemplo, transaminasas, quinasas, diquinasas, metiltransferasas, hidroximetiltransferasas, formyl-transferasas, forminotransferasas, carboxitransferasas, carbamoiltransferasas, amidino-transferasas, transaldolasas, transcetolasas, acetiltransferasas, aciltransferasas-palmitoiltransferasas, succiniltransferasas, maloniltransferasas, galetransferasas transferasas, sinapossiltransferasas, tiglioilo transferasas, 5 tetradecanosiltransferasas, hidroxicinamoiltransferasas, feruloiltransferasas, micoly transferasas, benzoiltransferasas, piperiltransferasas, trimetiltridecanoiltransferasa, miristoiltransferasas, cumarinas-transferasas, tiolasas, aminoaciltransferasas, fosforilasas, hexosyl-transferasas, pentasiltransferasas, sialiltransferasas, piridinilasas, difosforilasas, ciclitransferasas sulfurilasas, adenosinas, carboxiviniltransferasas, isopentosinas, aminocarboxipropy 10 transferasas, dimetilaliltransferasas, farnesiltransferasas, hexapreniltransferasas, desfosforilación de pentapreniltransferasas, pentapreniltransferasas, transferasas de nonaprenilo, geraniilgeraniiltransferasas, aminocarboxipropiltransferasas, oximino-transferasas, purinetransferasas, fosfodimutasas, fosfotransferasas, nucleotidtransferasas, polimerasas, colinefosfotransferasas, fosforilmutasas, sulfurotransferasas, sulfo-transferasas y CoA-transferasas;

(EC3) hidrolasas; por ejemplo, lipasas, esterases, amilasas, peptidasas, hidrolasas, lactonasas, deacilasas, 15 desacetilasas, pheoporidasas, depolimerasas, tiolesterasas, fosfatasas, difosfatasas, trifosfatasas, nucleotidasas, fitasas, fosfodiesterasas, fosfolipasas, sulfatasas, ciclasas, peptidasas, ribonucleasas, exonucleasas, endonucleasas, glicosidasas, nucleoidasas, glicosilasas, aminopeptidasas, dipeptidasas, carboxipeptidasas, metalocarboxipeptidasas, omega-peptidasas, endopeptidasas de serina, endopeptidasas de cisteína, endopeptidasas aspárticas, metaloendopeptidasas, endopeptidasas de treonina, aminasas, amidasas, desuccinilasas, desformasas, acilasas, 20 deimasas, desaminasas, dihidroasas, ciclohidrolasas nitradas, ATPasas, GTPasas, halidasas, deshalogenasas y sulfohidrolasas;

(EC 4) liasas; por ejemplo, descarboxilasas, carboxilasas, carboxiquinasas, aldolasas, epoxidasas, oxoácido-liasas, liasas de carbono-carbono, deshidratasas, hidratasas, sinasas, endolasas, exolasas, amoniaco-liasas, amidina-liasas, 25 aminas-liasas, liasas de carbono-azufre, liasas de carbono-haluro, liasas de fósforo-oxígeno y deshidroclorasas;

(EC 5) isomerasas, por ejemplo, isomerasas, racemasas, mutasas, tautomerasas, fosfomutasas, fosfoglucomutasas, aminomutasas, cicloisomerasa, ciclasas, topoisomerasas; y

(EC 6) ligasas; por ejemplo, sintetasas, tARN-ligasas, ácido-tiol ligasas, amidasas amida, peptidil-sinasas, 30 cicloligasas, carboxilasas, ADN-ligasas, ARN-ligasas y ciclasas.

Las clases más específicas de enzimas incluyen, sin limitación, sub-clases de oxidorreductasas, transferasas, liasas, isomerasas y ligasas, como se proporciona a continuación.

Las oxidorreductasas ejemplares incluyen, pero no se limitan a:

(EC 1.1) oxidorreductasas que actúan sobre el grupo CH-OH de donadores, y un aceptor;

(EC 1.2) oxidorreductasas de que actúan sobre el grupo aldehído u oxo de donadores, y un aceptor;

(EC 1.3) oxidorreductasas que actúan sobre el grupo CH-CH de donadores, y un aceptor;

35 (EC 1.4) oxidorreductasas que actúan sobre el grupo CH-NH<sub>2</sub> de donadores, y un aceptor;

(EC 1.5) oxidorreductasas que actúan sobre el grupo CH-NH de donadores, y un aceptor;

(EC 1.6) oxidorreductasas que actúan sobre NADH o NADPH, y un aceptor;

(EC 1.7) oxidorreductasas que actúan sobre otros compuestos nitrogenados como donadores, y un aceptor;

(EC 1.8) oxidorreductasas que actúan sobre un grupo azufre de donadores, y un aceptor;

40 (EC 1.9) oxidorreductasas que actúan sobre un grupo hemo de donadores, y un aceptor;

(EC 1.1) oxidorreductasas que actúan sobre difenoles y sustancias relacionadas como donadores, y un aceptor;

(EC 1.11) oxidorreductasas de que actúan sobre un peróxido como aceptor;

(EC 1.12) oxidorreductasas de que actúan sobre hidrógeno como donador, y un aceptor;

45 (EC 1.13) oxidorreductasas que actúan sobre donantes sencillos con incorporación de oxígeno molecular, incorporando uno o dos átomos de oxígeno;

(EC 1.14) oxidorreductasas que actúan sobre donantes apareados, con incorporación o reducción de oxígeno molecular, siendo el donante 2-oxoglutarato, NADH, NADPH, flavina reducida, flavoproteína, pteridina, proteína de hierro-azufre, ascorbato;

(EC 1.15) oxidorreductasas que actúan sobre radicales superóxido como aceptor;

(EC 1.16) oxidorreductasas oxidantes de iones metálicos, y un aceptor;

(EC 1.17) oxidorreductasas que actúan sobre grupos CH o CH<sub>2</sub>, y un aceptor;

(EC 1.18) oxidorreductasas que actúan sobre proteínas de hierro-azufre como donadores, y un aceptor;

(EC 1.19) oxidorreductasas que actúan sobre la flavodoxina reducida como donante, y un aceptor;

5 (EC 1.2) oxidorreductasas que actúan sobre fósforo o arsénico en donadores, y un aceptor; y

(EC 1.21) oxidorreductasas que actúan sobre X-H y Y-H para formar un enlace X-Y, y un aceptor; donde los aceptores para cada categoría donante pueden incluir, sin limitación: NAD, NADP, proteína hemo, oxígeno, disulfuro, quinona, una proteína de hierro-azufre, una flavina, un grupo nitrogenado, un citocromo, dinitrógeno y H<sup>+</sup>.

Las transferasas ejemplares incluyen, pero no se limitan a:

10 (EC 2.1) transferasas que transfieren grupos de un carbono;

(EC 2.2) transferasas que transfieren grupos aldehído o cetónicos;

(EC 2.3) aciltransferasas;

(EC 2.4) glucosiltransferasas;

(EC 2.5) transferasas que transfieren grupos alquilo o arilo, distintos de grupos metilo;

15 (EC 2.6) transferasas que transfieren grupos nitrogenados;

(EC 2.7) transferasas que transfieren grupos que contienen fósforo;

(EC 2.8) transferasas que transfieren grupos que contienen azufre; y

(EC 2.9) transferasas que transfieren grupos que contienen selenio.

Las hidrolasas ejemplares incluyen, pero no se limitan a:

20 (EC 3.1) hidrolasas que actúan sobre enlaces éster;

(EC 3.2) glicosilasas;

(EC 3.3) hidrolasas que actúan sobre enlaces de éter;

(EC 3.4) hidrolasas que actúan sobre enlaces peptídicos (peptidasas);

(EC 3.5) hidrolasas que actúan sobre enlaces de carbono-nitrógeno, distintos de enlaces peptídicos;

25 (EC 3.6) hidrolasas que actúan sobre anhídridos de ácido;

(EC 3.7) hidrolasas que actúan sobre enlaces carbono-carbono;

(EC 3.8) hidrolasas que actúan sobre enlaces haluro;

(EC 3.9) hidrolasas que actúan sobre enlaces de fósforo-nitrógeno;

(EC 3.1) hidrolasas que actúan sobre enlaces de azufre-nitrógeno;

30 (EC 3.11) hidrolasas que actúan sobre enlaces carbono-fósforo;

(EC 3.12) hidrolasas que actúan sobre enlaces de azufre-azufre; y

(EC 3.13) hidrolasas que actúan sobre enlaces carbono-azufre.

Las liasas ejemplares incluyen, pero no se limitan a:

(EC 4.1) liasas de carbono-carbono;

35 (EC 4.2) liasas de Carbono-oxígeno;

(EC 4.3) liasas de Carbono-nitrógeno;

(EC 4.4) liasas de Carbono-azufre;

(EC 4.5) liasas de carbono-haluro de; y

(EC 4.6) liasas de fósforo-oxígeno.

Las isomerasas ejemplares incluyen, pero no se limitan a:

(EC 5.1) racemasas y epimerasas;

(EC 5.2) cis-trans-Isomerasas;

5 (EC 5.3) isomerasas intramoleculares;

(EC 5.4) transferasas Intramoleculares (mutasas); y

(EC 5.5) liasas intramoleculares.

Las ligasas ejemplares incluyen, pero no se limitan a:

(EC 6.1) ligasas que forman enlaces de carbono-oxígeno;

10 (EC 6.2) ligasas que forman enlaces carbono-azufre;

(EC 6.3) ligasas que forman enlaces de carbono-nitrógeno;

(EC-6.4) ligasas que forman enlaces carbono-carbono;

(EC 6.5) ligasas que forman enlaces de éster fosfórico; y

(EC 6.6) ligasas que forman enlaces nitrógeno-metal.

15 Las isozimas (también conocidas como isoenzimas) son enzimas que difieren en la secuencia de aminoácidos, pero catalizan la misma reacción química. En algunos puntos en una ruta de interés, pueden estar presentes dos o más isozimas. Las isozimas pueden presentar diferentes parámetros cinéticos, o diferentes propiedades reguladoras.

Las enzimas involucradas en una ruta de interés o ruta asociada pueden también ser clasificadas de acuerdo con el papel de la enzima. Enzimas de implicación directa (clase 1) en una célula o lisado celular catalizan una reacción en la ruta. Es típico de rutas que tales enzimas directas son una de una cadena, donde un producto de una primera enzima es el sustrato de una segunda enzima, el producto de la segunda enzima es el sustrato de una tercera enzima, y así sucesivamente. Que eventualmente da como resultado el producto de interés. Enzimas de implicación indirecta (clase2) en una célula o un lisado celular que reacciona en una ruta asociada, usualmente en la producción de un sustrato utilizado en la ruta de interés. Puede ser una característica de una enzima en estas dos clases que la sobreproducción ("sobreexpresión") de la enzima es tóxica para la célula, incluso 2 veces, 3 veces o más sobreproducción. Tal toxicidad puede ser el resultado de la sobreproducción de un producto que es tóxico a altas concentraciones, o que la enzima desvía los recursos a una velocidad que impacta la fisiología normal de las células. La expresión de tales enzimas se beneficia de la acumulación selectiva modulada en un compartimiento separado con los métodos de la invención, tal como mediante el uso de un promotor inducible, con el fin de evitar tensiones indeseables en la célula.

Dentro de una ruta, las enzimas variarán en velocidad de renovación y la efectividad con la cual se produce un producto. Como resultado, ciertas enzimas en una ruta se convierten en la velocidad-limitante. Incrementar la concentración de enzimas limitantes de velocidad en una ruta (con relación a enzimas no limitantes de velocidad) permite un flujo incrementado a través de la ruta de interés (véase, por ejemplo, Zamboni et al Nature. Protocols (2009) 4: 878-892). Las enzimas no limitantes con frecuencia se asocian con la toxicidad cuando se produce sobre-producida, y por lo tanto las concentraciones disponibles de tales enzimas se modulan deseablemente por los métodos de la invención para incrementar selectivamente la acumulación de la actividad limitante de la velocidad en un punto de tiempo seleccionado y posiblemente también mientras que se secuestra a un compartimiento separado.

Una tercera clase de enzimas en un lisado celular o celular son enzimas competentes (clase3), que utilizan un sustrato o producto de la ruta de interés. Una característica de una enzima competitiva es que la cinética de la conversión del sustrato es suficientemente alta para que la presencia de la enzima disminuya el rendimiento global y/o la velocidad de producción del producto final deseado catalizado por la ruta de interés. Una célula normal puede requerir la expresión de enzimas competentes, y por lo tanto en lugar de detonar completamente la expresión de enzimas competentes, es deseable disminuir selectivamente la concentración de la enzima; ver, por ejemplo, la publicación PCT No. WO 2010/077806.

Por conveniencia de la denominación, una enzima en la ruta puede ser categorizada como una primera, enzima de entrada de ruta, o una enzima o enzimas posteriores. Por conveniencia, la enzima de entrada de ruta puede ser referida en la presente como E<sub>1</sub>, y las enzimas aguas abajo pueden ser numeradas consecutivamente, E<sub>2</sub>, E<sub>3</sub>, E<sub>n</sub>. Vías de interés para uso en los métodos de la presente invención normalmente comprenderán al menos una enzima, al menos dos enzimas, al menos tres enzimas, al menos cuatro enzimas, o más, por ejemplo, entre 1 y 50 enzimas, entre 1 y 40 enzimas, entre 1 y 30 enzimas, entre 1 y 20 enzimas, entre 1 y 10 enzimas, entre 1 y 5 enzimas, entre 1 y 2 enzimas, entre 2 y 50 enzimas, entre 2 y 40 enzimas, entre 2 y 30 enzimas, entre 2 y 20 enzimas, entre 2 y 10



enzimas Entre 2 y 5 enzimas, entre 2 y 4 enzimas, entre 5 y 50 enzimas, entre 5 y 40 enzimas, entre 5 y 30 enzimas, entre 5 y 20 enzimas, entre 5 y 10 enzimas, entre 5 y 8 enzimas, entre 10 y 50 enzimas, entre 10 y 40 enzimas, entre 10 y 30 enzimas, o entre 10 y 20 enzimas, inclusive.

5 Las enzimas en una ruta pueden ser naturales o modificadas para optimizar una característica particular de interés, por ejemplo, especificidad del sustrato, cinética de reacción, solubilidad y/o insensibilidad a la inhibición de la retroalimentación. Además, en algunos casos, el gen que expresa la enzima se optimizará para el uso de codones dentro de la célula huésped. En algunas realizaciones, la ruta completa comprende enzimas de un solo organismo, sin embargo, no se requiere, y se combinan enzimas de múltiples organismos contemplados. Para algunas finalidades, una ruta puede ser endógena a la célula huésped, pero esto tampoco es necesario, y una ruta o componentes completos de una ruta pueden introducirse en una célula huésped. En donde el sistema se proporciona en una celda intacta, generalmente el conjunto completo de enzimas de la ruta de interés estará presente en la célula. Para fines de producción libre de células, se pueden añadir una o más enzimas al lisado, o alternativamente puede ser producida por el lisado, para completar la ruta.

15 En el sistema de ruta, un primer sustrato ( $S_1$ ) por la enzima de entrada de la ruta, y se convierte en un primer producto, aunque se entenderá por un entendido en la técnica que una enzima puede actuar sobre más de un sustrato simultáneamente, y puede producir más de un producto, de manera que dos o más rutas puedan estar interconectadas en una sola enzima. El primer producto es un sustrato ( $S_2$ ) para la enzima de corriente abajo  $E_2$ , y se convierte en un segundo producto por  $E_2$ . En función de la complejidad de la ruta, el segundo producto puede ser el producto final (PF), o puede ser un sustrato ( $S_3$ ) para una tercera enzima corriente abajo ( $E_3$ ), y se convierte en un tercer producto por  $E_3$ , que puede ser un sustrato ( $S_4$ ) para una cuarta enzima. La enzima final en la ruta, que puede ser  $E_2$ ,  $E_3$ ,  $E_4$ , etc. produce el producto de interés (PF). Una característica de las rutas enzimáticas es que el producto de una enzima es el sustrato para la siguiente enzima. Los productos pueden ser estables o relativamente lábiles, pero en general, el producto final es suficientemente estable para que pueda ser aislado de la célula, lisado celular o mezcla de reacción. Las enzimas competentes utilizan un sustrato o producto de la ruta de interés, que puede incluir cualquiera de PF,  $S_1$ ,  $S_2$ ,  $S_3$  y/o  $S_4$ , y puede referirse como enzimas competentes ( $E_c$ ).

25 En algunas realizaciones de la invención, el sustrato inicial,  $S_1$ , es un metabolito central, o "mercancía" celular. Las rutas centrales de metabolismo incluyen glicólisis y el ciclo de ácido cítrico. Tales compuestos de  $S_1$  no son generalmente específicos para la ruta de interés, pero son compuestos ampliamente encontrados en diversas células y son sustratos para múltiples enzimas y rutas. Los ejemplos de sustratos de mercancía incluyen, sin limitación, glucosa, ATP, piruvato, fosfoenol piruvato, y similares. Una enzima de entrada de ruta,  $E_1$ , puede convertir un sustrato de mercancía a un producto en que es un sustrato selectivo para uno o un número relativamente pequeño de enzimas.

30 En general, una enzima de entrada clave se define como una que realiza la primera etapa comprometida en una ruta a un producto de interés. Esta etapa generalmente involucra el compromiso bioquímico de un compuesto a la ruta de un producto de interés. Los ejemplos de enzimas de entrada clave incluyen, pero no se limitan a, los expuestos en la tabla 2.

Tabla 2: lista Ejemplar de enzimas de ruta de entrada clave

Enzima(s) de entrada clave	Ruta biosintética	Productos ejemplares	Enzima <i>E. coli</i>
amidofosforibosil transferasa	biosíntesis de purina	GMP, GDP, GTP, dGDP, dGTP, AMP, ADP, ATP, dADP, dATP, monofosfato de inosina	PurF
orotato fosforribosiltransferasa	biosíntesis de pirimidina	UMP, UDP, UTP, CDP, CTP	PyrE
2-dehidro-3-deoxifosfoheptonato de aldolasa	biosíntesis de corismato	Shikimate, tirosina, fenilalanina, triptófano	AroE,F,G
fosforribosiltransferasa HisG	biosíntesis de histidina	Histidina	HisG
acetolactato/acetohidroxibutanoato sintasa	biosíntesis de isoeucina, leucina, valina	Isoleucina, leucina, valina	IlvH,M,N
UDP-N-acetilglucosamina aciltransferasa	biosíntesis de lipopolisacáridos	Lípido A disacárido	LpxA

aspartato aminotransferasa	biosíntesis de lisina, treonina y metionina	Lisina, treonina, metionina	AspC
arginina descarboxilasa	biosíntesis de putrescina	Putrescina	SpeA
GTP ciclohidroasa I	biosíntesis de tetrahidrofolato	Tetrahidrofolato	FolE
acetil-CoA carboxilasa	biosíntesis de ácidos grasos	Malonil-CoA	AccA,B,C,D

Un ejemplo no limitativo específico de una ruta, proporcionado con fines ilustrativos, es la ruta para la síntesis de ácido shikímico (véase figura 4). En esta ruta, por ejemplo, una reacción entre los compuestos de mercancía celular fosfoenolpiruvato (S<sub>1A</sub>) y eritrosa-4-fosfato (S<sub>1B</sub>) es catalizada por la enzima DAHP Sintasa (E<sub>1</sub>) para formar 3-desoxi-D-arabinosheptulosa-7-fosfato (DAHP). DAHP (S<sub>2</sub>) se transforma en 3-deshidroquinato (3-DHQ) por la segunda enzima en la ruta, DHQ sintasa (E<sub>2</sub>). Se deshidrata 3-DHQ (S<sub>3</sub>) a 3-deshidrosikimato por la tercera enzima en la ruta, 3-DHQ deshidratasa (E<sub>3</sub>). El 3-deshidrosikimato (S<sub>4</sub>) se reduce al ácido shikímico (PF) por la cuarta enzima en la ruta, shikimato deshidrogenasa (E<sub>4</sub>), utilizando NADPH como cofactor. Las enzimas de la ruta son conocidas en la técnica y se han caracterizado en un número de organismos, incluyendo, por ejemplo, E Coli, en la que las enzimas son codificadas por los loci genéticos como sigue: DAHP sintasa (aroG, aroF, aroH); DHQ sintasa (aroB); 3-DHQ deshidratasa (aroD); shikimato deshidrogenasa (aroE); véase, por ejemplo, la publicación PCT N° WO2010/074760.

#### Métodos de producción

La producción de alto rendimiento de un producto de interés se logra proporcionando una célula en la cual se expresan enzimas citoplásmicas que comprenden una ruta de interés, por ejemplo, a niveles fisiológicamente normales, o a niveles mayores que fisiológicamente normales; y en donde al menos una enzima clave de la ruta es (a) expresada a altos niveles y (b) reubicada a un compartimiento distinto del compartimiento de origen natural. En algunas realizaciones la enzima clave es secuestrada en el periplasma. La enzima clave controla el flujo a través de la ruta de interés, y puede ser una enzima de entrada de ruta y/o una enzima limitante de la velocidad. Una contraparte natural a la(s) enzima(s) clave se expresa usualmente en niveles normales en el citoplasma. Durante el cultivo celular puede ser deseable controlar los componentes del medio de crecimiento de las células para evitar la exposición de la enzima secuestrada periplásmica a condiciones que pueden disminuir su actividad, por ejemplo, exposición a metales y similares. Por ejemplo, se ha encontrado que la DAHP sintasa en la ruta del ácido shikímico puede desactivarse a través de la oxidación catalizada con cobre, y por lo tanto es deseable modificar las condiciones de cultivo aumentando la concentración de manganeso y de metales de magnesio en el medio de crecimiento para superar el cobre disponible (véase, por ejemplo, Bauerle *et al.*, *J Bacteriol.* (1999) 181: 1636-1642; y Stadtman *et al.* *J Biol. Chem* (1991) 266: 2005-2008). En otras realizaciones, el cofactor (s) se proporciona o se alteran concentraciones de co-factor(es) en el medio de crecimiento para mejorar la activación enzimática en el periplasma u otro sitio reubicado de enzima.

Para fines de producción, se utiliza un lisado de la célula, en donde la enzima periplásmicamente secuestrada se pone en contacto operable con las enzimas de la ruta de interés expresada en el citoplasma. Las células se lisan por cualquier método conveniente que mantenga sustancialmente la actividad enzimática, por ejemplo, sonicación, prensa francesa, y similares como se conoce en la técnica. El lisado se puede fraccionar, secar la materia particulada, o se puede usar en ausencia de etapas de procesamiento adicionales. El lisado celular se puede combinar adicionalmente con uno o más sustratos, enzimas, nutrientes, co-factores, tampones, agentes reductores y/o sistemas generadores de ATP, etc. según se requiera para la actividad enzimática. Tal sistema, en ciertas realizaciones, puede referirse en la presente como un "sistema libre de células", es decir, un sistema aislado que contiene un lisado o extracto celular diseñado expresamente para sintetizar una enzima o una cascada de enzimas que, cuando actúa en una secuencia dada (por ejemplo, en una ruta enzimática) y proporciona sobre un sustrato determinado, resultados en la generación preferencial de un producto, el compuesto de interés. Un compuesto de interés es típicamente una entidad química (por ejemplo, una molécula orgánica pequeña), que se puede usar como ingrediente farmacéutico activo (API), precursor químico o intermediario.

Como se utiliza en la presente, un "sustrato" es un compuesto o mezcla de compuestos capaces de proporcionar los elementos necesarios para sintetizar un compuesto de interés.

Como se utiliza en la presente, una "molécula orgánica pequeña" o "molécula pequeña" se refiere a una molécula orgánica con un peso molecular inferior a 800 g/mol (por ejemplo, menos de 700 g/mol, menos de 600 g/mol, menos

de 500 g/mol, menos de 400 g/mol, menos de 300 g/mol, menos de 200 g/mol, menos de 100 g/mol, entre 50 y 800 g/mol, inclusive, entre 100 y 800 g/mol, inclusive, o entre 100 y 500 g/mol, inclusive). En ciertas realizaciones, la molécula orgánica pequeña es un agente terapéuticamente activo tal como un fármaco (por ejemplo, una molécula orgánica pequeña aprobada por la Administración de Alimentos y Fármacos de los Estados Unidos como se proporciona en el Código de Regulaciones Federales (CFR)). La molécula orgánica pequeña puede comprender también un metal. En este caso, la molécula orgánica pequeña es también referida como una "molécula organometálica pequeña".

Como se utiliza en la presente, un "equivalente reductor" o "agente reductor" es una especie química que transfiere el equivalente de un electrón en una reacción redox. Los ejemplos de equivalentes reductores son un solo electrón (por ejemplo, en reacciones que involucran iones metálicos), un átomo de hidrógeno (que consiste en un protón y un electrón), y un ion hidruro (:H-) que lleva dos electrones (por ejemplo, en reacciones que implican NAD). Un "aceptor equivalente reductor" es una especie química que acepta el equivalente de un electrón en una reacción redox.

Como se utiliza en la presente, un "sistema de regeneración de trifosfato de adenosina" o "sistema de regeneración de ATP es un sistema químico o bioquímico que convierte la adenosina, AMP y ADP en ATP. Ejemplos de sistemas de regeneración de ATP incluyen aquellos que involucran metabolismo de glucosa, metabolismo de glutamato y fotosíntesis.

Los lisados de células de diferentes fondos genéticos (por ejemplo, previamente alterados o genéticamente modificados por ingeniería genética) o especies, o que se preparan por diferentes estrategias, pueden ser mezcladas y simultáneamente o en forma continua utilizado secuencialmente en un bioproceso con el lisado celular de la invención. El lisado puede estar libre o inmovilizado, y puede reutilizarse o desecharse en cada etapa del proceso. Por ejemplo, en ciertas realizaciones, el lisado celular es un lisado de un organismo de *E. coli* diseñado para sobreexpresar una o más enzimas en la ruta de interés. En ciertas realizaciones, el lisado celular es una combinación de diferentes lisados celulares, por ejemplo, una combinación de dos, tres, cuatro, cinco, seis, siete, ocho, nueve o diez lisados celulares diferentes, obtenidos de dos, tres, cuatro, cinco, seis, siete, ocho, nueve, o diez diferentes organismos de *E. coli* diferentes cada uno diseñado para sobreexpresar una o más enzimas en la ruta de interés.

Los métodos de la invención proporcionan altos rendimientos del producto deseado, cuyo rendimiento es mayor que el rendimiento que se puede lograr con un hospedador microbiano nativo. También se puede incrementar la productividad (es decir, velocidad de producción por unidad de volumen o biomasa). En una realización de la invención, el rendimiento del producto es al menos aproximadamente 2 veces por encima de la velocidad basal, al menos aproximadamente 5 veces por encima de la velocidad basal, al menos aproximadamente 10 veces por encima de la velocidad basal, al menos aproximadamente 25 veces por encima de la velocidad basal, por lo menos aproximadamente 50 veces por encima de la velocidad basal, o más, por ejemplo, entre aproximadamente 2 veces y aproximadamente 100 veces por encima de la velocidad basal. En ciertas realizaciones, la velocidad de producción del producto utilizando los métodos de la invención es de 0,1 a 20 gramos de producto/L/h.

Se pueden adaptar inóculos diferentes a condiciones diferentes (por ejemplo, dos lotes cultivados en dos diferentes fuentes de carbono) o pueden tener diferentes genotipos y luego mezclarse para llevar a cabo el proceso (por ejemplo, para obtener el consumo simultáneo de una mezcla de fuentes de carbono o procesamiento secuencial de un metabolito a través de una ruta dividida en dos lotes separados de células). Un proceso también puede llevarse a cabo secuencialmente permitiendo que un conjunto de reacciones proceda en un recipiente y luego transfiriendo el sobrenadante a un segundo recipiente.

Las reacciones pueden utilizar un reactor a gran escala, pequeña escala, o pueden multiplexarse para realizar una pluralidad de síntesis simultáneas. Las reacciones continuas utilizarán un mecanismo de alimentación para introducir un flujo de reactivos, y puede aislar el producto final como parte del proceso. Los sistemas discontinuos también son de interés, en donde se pueden introducir reactivos adicionales con el tiempo para prolongar el período de tiempo para la síntesis activa. Un reactor se puede ejecutar en cualquier modo tal como lote, lote extendido, semi-discontinuo, semi-continuo, alimentado-discontinuo, y continuo, y que se seleccionará de acuerdo con el propósito de la solicitud.

Las reacciones pueden ser de cualquier volumen, ya sea en una escala pequeña (por ejemplo, por lo menos aproximadamente 1 ml y no más de aproximadamente 15 ml) o en una reacción de escala (por ejemplo, donde el volumen de reacción es de al menos aproximadamente 15 ml, usualmente al menos aproximadamente 50 ml, más usualmente al menos aproximadamente 100 ml, y puede ser 500 ml, 1000 ml, o mayor hasta muchos miles de litros de volumen). Las reacciones se pueden llevar a cabo a cualquier escala.

Se pueden incluir sales y tampones diversos, donde las especies iónicas se optimizan típicamente con respecto a la producción de productos. Cuando se cambia la concentración de un componente particular del medio de reacción, se puede cambiar de manera acorde otro componente. Además, los niveles de concentración de los componentes en el reactor pueden variar con el tiempo. El ajustador del potencial de oxidación/reducción de tiol/disulfuro puede ser ditiotreitól, ácido ascórbico, glutatión y/o sus formas oxidadas. También se pueden usar otros ajustadores del potencial redox general.

En un modo de operación semi-continua, el reactor puede ser operado en diálisis, lote de diafiltración o modo de carga-carga. Una solución de alimentación puede suministrarse al reactor a través de la misma membrana o una

unidad de inyección separada. El producto sintetizado se acumula en el reactor, y luego se aísla y purifica de acuerdo con el método habitual para la purificación después de la finalización de la operación del sistema. Alternativamente, el producto puede ser retirado durante el proceso ya sea en un modo continuo o discontinuo de la opción de devolver parte o todos los compuestos restantes al reactor.

- 5 En donde existe un flujo de reactivos, la dirección del flujo de líquido puede ser perpendicular y/o tangencial a una membrana. El flujo tangencial es efectivo para reciclar el ATP y para evitar el taponamiento de la membrana y puede sobreponerse en un flujo perpendicular. El flujo perpendicular a la membrana puede ser provocado o efectuado por una bomba de presión positiva o una bomba de succión de vacío o aplicando presión de transmembrana utilizando otros métodos conocidos en la técnica. La solución en contacto con la superficie exterior de la membrana puede cambiarse cíclicamente, y puede estar en un flujo tangencial constante con respecto a la membrana. El reactor puede agitarse interna o externamente.

10 La cantidad de producto producido en una reacción puede medirse de diversas maneras, por ejemplo, mediante ensayos enzimáticos que producen un producto coloreado o fluorométrico o por métodos de HPLC. En ciertas realizaciones, el producto se mide utilizando un ensayo que mide la actividad o concentración del producto particular que se produce.

### Ejemplos

Los siguientes ejemplos se exponen para proporcionar los entendidos en la técnica con una descripción completa y descripción de cómo hacer y usar la presente invención, y no se pretende que limiten el alcance de la invención o para representar que los experimentos a continuación son todos o los únicos experimentos realizados. Se han realizado esfuerzos para asegurar la precisión con respecto a los números utilizados (por ejemplo, Cantidades, temperatura y similares), pero pueden estar presentes algunos errores y desviaciones experimentales. A menos que se indique lo contrario, las partes son partes en peso, peso molecular es peso molecular promedio en peso, temperatura en grados Celsius, y la presión se encuentra en o cerca de la presión atmosférica.

#### Ejemplo 1 Producción de ácido shikímico

25 El ácido shikímico es un intermediario en la ruta biosintética del chorisato, donde la enzima de entrada clave es 2-deshidro-3-desoxifosfoheptonato aldolasa (3- desoxi-D-arabino-heptuloso-7-fosfato, DAHP, sintasa). La DAHP sintasa cataliza la primera etapa comprometida en la producción de shikimato convirtiendo los metabolitos centrales fosfoenolpiruvato (PEP) y eritrosa-4-fosfato (E4P) a DAHP. En *E. Coli*, existen tres enzimas de DAHP sintasa-AroG, AroE, y AroF-codificados por genes *aroG*, *aroE* y *aroF*, respectivamente. Es común que las versiones resistentes a la retroalimentación de estas enzimas (Kikuchi et al., *Appl. Environ. Microbiol.* (1997) 63: 761; Ray et al., *J Bacteriol.* (1988) 170: 5500; Weaver y Herrmann, *J. Bacteriol.* (1990) 172:6581) se utilicen para asegurar la máxima actividad.

30 En este ejemplo, se modifica un gen DAHP sintasa para contener diversas secuencias de señal periplásmica (péptidos líder periplásmicas) dirigir la enzima al periplasma. Optimización de la expresión, evaluación de diversas sintasas DAHP, y la identificación y eliminación del sitio de la proteasa periplásmica se utilizan para tratar riesgos potenciales y retos asociados con la expresión periplásmica objetivo. Expresión de DAHP sintasa activa en el periplasma, y acoplado con la expresión de citoplásmica de los genes de la ruta corriente abajo para demostrar un crecimiento más robusto de la cepa modificada post-inducción (con relación a una ruta de genes de selección que sobre expresa la solamente en el citoplasma), seguido por demostración de un flujo incrementado al producto de interés después de la lisis en una reacción activa, libre de células.

40 El AroG se dirige para la expresión en el periplasma de *E. coli*. Las secuencias de ADN que codifican varios péptidos líder periplásmicos se añaden al gen *aroG* a través de la amplificación por PCR para crear una biblioteca de secuencias de ADN que codifican AroG con diferentes péptidos líder periplásmicos en la región N-terminal de la proteína (biblioteca de PerS-AroG). Alternativamente, se crean secuencias de ADN que codifican para la biblioteca de PerS-AroG mediante síntesis de ADN. Se analizan varias secuencias de señal periplásmica para determinar que es más eficiente para producir el nivel más alto de enzima activa en el periplasma. Las guías periplásmicas adecuadas se exponen en las tablas 4 y 5 del ejemplo 2. Las secuencias de ADN que codifican para la biblioteca de PerS-AroG se insertan en vectores de expresión adecuados para crear una biblioteca de vectores de expresión de PerS-AroG. La biblioteca de vectores de expresión se utiliza para transformar una cepa adecuada de *E. Coli* que se recubre y se criba para la expresión de PerS-AroG y localización periplásmica de AroG. La construcción seleccionada se puede optimizar adicionalmente para la expresión mediante la prueba de una pluralidad de promotores y sitios de unión a ribosomas.

45 Por ejemplo, las secuencias de ADN que codifican para la biblioteca de PerS-AroG se clonan en un vector pDuet para la expresión inducible del promotor T7/*lacO*. La cepa de *E. coli* BL21 (DE3), o cepa similar que expresa la polimerasa T7, se transforma con plásmidos que contienen la biblioteca PerS-AroG. La modificación de la expresión se logra mediante el uso de niveles variables del inductor isopropil- $\beta$ -D-1-tiogalactopiranosido (IPTG), a través del uso de promotores variantes y sitios de unión a ribosomas, así como a través del uso de variación del número de copias entre diferentes vectores pDuet. También se pueden utilizar otros plásmidos, sistemas de expresión, o cepas familiares para los entendidos en la técnica.

Varias cepas se crean con versiones de AroG-expresadas por pDuet con diversas guías periplásmicas, y sin líder periplásmico. Cultivo de la densidad óptica intermedia en medio definido rico antes de la inducción de la expresión con IPTG 0,05-1 mM. La expresión se induce durante varias horas para permitir la acumulación de DAHP sintasa en el periplasma. La DAHP sintasa periplásmica se extrae utilizando choque osmótico, u otros métodos conocidos por los entendidos en la técnica. Verificación de la expresión de proteína de longitud completa se determina mediante electroforesis en gel de proteína desnaturizante con estándares apropiados. Se pueden usar diversos métodos para optimizar la expresión o plegamiento de la DAHP sintasa periplásmicamente dirigida (u otras enzimas). Estos incluyen, pero no se limitan a, los siguientes: i) la optimización de la expresión mediante el uso de niveles de IPTG variables, diferentes orígenes de replicación de plásmidos, y/o modificación de RBS y/o promotores, ii) la identificación y eliminación de sitios conocidos para proteasas periplásmicas específicas a través de sustituciones conservadoras de aminoácidos, y iii) el uso de enzimas ortólogos. Los datos indican la expresión periplásmica de AroG de longitud completa cuando se usan secuencias de señal periplásmica de OmpA y STII, como se muestra en la figura 1.

La actividad específica de la DAHP sintasa periplásmicamente dirigida se determina en la célula entera o en el extracto periplásmico utilizando un ensayo espectrofotométrico continuo que monitoriza la absorbancia a 232 nm para medir la conversión de PEP (con E4P) a DAHP. Por ejemplo, la célula entera o extracto periplásmico contiene Tris-HCl 10 mM (pH 7,5) con fosfato potásico 35 mM (pH 7,0) y PEP-K 500 µM para estabilizar la proteína. Antes del ensayo, el extracto se hace pasar a través de una columna Sephadex G-25 equilibrada con la misma solución tampón para eliminar aminoácidos y otras moléculas menores de 5 kDa. Se añade un microlitro de extracto purificado a 99 microlitros de mezcla de reacción (100 µM PEP-K, 300 µM de E4P-Na, 10 mM de 1,3-bis [tris(hidroximetil)metilamino]propano, MgCl<sub>2</sub> 10 µM, pH 7,0) y la absorbancia a 232 nm se monitoriza en el transcurso de 0,5-2 horas. Los extractos de control apropiados de cepas que no expresan AroG, así como las reacciones de control realizadas sin E4P, se incluyen para la normalización. La concentración de proteína total en la célula entera o extracto periplásmico se determina utilizando un ensayo de Bradford estándar familiar para los entendidos en la técnica. La fracción de la célula entera o proteína periplásmica que es AroG se determina mediante el análisis de imágenes de geles de poliacrilamida teñidos con coomassie de células enteras o extracto periplásmico.

Debido a que es conocido que Cu<sup>++</sup> provoca la inactivación irreversible de mediciones DAHP sintasa (Park y Bauerle, *J. Bacteriol* (1999) 181:1636) tales como limitar el contenido de Cu<sup>++</sup> del medio de crecimiento e incrementar las concentraciones de otros cationes divalentes tales como Mn<sup>++</sup> para favorecer la activación completa de la enzima, así como para conservar su actividad. Demostración de la actividad de un AroG periplásmicamente expresado, enzimas adicionales en la ruta bioquímica para el ácido shikímico útiles para proporcionar precursores de ruta u otros sustratos de ruta (listados a continuación) se subclonan en los vectores pDuet para la sobreexpresión citoplasmática en *E. Coli* BL21 (DE3) como se describió anteriormente.

Tabla 3.

EC#	Enzima	<i>E. coli</i>	Accesión de Genbank N°.
2.2.1.1	transcetolasa	TktA	AAT48155.1
4.2.3.4	deshidroquinato sintasa	AroB	AAC76414.1
4.2.1.10	deshidroquinato deshidratasa	AroD	AAC74763.1
1.1.1.25	shikimato deshidrogenasa	AroE	AAC76306.1

La sobreexpresión de una o varias de estas enzimas ha demostrado mejorar la producción de ácido shikímico *in vivo* (Patnaik y Liao, *Appl. Environ. Microbiol.* (1994) 60: 3903; Flores *et al. Nat. Biotechnol.* (1996) 14: 620; Herrmann, *Plant Physiol.* (1995) 107: 7; Bongaerts *et al. Met. Eng.* (2001) 3: 289; Kramer *et al. Met. Eng.* (2003) 5: 277). En el caso de la transcetolasa (TktA), una enzima con el propósito de incrementar el suministro de un precursor de ruta (4-fosfato eritrosa), la enzima será exportada al periplasma si se observa que la sobreexpresión de la transqueasa es perjudicial para el crecimiento del organismo. En un ejemplo similar, el producto génico de la transcripción de nucleótido soluble (gen SthA) también puede evaluarse para la sobreexpresión ya sea en el citoplasma o en el periplasma para evaluar el efecto de su sobreexpresión citoplásmica sobre el crecimiento. Adicionalmente, se incluye un vector que contiene una DAHP sintasa citoplásmicamente dirigida para servir como control de crecimiento celular. Cada vector pDuet puede expresar dos genes de promotores individuales, asegurando la máxima expresión de todas las proteínas con hasta tres vectores. La disposición policistrónica de genes para expresar todos los genes de un único vector si es necesario para un crecimiento celular mejorado.

5 Se realiza la electroforesis en gel de proteína desnaturalizante con patrones apropiados para asegurar la expresión de proteína de longitud completa en el citoplasma. Tras la confirmación de la expresión de la proteína de longitud completa, los plásmidos se cotransforman en BL21(DE3) y se seleccionan co-transformantes sobre medios antibióticos apropiados para asegurar la presencia de todos los plásmidos. Como se ha descrito anteriormente, las células se hacen crecer en medio definido rico a densidad óptica intermedia, y la sobreexpresión se induce mediante la adición de IPTG 0,05-1 mM. La medición espectrofotométrica de la densidad óptica del cultivo a intervalos definidos para determinar el tiempo de generación para células que sobreexpresan TktA, AroB, Ard y AroE en el citoplasma junto con DAHP sintasa sobre-expresada en el citoplasma (control) o periplasma.

10 En diversos puntos de tiempo después de la inducción, las células son cosechadas, lisadas mediante el uso de un homogeneizador de alta presión, y se mezcla con glucosa, glutamato y otros sustratos que se van a usar en una reacción libre de células para la producción de shikimato. Los niveles de shikimato (e intermedios que incluyen 3-deshidroquinato y 3-deshidroshikimato) se miden mediante HPLC utilizando métodos familiares para los entendidos en la materia. La velocidad y grado de crecimiento, así como los niveles y velocidad de producción de shikimato en el lisado se comparan cuando DAHP sintasa se expresa en el periplasma con relación a los obtenidos cuando se expresa DAHP sintasa en el citoplasma.

**Ejemplo 2. Efectos de Crecimiento y actividad de periplásmicamente de expresión AroG**

20 *Expresión periplásmica de DAHP sintasa.* Se construye una biblioteca de plásmidos que contiene el gen que codifica AroG (Genbank Acc. No AAC73841.1) modificado con diversas secuencias de señal periplásmica que dirigen la enzima al periplasma. Las secuencias de ADN de los cebadores usados para construir las secuencias de codificación para las guías periplásmicas probadas se exponen en las tablas 4 y 5. Las secuencias de ADN que codifican para un conjunto de secuencias de señal periplásmica se agregan al gen *aroG* a través de la amplificación de PCR utilizando los siguientes cebadores:

Tabla 4. Cebadores utilizados para agregar señales de orientación periplásmica a <i>aroG</i>		
Guía	Cebador	Secuencia
ninguna	F 5'	gcaattcgggtctcccatgaattatcagaacgacgatttacgcatc (SEQ ID NO:11)
	R 3'	gaattcggcgccgcttaccgacgacgcttttac (SEQ ID NO:12)
OmpA	F 5'	gcaattcgggtctcccatgaaaaaaaaacgcaattgcgatagcg (SEQ ID NO:13)
	R 3'	gaattcggcgccgcttaccgacgacgcttttac (SEQ ID NO:14)
StII	F 5'	gcaattcgggtctcccatgaaaaaaaaatattgctttctgctcg (SEQ ID NO:15)
	R 3'	gaattcggcgccgcttaccgacgacgcttttac (SEQ ID NO:16)
DsbA	F 5'	gcaattcgggtctcccatgaaaaagattggctggcgctg (SEQ ID NO:17)
	R 3'	gaattcggcgccgcttaccgacgacgcttttac (SEQ ID NO:18)
MalE	F 5'	gcaattcgggtctcccatgaaaaataaaaacaggtgcacgcatcc (SEQ ID NO:19)
	R 3'	gaattcggcgccgcttaccgacgacgcttttac (SEQ ID NO:20)
PhoA	F 5'	gcaattcgggtctcccatgaaacaaagcactattgcactggc (SEQ ID NO:21)
	R 3'	gaattcggcgccgcttaccgacgacgcttttac (SEQ ID NO:22)
SfmC	F 5'	gcaattcgggtctcccatgatgactaaaataaagtattgatgctc (SEQ ID NO:23)
	R 3'	gaattcggcgccgcttaccgacgacgcttttac (SEQ ID NO:24)

Tabla 5. Cebador inverso para la adición de la etiqueta 6xHis del término C a todas las construcciones

Guía	Secuencia
R 3'	GAATGCGGCCGCTTAGTGGTGATGATGGTGGTATGCCCGCGACGCGCTTTTAC (SEQ ID NO:25)

Se digieren las construcciones de aroG PCR-amplificadas utilizando los cebadores en las tablas 4 y 5 con Bsal/NotI y se subclonaron en un vector pDuet Digerido con NcoI-NotI para la expresión Inducible. Cepa BL21 (DE3) de *E. coli* se transforma con plásmidos que contienen el AroG subclonado, periplásticamente-dirigido. La modificación de la expresión se logra mediante el uso de niveles variables de IPTG, así como a través del uso de variación del número de copias entre diferentes vectores pDuet. También se pueden utilizar otros plásmidos, sistemas de expresión, o cepas familiares para los entendidos en la técnica.

Cultivos de células madre de trabajo congelado de las siguientes cepas:

- BL21 (DE3): pACYC-Duet 1 (control del vector vacío)
- BL21 (DE3): pACYC-AroG (sin secuencia de señal periplásmica)
- BL21 (DE3): pACYC-OmpA-AroG (que contiene AroG Con una secuencia de señal periplásmica de OmpA)
- BL21 (DE3): pACYC-STII-AroG (que contiene AroG con secuencia de señal periplásmica de STII)

se inocularon a una densidad óptica a 600 nm de 0,0025 en 250 ml de medio definido en EZ Rich (Neidhardt et al. *J. Bacteriol.* (1974) 119:736) que contiene 34 µg/mL cloramfenicol. Después del crecimiento a 37° C a OD600 0,6, se añadió IPTG 0,1 mM para inducir la expresión de la proteína. El crecimiento y la inducción se llevaron a cabo después durante 16 h a 25° C.

Las fracciones de células enteras, periplásmicas y citoplásmicas se obtuvieron utilizando métodos familiares para los entendidos en la técnica (por ejemplo, véase *Chen et al. Biochem. Eng. J* (2004) 19: 211; Soares *et al. Prot. Eng.* (2003) 16: 1131). Específicamente, se obtuvieron fracciones de células enteras recogiendo 12 ml de cultivo a 3000 x g durante 30 min. a 4° C. Los sedimentos celulares se resuspendieron en 12 ml de Tris-HCl 1 mM, pH 7,0. Las células resuspendidas se lisaron en dos pasadas a través de un homogeneizador de alta presión EmulsiFlex-C3 (Avestin, Canada) a 15000-17000 psi. Para eliminar los restos celulares, las muestras se centrifugaron a 21.000 x g durante 15 min. Se obtuvieron fracciones periplásmicas y citoplásmicas como sigue: se cosecharon 200 ml de cultivo a 3000 x g durante 30 min a 4° C. Los sedimentos celulares se resuspendieron suavemente 4°C en 2 ml de Tris-HCl 1 mM de pH 7,0, después se centrifugó a 3000 x g durante 30 minutos. El sobrenadante se utilizó como fracción periplásmica y la pella se procesó adicionalmente para obtener la fracción citoplásmica. La resuspensión vigorosa del sedimento en 11 ml de Tris-HCl 50 mM y NaCl 50 mM, seguido por aplicación de lisis y clarificación como se describe para el extracto de células enteras, produjo la fracción citoplásmica.

Verificación de la expresión de proteína de longitud completa se determina mediante electroforesis en gel de proteína desnaturizante con estándares apropiados. Específicamente, 19,5 µl de tampón que funciona con SDS-PAGE que contiene ditiotreitól (DTT) 0,04 M) a 39 µl de extracto, después se incubó 5 min a 99° C las muestras se realizaron en geles de Bis-Tris al 10% a 200 V durante 55 minutos. Las transferencias Western se realizaron transfiriendo proteínas sobre una membrana de nitrocelulosa utilizando el módulo de transferencia XCell IITM a 30 V durante 90 minutos. Las membranas se lavaron luego dos veces con PBS (solución salina tamponada con fosfato) durante 5 min seguido de 1 h de etapa de bloqueo con PBS y 1% de leche seca no grasa (temperatura ambiente en un agitador orbital). Se diluyó un C-terminal anti-His-HRP (peroxidasa de rábano picante) anticuerpo (Invitrogen) a 1:5000 en tampón de bloqueo y se incubó con la membrana lavada para las proteínas 1 h que contenían Se observó proteínas que contenían un marcador His en el extremo C-terminal en la que membrana de nitrocelulosa después de la incubación en solución de transferencia-transferencia de TMB (Invitrogen). Los datos indican la expresión periplásmica de AroG de longitud completa cuando se usan secuencias de señal periplásmica de OmpA y STII, como se muestra en la figura 1.

*Actividad de DAHP sintasa expresada periplásticamente.* Cultivos de BL21 (DE3) Que expresa OmpA-aroG, o que contenían un control del vector vacío pACYC en 50 ml de medio definido rico en EZ suplementado con 50 µM de MnCl<sub>2</sub> a 37° C. Los cultivos se indujeron con IPTG 0,1 mM cuando la DO600 alcanzó 0,3 y se hizo crecer durante 3 h adicionales a 30° C. Los cultivos se cosecharon por centrifugación a 3000 x g durante 20 min seguido por resuspensión en 13 ml de tampón de fosfato potásico 35 mM, pH 7, y PEP 0,5 mM. Las células se lisaron mediante homogenización a 15000 psi. La Figura 2 contiene datos de crecimiento de estas cepas, lo que indica que la expresión periplásmica de AroG no tiene ningún efecto negativo en el crecimiento celular.

La actividad específica de la DAHP sintasa dirigida periplásticamente se determina utilizando un ensayo espectrofotométrico continuo que controla la absorbancia a 232 nm para medir la conversión de PEP (con E4P) a DAHP. Los ensayos de actividad de la sintasa de DAHP se realizaron en fracciones de proteínas de extracto de células completas de cepas BL21(DE3) que expresan OmpA-aroG, o que contienen un control de vector vacío pACYC. Las fracciones de proteínas se purificaron por filtración en gel utilizando columnas SpinTrap G-25 PD (GE Healthcare). Las mezclas de reacción contenían 100 µM de PEP, 300 µM E4P, tampón de bis-tris propano 10 mM (pH

7), 50  $\mu\text{M}$   $\text{MnCl}_2$  y fracción de proteína 50  $\mu\text{g}/\text{ml}$ . Las reacciones se incubaron a 25°C. La actividad específica de la sintasa de DAHP expresada periplasmáticamente se muestra en la Figura 3.

### Ejemplo 3. Producción libre de células de isobutanol y/o 1-butanol.

5 Los métodos actuales para la producción de isobutanol en *E. coli* se basan en la sobreexpresión de las enzimas de *E. coli* de la biosíntesis de valina IlvI, H, C, D en concierto con la sobreexpresión de dos enzimas heterólogas: la enzima alcohol deshidrogenasa 2 de *S. cerevisiae* (ADH2, GenBank AAA34411.1) y la enzima descarboxilasa de 2-cetoácido de *L. lactis* (KivD, GenBank CAG34226.1) (véase, por ejemplo, Atsumi et al, Nature (2008) 451: 86). De forma similar, la producción de 1-butanol requiere la sobreexpresión de las mismas dos enzimas heterólogas (ADH, KivD) combinadas con la sobreexpresión de IlvA y las enzimas de LeuABCD de la biosíntesis de isoleucina y leucina en *E. coli* (*Atsumi ibid*). La acumulación de alcoholes superiores (por ejemplo, isobutanol, 1-butanol, n-butanol) es tóxico a niveles muy bajos, 2% (p/v) (*Atsumi ibid*; Reyes et al, Plos One (2011) 6: e17678) que produce un crecimiento celular deficiente y un producto pobre títulos cuando estas rutas son activas en el crecimiento de *E. coli*.

15 La reubicación periplásmica de la enzima clave que desvía el flujo de precursores de la biosíntesis de aminoácidos a isobutanol o 1-butanol eliminaría la acumulación del producto durante el crecimiento celular, y permite la post-lisis de la producción libre de células en una cepa diseñada para sobreexpresar enzimas de ruta como se describe en Atsumi et al., Nature (2008) 451: 86. Se creó una biblioteca de la enzima de entrada clave, KivD, con diversas secuencias de señal periplásmica siguiendo los métodos descritos en los ejemplos 1 y 2. Posteriormente a la selección del miembro de biblioteca que exhibe la expresión periplásmica más eficiente, y la verificación de la actividad, se modificaría una cepa diseñada para producir isobutanol y 1-butanol, como se describe en lo que antecede, para incluir KivD-expresada periplásmicamente. El crecimiento metabólicamente saludable de *E. coli* diseñado para producir isobutanol y 1-butanol, a medida que la ruta sería inactiva con una KivD periplásmicamente expresado. Tras la lisis celular, se combinó el contenido periplásmico y citoplasmático se combinaría activando el isobutanol y la producción de 1-butanol a partir de la glucosa.

### Ejemplo 4. Procedimiento para la producción libre de isoprenoides y terpenos

25 La sobreproducción de isoprenoides en *E. coli* requiere uno de dos métodos generales: (1) el uso de la desoxi-xilulosa -5 fosfato de *E. coli* nativa (DXP) ruta, o (2) el uso del mevalonato no-nativo (MEV)-dependiente (véase, por ejemplo, Martin et al., Nat. Biotechnol. (2003) 21: 796-802). El amorfadieno, el precursor para la atermal ina terpenoide antimalaria ha sido producido por la ruta MEV en *E. coli* mediante sobre-expresión de los genes nativos *atoB*, *idi*, *ispA*, así como los genes para los *S. cerevisiae*. El hidrometilglutaril (HMG)-CoA sintasa (ERG13, GenBank CAA90557.1), una HMG-CoA reductasa truncada (HMGR, GenBank CAA86503.1), mevalonato quinasa (ERG 12, GenBank CAA39359.1), quinasa de fosfomevalonato (ERG8, GenBank CAA90191.1), pirofosfato de mevalonato descarboxilasa (MVD1, GenBank CAA66158.1) y una versión de la *Artemisia annua amorfadieno sintasa* (ADS, GenBank AAK15697.1) codón-optimizado para la expresión en levadura (véase, por ejemplo, Martin et al., Nat. Biotechnol. (2003) 21: 796-802 y Figura 5).

35 Los intermediarios de difosfato de prenilo y la HMG-CoA en la ruta isoprenoide descrita son, sin embargo, tóxico y se ha demostrado que se acumulan si la actividad de la sintasa (ADS para este terpenoide) y otras enzimas en la ruta están desequilibradas (véase, por ejemplo, Martin et al., Nat. Biotechnol. (2003) 21: 796-802; Withers et al. Appl. Environ. Microbiol. (2007) 73: 6277-6283). La reubicación de la enzima de entrada clave de esta ruta, AtoB, evitaría la necesidad de una expresión de sintonización fina con el fin de evitar estos problemas de toxicidad durante el crecimiento celular. Específicamente, una biblioteca de la enzima de entrada clave, AtoB, con diversas secuencias de señal periplásmica se crearon siguiendo los métodos descritos en los ejemplos 1 y 2. después de La selección del miembro de biblioteca que exhibe la expresión periplásmica más eficiente, y la verificación de la actividad, una cepa de *E. coli* diseñada para producir isoprenoides (véase, por ejemplo, Martin *ibid*) se modificaría con AtoB periplásmicamente expresado.

45 El crecimiento metabólicamente saludable de esta cepa se lograría, ya que la ruta sería inactiva con un AtoB periplásmicamente expresado. En la lisis celular, el contenido periplásmico y citoplasmático se combinaría activando la producción isoprenoide a partir de glucosa.

### Ejemplo 5. Producción libre de células de poli-3-hidroxi-butarato

50 El *E. coli* se manipuló metabólicamente para producir poli-3-hidroxi-butarato (PHB), un bloque de construcción de biopolímero importante, utilizando una ruta de tres etapas a partir de acetil-CoA (Tyo et al.). Las tres enzimas heterólogas involucradas son la R-eutropha beta-cetotiolasa (PhaA, GenBank CAJ92573.1) y acetooacetil-CoA reductasa (PhaB, GenBank AAA21973.1) y el *Allochromatium vinosum* PHB sintasa subunidad PhaE, Gen Bank ABK60192.1; subunidad PhaC, GenBank ABK60193.1).

55 Cuando esta ruta es activa en *E. coli*, la velocidad de crecimiento se relaciona inversamente con el flujo de PHB debido a la desviación del carbono de la biomasa y/o la acumulación de grandes, gránulos de PHB tóxicos en el citoplasma (véase, por ejemplo, Tyo et al. *Metabolic Engineering* (2010) 12: 187-195). La reubicación de la enzima de



- 5 entrada clave de esta ruta, PhaA, eliminaría los problemas de toxicidad durante el crecimiento celular. Una biblioteca de la enzima de entrada clave, PhaA, con diversas secuencias de señal periplásmica se crearon siguiendo los métodos descritos en los ejemplos 1 y 2. Posteriormente a la selección del miembro de biblioteca que exhibe la expresión periplásmica más eficiente, y la verificación de la actividad, una cepa de *E. coli* manipulada para producir PHB (véase, por ejemplo, Tyo *ibid*) se modificaría con PhaA periplásmicamente expresado. El crecimiento metabólicamente saludable de esta cepa se lograría, ya que la ruta sería inactiva con un PhaA periplásticamente expresado. Tras la lisis celular, el contenido periplásmico y citoplasmático se combinaría activando la producción de PHB a partir de glucosa.

Otras realizaciones

- 10 Como se utiliza en la presente, "a" o "un" significa "un" al menos uno "o" uno o más" a menos que se indique lo contrario.

A menos que se defina de otra manera, todos los términos técnicos y científicos utilizados en la presente tienen el mismo significado que la que se entiende comúnmente por un entendido en la técnica a la cual pertenece esta invención.

## REIVINDICACIONES

1. Un método para producir un producto de una ruta biosintética de interés, el cual comprende:

(a) células bacterianas en crecimiento que expresan enzimas de una ruta biosintética de interés para la producción de un producto, en donde al menos una de las enzimas:

5 (i) es una enzima que controla el flujo metabólico en la ruta biosintética y

(ii) es una enzima que se modifica genéticamente para contener una secuencia peptídica de orientación periplásmica y se reubica en el periplasma de la célula;

10 produciendo así células bacterianas que contienen al menos una enzima modificada que es secuestrada en el espacio periplásmico, en donde la modificación genética y el secuestro periplásmico no tienen efecto negativo sobre el crecimiento de las células bacterianas;

(b) lisar las células bacterianas, produciendo así un lisado celular que contiene las enzimas de la ruta biosintética y la al menos una enzima secuestrada con periplasma genéticamente modificada; y

(c) combinar el lisado celular con una o más sustancias seleccionadas de sustratos, enzimas, nutrientes, cofactores, tampones, agentes reductores y sistemas generadores de ATP;

15 (d) incubar el lisado celular durante un período de tiempo en condiciones suficientes para producir enzimáticamente dicho producto,

en el que el producto se selecciona del grupo que consiste en antibióticos; biosurfactantes; combustibles biológicos, aminoácidos, ácidos orgánicos, ácidos grasos, alcoholes, polioles, sabores, fragancias, nucleótidos, vitaminas, pigmentos, azúcares, polisacáridos, biopolímeros, plásticos, isoprenoides, terpenos y metabolitos celulares.

20 2. El método de la reivindicación 1, en el que la enzima que se modifica genéticamente se selecciona del grupo que consiste en una oxidoreductasa (EC 1), una transferasa (EC 2), una hidrolasa (EC 3), una liasa (EC 4), una isomerasa (EC 5) y una ligasa (EC 6).

25 3. El método de reivindicación 1 o 2, en el que el producto de la ruta biosintética es un metabolito celular seleccionado del grupo compuesto por ácido 2,3-dihidroxi benzoico, 2-cetoglutarato, 3-fosfoglicerato, 4-hidroxibenzoato, 6-fosfogluconato, acetoacetil-CoA, acetil-CoA, acetilfosfato, adenina, adenosina, fosfosulfato de adenosina, difosfato de adenosina (ADP), glucosa ADP, alanina, monofosfato de adenosina (AMP), antranilato, arginina, asparagina, aspartato, trifosfato de adenosina (ATP), aspartato de carbamilo, parafosfato de carburo de silicio, cis-aconito, citrato, citrulina, monofosfato de citosina (CMP), Coenzima A, trifosfato de citosina (CTP), AMP cíclico, citidina, citosina, monofosfato de deoxiadenosina (dAMP), trifosfato de deoxiadenosina (dATP), monofosfato de deoxicitosina (dCTP), deoxiadenosina, desoxiguanosina, desoxirribosa-5-fosfato, monofosfato de deoxiguanosina (dGMP), dihidroorotato, fosfato de dihidroxiacetona, difosfato de deoxitimidina (dTDP), eritrosa-4-fosfato, trifosfato de deoxitimidina (dTTP), dinucleótido de flavina adenina (FAD), mononucleótido de flavina, fructosa-1,6-bisfosfato, fructosa-6-fosfato, fumarato, difosfato de guanosina (GDP), gluconato, gluconolactona, glucosamina-6-fosfato, glucosa-6-fosfato, glucosa-1-fosfato, glutamato, glutamina, glutationa, disulfuro de glutationa, gliceraldehído-3-fosfato, glicerato, glicerol-3-fosfato, guanosina monofosfato (GMP), trifosfato de guanosina (GTP), guanina, guanosina, histidina, histidinol, homocisteína, difosfato de inosina, monofosfato de inosina, trifosfato de inosina, isoleucina, lisina, malato, malonil-CoA, metionina, mio-inositol, N-acetil-glucosamina-1-fosfato, N-acetil-ornitina, dinucleótido de nicotinamida adenina (NAD<sup>+</sup>), nicotinamida adenina dinucleótido hidrato (NADH), +), nicotinamida adenina dinucleótido fosfato (NADP<sup>+</sup>), +), nicotinamida adenina dinucleótido hidrógeno-fosfato (NADPH), ornitina, oxaloacetato, fenilalanina, fenilpiruvato, fosfoenolpiruvato, prolina, propionil-CoA, pirofosfato de fosforibosilo (PRPP), piruvato, quinolinato, riboflavin, ribosa-5-fosfato, ribulosa-5-fosfato, S-adenosil-L-metionina, serina, ácido shikímico, shikimato, succinato, succinil-CoA, treonina, triptófano, tirosina, difosfato de uridina (UDP), UDP-glucosa, UDP-glucuronato, UDP-N-acetilglucosamina, uridina, trifosfato de uridina (UTP), valina y xilulosa-5-fosfato

45 4. El método de la reivindicación 3, en el que el metabolito celular es ácido shikímico o shikimato.

5. El método de la reivindicación 3, en el que las enzimas de la ruta biosintética se seleccionan del grupo que consiste en 3-deoxi-D-arabinoheptulosonato 7-fosfato sintetasa, 3-deshidroquinato sintetasa, 3-deshidroquinato deshidratasa y shikimato deshidrogenasa.

50 6. El método de la reivindicación 1, que comprende además combinar el lisado celular con uno o más lisados celulares adicionales.

7. El método de la reivindicación 1, en el que la al menos una enzima que está modificada genéticamente es una enzima de entrada a la ruta biosintética o una enzima limitante de la velocidad.

8. El método de la reivindicación 1, en el que la al menos una enzima que está modificada genéticamente es una enzima que aumenta la velocidad de dicho sustrato o cofactor suministrado a la ruta biosintética.
9. El método de la reivindicación 1, en el que una contraparte nativa de la enzima modificada genéticamente se expresa a niveles citoplásmicos normales.
- 5 10. El método de la reivindicación 9, en el que se elimina la contraparte nativa de la enzima modificada genéticamente.
11. El método de la reivindicación 1, en el que la enzima modificada genéticamente se sobreexpresa en la célula bacteriana.
- 10 12. El método de la reivindicación 1, en el que al menos dos enzimas en la ruta de interés se modifican genéticamente para comprender una secuencia de direccionamiento periplásmico peptídico.
13. El método de la reivindicación 1, en el que las células bacterianas se cultivan en un medio de crecimiento que se ha modificado mediante la adición o mejora de un factor que aumenta o conserva la actividad de la enzima modificada genéticamente.
- 15 14. El método de la reivindicación 1, en el que la secuencia de orientación periplásmica se selecciona del grupo que consiste en:
- MKIKTGARILALSALTTMMFSASALA (SEQ ID NO:1),
- MKQSTIALALLPLLFTPVTKA (SEQ ID NO:2),
- MMITLRKLPLAVAVAAGVMSAQAMA (SEQ ID NO:3),
- MNKKVLTLSAVMASMLFGAAHA (SEQ ID NO:4),
- 20 MKYLLPTAAAGLLLLAAQPAMA (SEQ ID NO:5),
- MKKIWLALAGLVLAFSASA (SEQ ID NO:6),
- MMTKIKLLMLIIFYLIISASAHA (SEQ ID NO:7),
- MKQALRVAFGFLILWASVLHA (SEQ ID NO:8),
- MRVLLFLLLSLFPALFA (SEQ ID NO:9), y
- 25 MANNDLFQASRRRFLAQLGGLTVAGMLGPSLLTPRRATA (SEQ ID NO:10).

Figura 1

Transferencia Western de AroG con marcador H (expresado de médula pET)

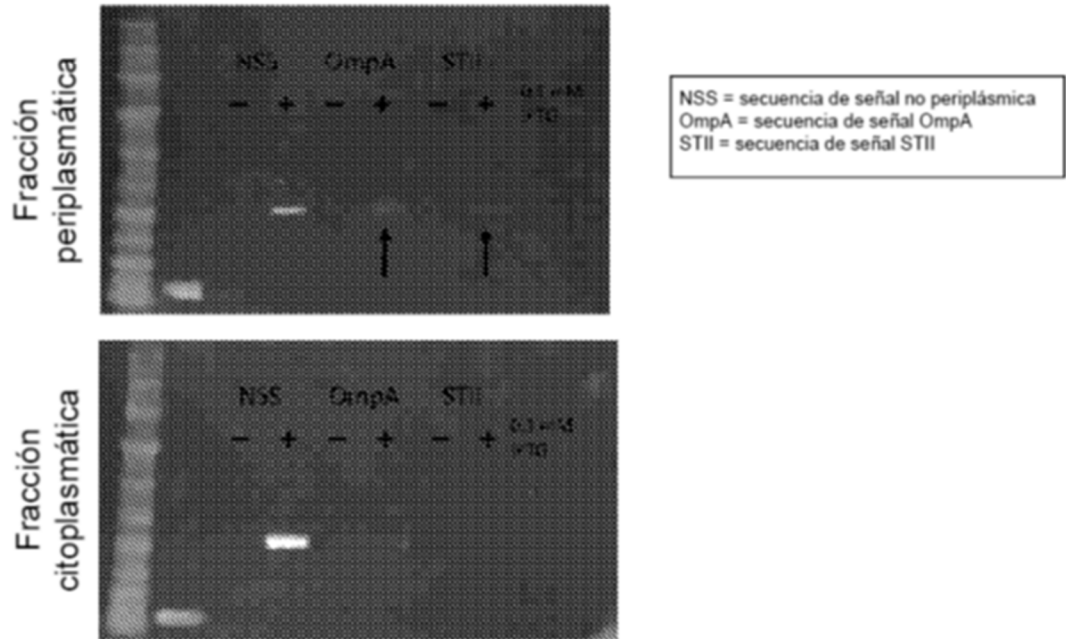


Figura 2

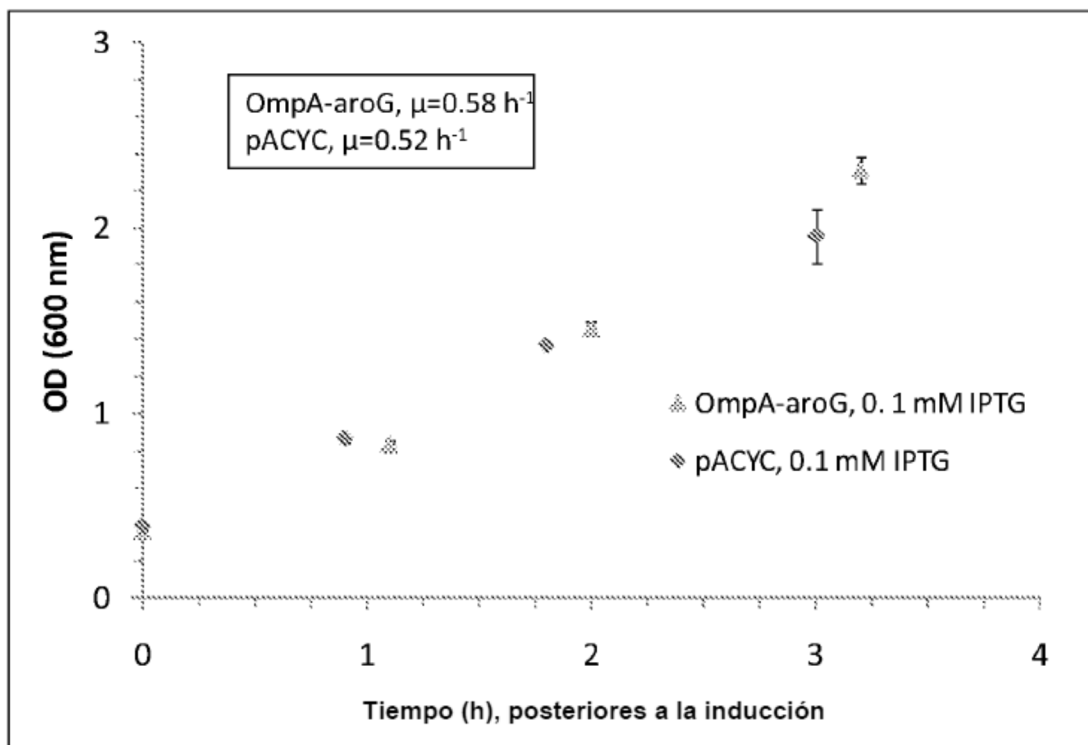


Figura 3

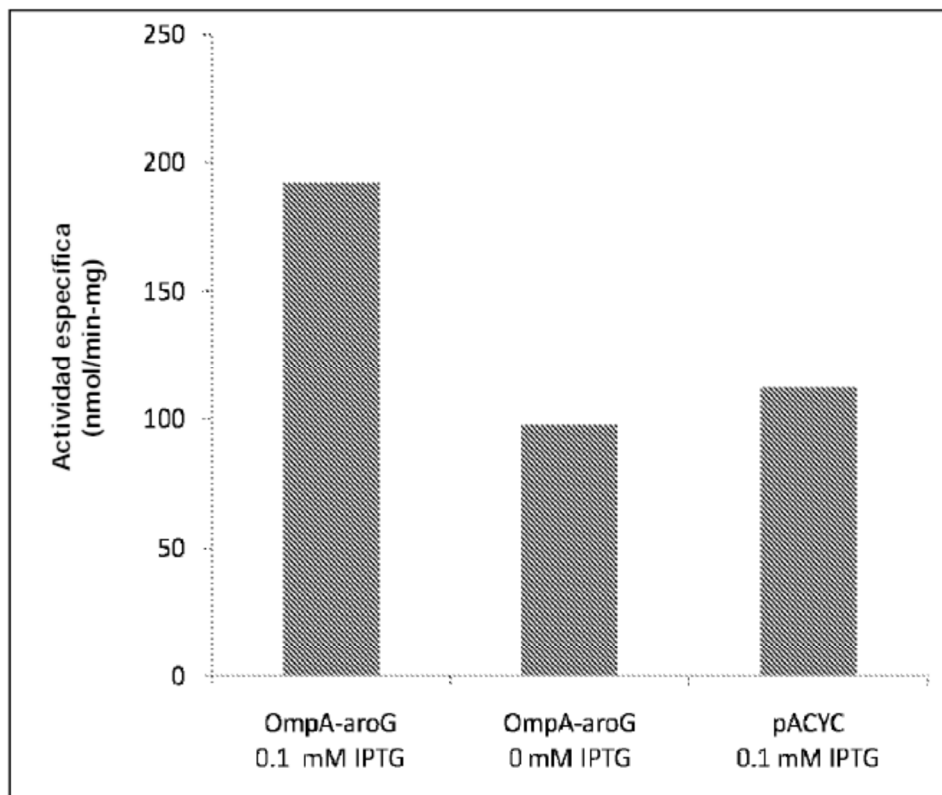


Figura 4

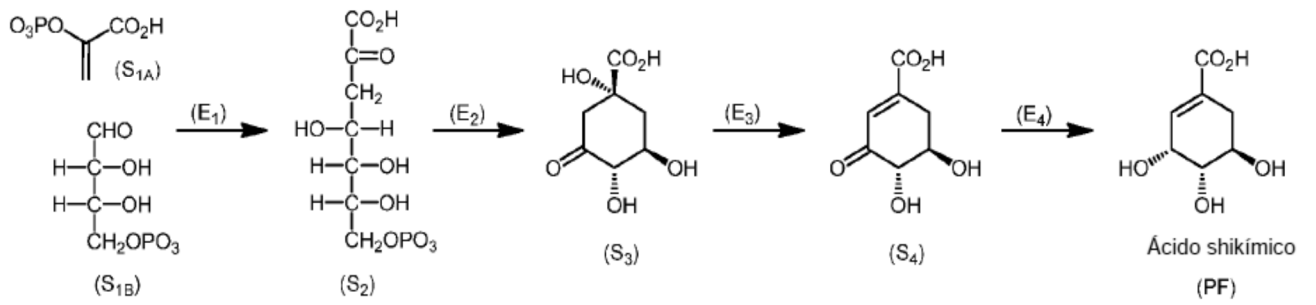


Figura 5

