

19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 839 424**

51 Int. Cl.:

**C12N 9/50** (2006.01)

**C12N 15/52** (2006.01)

**C12N 15/70** (2006.01)

**C12P 7/42** (2006.01)

**C12P 19/30** (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **31.08.2011 E 17163812 (5)**

97 Fecha y número de publicación de la concesión europea: **07.10.2020 EP 3219796**

54 Título: **Métodos para control de flujo en vías metabólicas mediante manipulación de proteasas**

30 Prioridad:

**31.08.2010 US 378828 P**

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

**05.07.2021**

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 2 839 424 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 control de flujo en vías metabólicas mediante manipulación de proteasas

Solicitudes relacionadas

5 La presente solicitud reivindica la prioridad bajo 35 U.S.C. § 119(e) a la solicitud provisional de EE. UU., U.S.S.N. 61/378.828, presentada el 31 de agosto de 2010.

Antecedentes de la invención

10 Se ha comprobado que la producción de sustancias químicas mediante vías enzimáticas sintéticas en hospedantes microbianos es útil para muchas clases importantes de moléculas, incluidos isoprenoides, poliquétidos, péptidos no ribosomales, bioplásticos y bloques de construcción químicos. Debido a la modularidad inherente de la información biológica, la biología sintética exhibe el gran potencial de expandir incluso más esta lista de compuestos producidos en forma microbiana. Incluso incorporando una nueva vía bioquímica en la red metabólica de una célula hospedante se pueden alterar los sutiles mecanismos reguladores que ha desarrollado la célula durante milenios. De hecho, la producción final de un compuesto a menudo está limitada por efectos perjudiciales sobre el metabolismo de la célula genomanipulada que son difíciles de predecir debido a la comprensión limitada de las interacciones complejas que ocurren dentro de la célula. El consumo desregulado de recursos celulares, la carga metabólica de producción de proteínas heterólogas y la acumulación de intermedios de las vías/productos inhibidores o tóxicos para el hospedante son todas cuestiones importantes que pueden limitar el rendimiento general.

20 El concepto de ingeniería metabólica ha surgido para cumplir este propósito, que se puede definir como la modificación deliberada de redes metabólicas y celulares empleando diversas técnicas experimentales para lograr las metas deseadas. Lo que distingue a la ingeniería metabólica de la ingeniería genética y el perfeccionamiento de las cepas es que considera la red metabólica y otras redes como un todo para identificar dianas a manipular. En este sentido, el flujo metabólico es un concepto esencial en la práctica de la ingeniería metabólica. Si bien los niveles de expresión de genes y las concentraciones de proteínas y metabolitos en la célula pueden proporcionar indicios en cuanto al estado de la red metabólica, presentan limitaciones inherentes para describir perfectamente el fenotipo celular debido a la falta de información sobre las correlaciones entre estos componentes celulares. Los flujos metabólicos representan las tasas de reacción en las vías metabólicas y sirven para integrar estos factores a través de un marco matemático. Por lo tanto, los flujos metabólicos se pueden considerar una forma de representar el fenotipo de la célula como resultado de interacciones entre varios componentes de la célula; el flujo metabólico observado refleja las consecuencias de transcripción interconectada, traducción y reacciones enzimáticas que incorporan regulaciones complejas.

35 La síntesis libre de células ofrece ventajas frente a los métodos de producción *in vivo*. Notablemente, la falta de una pared celular *in vitro* permite el control directo del entorno de síntesis. El potencial de reacciones redox (oxidación-reducción), el pH o la fuerza iónica pueden alterar directamente, e incluso más ampliamente, los intervalos no fisiológicos, debido a la falta de crecimiento celular o necesidad de viabilidad celular. Asimismo, la recuperación directa de productos se puede lograr más fácilmente. Los sistemas libres de células pueden dirigir la mayoría, si no todos, los recursos metabólicos de la célula a la producción exclusiva de una vía. Las presentes composiciones y métodos abordan estas cuestiones.

Compendio de la invención

40 La invención se describe en las reivindicaciones anejas. Se dan a conocer sistemas, composiciones y métodos para controlar el flujo de las vías metabólicas a través de la modificación de enzimas diana que compiten con enzimas implicadas en una vía de interés, incluida la modificación para mantener o alterar concentraciones celulares de las enzimas diana durante una fase de crecimiento celular, seguida por manipulación para reducir concentraciones de las enzimas diana durante una fase de producción, cuando se produce el producto de la vía de interés. La fase de crecimiento celular implica células intactas, mientras que la fase de producción se puede efectuar con lisados de dichas células.

50 En los sistemas, composiciones y métodos que se dan a conocer en este documento, las enzimas diana (p. ej., enzimas que compiten para sustratos de enzimas en la vía de interés) son focalizadas para escisión por una proteasa específica del sitio. Dichas enzimas diana son indeseables durante la fase de producción de la vía metabólica. La enzima(s) diana se puede modificar genéticamente para comprender por lo menos un sitio de escisión (también se denomina en este documento sitio de reconocimiento de proteasas específicas del sitio) para una proteasa específica del sitio o puede tener un sitio de escisión natural. El sitio de escisión típicamente se localiza de manera tal que la actividad enzimática se vea mínimamente afectada y que cuando la enzima diana se escinda, esté inactivada.

55 La correspondiente proteasa específica del sitio, es decir, la proteasa que reconoce el sitio de escisión en la enzima diana, se modifica genéticamente para incorporar una secuencia de péptidos que proporciona la reubicación de la proteasa específica del sitio en un compartimento celular o extracelular distinto del compartimento natural. Cuando la proteasa específica del sitio es reubicada durante la fase de crecimiento celular del ciclo celular, no está en

contacto con, y por lo tanto no es capaz de escindir, la enzima diana. Cuando la célula se lisa antes de la fase de producción de metabolitos, la proteasa específica del sitio se pone en contacto con y escinde la enzima diana, inactivando de este modo la enzima diana.

5 En algunas realizaciones, la proteasa específica del sitio es aislada en el periplasma. En otras realizaciones, la proteasa específica del sitio se segrega en el medio de crecimiento.

La proteasa específica del sitio puede ser endógena a la célula, o la secuencia codificante de la proteasa específica del sitio puede introducirse en la célula para expresión.

10 En un aspecto, se dan a conocer en este documento células genéticamente modificadas para comprender (i) una secuencia que codifica una proteasa específica del sitio que comprende una secuencia de reubicación; y (ii) una secuencia codificante para al menos una enzima diana, que compite con una enzima implicada en una vía de interés, en donde la secuencia que codifica la enzima diana comprende un sitio de reconocimiento de proteasas específicas del sitio. En algunas realizaciones, la secuencia que codifica la proteasa específica del sitio se enlaza operativamente a un promotor inducible.

15 En algunas realizaciones, la secuencia de reubicación es una secuencia de direccionamiento periplásmica. En algunas realizaciones, la secuencia se selecciona del grupo que consiste en:

MKIKTGARILALSALTTMMFSASALA (SEQ ID NO:1),

MKKTAIAlAVALAGFATVAQA (SEQ ID NO:2)

MKQSTIALALLPLLFTPVTKA (SEQ ID NO:3),

MMITLRKLPLAVAVAAGVMSAQAMA (SEQ ID NO:4),

20 MNKKVLTLSAVMASMLFGAAHA (SEQ ID NO:5),

MKYLLPTAAAGLLLLAAQPAMA (SEQ ID NO:6),

MKKNIASFLLASMFVFSIATNAYA (SEQ ID NO:7)

MKKIWLALAGLVLAFSASA (SEQ ID NO:8),

MMTKIKLLMLIIFYLIISASAHA (SEQ ID NO:9),

25 MKQALRVAFGFLILWASVLHA (SEQ ID NO: 10), y

MRVLLFLLLSLFLPAFS (SEQ ID NO:11).

En algunas realizaciones, la secuencia que codifica una proteasa específica del sitio codifica una proteasa específica del sitio seleccionada del grupo que consiste en TEV N1a, HRV 3C, Enterocinasa, Factor Xa y Trombina.

30 En determinadas realizaciones, la secuencia codificante de la enzima diana se modifica genéticamente para comprender el sitio de reconocimiento de proteasas específicas del sitio.

En algunas realizaciones, la enzima diana compite con una enzima que aumenta la tasa de precursor suministrado a la vía de interés. En algunas realizaciones, la enzima diana compite con una enzima clave de entrada a la vía.

35 En determinadas realizaciones, el gen que codifica la enzima de contrapartida nativa se reemplaza con la secuencia de ADN que codifica la enzima diana. En determinadas realizaciones, el gen que codifica la enzima de contrapartida nativa se desactiva (p. ej., se inactiva o reemplaza con una molécula de ADN inactiva). En algunas realizaciones, la enzima diana se sobreexpresa en la célula.

En realizaciones particulares, la proteasa específica del sitio o la enzima diana está presente o bien en un vector episomal o en un cromosoma.

40 En determinadas realizaciones, la vía de interés es la síntesis de: a) un antibiótico; b) un biotensioactivo; c) un aminoácido; d) un ácido orgánico; e) un ácido graso; f) un alcohol o poliol; g) un saborífero o fragancia; h) un nucleótido; i) una vitamina; j) un pigmento; k) un azúcar o polisacárido; l) un biopolímero o plástico; m) un metabolito de *E. coli*; n) ácido shikímico y/o shikimato; o) un isoprenoide o terpeno; p) poli-3-hidroxibutirato; o q) isobutanol y/o 1-butanol.

45 En algunas realizaciones, la vía de interés es la síntesis de ácido shikímico y/o shikimato, un isoprenoide o terpeno, poli-3-hidroxibutirato, isobutanol y/o 1-butanol.

En algunas realizaciones, el medio de crecimiento celular se ha modificado por adición o potenciación de un factor que aumenta o preserva la actividad de la enzima diana.

En un aspecto, se dan a conocer en este documento lisados celulares de cualquiera de las células descritas en la presente invención.

5 En otro aspecto, se dan a conocer en este documento sistemas para producir un producto de una vía de interés, en donde los sistemas comprenden uno o más lisados de cualquiera de las células descritas en la presente invención, y opcionalmente uno o más sustratos, enzimas, nutrientes, co-factores, tampones, agentes reductores y sistemas generadores de ATP.

10 Incluso en otro aspecto, se dan a conocer en este documento métodos para producir un producto de una vía de interés, en donde el método comprende desarrollar una cualquiera de las células descritas en la presente memoria hasta una densidad celular deseada, lisando por lo menos una porción de las células, y opcionalmente combinando el lisado con uno o más sustratos, enzimas, nutrientes, co-factores, tampones, agentes reductores y/o sistemas generadores de ATP, en donde las enzimas en la vía de interés catalizan la producción del producto, y la proteasa específica del sitio escinde la enzima diana. En algunas realizaciones, el método comprende además combinar el lisado con uno o más lisados adicionales.

15 Incluso en otro aspecto, se dan a conocer en este documento vectores que comprenden una secuencia que codifica proteasas específicas del sitio que comprende una secuencia de reubicación, en donde, opcionalmente, la secuencia codificante está operativamente enlazada a un promotor inducible.

#### Breve descripción de los dibujos

20 La Figura 1 ilustra imágenes de geles de electroforesis en gel de poliacrilamida (PAGE) teñidas con Coomassie de muestras tomadas después de incubaciones de escisión usando extractos celulares obtenidos de cultivos control y cultivos de expresión de proteasas específicas del sitio con y sin inducción de expresión de proteasas específicas del sitio.

La Figura 2 ilustra imágenes de una inmunotransferencia de muestras tomadas después de incubaciones de escisión que usan extractos celulares obtenidos de cultivos control, de cultivos de enterocinasas específicas del sitio dirigidas a nivel periplásmico, además de enterocinasa recombinante purificada.

#### 25 Descripción detallada de las realizaciones

30 Los sistemas, composiciones y métodos descritos en este documento se basan en la idea de que las células genéticamente modificadas se pueden genomanipular para controlar el flujo metabólico a través de la modificación genética de enzimas diana que compiten con enzimas implicadas en una vía de interés, incluida la modificación genética para mantener o alterar concentraciones celulares de las enzimas diana durante una fase de crecimiento celular, seguida de manipulación para reducir las concentraciones de las enzimas diana activas durante la fase de producción, en donde se produce el producto de la vía de interés. Durante la fase de producción, la inactivación de la enzima diana resulta de la escisión de la enzima por una proteasa específica del sitio. La proteasa específica del sitio puede ser una proteasa endógena o exógena. En determinadas realizaciones, la enzima diana se modifica genéticamente para comprender un motivo (p. ej., un sitio de reconocimiento de proteasas específicas del sitio) que es escindido por una proteasa específica del sitio.

35 En algunas realizaciones, la expresión de la proteasa específica del sitio HRV3C es inducible. En determinadas realizaciones, la proteasa específica del sitio se modifica genéticamente para ser reubicada en un compartimiento celular o extra-celular (p. ej., un compartimiento celular o extra-celular) en donde la proteasa específica del sitio no ocurre naturalmente, de forma tal que la proteasa específica del sitio no inactive la enzima diana al ser reubicada. En determinadas realizaciones, la proteasa específica del sitio se modifica genéticamente para incluir una secuencia de péptidos que proporciona el direccionamiento periplásmico de la proteasa, es decir en donde la proteasa es aislada en el periplasma de la célula.

40 En determinadas realizaciones, la enzima diana es una que compite con una enzima implicada en una vía de interés. En algunas realizaciones, la enzima diana compite con una enzima de entrada a la vía. En determinadas realizaciones, la enzima diana compite con una enzima limitante de la tasa. En determinadas realizaciones, la enzima diana compite con una enzima que aumenta la tasa de precursor suministrada a la vía de interés o que suministra otro sustrato o cofactor requerido.

45 En algunas realizaciones, el gen que codifica la enzima de contrapartida nativa se reemplaza con la secuencia de ADN que codifica la enzima diana. En determinadas realizaciones, el gen que codifica la enzima de contrapartida nativa se desactiva (p. ej., se inactiva o reemplaza con ADN artificial). En determinadas realizaciones, la enzima diana se sobreexpresa (p. ej., la expresión de la enzima aumenta por encima de los niveles de expresión de enzimas fisiológicamente normales) en la célula.

50 En determinadas realizaciones, la enzima diana y/o la proteasa específica del sitio está presente o bien en un vector episomal o en un cromosoma.

En determinadas realizaciones, el medio de crecimiento celular se ha modificado por adición o potenciación de un factor (p. ej., un nutriente, co-factor, agente reductor) que aumenta o preserva la actividad de la enzima diana.

5 En un aspecto, se da a conocer una célula genéticamente modificada, que comprende una proteasa específica del sitio modificada para comprender una secuencia de reubicación, p. ej., una secuencia de direccionamiento periplásmica. La célula puede además comprender una enzima diana que ha sido genéticamente modificada para introducir una secuencia de reconocimiento para la proteasa específica del sitio, que inactiva la enzima diana tras la escisión.

10 En otro aspecto, se dan a conocer los lisados de dicha célula genéticamente modificada. El lisado puede combinarse con uno o más sustratos, nutrientes, cofactores, tampones, agentes reductores, sistemas generadores de ATP y/u otros agentes para generar un sistema libre de células para producir un producto de interés.

15 En otro aspecto, se dan a conocer métodos para producir un producto de interés, en donde el método comprende desarrollar células que expresan una proteasa específica del sitio genéticamente modificada para incorporar una secuencia de péptidos que proporciona su reubicación, en donde una enzima diana que comprende una secuencia de reconocimiento para la proteasa específica del sitio se expresa en el citoplasma de la célula, lisando las células para poner en contacto la proteasa específica del sitio con la enzima diana, y producir el producto de la vía en un sistema libre de células que comprende el lisado. Se pueden añadir sustratos, nutrientes, cofactores, tampones, agentes reductores, sistemas generadores de ATP y/u otros agentes adicionales al sistema libre de células. En una realización, el producto de interés se produce a una alta tasa de flujo.

20 Incluso en otro aspecto, se dan a conocer métodos para producir un producto de interés, en donde el método comprende desarrollar células genéticamente modificadas para sobreexpresar por lo menos una proteasa específica del sitio diana periplásmica, lisando las células y produciendo el producto de la vía de interés en un sistema libre de células que comprende el lisado. Se pueden añadir sustratos, nutrientes, cofactores, tampones, agentes reductores, sistemas generadores de ATP y/u otros agentes adicionales al sistema libre de células.

#### Métodos de producción

25 La producción de alto rendimiento de un producto de interés se logra proporcionando una célula en la que las enzimas citoplásmicas que comprenden una vía de interés se expresan, p. ej., en niveles fisiológicamente normales, o niveles superiores a los fisiológicamente normales; y en donde por lo menos una enzima diana implicada en la vía (a) se expresa en niveles normales y (b) comprende un reconocimiento de inactivación para una proteasa específica del sitio. En algunas realizaciones, la proteasa específica del sitio se expresa en la célula, pero se reubica en un sitio  
30 en donde no entra en contacto con la enzima diana.

Durante el cultivo celular, puede ser conveniente controlar los componentes del medio de crecimiento de las células con el fin de evitar la exposición de la proteasa aislada periplásmica a condiciones que afectan la actividad, p. ej., exposición a metales y similares. Por ejemplo, se ha descubierto que DAHP sintasa en la vía de ácido shikímico, se puede inactivar a través de la oxidación catalizada con cobre, y por lo tanto es conveniente modificar las  
35 condiciones de cultivo aumentando la concentración de cofactores, tales como metales manganeso y magnesio, en el medio de crecimiento para desplazar el cobre disponible (véanse, p. ej., Bauerle et al., J. Bacteriol. (1999) 181:1636-1642; y Stadtman et al., J. Biol. Chem. (1991) 266:2005-2008). Por lo tanto, en ciertas realizaciones, se dan a conocer uno o más cofactores, o se alteran las concentraciones de uno o más cofactores en el medio de crecimiento para potenciar la activación enzimática en el periplasma u otro sitio de enzimas reubicadas.

40 Para propósitos de producción, se utiliza un lisado de la célula, en donde la proteasa específica del sitio aislada en forma periplásmica se pone en contacto operable con las enzimas de la vía de interés expresadas en el citoplasma. Las células se lisan por cualquier método conveniente que prácticamente mantenga la actividad enzimática, p. ej., sonicación, French press y similares, como se conoce en la técnica. El lisado se puede fraccionar, la materia particulada se puede centrifugar o se puede usar en ausencia de pasos de procesamiento adicionales. El lisado  
45 celular puede además combinarse con uno o más sustratos, enzimas, nutrientes, cofactores, tampones, agentes reductores y/o sistemas generadores de ATP, según lo requiera la actividad enzimática. Dicho sistema, en ciertas realizaciones, se puede denominar en este documento "sistema libre de células", p. ej., un sistema aislado que contiene un lisado celular o extracto expresamente manipulado para sintetizar una enzima o cascada de enzimas que, cuando actúan en una frecuencia determinada (p. ej., en una vía enzimática) y proporción sobre un sustrato  
50 determinado, resulta en la generación preferencial de un producto, el compuesto de interés. Un compuesto de interés puede ser una entidad química (p. ej., una molécula orgánica pequeña), que se puede usar como ingrediente farmacéutico activo (API), precursor químico o intermedio.

Tal como se emplea en este documento, un "sustrato" es un compuesto o mezcla de compuestos capaz de proporcionar los elementos requeridos necesarios para sintetizar un compuesto de interés.

55 Tal como se emplea en este documento, una "molécula orgánica pequeña" o "molécula pequeña" se refiere a una molécula orgánica con un peso molecular de menos de 800 g/mol (p. ej., 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 determinadas

realizaciones, la molécula orgánica pequeña es un agente terapéuticamente activo tal como un fármaco (p. ej., una molécula orgánica pequeña aprobada por la Administración de Medicamentos y Alimentos de EE. UU. (U.S. Food and Drug Administration) según el Código de Regulaciones Federales (CFR)). La molécula orgánica pequeña puede además comprender un metal. En este caso, la molécula orgánica pequeña también se denomina "molécula organometálica pequeña".

Tal como se emplea en este documento, un "equivalente reductor" o "agente reductor" es una especie química que transfiere el equivalente de un electrón en una reacción rédox. Los ejemplos de equivalentes reductores son un electrón único (por ejemplo, en reacciones que implican iones metálicos), un átomo de hidrógeno (que consiste en un protón y un electrón) y un ion de hidruro (:H<sup>-</sup>) que porta dos electrones (por ejemplo, en reacciones que implican nicotinamida adenina dinucleótido (NAD<sup>+</sup>)). Un "aceptor equivalente de reducción" es una especie química que acepta el equivalente de un electrón en una reacción rédox.

Tal como se emplea en este documento, un "sistema de regeneración de adenosina trifosfato" o "sistema de regeneración de ATP" es un sistema químico o bioquímico que convierte adenosina, AMP y ADP en ATP. Los ejemplos de sistemas de regeneración de ATP incluyen aquellos que implican metabolismo de glucosa, metabolismo de glutamato y fotosíntesis.

Los lisados de células de diferentes orígenes genéticos (p. ej., previamente alterados o genomanipulados) o especies, o que se preparan por diferentes estrategias, se pueden mezclar y usar simultánea o secuencialmente en un proceso con el lisado celular. El lisado puede estar libre o inmovilizado, y se puede reutilizar o desechar en cada etapa del proceso. Por ejemplo, en determinadas realizaciones, el lisado celular es un lisado de un organismo de *E. coli* manipulado para sobreexpresar una o más enzimas en la vía de interés. En determinadas realizaciones, el lisado celular es una combinación de diferentes lisados celulares, p. ej., una combinación de dos, tres, cuatro, cinco, seis, siete, ocho, nueve o diez lisados celulares diferentes, obtenidos de dos, tres, cuatro, cinco, seis, ocho, nueve o diez organismos diferentes de *E. coli*, cada uno manipulado para sobreexpresar una o más enzimas en la vía de interés.

Los métodos descritos en este documento dan a conocer altos rendimientos del producto deseado, en donde el rendimiento es mayor que el rendimiento que se puede lograr con un hospedante microbiano nativo. La productividad (es decir, la tasa de producción por unidad de volumen o biomasa) también se puede aumentar. En una realización, el rendimiento del producto está por lo menos aproximadamente 2 veces encima de la tasa basal, por lo menos aproximadamente 5 veces encima de la tasa basal, por lo menos aproximadamente 10 veces encima de la tasa basal, por lo menos aproximadamente 25 veces encima de la tasa basal, por lo menos aproximadamente 50 veces encima de la tasa basal o más, p. ej., entre aproximadamente 2 veces y aproximadamente 100 veces encima de la tasa basal. En determinadas realizaciones, la tasa de rendimiento del producto que usa los métodos inventivos está entre aproximadamente 0,1 y 20 gramos de producto/L/h.

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

Las reacciones pueden utilizar un reactor a gran escala, pequeña escala, o se pueden multiplexar para efectuar una pluralidad de síntesis simultáneas. Las reacciones continuas usarán un mecanismo de alimentación para introducir un flujo de reactivos, y pueden aislar el producto final como parte del proceso. Los sistemas de lotes son también de interés, en donde se pueden introducir reactivos adicionales con el transcurso del tiempo para prolongar el período de tiempo de síntesis activa. Se puede operar un reactor en cualquier modo tal como lote, lote extendido, semi-lote, semi-continuo, lote alimentado y continuo, y se seleccionará según el propósito de aplicación.

Las reacciones pueden ser de cualquier volumen, o bien en una escala pequeña (p. ej., usualmente por lo menos aproximadamente 1 ml y no más de aproximadamente 15 ml) o en una reacción ascendente (p. ej., en donde el volumen de reacción es por lo menos aproximadamente 15 ml, usualmente por lo menos aproximadamente 50 ml, más usualmente por lo menos aproximadamente 100 ml, y puede ser 500 ml, 1000 ml, o más hasta un volumen de varios miles de litros). Las reacciones se pueden llevar a cabo a cualquier escala.

Se pueden incluir diversas sales y tampones, en donde las especies iónicas se pueden optimizar con respecto a la producción del producto. Al cambiar la concentración de un componente particular del medio de reacción, se puede cambiar en consecuencia otro componente. Además, los niveles de concentración en el reactor se pueden variar con el tiempo. El ajustador del potencial de oxidación/reducción de tiol/disulfuro puede ser ditiotreitól, ácido ascórbico, glutatión y/u otras formas oxidadas. Otros ajustadores con potencial de rédox general también se pueden utilizar.

En un modo de operación semi-continuo, el reactor puede operar en modo de diálisis, lote de diafiltración o lote alimentado. Se puede suministrar una disolución de alimentación 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 usual para purificación después de completar la operación del sistema. Alternativamente, el producto se puede eliminar durante el proceso o bien en un modo continuo o discontinuo con la opción de retornar parte o todos los compuestos remanentes al reactor.

5 Si hay un flujo de reactivos, la dirección del flujo de líquido puede ser perpendicular y/o tangencial a una membrana. El flujo tangencial es eficaz para reciclar ATP y para prevenir la obstrucción de las membranas, y se puede superponer en flujo perpendicular. El flujo perpendicular a la membrana se puede causar o efectuar con una bomba de presión positiva o una bomba de succión al vacío, o aplicando presión trans-membrana usando otros métodos conocidos en la técnica. La disolución en contacto con la superficie exterior de la membrana se puede cambiar cíclicamente, y puede estar en un flujo tangencial constante con respecto a la membrana. El reactor se puede agitar  
10 interna o externamente.

La cantidad de producto producido en una reacción se puede medir de diversas formas, por ejemplo, por ensayos enzimáticos que producen un producto coloreado o fluorométrico o por métodos de cromatografía de líquidos de alto rendimiento (HPLC). En determinadas realizaciones, el producto se mide utilizando un ensayo que mide la actividad o concentración del producto particular que se está produciendo.

15 Vías de interés

Tal como se emplea en este documento, la expresión "vía enzimática" o "vía de interés" se refiere a un sistema celular o acelular (libre de células) para convertir un sustrato a un producto de interés, en donde el sistema comprende una pluralidad de enzimas y puede además comprender sustratos accionados por una o más de las  
20 enzimas, productos de reacciones catalizadas por las enzimas, cofactores 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 vías metabólicas y se han descrito en sistemas microbianos, y están accesibles en bases de datos públicas; véanse, p. ej., Smolke, Ed., *The Metabolic Pathway Engineering Handbook: Tools and Applications*, CRC Press, Nueva York (2009); Stephanopoulos, Nielsen, and Aristidou, Eds., *Metabolic Engineering: Principles and Methodology*, Academic Press, Nueva York (1998); Greenberg, *Metabolic Pathways: Energetics, Tricarboxylic Acid Cycle, and Carbohydrates*, Academic Press, Nueva York (1967); y la serie de múltiples volúmenes de D.M. Greenberg, titulada *Metabolic pathways*, Volúmenes 1-7.

Las vías de interés incluyen, por ejemplo, vías implicadas en la biosíntesis de carbohidratos, aminoácidos, ácidos nucleicos, esteroides, ácidos grasos y productos naturales, y abarcan la síntesis de compuestos químicos y materiales como, entre otros:

30 a) antibióticos; p. ej., actinomicina, bleomicina, rifamicina, cloranfenicol, carbapenemas, tetraciclina, lincomicina, eritromicina, estreptomina, ciclohexamida, puromicina, cicloserina, bacitracina, penicilina, cefalosporina, vancomicina, polimixina y gramicidina;

b) biotensioactivos; p. ej., ramnolípidos, soforolípidos, glicolípidos y lipopéptidos;

c) combustibles biológicos, p. ej., bioetanol, biodiésel y biobutanol;

35 d) aminoácidos, p. ej., L-glutamato, L-lisina, L-fenilalanina, ácido L-aspártico, L-isoleucina, L-valina, L-triptófano, L-prolina (hidroxiprolina), L-treonina, L-metionina, L-tirosina y D- $\alpha$ -hidroxifenilglicina;

e) ácidos orgánicos; p. ej., ácido cítrico, ácido láctico, ácido glucónico, ácido acético, ácido propiónico, ácido succínico, ácido fumárico y ácido itacónico;

f) ácidos grasos, p. ej., ácido araquidónico, ácido graso poliinsaturado (PUBA) y ácido  $\alpha$ -linoleico;

40 g) alcoholes y polioles; p. ej., glicerol, manitol, eritritol, xilitol, poli-3-hidroxi-butirato, isobutanol y 1-butanol;

h) saporíferos y fragancias, p. ej., vainilla, benzaldehído, dihidroxiacetona, 4-(R)-decanólido y 2-actil-1-pirrolina;

i) nucleótidos; p. ej., ácido 5'-guanílico y ácido 5'-inosínico;

j) vitaminas; p. ej., vitamina C, vitamina F, vitamina B2, provitamina D2, vitamina B12, ácido fólico, nicotinamida, biotina, ácido 2-ceto-L-gulónico y provitamina Q10;

45 k) pigmentos; p. ej., astaxantina,  $\beta$ -caroteno, leucopeno, monascorubrina y rubropunctatina;

l) azúcares y polisacáridos; p. ej., ribosa, sorbosa, xantano, gelano y dextrano; y

m) biopolímeros y plásticos, p. ej., polihidroxialcanoatos (PHA), ácido poli- $\gamma$ -glutámico y 1,3-propanodiol.

Otros ejemplos de vías de interés incluyen la síntesis de varios metabolitos de *E. coli*. Un "metabolito" es cualquier sustancia utilizada o producida durante el metabolismo (p. ej., una enzima, sustrato o producto). En la presente  
50 memoria, un metabolito es a menudo, aunque no siempre, el producto de una enzima en la vía de interés. Los

metabolitos de *E. coli* ilustrativos incluyen, aunque sin limitarse a ello, ácido 2,3-dihidroxibenzoico, 2-cetoglutarato, 3-fosfoglicerato, 4-hidroxibenzoato, 6-fosfogluconato, acetoacetil-CoA, acetil-CoA, acetilfosfato, adenina, adenosina, adenosina fosfosulfato, 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, desoxirribosa-5-P, dGMP, dihidroorotato, dihidroxiacetona fosfato, dTDP, dTTP, eitrosa-4-fosfato, FAD, flavina mononucleótido, fructosa-1,6-bisfosfato, fructosa-6-fosfato, fumarato, GDP, gluconato, gluconolactona, glucosamina-6-fosfato, glucosa-6-fosfato, glucosa-1-fosfato, glutamato, glutamina, glutatión, glutatión disulfuro, gliceraldehído-3-fosfato, glicerato, glicerol-3-fosfato, GMP, GTP, guanina, guanosina, histidina, histidinol, homocisteína, inosina difosfato, inosina monofosfato, inosina trifosfato, isoleucina, lisina, malato, malonil-CoA, metionina, mio-inositol, N-Acetil-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, shikimato, succinato, succinil-CoA, treonina, triptófano, tirosina, UDP, UDP-glucosa, UDP-glucuronato, UDP-N-acetilglucosamina, uridina, UTP, valina y xilulosa-5 -fosfato.

En determinadas realizaciones, la vía de interés proporciona la síntesis de ácido shikímico y/o shikimato (el shikimato es la forma aniónica de ácido shikímico) y sus intermedios sintéticos, un isoprenoide o terpeno (p. ej., amorfadieno, farneseno, licopeno, astaxantina, vitamina A, mentol, beta-caroteno), poli-3-hidroxibutirato, isobutanol y 1-butanol.

Se puede catalizar un número de reacciones por enzimas en una vía de interés. Las clases generales de enzimas, que se pueden identificar por número de clasificación de la enzima, provisto entre paréntesis, incluyen, aunque sin limitarse a ello:

(EC 1) oxidorreductasas; p. ej., deshidrogenasas, oxidasas, reductasas, oxidorreductasas, sintasas, oxigenasas, monooxigenasas, dioxygenasas, lipoxigenasas, hidrogenasas, tránshidrogenasas, peroxidasas, catalasas, epoxidasas, hidroxilasas, demetilasas, desaturasas, dismutasas, hidroxiltransferasas, deshalogenasas y desyodinasas;

(EC2) transferasas; p. ej., transaminasas, cinasas, dicinasas, metiltransferasas, hidroximetiltransferasas, formiltransferasas, formiminotransferasas, carboxitransferasas, carbamoiltransferasas, amidinotransferasas, transaldolasas, transcetolasas, acetiltransferasas, aciltransferasas palmitoiltransferasas, succiniltransferasas, maloniltransferasas, galoiltransferasas, sinapoiltransferasas, tigloiltransferasas, tetradecanoiltransferasas, hidroxicinnamoiltransferasas, feruloiltransferasas, miciloiltransferasas, benzoiltransferasas, piperoiltransferasas, trimetiltridecanoiltransferasa, miristoiltransferasas, coumaroiltransferasas, tiolasas, aminoaciltransferasas, fosforilasas, hexosiltransferasas, pentosiltransferasas, sialiltransferasas, piridinilasas, difosforilasas, ciclotransferasas, sulfurilasas, adenosiltransferasas, carboxiviniltransferasas, isopenteniltransferasas, aminocarboxipropiltransferasas, dimetilaliltransferasas, farnesiltransferasas, hexapreniltransferasas, decaprenilcistransferasas, pentapreniltransferasas, nonapreniltransferasas, geranilgeraniltransferasas, aminocarboxipropiltransferasas, oximinotransferasas, purinatransferasas, fosfodismutasas, fosfotransferasas, nucleotidiltransferasas, polimerasas, colinafosfotransferasas, fosforilmutasas, sulfurtransferasas, sulfotransferasas y CoA-transferasas;

(EC3) hidrolasas; p. ej., lipasas, esteratas, amilasas, peptidasas, hidrolasas, lactonasas, desacilasas, desacetilasas, feoforbidasas, depolimerasas, tiolesteratas, fosfatasas, difosfatasas, trifosfatasas, nucleotidasas, fitasas, fosfodiesteratas, fosfolipasas, sulfatasas, ciclasas, oligonucleotidasas, ribonucleasas, exonucleasas, endonucleasas, glicosidasas, nucleosidasas, glicosilasas, aminopeptidasas, dipeptidasas, carboxipeptidasas, metalocarboxipeptidasas, omega-peptidasas, serina endopeptidasas, cisteína endopeptidasas, endopeptidasas aspárticas, metaloendopeptidasas, treonina endopeptidasas, aminosas, amidasas, desuccinilasas, deformilasas, acilasas, desiminasas, desaminasas, dihidrolasas, ciclohidrolasas, nitrilasas, ATPasas, GTPasas, halidasas, deshalogenasas y sulfhidrolasas;

(EC 4) liasas; p. ej., descarboxilasas, carboxilasas, carboxicinasas, aldolasas, epoxilasas, oxoácido-liasas, liasas carbono-carbono, deshidratasas, hidratasas, sintasas, endoliasas, exoliasas, amoníaco-liasas, amidina-liasas, amine-liasas, liasas carbono-azufre, liasas carbono-haluro, liasas fósforo-oxígeno y deshidroclorinasas;

(EC 5) isomerasas; p. ej., isomerasas, racemasas, mutasas, tautomeratas, fosfomutasas, fosfoglucomutasas, aminomutasas, cicloisomerasa, ciclasas, topoisomerasas; y

(EC 6) ligasas; p. ej., sintetatas, tNRA-ligasas, ligasas ácido-tiol, amida sintasas, péptido sintasas, cicloligasas, carboxilasas, ADN-ligasas, ARN-ligasas y ciclasas.

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

Las oxidorreductasas ilustrativas incluyen, aunque sin limitarse a ello:

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

## ES 2 839 424 T3

- (EC 1.2) oxidorreductasas que actúan sobre el grupo de donantes aldehído u oxo, y un aceptor;
- (EC 1.3) oxidorreductasas que actúan sobre el grupo de donantes CH-CH, y un aceptor;
- (EC 1.4) oxidorreductasas que actúan sobre el grupo de donantes CH-NH<sub>2</sub>, y un aceptor;
- (EC 1.5) oxidorreductasas que actúan sobre el grupo de donantes CH-NH, y un aceptor;
- 5 (EC 1.6) oxidorreductasas que actúan sobre NADH o NADPH, y un aceptor;
- (EC 1.7) oxidorreductasas que actúan sobre otros compuestos de nitrógeno como donantes, y un aceptor;
- (EC 1.8) oxidorreductasas que actúan sobre un grupo de donantes azufre, y un aceptor;
- (EC 1.9) oxidorreductasas que actúan sobre un grupo de donantes hemo, y un aceptor;
- (EC 1.1) oxidorreductasas que actúan sobre difenoles y sustancias relacionadas como donantes, y un aceptor;
- 10 (EC 1.11) oxidorreductasas que actúan sobre un peróxido como aceptor;
- (EC 1.12) oxidorreductasas que actúan sobre hidrógeno como donante, y un aceptor;
- (EC 1.13) oxidorreductasas que actúan sobre donantes individuales con incorporación de oxígeno molecular, que incorpora uno o dos átomos de oxígeno;
- (EC 1.14) oxidorreductasas que actúan sobre donantes apareados, con incorporación o reducción de oxígeno molecular, en donde el donante es 2-oxoglutarato, NADH, NADPH, flavina reducida, flavoproteína, pteridina, proteína de hierro-azufre, ascorbato;
- 15 (EC 1.15) oxidorreductasas que actúan sobre radicales de superóxido como aceptor;
- (EC 1.16) oxidorreductasas que oxidan iones metálicos, y un aceptor;
- (EC 1.17) oxidorreductasas que actúan sobre grupos CH o CH<sub>2</sub>, y un aceptor;
- 20 (EC 1.18) oxidorreductasas que actúan sobre proteínas de hierro-azufre como donantes, y un aceptor;
- (EC 1.19) oxidorreductasas que actúan sobre flavodoxina reducida como donante, y un aceptor;
- (EC 1.2) oxidorreductasas que actúan sobre donantes de fósforo o arsénico en donantes, 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; en donde los aceptores para cada categoría de donante pueden incluir, aunque sin limitarse a ello: NAD, NADP, proteína hemo, oxígeno, disulfuro, quinona, una proteína hierro-azufre, una flavina, un grupo nitrógeno, un citocromo, dinitrógeno y H<sup>+</sup>.
- 25
- Las transferasas ilustrativas incluyen, aunque sin limitarse a ello:
- (EC 2.1) transferasas que transfieren grupos on-carbono;
- (EC 2.2) transferasas que transfieren grupos aldehído o cetónicos;
- 30 (EC 2.3) Aciltransferasas;
- (EC 2.4) Glicosiltransferasas;
- (EC 2.5) transferasas que transfieren grupos alquilo o arilo, distintos de grupos metilo;
- (EC 2.6) transferasas que transfieren grupos nitrogenados;
- (EC 2.7) transferasas que transfieren grupos que contienen fósforo;
- 35 (EC 2.8) transferasas que transfieren grupos que contienen azufre; y
- (EC 2.9) transferasas que transfieren grupos que contienen selenio.
- Las hidrolasas ilustrativas incluyen, aunque sin limitarse a ello:
- (EC 3.1) hidrolasas que actúan sobre enlaces éster;
- (EC 3.2) Glicosilasas;

(EC 3.3) hidrolasas que actúan sobre enlaces éter;

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

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

(EC 3.6) hidrolasas que actúan sobre anhídridos ácidos;

5 (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 fósforo-nitrógeno;

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

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

10 (EC 3.12) hidrolasas que actúan sobre enlaces azufre-azufre; e

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

Las liasas ilustrativas incluyen, aunque sin limitarse a ello:

(EC 4.1) Liasas carbono-carbono;

(EC 4.2) Liasas carbono-oxígeno;

15 (EC 4.3) Liasas carbono-nitrógeno;

(EC 4.4) Liasas carbono-azufre;

(EC 4.5) Liasas carbono-haluro; y

(EC 4.6) Liasas fósforo-oxígeno.

Las isomerasas ilustrativas incluyen, aunque sin limitarse a ello:

20 (EC 5.1) Racemasas y epimerasas;

(EC 5.2) cis-trans-Isomerasas;

(EC 5.3) Isomerasas intramoleculares;

(EC 5.4) Transferasas intramoleculares (mutasas); y

(EC 5.5) Liasas intramoleculares.

25 Las ligasas ilustrativas incluyen, aunque sin limitarse a ello:

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

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

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

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

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

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

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

35 Las enzimas implicadas en una vía de interés o vía asociada pueden también clasificarse de acuerdo con la función de la enzima. Las enzimas de implicancia directa (clase 1) en una célula o lisado celular catalizan una reacción en la vía. Es típico de las vías que dichas enzimas directas sean una de una cadena, en 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 resulta en el producto de interés. Las enzimas de

implicancia indirecta (clase 2) en una célula o lisado celular reaccionan en una vía asociada, usualmente en la producción de un sustrato utilizado en la vía de interés.

Dentro de una vía, las enzimas suelen variar en tasa de recambio y en la efectividad con la cual se produce un producto. En consecuencia, ciertas enzimas en una vía se tornan limitantes de la tasa. Aumentar la concentración de enzimas limitantes de la tasa en una vía (en relación a las enzimas no limitantes de la tasa) permite el aumento del flujo a través de la vía de interés (véase, p. ej., Zamboni et al. Nature Protocols (2009) 4:878-892). A menudo, las tasas limitantes de la tasa se asocian con toxicidad cuando se producen excesivamente, y por lo tanto las concentraciones disponibles de dichas enzimas convenientemente se modulan. Véase, p. ej., la publicación PCT núm. WO 2010/077806.

Una tercera clase de enzimas en una célula o lisado celular son enzimas competitivas en una célula o lisado celular (clase 3), que utilizan un sustrato o producto de la vía de interés. Una característica de una enzima competitiva es que la cinética de conversión del sustrato es lo suficientemente alta de forma tal que la presencia de la enzima disminuye el rendimiento general y/o la tasa de producción del producto final deseado catalizado por la vía de interés. Una célula normal puede requerir la expresión de enzimas competitivas, y, por lo tanto, en lugar de inactivar la expresión de células competitivas completamente, es conveniente reducir selectivamente la concentración de la enzima activa por los métodos que se describen en este documento.

Para practicidad en cuanto a denominación, una enzima en la vía se puede categorizar como una primera enzima de entrada a la vía, o una enzima o enzimas en una fase posterior. Para fines prácticos, la enzima de entrada a la vía se puede denominar en este documento  $E_1$ , y las enzimas de fases posteriores se pueden numerar en forma consecutiva,  $E_2$ ,  $E_3$ , ...  $E_n$ . Las vías de interés para uso en los métodos descritos en este documento por lo general comprenderán por lo menos una enzima, por lo menos dos enzimas, por lo menos tres enzimas, por lo menos cuatro enzimas o más, p. ej., 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.

Las enzimas en una vía pueden ocurrir naturalmente o se pueden modificar para optimizar una característica particular de interés, p. ej., especificidad del sustrato, cinética de reacción, solubilidad y/o insensibilidad a la inhibición de retroalimentación. A su vez, en algunos casos, el gen que expresa la enzima se optimizará para uso de codones dentro de la célula hospedante. En algunas realizaciones, la vía completa comprende enzimas de un solo organismo, no obstante, esto no se requiere, y también se contempla combinar enzimas de múltiples organismos. Para algunos propósitos, una vía puede ser endógena a la célula hospedante, pero esto no se requiere, y se puede introducir una vía completa o componentes de una vía en una célula hospedante. Si el sistema se provee en una célula intacta, el conjunto completo de enzimas de la vía de interés puede estar presente en la célula. Para los fines de producción libre de células, se pueden añadir una o más enzimas al lisado, o alternativamente pueden ser producidas por el lisado, como para completar la vía. La enzima se puede aislar y/o purificar.

En el sistema de la vía, un primer sustrato ( $S_1$ ) es accionado por la enzima de entrada a la vía, y se convierte a un primer producto, aunque el experto en la técnica comprenderá que una enzima puede accionar más de un sustrato simultáneamente, y puede producir más de un producto, de forma tal que dos o más vías puedan interconectarse en una sola enzima. El primer producto es un sustrato ( $S_2$ ) para una enzima de una fase posterior  $E_2$ , y es convertido a un segundo producto por  $E_2$ . Dependiendo de la complejidad de la vía, el segundo producto puede ser el producto final (PF), o puede ser un sustrato ( $S_3$ ) para una tercera enzima de una fase posterior ( $E_3$ ), y es convertido a un tercer producto por  $E_3$ , que puede ser un sustrato ( $S_4$ ) para una cuarta enzima. La enzima final en la vía, que puede ser, por ejemplo,  $E_2$ ,  $E_3$  o  $E_4$ , produce el producto de interés (PF). Es una característica de las vías enzimáticas que el producto de una enzima sea el sustrato para la enzima siguiente.

Los productos pueden ser estables o relativamente lábiles, pero en algunos casos el producto final es lo suficientemente estable como para ser aislado de la célula, lisado celular o mezcla de reacción. Las enzimas competitivas utilizan un sustrato o producto de la vía de interés, que puede incluir uno cualquiera de PF,  $S_1$ ,  $S_2$ ,  $S_3$  y/o  $S_4$ , y se puede denominar enzimas competitivas ( $E_c$ ).

En algunas realizaciones, el sustrato inicial,  $S_1$ , es un metabolito central, o "commodity" celular. Las vías centrales de metabolismo incluyen glicólisis y el ciclo del ácido cítrico. Dichos compuestos  $S_1$  pueden no ser específicos de la vía de interés, pero son compuestos que se hallan ampliamente en varias células y son sustratos para múltiples enzimas y vías. Los ejemplos de sustratos commodity incluyen, entre otros, glucosa, ATP, piruvato, fosfoenol piruvato y similares. Una enzima de entrada a la vía,  $E_1$ , puede convertir un sustrato commodity a un producto que es un sustrato selectivo para una o un número relativamente pequeño de enzimas. En algunas realizaciones, los compuestos  $S_1$  o commodities celulares se añaden al lisado celular.

En algunas realizaciones, una enzima de entrada clave se define como una que realiza la primera etapa en una vía a un producto de interés. Esta etapa puede implicar el compromiso bioquímico de un compuesto a la vía de un producto de interés. Los ejemplos de enzimas de entrada clave incluyen, entre otros, aquellos expuestos en la Tabla 1.

Tabla 1: Lista ilustrativa de enzimas clave de entrada a la vía			
Enzima(s) clave de entrada	Vía biosintética	Productos ilustrativos	enzima <i>E. coli</i>
amidofosforibosil transferasa	biosíntesis de purina	GMP, GDP, GTP, dGDP, dGTP, AMP, ADP, ATP, dADP, dATP, inosina monofosfato	PurF
orotato fosforibosiltransferasa	biosíntesis de pirimidina	UMP, UDP, UTP, CDP, CTP	PyrE
2-deshidro-3-desoxifosfoheptonato aldolasa	biosíntesis de corismato	Shikimato, Tirosina, Fenilalanina, Triptófano	AroE,F,G
fosforibosiltransferasa HisG	biosíntesis de histidina	Histidina	HisG
acetolactato / acetohidroxibutanoato sintasa	biosíntesis de isoleucina, leucina, valina	Isoleucina, Leucina, Valina	IlvH,M,N
UDP-N-acetilglucosamina aciltransferasa	biosíntesis de lipopolisacárido	Disacárido de lípidos A	LpxA
aspartato aminotransferasa	biosíntesis de lisina, treonina y metionina	Lisina, Treonina, Metionina	AspC
arginine descarboxilasa	biosíntesis de putrescina	Putrescina	SpeA
GTP ciclohidrolasa I	biosíntesis de tetrahidrofolato	Tetrahidrofolato	FolE
acetil-CoA carboxilasa	biosíntesis de ácido graso	Malonil-CoA	AccA,B,C,D

Un ejemplo no limitativo específico de una vía, provisto para propósitos ilustrativos, es la vía para la síntesis de ácido shikímico. En esta vía, por ejemplo, una reacción entre los compuestos commodity celulares fosfoenolpiruvato ( $S_{1A}$ ) y eritrosa-4-fosfato ( $S_{1B}$ ) es catalizada por la enzima DAHP sintasa ( $E_1$ ) para formar 3-desoxi-D-arabinoseheptulosa-7-fosfato (DAHP). DAHP ( $S_2$ ) se transforma a 3-deshidroquinato (3-DHQ) por la segunda enzima en la vía, DHQ sintasa ( $E_2$ ). 3-DHQ ( $S_3$ ) es deshidratado a 3-deshidros hikimato por la tercera enzima en la vía, 3-DHQ deshidratasa ( $E_3$ ). 3-deshidros hikimato ( $S_4$ ) se reduce a ácido shikímico (PF) por la cuarta enzima en la vía, shikimato deshidrogenasa ( $E_4$ ), usando NADPH como cofactor. Las enzimas de la vía se conocen en la técnica y se han caracterizado en un número de organismos, incluida, por ejemplo, *E. coli*, en donde las enzimas son codificadas por los locus genéticos de la siguiente manera: DAHP sintasa (aroG, aroF, aroH); DHQ sintasa (aroB); 3-DHQ deshidratasa (aroD); shikimato deshidrogenasa (aroE); véase, p. ej., la publicación PCT núm. WO2010/074760.

#### Flujo metabólico

"Flujo" o "flujo metabólico" se refiere a la tasa en que las moléculas pasan por una vía o reacción de interés. Entre los factores del flujo de control están la tasa de catálisis de enzimas en la vía, la disponibilidad de sustrato, la concentración de enzimas en una célula y/o la proximidad de enzimas en una vía.

Si bien se desea una alta tasa de flujo a través de una vía de interés, al mismo tiempo puede generar problemas de toxicidad si un producto no normalmente acumulado en altos niveles en la célula se produce en una alta tasa en relación con aquella que ocurre bajo condiciones normales. Se entiende que una alta tasa de flujo es específica de la vía, y se refiere a la concentración del producto de la vía durante el transcurso del tiempo, como por ejemplo, la producción de un producto a una tasa de aproximadamente 0,1 gramos (g) a aproximadamente 20 g de producto/L/h.

En algunas realizaciones, la tasa de producción de producto es aproximadamente 0,5, aproximadamente 1, aproximadamente 2, aproximadamente 5, aproximadamente 10 o aproximadamente 15 gramos de producto/L/h.

Los métodos para determinar las tasas de flujo se conocen y estudian en la técnica; véanse, p. ej., Wiechert et al., *Metab. Ing.* (2001) 3:265-283; Wiechert et al., *Metab. Ing.* (2001)3:195-206; y textos de ingeniería metabólica tales como Lee and Papoutsakis, Eds., *Metabolic Engineering*, Marcel Dekker, Inc. Nueva York (1999); Stephanopoulos, Nielsen, and Aristidou, Eds., *Metabolic Engineering: Principles and Methodology*, Academic Press, Nueva York (1998); y Nielsen and Eggeling, Eds., *Metabolic Engineering*, Springer, Londres (2001). El flujo se puede calcular a partir de cantidades mensurables usando técnicas tales como análisis de flujo metabólico (MFA), por ejemplo, por medición directa de la tasa de conversión de sustrato isotópicamente marcado.

#### 10 Proteasa específica del sitio

Tal como se emplea en este documento, la expresión se refiere a una proteasa, por ejemplo, una endoproteasa, que se escinde selectivamente en un motivo de aminoácidos específico. En algunas realizaciones, el motivo de aminoácidos es un motivo de por lo menos 4 residuos de aminoácidos para reducir la escisión de proteína de fondo, y puede ser un motivo de por lo menos 5 aminoácidos, por lo menos 6 aminoácidos o más. Los expertos en la técnica conocen dichas proteasas, e incluyen, sin limitación, proteasa del virus del tabaco (ENLYFQ<sup>G/s</sup>); proteasa del virus de la fiebre amarilla (GARR<sup>G/s</sup>); trombina (LVPRGS); y Factor Xa (I<sup>F</sup>/DGR).

En determinadas realizaciones, se describen en esta memoria métodos para generar proteasas específicas del sitio que tienen una secuencia de reubicación periplásmica. La síntesis de genes que codifican cada proteasa específica del sitio permite la optimización de codones para expresión en células de *E. coli*. La reacción en cadena de la polimerasa (PCR), u otros métodos familiares para los expertos en la técnica, se pueden usar para la adición de la secuencia de reubicación periplásmica. Los genes se sub-clonan en vectores que permiten expresión controlada (p. ej., vectores que permiten la expresión controlada usando el sistema de inducción T7, como un vector de la serie pET).

Nombre	MEROPS	UniProt	Sitio específico
TEV Nia	C04.004	P04517 (2036-2279)	E-N-L-Y-F-Q↓G (SEQ ID NO:12)
HRV 3C	C03.007	P04936 (1508-1690)	L-E-V-L-F-Q↓G-P (SEQ ID NO:13)
Enterocinasa	S01.156	P98072 (801-1035) bovino	D-D-D-D-K↓ (SEQ ID NO:14)
Factor Xa	S01.216	P00743 (234-468) bovino	I-E-G-R↓ (SEQ ID NO:15)
Trombina (flla)	S01.217	P00735 (367-623) bovino	L-V-P-R↓G (SEQ ID NO:16)

En determinadas realizaciones, se introduce una secuencia codificante de proteasa específica del sitio HRV 3V en la célula microbiana. La secuencia codificante de proteasa específica del sitio puede estar operativamente enlazada a un promotor constitutivo o regulable, p. ej., un promotor inducible, que se puede denominar en este documento constructo de expresión de proteasas específicas del sitio. El constructo de expresión de proteasas específicas del sitio se puede proveer en un vector episomal, p. ej., un plásmido, YAC, BAC o vector vírico, como se conoce en la técnica. Alternativamente, el constructo de expresión de proteasas específicas del sitio se puede integrar en el cromosoma de la célula hospedante microbiana.

En algunas realizaciones, la tasa de expresión de la proteasa específica del sitio se ajusta experimentalmente para optimizar la eficiencia de exportación de la proteasa fuera de la célula o hacia el periplasma de la célula. La translocación deficiente puede provocar una capacidad insuficiente de la maquinaria de exportación. Los métodos para ajustar la tasa de expresión incluyen, entre otros, modificación del número de copias del plásmido que porta el gen que codifica la proteasa específica del sitio que se ha de exportar al periplasma. Los replicones conocidos y utilizados en la técnica incluyen P15A (10 copias/célula), ColA (30 copias/célula), ColE1 (40 copias/célula) y RSF1030 (>100 copias/célula). El sitio de unión al ribosoma en 5' UTR del gen que codifica la proteasa específica del sitio que se ha de exportar al periplasma se puede modificar, en donde se puede crear una biblioteca de sitios de unión al ribosoma con fortalezas variables y ensayar para uso en las realizaciones descritas en la presente invención, véanse *Nature Biotechnology* 27: 946-950, 2009; *Nature Biotechnology* 14:629, 1996. La región del promotor en dirección 5' del gen que codifica la proteína, p. ej., una proteasa específica del sitio, que se ha de exportar se puede modificar para ajustar la tasa de transcripción, en donde se puede crear una biblioteca de regiones promotoras con

fortalezas variables y ensayar para uso en las realizaciones descritas en la presente invención, véase PNAS 102:12678-12683, 2005; BMC Biotechnology 7:34, 2007.

#### Sitio de escisión de proteasas

5 Como se describió anteriormente, una proteasa específica del sitio escinde enzimas diana en un modo específico de la secuencia, p. ej., en ENLYFQ<sup>G/S</sup>, GARR<sup>G/S</sup>, LVPRGS o I<sup>E/D</sup>GR. Los métodos descritos en la presente memoria pueden utilizar, en parte, una modificación de secuencias genéticas diana para preparar una enzima diana lábil para digestión de proteasa específica del sitio y consecuente inactivación. Asimismo, la proteasa específica del sitio se puede seleccionar de forma tal que las enzimas de la vía de interés no sean escindidas por la proteasa.

10 La enzima dirigida para escisión de proteasas específicas del sitio puede ser una enzima que compita con una enzima en la vía de interés, p. ej., una enzima que utilice un sustrato o producto de la vía y al hacerlo reduzca el flujo a través de la vía o reduzca la producción del producto deseado. La enzima dirigida puede ser alterada por una inserción o sustitución de aminoácidos para generar un sitio de escisión de proteasas específicas del sitio, por ejemplo, por una sustitución neutra con cambios conservadores o sin cambios de aminoácidos. Por "sustitución neutra" se entiende que el aminoácido en una posición dentro del motivo (p. ej., sitio de reconocimiento de proteasas específicas del sitio) se sustituye con un aminoácido que mantiene la actividad de la enzima diana. Dicha secuencia de aminoácidos se puede denominar "secuencia lábil de proteasas". En determinadas realizaciones, el sitio para escisión se selecciona de modo tal que, después de la escisión, la actividad de la enzima diana se destruye. En algunas realizaciones, la actividad de la enzima no se destruye por completo (p. ej., 100% inactiva, o se reduce al 100%), sino que se reduce al 99%, aproximadamente 90%, aproximadamente 80%, aproximadamente 70%,  
20 aproximadamente 60% o aproximadamente 50%. En algunas realizaciones, se considera que una enzima diana está inactiva si el nivel de actividad enzimática es tal que la enzima diana no es capaz de competir con enzimas en la vía de interés.

25 Una enzima diana se puede seleccionar por análisis de la vía para identificar enzimas que desvían/desaceleran el flujo hacia fuera del producto de interés o que catalizan una reacción secundaria que desvía el flujo de un producto de interés. Se puede determinar primero si estas actividades enzimáticas se pueden eliminar de la cepa hospedante (p. ej., a través de inactivación genómica) sin afectar adversamente el crecimiento celular. En casos en los que la eliminación de la actividad enzimática perjudicaría sustancialmente o afectaría en algún modo negativamente la salud metabólica del hospedante, la enzima(s) que realiza esta actividad se dirige para degradación por una proteasa específica del sitio reubicada. De esta manera, se puede incrementar el flujo hacia el producto de interés.

30 En algunas realizaciones, se crea inicialmente una biblioteca de secuencias de enzimas diana con sitios de escisión de proteasa en posiciones variables para determinar el sitio de reconocimiento específico del sitio óptimo y/o la localización del sitio. Las posiciones se pueden seleccionar en base a la estructura (p. ej., estructura terciaria) de la enzima diana, por ejemplo, insertando el sitio de reconocimiento específico del sitio en regiones bucle expuestas en la superficie. Si no está disponible la información estructural para la enzima diana, el posicionamiento del sitio de reconocimiento se puede determinar comparando secuencias de aminoácidos de homólogos de enzimas diana e identificando regiones hidrófilas con alta variabilidad y longitud de secuencias. Los residuos hidrófilos se exponen comúnmente en la superficie de la enzima, y la alta variabilidad y longitud de las secuencias puede indicar que la región no es crítica para la función de la enzima y puede tolerar la modificación de las secuencias. Por ejemplo, véanse Journal of Virology 78:5079, 2004; patente de EE. UU. 7223390, que describen que se puede usar digestión proteolítica limitada seguida por co-fraccionamiento cromatográfico y secuenciación de polipéptidos/espectroscopia de masas N-terminal para determinar aquellos sitios candidatos que sean accesibles para digestión de proteasas y no esenciales para formar una estructura de enzimas/proteínas integral.

35 La modificación de una secuencia de aminoácidos de la enzima diana para introducir un sitio de reconocimiento específico del sitio se puede obtener por métodos de ingeniería de genes conocidos en la técnica, por ejemplo, usando métodos basados en PCR y basados en ligadura, o por métodos de modificación genómica conocidos por el experto en la técnica. Véanse, por ejemplo, Phusion site-directed mutagenesis kit (New England Biolabs); Journal of Virological Methods 149:85, 2008; Nucleic Acids Research 32:e174, 2004; Journal of Bacteriology 180: 2063, 1998; Proceedings of the National Academy of Sciences of the United States of America 97:6640, 2000.

#### Secuestro periplásmico

50 En algunas realizaciones, la reubicación de proteasas específicas del sitio es en el espacio periplásmico de la célula. En dichos aspectos, se describen en este documento métodos para generar células, lisados y sus usos, mediante los cuales se modifica genéticamente una proteasa para incorporar una secuencia de reubicación de péptidos que proporciona el direccionamiento periplásmico de la proteasa. Las secuencias de reubicación periplásmica (también denominadas en este documento secuencias de péptidos de direccionamiento, señales diana, péptidos de señalización o secuencias de señalización), en algunas realizaciones, se encuentran en el término N de las proteínas segregadoras de bacterias. Varían en longitud de aproximadamente 15 a aproximadamente 70 aminoácidos. Las secuencias primarias de aminoácidos también varían, pero pueden tener una estructura general común que incluya las siguientes partes: (i) una parte N-terminal que tiene una longitud variable y que porta, por ejemplo, una carga positiva neta; (ii) un núcleo hidrófobo central de aproximadamente 6 a aproximadamente 15

aminoácidos; y (iii) aproximadamente 4 a aproximadamente 6 aminoácidos que definen el sitio de escisión para peptidasas de señalización.

Las secuencias de reubicación periplásmica adecuadas para uso en este documento pueden derivar de una proteína que se segrega en una bacteria gramnegativa. La proteína segregada puede ser codificada por la bacteria o por un bacteriófago que infecta la bacteria. Los ejemplos de fuentes bacterianas gramnegativas de proteínas segregadas incluyen, aunque sin limitarse a ello, miembros de los géneros *Escherichia*, *Pseudomonas*, *Klebsiella*, *Salmonella*, *Caulobacter*, *Metilomonas*, *Acetobacter*, *Achromobacter*, *Acinetobacter*, *Aeromonas*, *Agrobacterium*, *Alcaligenes*, *Azotobacter*, *Burkholderia*, *Citrobacter*, *Comamonas*, *Enterobacter*, *Erwinia*, *Rhizobium*, *Vibrio* y *Xanthomonas*.

Existen tres vías para translocación de proteínas en las membranas periplásmicas: (i) vías de translocación dependientes de SecB, (ii) partícula de reconocimiento de señales (SRP) y (iii) translocación de arginina gemela (TAT). Las vías de translocación dependiente de SecB y partícula de reconocimiento de señales usan ambas el translocón SecYEG. La vía de translocación de arginina gemela usa el complejo TatABCE. La translocación dependiente de SecB es la más comúnmente utilizada, pero esta vía no es capaz de transportar proteínas plegadas. El plegado citoplásmico rápido puede necesitar el uso de la vía TAT. Los ejemplos de proteínas segregadas bacterianas que tienen secuencias de reubicación periplásmica incluyen, aunque sin limitarse a ello, proteínas codificadas por los siguientes genes: *ompA*, *genelll*, fosfatasa alcalina de *E. coli*, *lamB*, *malE*, *secE*, *secY* y *prlA-4*. El experto en la técnica puede identificar la secuencia de reubicación periplásmica localizada en el término N de cada una de estas proteínas y otras proteínas segregadoras bacterianas. El experto en la técnica también sabe que se pueden efectuar algunas sustituciones, adiciones y/o eliminaciones de aminoácidos en secuencias de reubicación periplásmica a la vez que se retiene la función de direccionamiento periplásmico. Por consiguiente, una secuencia de reubicación periplásmica funcional descrita en este documento puede ser una secuencia totalmente natural o modificada.

En algunas realizaciones, la proteasa específica del sitio se reubica en una ubicación extracitoplásmica. Las señales para segregación en células bacterianas se conocen en la técnica, por ejemplo, translocónes sec y Tat en *B. subtilis* (véanse para revisión Buist et al. (2006) *Microbiology* 152(Pt 10):2867-74; y Tjalsma et al. (2004) *Microbiol Mol Biol Rev.* 68(2):207-33). La exportación de ClyA de células bacterianas también se conoce en la técnica, por ejemplo, véase Ludwig et al. (2010) *J Bacteriol.* 192(15):4001-11.

En otras realizaciones, la proteasa específica del sitio se reubica en un compartimiento nuclear. Las señales de péptidos para reubicación nuclear se conocen en la técnica, por ejemplo, el péptido de antígenos de células T grandes SV40: PKKKRKV (véase Kalderon et al. (1984) *Cell* 39 (3 Pt 2): 499-509); y el péptido de nucleoplasmina: KR[PAATKKAGQA]KKKK (véase Dingwall et al. (1988) *J Cell Biol.* 107 (3): 841-9). En otras realizaciones, la ubicación se puede realizar en la membrana tilacoide, véanse, por ejemplo, Schnell (1998) *Annual Review of Plant Physiology and Plant Molecular Biology* Vol. 49: 97-126; Ko and Cashmore (1989) *EMBO J.* 8(11): 3187-3194; de Boer et al. (1991) *EMBO J.* 10(10): 2765-2772; y Mackle and Zilinskas (1994) *J. Bacteriol.* 176(7):1857-1864.

Las proteínas de fusión descritas en este documento pueden comprender una secuencia de reubicación periplásmica y/o una proteasa específica del sitio. En determinadas realizaciones, la secuencia de reubicación periplásmica óptima para la proteasa específica del sitio se determina en forma empírica a partir de una selección de dichas secuencias. La eficiencia de secreción de la proteasa específica del sitio puede depender de varios parámetros, incluida la secuencia de reubicación, la proteasa que se esté reubicando, la cepa de la célula hospedante, el nivel de expresión de la proteasa y otros parámetros. Por ejemplo, una biblioteca de genes modificados con regiones 5' variables que codifican diferentes secuencias de ubicación periplásmicas se pueden crear usando PCR u otros métodos conocidos por el experto en la materia. Esta biblioteca se subclona en un vector que permite la expresión controlada (p. ej., un vector que permite la expresión controlada usando el sistema de inducción T7 tal como un vector de la serie pET), y se ensaya para eficiencia de exportación, así como también actividad de la proteína (véanse *J. Biol. Chem.* 267:4882, 1992; *J. Biol. Chem.* 267:12375, 1992; US2007/0111283 A1; y *Biotechnology Advances* 23:177, 2005).

Los ejemplos de secuencias de reubicación periplásmica incluyen, sin limitación:

Tabla 3. Ejemplos de secuencias de reubicación periplásmica				
Nombre	Péptido de señalización	kDa	Vía	Origen
MalE	MKIKTGARILALSALTTMMFSASALA (SEQ ID NO:1)	2,7	Sec	<i>E. coli</i>
OmpA	MKKTAIAlAVALAGFATVAQA (SEQ ID NO:2)	2,0		<i>E. coli</i>
PhoA	MKQSTIALALLPLLFTPVTKA (SEQ ID NO:3)	2,3	Sec	<i>E. coli</i>

Tabla 3. Ejemplos de secuencias de reubicación periplásmica				
Nombre	Péptido de señalización	kDa	Vía	Origen
LamB	MMITLRKPLAVAVAAGVMSAQAMA (SEQ ID NO:4)	2,6	Sec	<i>E. coli</i>
MglB	MNKKVLTLSAVMASMLFGAAHA (SEQ ID NO:5)	2,4	Sec	<i>E. coli</i>
PelB	MKYLLPTAAAGLLLLAAQPAMA (SEQ ID NO:6)	2,2	Sec	<i>E. carotovora</i>
STII	MKKNIAFLLASMFVFSIATNAYA (SEQ ID NO:7)	2,6		<i>E. coli</i>
DsbA	MKKIWLALAGLVLAFSASA (SEQ ID NO:8)	2,0	SRP	<i>E. coli</i>
SfmC	MMTKIKLLMLIIFYLIISASAHA (SEQ ID NO:9)	2,6	SRP	<i>E. coli</i>
TolB	MKQALRVAFGFLILWASVLHA (SEQ ID NO:10)	2,4	SRP	<i>E. coli</i>
TorT	MRVLLFLLLSLFMLPAFS (SEQ ID NO:11)	2,1	SRP	<i>E. coli</i>

Un sitio de escisión se ubica opcionalmente entre la secuencia de reubicación y la proteasa específica del sitio para permitir la separación. El sitio de escisión puede ser cualquier sitio que pueda usarse en la separación de la secuencia de reubicación y la secuencia codificante de la proteasa específica del sitio. Se puede usar cualquier sitio de escisión que haga uso de algún método para escisión de proteínas. Las secuencias de reubicación periplásmica de *E. coli* pueden contener dentro de su secuencia de señalización un motivo reconocido por la peptidasa líder, Lep, para procesamiento y escisión de la secuencia de señalización. Otros métodos que se han usado exitosamente y que conocen los expertos en la técnica incluyen métodos de escisión de proteasas, p. ej., trombina, factor Xa proteasa y otras endo peptidasas, como tripsina. Los genes que codifican la proteína de fusión se pueden sintetizar para incluir un sitio de escisión para una de estas proteasas entre la secuencia de reubicación y la secuencia de proteasas. Otro sistema para fusión y escisión es el sistema de dominio de unión inteína/quitina (S. Chong, et al., Gene, Vol. 192, 271-281 (1997)). Este sistema hace uso de las propiedades de autoescisión de las proteínas inteína.

#### Ácidos nucleicos, polipéptidos y células

Los ácidos nucleicos utilizados que se describen en este documento, o bien ARN, ARNi, ácido nucleico antisentido, ADNc, ADN genómico, vectores, cromosomas artificiales, virus o sus híbridos pueden aislarse de una diversidad de fuentes, genomanipularse, ampliarse y/o expresarse/generarse en forma recombinante. Una molécula de ácido nucleico o moléculas de ácido nucleico que codifican cualquiera de las enzimas diana o proteasas específicas del sitio asociadas con las realizaciones de la presente invención se pueden introducir en una célula o células empleando los métodos y técnicas conocidos en el campo. Por ejemplo, se pueden introducir moléculas de ácido nucleico en células por protocolos estándar tales como transformación, incluida transformación química y electroporación, transducción y bombardeo de partículas. También se puede lograr la expresión de una molécula(s) de ácido nucleico que codifica una enzima diana o una proteasa específica del sitio, integrando la molécula de ácido nucleico en el genoma de la célula. La molécula(s) de ácido nucleico se puede integrar en el ADN genómico de una célula usando técnicas convencionales conocidas en el campo. Los polipéptidos recombinantes generados a partir de estos ácidos nucleicos se pueden aislar individualmente o clonar y ensayar para una actividad deseada. "Polipéptido" y "proteína", tal como se emplean en esta memoria, abarcan enzimas diana y proteasas específicas del sitio. Se puede utilizar cualquier sistema de expresión recombinante, incluidos, aunque sin limitarse a ello, sistemas de expresión de células bacterianas, mamíferas, de levadura, insectos o plantas. Los ácidos nucleicos empleados como se describe en la presente invención se pueden sintetizar *in vitro* por técnicas de síntesis química conocidas, como se describe, p. ej., en Adams et al., J. Am. Chem. Soc. (1983) 105:661; Belousov 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 en la patente de EE. UU. 4.458.066.

Las células hospedantes para ingeniería de vías incluyen una diversidad de microorganismos heterotróficos y autotróficos, incluidos, aunque sin limitarse a ello, bacterias, hongos y protozoos. En determinadas realizaciones, las células hospedantes incluyen aquellas para las cuales un polipéptido se puede dirigir a un compartimento celular o a compartimentos celulares conocidos. En algunas realizaciones, la célula es cualquier tipo de célula que expresa en forma recombinante cualquiera de uno o más de los ácidos nucleicos (p. ej., secuencias que codifican enzimas diana

o proteasas específicas del sitio) descritos en este documento. Dichas células incluyen células procariotas y eucariotas. En algunas realizaciones, la célula es una célula bacteriana, 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.*, *Microlunatus spp.*, *Geobacter spp.*, *Geobacillus spp.*, *Arthrobacter spp.*, *Flavobacterium spp.*, *Serratia spp.*, *Saccharopolispora spp.*, *Thermus spp.*, *Stenotrophomonas spp.*, *Chromobacterium spp.*, *Sinorhizobium spp.*, *Saccharopolispora spp.*, *Agrobacterium spp.* y *Pantoea spp.* La célula bacteriana puede ser una célula gramnegativa tal como una célula de *Escherichia coli* (*E. coli*), o una célula grampositiva tal como una especie de *Bacillus*. En otras realizaciones, la célula es una célula fúngica tal como células de levadura, p. ej., *Saccharomyces spp.*, *Schizosaccharomyces spp.*, *Pichia spp.*, *Paffia spp.*, *Kluyveromyces spp.*, *Candida spp.*, *Talaromyces spp.*, *Brettanomyces spp.*, *Pachysolen spp.*, *Debaryomyces spp.*, *Yarrowia spp.* y cepas de levadura poliploides industriales. Otros ejemplos no limitativos de hongos incluyen *Aspergillus spp.*, *Pennicillium spp.*, *Fusarium spp.*, *Rhizopus spp.*, *Acremonium spp.*, *Neurospora spp.*, *Sordaria spp.*, *Magnaporthe spp.*, *Allomyces spp.*, *Ustilago spp.*, *Botrytis spp.* y *Trichoderma spp.* En otras realizaciones, la célula es una célula de alga, una célula vegetal o una célula mamífera. Se ha de apreciar que algunas células compatibles con las realizaciones descritas en este documento pueden expresar una copia endógena de uno o más de los genes descritos en la presente memoria, además de una copia recombinante. Las especies de interés incluyen, sin limitación, *S. cerevisiae*, *E. coli*, especies de *Pseudomonas*, especies de *Klebsiella* y especies de *Synechocystis*. Para evitar la degradación indeseada de la proteasa específica del sitio reubicada, la cepa hospedante puede modificarse para eliminar varias proteasas compartimentales (p. ej., proteasas periplásmicas) y/o para aumentar con proteínas tales como chaperonas y maturasas con el fin de ayudar al plegado de proteínas. Dichas modificaciones y aumentos emplean métodos conocidos por los expertos en la técnica, véanse, p. ej., las patentes de EE. UU. 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 descritos en este documento se expresan en forma recombinante en células bacterianas. Las células bacterianas en este documento se pueden cultivar en medios de cualquier tipo (ricos o mínimos) y en cualquier composición. En algunas realizaciones, las células se cultivan en medio mínimo. Como entendería el experto en la técnica, la optimización de rutina permitiría el uso de una diversidad de tipos de medios. El medio seleccionado se puede enriquecer con diversos componentes adicionales. Algunos ejemplos no limitativos de componentes de enriquecimiento incluyen glucosa, antibióticos, isopropil  $\beta$ -D-1-tiogalactopiranosido (IPTG), tetraciclina o anhidro-tetraciclina (aTc) para inducción de genes y ATCC Trace Mineral Supplement. De modo similar, otros aspectos del medio y condiciones de crecimiento de las células de las realizaciones descritas en la presente invención se pueden optimizar con experimentación de rutina. El pH y la temperatura son ejemplos no limitativos de factores que se pueden optimizar. En algunas realizaciones, la concentración y la cantidad del componente de enriquecimiento se pueden optimizar. En algunas realizaciones, se optimizan la frecuencia con la que se enriquece el medio con uno de los componentes de enriquecimiento y la cantidad de tiempo en que se cultiva el medio.

Las técnicas para la manipulación de ácidos nucleicos, p. ej., subclonación, sondas de etiquetado (p. ej., etiquetado con cebadores aleatorios que usan polimerasa Klenow, traducción de Nick, ampliación), secuenciación, hibridación y similares se describen en la bibliografía científica y de patentes, véanse, p. ej., Sambrook, Ed., *Molecular Cloning: A Laboratory Manual* (2ª Ed.) Vol. 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, Part I. Theory and Nucleic Acid Preparation*, Elsevier, N.Y. (1993).

Se ha de apreciar que los genes que codifican enzimas diana y proteasas específicas del sitio asociados con las realizaciones descritas en este documento se pueden obtener de una diversidad de fuentes. Como sabrá el experto en la técnica, existen genes homólogos para estas enzimas diana y proteasas específicas del sitio en muchas especies, y se pueden identificar mediante búsquedas de homología, por ejemplo, a través de una búsqueda de proteínas BLAST, disponible en el sitio de internet de NCBI ([ncbi.nlm.nih.gov](http://ncbi.nlm.nih.gov)). Los genes que codifican estas enzimas diana y proteasas específicas del sitio se pueden ampliar por PCR de ADN de cualquier fuente que contenga la enzima diana determinada y la proteasa específica del sitio, por ejemplo, usando cebadores degenerados, como entendería el experto en la técnica. En algunas realizaciones, el gen que codifica una enzima diana determinada y una proteasa específica del sitio puede ser sintético (artificial), por ejemplo, ADN sintetizado a partir de azúcares, compuestos basados en nitrógeno, fosfatos y otros compuestos/reactivos requeridos para síntesis de ADN. Cualquier medio para obtener los genes que codifican las enzimas diana y proteasas específicas del sitio analizado en este documento es compatible con los aspectos de las realizaciones descritas en la presente invención.

Las realizaciones descritas en la presente invención también dan a conocer polipéptidos y proteínas aislados (p. ej., enzimas diana y proteasas específicas del sitio) codificados por los ácidos nucleicos. Dichos polipéptidos son útiles, por ejemplo, solos o como proteínas de fusión. Los polipéptidos asociados con las células, lisados celulares y métodos descritos en este documento se pueden aislar de muestras biológicas, incluidos homogeneizados celulares o de tejido, y también se pueden expresar en forma recombinante en una diversidad de sistemas de expresión de procariotas y eucariotas construyendo un vector de expresión apropiado para el sistema de expresión, introduciendo

el vector de expresión en el sistema de expresión, y aislando el polipéptido o la proteína que se expresa en forma recombinante, p. ej., una enzima diana o una proteasa específica del sitio. Los polipéptidos también se pueden sintetizar químicamente usando métodos bien consolidados de síntesis de péptidos. En algunas realizaciones, se pueden añadir enzimas y/o proteasas aisladas directamente a los lisados celulares durante la fase de producción en la producción de metabolitos.

Se puede utilizar una variedad de metodologías conocidas por el experto en la materia para obtener polipéptidos aislados asociados con las realizaciones que se describen en este documento. El polipéptido se puede purificar de las células por medios cromatográficos o reconocimiento inmunológico. Alternativamente, se puede introducir un vector de expresión en las células para causar la producción del polipéptido. En otro método, se pueden microinyectar transcritos de ARNm o de algún modo introducir en las células para causar 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 utilizar para producir polipéptidos. Los expertos en la técnica pueden seguir fácilmente métodos conocidos para aislar polipéptidos. Estos incluyen, entre otros, inmunocromatografía, HPLC, cromatografía de exclusión de tamaño, cromatografía de intercambio iónico y cromatografía de afinidad inmune.

La expresión de las moléculas, como las enzimas diana y las proteasas específicas del sitio, que se describe en este documento, se puede determinar usando métodos de rutina conocidos por los expertos en la técnica. Estos métodos incluyen, aunque sin limitarse a ello: ampliación de ARN directa, transcripción inversa de ARN a ADNc, RT-PCR en tiempo real, ampliación de ADNc, hibridación y métodos de ensayo basados en inmunología, que incluyen, entre otros, inmunohistoquímica, ensayo de captura sándwich de anticuerpos, ELISA y ensayo por inmunoadsorción ligado a enzimas de puntos (ensayo EliSpot). Por ejemplo, la determinación de la presencia del nivel de moléculas de ácido nucleico en una célula se puede llevar a cabo mediante cualquier ensayo de determinación de ácido nucleico estándar, incluida la reacción en cadena de la polimerasa (PCR), o ensayando con sondas de hibridación etiquetadas. Dichos métodos de hibridación incluyen, aunque sin limitarse a ello, técnicas de micromatrices.

Las realizaciones descritas en este documento implican, por lo tanto, en un aspecto, enzimas diana y proteasas específicas del sitio, genes que codifican esas enzimas y proteasas, modificaciones funcionales y variantes de lo antedicho, además de usos asociados.

Las realizaciones descritas en este documento también incluyen secuencias codificantes degeneradas que incluyen codones alternativos a aquellos presentes en los materiales nativos. Por ejemplo, los residuos serina son codificados por los codones TCA, AGT, TCC, TCG, TCT y AGC. Cada uno de los seis codones es equivalente para los fines de codificar un residuo serina. Por consiguiente, será obvio para el experto en la materia que se puede emplear cualquier triplete de nucleótidos que codifique serina para dirigir el aparato de síntesis de proteínas, *in vitro* o *in vivo*, para incorporar un residuo erina a un polipéptido extensible. De modo similar, las secuencias de nucleótidos triples que codifican otros residuos de aminoácidos incluyen, aunque sin limitarse a ello: CCA, CCC, CCG y CCT (codones prolina); CGA, CGC, CGG, CGT, AGA y AGG (codones arginina); ACA, ACC, ACG y ACT (codones treonina); AAC y AAT (codones asparagina); y ATA, ATC y ATT (codones isoleucina). Se pueden codificar otros residuos de aminoácidos de modo similar por múltiples secuencias de nucleótidos. Por lo tanto, las realizaciones descritas en este documento degeneran ácidos nucleicos que difieren de los ácidos nucleicos aislados biológicamente en la secuencia de codones debido a la degeneración del código genético. Las realizaciones también abarcan la optimización de codones para ajustarse al uso óptimo de los codones de una célula hospedante, p. ej., una célula de *E. coli* u otra célula bacteriana.

Las realizaciones descritas en la presente invención también dan a conocer moléculas de ácido nucleico modificadas que incluyen adiciones, sustituciones y eliminaciones de uno o más nucleótidos. En algunas realizaciones, estas moléculas de ácido nucleico modificadas y/o los polipéptidos que codifican retienen por lo menos una actividad o función de la molécula de ácido nucleico no modificada y/o los polipéptidos, como la actividad enzimática (p. ej., proteasa). En determinadas realizaciones, las moléculas de ácido nucleico modificadas codifican polipéptidos modificados, por ejemplo, polipéptidos que tienen sustituciones de aminoácidos conservadoras como se describe en otra parte de la presente invención. Las moléculas de ácido nucleico modificadas están estructuralmente relacionadas a las moléculas de ácido nucleico no modificadas, y en algunas realizaciones, están estructuralmente relacionadas a las moléculas de ácido nucleico no modificadas de forma tal que las moléculas de ácido nucleico modificadas se hibridan bajo condiciones rigurosas conocidas por el experto en la técnica.

Por ejemplo, se pueden preparar moléculas de ácido nucleico modificadas que codifican polipéptidos que tienen un solo cambio de aminoácidos. 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 descrito en la presente invención. Asimismo, se pueden preparar moléculas de aminoácidos que codifican polipéptidos que tienen dos cambios de aminoácidos, que tienen, p. ej., 2-6 cambios de nucleótidos. Numerosas moléculas de ácido nucleico modificadas como estas serán fácilmente contempladas por el experto en la técnica, incluidas, por ejemplo, sustituciones de nucleótidos en codones que codifican aminoácidos 2 y 3, 2 y 4, 2 y 5, 2 y 6, y así sucesivamente. En el ejemplo precedente, cada combinación de dos aminoácidos se incluye en el conjunto de moléculas de aminoácidos modificadas, además de todas las sustituciones de nucleótidos que codifican las sustituciones de aminoácidos. También se pueden preparar moléculas de ácido nucleico adicionales que codifican polipéptidos que tienen sustituciones adicionales (p. ej., 3 o más), adiciones o eliminaciones (p. ej., por introducción

de un codón finalizador o un sitio(s) de empalme) como contemplará fácilmente el experto en la técnica. Cualquiera de los ácidos nucleicos o polipéptidos anteriores se puede ensayar por experimentación de rutina para retención de relación estructural o actividad, p. ej., actividad enzimática, a los ácidos nucleicos y/o polipéptidos descritos en la presente memoria.

- 5 En determinadas realizaciones, se contemplan variantes de las enzimas diana y proteasas específicas del sitio que se describen en este documento. Tal como se emplea en este documento, una "variante" de una enzima diana o proteasa específica del sitio es una proteína que contiene una o más modificaciones a la secuencia de aminoácidos primaria de la enzima diana o proteasa específica del sitio. Las modificaciones que crean una variante se pueden realizar a una enzima diana o proteasa específica del sitio, por ejemplo, para reducir o eliminar una actividad de la enzima diana en presencia de la proteasa específica del sitio o para alterar la distribución celular de la proteasa específica del sitio.

15 Las modificaciones a una enzima diana o una proteasa específica del sitio se pueden realizar al ácido nucleico que codifica la proteína, y pueden incluir eliminaciones, mutaciones puntuales, truncamientos, sustituciones de aminoácidos y adiciones de restos aminoácidos o no aminoácidos. Alternativamente, se pueden efectuar modificaciones directamente al polipéptido(s) o la proteína, como por escisión, adición de una molécula enlazadora, adición de un resto detectable, como biotina, adición de un ácido graso, y similares. Las modificaciones también abarcan proteínas de fusión. El experto en la materia conocerá los métodos para pronosticar el efecto sobre la conformación de la proteína de un cambio en la secuencia de aminoácidos, y puede por ende "designar" una enzima diana o proteasa específica del sitio variante de acuerdo con esos métodos conocidos. Un ejemplo de dicho método es descrito por Dahiyat y Mayo en Science 278:82 87, 1997, mediante el cual se pueden diseñar proteínas *de novo*. El método se puede aplicar a una proteína conocida para variar solamente una porción de la secuencia de la proteína. Al aplicar los métodos de cómputos de Dahiyat y Mayo, pueden proponerse variantes específicas de una proteína y ensayarse para determinar si la variante retiene una conformación deseada.

25 En algunas realizaciones, las variantes incluyen enzimas diana y proteasas específicas del sitio que son modificadas especialmente para alterar una característica de la proteína no relacionada con su actividad fisiológica deseada. Por ejemplo, se pueden sustituir o eliminar residuos cisteína para prevenir enlaces disulfuro indeseados. De modo similar, se pueden cambiar ciertos aminoácidos para potenciar la expresión de una proteína eliminando la proteólisis por proteasas en un sistema de expresión (p. ej., residuos de aminoácidos dibásicos en sistemas de expresión de levaduras en donde está presente la actividad de proteasa KEX2).

30 Las mutaciones de un ácido nucleico que codifica una enzima diana o una proteasa específica del sitio pueden preservar el marco de lectura del aminoácido de la secuencia codificante, y en algunos casos no crean regiones en el ácido nucleico que probablemente hibridan para formar estructuras secundarias, como horquillas o bucles, que pueden ser perjudiciales para la expresión de la proteína variante.

35 Las mutaciones se pueden efectuar seleccionando una sustitución de aminoácidos, o por mutagénesis aleatoria de un sitio seleccionado en un ácido nucleico que codifica la enzima diana o la proteasa específica del sitio. Las proteínas variantes luego se expresan y ensayan para una o más actividades a fin de determinar qué mutación proporciona una enzima o proteasa variante con las propiedades deseadas, p. ej., escisión de la proteasa específica del sitio. Se pueden efectuar otras mutaciones a las variantes (o a proteínas no variantes) que son silenciosas en cuanto a la secuencia de aminoácidos de la proteína, pero que proporcionan codones para traducción eficiente en un hospedante particular. El experto en la técnica conoce ejemplos de codones para traducción eficiente de un ácido nucleico en, p. ej., *E. coli*. Se pueden efectuar incluso otras mutaciones a las secuencias no codificantes de un gen o clon de ADNc para potenciar la expresión de la proteína. La actividad de las variantes de proteínas se puede ensayar clonando el gen que codifica la proteína variante en un vector de expresión bacteriano o mamífero, introduciendo el vector en una célula hospedante apropiada, expresando la proteína variante y ensayando una capacidad funcional de la proteína descrita en este documento.

50 El experto en la materia también sabrá que se pueden efectuar sustituciones conservadoras de aminoácidos en proteínas para proporcionar variantes funcionalmente equivalentes de las enzimas diana y proteasas específicas del sitio precedentes, es decir, las variantes retienen las capacidades funcionales de las proteínas. Tal como se emplea en este documento, una "sustitución de aminoácidos conservadora" se refiere a una sustitución de aminoácidos que no altera la carga relativa o las características de tamaño de la proteína en la cual se efectúa la sustitución de aminoácidos. Se pueden preparar variantes de acuerdo con los métodos para secuenciación de proteínas conocidos por los expertos en la materia, como aquellos que se hallan en referencias que compilan dichos métodos, p. ej., Molecular Cloning: A Laboratory Manual, J. Sambrook, et al., eds., Segunda edición, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, Nueva York, 1989 o Current Protocols in Molecular Biology, F.M. Ausubel, et al., eds., John Wiley & Sons, Inc., Nueva York. Las variantes funcionalmente equivalentes ilustrativas de la proteína incluyen sustituciones de aminoácidos conservadoras en las secuencias de aminoácidos de las enzimas y proteasas descritas en este documento. Las sustituciones conservadoras de aminoácidos incluyen sustituciones efectuadas 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 determinadas realizaciones, menos de todos los aminoácidos se pueden cambiar cuando se preparan enzimas y proteasas variantes. Si se sabe qué residuos de aminoácidos en particular confieren función y se requiere esa función, dichos aminoácidos no se reemplazarán, o alternativamente, se reemplazarán por sustituciones de aminoácidos conservadoras. En algunas realizaciones, se cambian 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 residuos al preparar proteínas variantes. En algunas realizaciones, se realiza el menor número de sustituciones. Por lo tanto, un método para generar enzimas o proteasas variantes consiste en sustituir un aminoácido sencillo particular por todos los otros aminoácidos, luego determinar la actividad de la variante, luego repetir el proceso con una o más de las enzimas o proteínas que tienen la mejor actividad.

Se pueden efectuar sustituciones de aminoácidos conservadoras en la secuencia de aminoácidos de la proteína para producir variantes funcionalmente equivalentes de la proteína, por alteración de un ácido nucleico que codifica una enzima diana o una proteasa específica del sitio. Dichas sustituciones se pueden efectuar mediante una diversidad de métodos conocidos por los expertos en la técnica. Por ejemplo, se pueden realizar sustituciones de aminoácidos por mutación dirigida por PCR, mutagénesis dirigida al sitio de acuerdo con el método de Kunkel (Kunkel, Proc. Nat. Acad. Sci. U.S.A. (1985) 82: 488-492), o por síntesis química de un gen que codifica una proteína. Por ejemplo, en algunas realizaciones, un ácido nucleico que codifica una enzima diana se modifica para comprender un sitio de escisión de una proteasa específica del sitio, y la variante de la enzima expresada aún retiene la actividad funcional de la enzima no modificada.

#### Vectores y constructos de expresión

Los vectores útiles para la transformación de un fragmento de ADN aislado que codifica una enzima diana y/o una proteasa específica del sitio en células hospedantes adecuadas (p. ej., células de *E. coli*) se conocen en la técnica. Tal como se emplea en este documento, un "vector" puede ser cualquiera de un número de ácidos nucleicos en los cuales se puede insertar una secuencia o secuencias deseadas (p. ej., secuencias que codifican enzimas diana o proteasas específicas del sitio) por restricción y ligadura para transporte entre diferentes entornos genéticos o para expresión en una célula hospedante. Los vectores están típicamente compuestos por ADN, aunque también están disponibles los vectores de ARN. Los vectores incluyen, aunque sin limitarse a ello, plásmidos, fósmidos, fagémidos, genomas de virus y cromosomas artificiales. El vector puede contener secuencias que dirigen la transcripción y traducción del gen relevante, un marcador seleccionable y secuencias que permiten la replicación autónoma o la integración cromosómica. Los vectores adecuados comprenden una región 5' del gen que aloja los controles del inicio de la transcripción y una región 3' del fragmento de ADN que controla la finalización de la transcripción. También se pueden emplear vectores que promueven la integración del gen quimérico que codifica una proteína de fusión en el genoma de la célula hospedante. Dichos vectores pueden ser para integración aleatoria, integración dirigida al sitio, o para recombinación homóloga. Un vector puede tener características que permiten tipos de un solo cruce o de dobles cruces de recombinación homóloga. Se pueden integrar una o múltiples copias en el genoma de una célula hospedante (p. ej., célula bacteriana o de levadura).

Un vector de clonación es capaz de replicarse en forma autónoma o integrada en el genoma en una célula hospedante. Un vector de clonación se caracteriza además por uno o más sitios de restricción de endonucleasas en los cuales el vector se puede cortar (p. ej., vía digestión enzimática) en un modo determinable y en donde se puede ligar la secuencia de ADN deseada. De esta manera, el nuevo vector recombinante retiene su capacidad de replicarse en la célula hospedante. Cuando se emplean vectores plasmídicos, la replicación de la secuencia deseada puede ocurrir muchas veces a medida que el plásmido aumenta en número de copias dentro del hospedante o solo una vez por hospedante antes del que el hospedante se reproduzca por mitosis. Cuando se usan vectores 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 por restricción y ligadura, de manera tal que se una en forma operativa a secuencias reguladoras y pueda expresarse como un transcrito de ARN. Los vectores pueden además contener una o más secuencias marcadoras adecuadas para uso en la identificación de células que han sido o no transformadas o transfectadas con el vector. Los marcadores incluyen, por ejemplo, genes que codifican proteínas que aumentan o disminuyen o bien la resistencia o la sensibilidad a los antibióticos u otros compuestos, genes que codifican enzimas cuyas actividades son detectables por ensayos convencionales conocidos en la técnica (p. ej.,  $\beta$ -galactosidasa, luciferasa o fosfatasa alcalina), y genes que afectan visiblemente el fenotipo de células transformadas o transfectadas, hospedantes, colonias o placas (p. ej., proteína fluorescente verde). Los vectores incluyen aquellos capaces de replicación autónoma y expresión de productos génicos estructurales presentes en los segmentos de ADN a los que están operativamente unidos.

Tal como se emplea en este documento, se dice que una secuencia codificante y secuencias reguladoras están "operativamente" enlazadas o unidas cuando están covalentemente enlazadas en forma tal como para disponer la expresión o transcripción de la secuencia codificante bajo la influencia o el control de las secuencias reguladoras. Si se desea que las secuencias codificantes se traduzcan a una proteína funcional, se dice que dos secuencias de ADN están operativamente unidas si la inducción de un promotor en las secuencias reguladoras 5' resulta en la transcripción de la secuencia codificante, y si la naturaleza del enlace entre las dos secuencias de ADN no (1) resulta en la introducción de una mutación con desplazamiento del marco, (2) interfiere con la capacidad de la región promotora de dirigir la transcripción de las secuencias codificantes, o (3) interfiere con la capacidad del

correspondiente transcrita de ARN de traducirse a una proteína. Por consiguiente, una región promotora estaría operativamente unida a una secuencia codificante si la región promotora fuese capaz de efectuar la transcripción de la secuencia de ADN de forma tal que el transcrita resultante pudiese traducirse, por ejemplo, a la proteasa específica del sitio deseada.

5 Cuando la molécula de ácido nucleico que codifica cualquiera de las enzimas diana o proteasas específicas del sitio descritas en este documento se expresa en una célula, se puede utilizar una diversidad de secuencias de control de la transcripción (p. ej., secuencias promotor/potenciador) para dirigir su expresión. El promotor puede ser un promotor nativo, es decir, el promotor del gen en su contexto endógeno, que proporciona regulación normal de expresión del gen. En algunas realizaciones, el promotor puede ser constitutivo, es decir, el promotor no está regulado, permitiendo la transcripción continua de su gen asociado. También se puede emplear una variedad de promotores condicionales, como promotores controlados por la presencia o ausencia de una molécula. Por ejemplo, algunos promotores inducibles se activan con sustancias químicas tales como isopropil  $\beta$ -D-1-tiogalactopiranosido (IPTG) o Tetraciclina (Tet).

15 La naturaleza precisa de las secuencias reguladoras necesarias para la expresión de genes puede variar entre especies y tipos de células, y puede incluir, según sea necesario, secuencias 5' no transcritas y 5' no traducidas implicadas con el inicio de la transcripción y la traducción, respectivamente, como caja TATA, secuencia caperuza, secuencia CAAT y similares. En particular, dichas secuencias reguladoras en 5' no transcritas incluirán una región promotora que incluye una secuencia promotora para control de la transcripción del gen operativamente unido. Las secuencias reguladoras pueden además incluir secuencias potenciadoras o secuencias activadoras en dirección 5' según se desee. Los vectores pueden opcionalmente incluir secuencias de señalización o líderes 5'. La elección y el diseño de un vector apropiado están dentro de la capacidad y el criterio del experto en la técnica.

20 Los vectores de expresión que contienen todos los elementos necesarios para expresión se comercializan y los conoce el experto en la técnica. Véase, p. ej., Sambrook et al., *Molecular Cloning: A Laboratory Manual*, Segunda edición, Cold Spring Harbor Laboratory Press, 1989. Las células se genomanipulan introduciendo en las células ADN heterólogo (o ARN). Ese ADN heterólogo (o ARN) se dispone bajo control operable de elementos de transcripción para permitir la expresión del ADN heterólogo en la célula hospedante. En algunas realizaciones, se pueden clonar dos o más ácidos nucleicos en el mismo vector de expresión o plásmido.

25 Los métodos descritos en este documento pueden hacer uso de expresión constitutiva o regulada de varias secuencias de ácido nucleico que codifican las enzimas diana y/o proteasas específicas del sitio. La expresión se puede regular por varios indicios, por ejemplo, inducción de sustancias químicas (p. ej., IPTG), cambio de la fase de crecimiento, agotamiento de un nutriente, cambios de temperatura (p. ej., promotor de choque térmico) y/o luz. En algunas realizaciones, los promotores inducibles se regulan por la presencia de un agente de inducción, por ejemplo, una sustancia química tal como lactosa, arabinosa o tetraciclina, como se conoce en la técnica. Si se indica expresión de "alto nivel", la concentración de la proteína expresada en la célula puede estar por lo menos aproximadamente 2 veces encima de los niveles basales, por lo menos 10 veces encima de los niveles basales, por lo menos 25 veces encima de los niveles basales, por lo menos 50 veces encima de los niveles basales, o más, p. ej., entre aproximadamente 2 veces y aproximadamente 100 veces encima de los niveles basales.

30 Puede ser conveniente ajustar en forma experimental la tasa de expresión para optimizar la eficiencia de la célula o la exportación del compartimiento celular de las proteasas específicas del sitio que se describen en este documento. La translocación deficiente, por ejemplo, al periplasma de una célula puede producir la capacidad insuficiente de la maquinaria de exportación. Los métodos para ajustar la tasa de expresión de la proteasa específica del sitio incluyen, sin limitación, modificación del número de copias del plásmido que porta el gen que codifica la proteasa que se ha de exportar al periplasma. Los replicones conocidos y utilizados en la técnica incluyen P15A (10 copias/célula), ColA (30 copias/célula), ColE1 (40 copias/célula) y RSF1030 (>100 copias/célula). El sitio de unión al ribosoma en 5' UTR del gen que codifica la proteasa que se ha de exportar al periplasma se puede modificar, en donde se puede crear y ensayar una biblioteca de sitios de unión al ribosoma con fortalezas variables; véanse, p. ej. Salis et al., *Nature Biotechnology* (2009) 27: 946-950; y Simmons et al., *Nature Biotechnology* (1996) 14:629-634. La región promotora en dirección 5' del gen que codifica la proteasa que se ha de exportar se puede modificar para ajustar la tasa de transcripción, en donde se puede crear y ensayar una biblioteca de regiones promotoras con fortalezas variables; véanse, p. ej., Alper et al., *PNAS* (2005) 102:12678-12683; y De Mey et al., *BMC Biotechnology* (2007) 7:34.

#### Expresión inducible

35 Los métodos descritos en este documento pueden hacer uso de expresión regulada de varias secuencias codificantes. La expresión se puede regular por varios indicios, por ejemplo, inducción por sustancias químicas, cambio de la fase de crecimiento, agotamiento de un nutriente, cambio de temperatura, y luz. En algunas realizaciones, los promotores inducibles son regulados por la presencia de un agente inductor, p. ej., una sustancia química tal como lactosa, arabinosa y tetraciclina, conocida en la técnica.

Los vectores de expresión y clonación usualmente contienen un promotor que es reconocido por el organismo hospedante y está operativamente enlazado a la secuencia codificante de interés. Los promotores son secuencias no traducidas localizadas en dirección 5' (5') al codón de inicio de un gen estructural que controla la transcripción y

traducción de una secuencia de ácido nucleico particular a la que están operativamente enlazados. Dichos promotores pueden categorizarse dentro de dos clases, inducibles y constitutivos. Los promotores inducibles son promotores que inician niveles en aumento de transcripción de ADN bajo su control en respuesta a algún cambio en las condiciones de cultivo, p. ej., la presencia o ausencia de un nutriente o un cambio de temperatura. En este momento, se conoce una gran cantidad de promotores reconocidos por una diversidad de células hospedantes potenciales. Si bien se puede usar el promotor nativo, en algunas realizaciones, se usan promotores heterólogos, ya que permiten una mayor transcripción de la enzima diana o la proteasa específica del sitio y mayores rendimientos.

Los promotores adecuados para uso con hospedantes procariontes incluyen los sistemas promotores  $\beta$ -lactamasa y lactosa, fosfatasa alcalina, un sistema promotor de triptófano (*trp*), y numerosos promotores híbridos como el promotor *tac*. No obstante, otros promotores bacterianos conocidos son también adecuados, p. ej., el promotor *lacI*, el promotor T3, el promotor T7, el promotor de arabinosa, el promotor *gpt*, el promotor lambda PR, el promotor lambda PL, los promotores de operones que codifican enzimas glicolíticas tales como 3-fosfoglicerato cinasa (PGK), y el promotor de fosfatasa ácido. Sus secuencias de nucleótidos se han publicado, permitiendo así que un trabajador experimentado las ligue operativamente a una secuencia de interés usando enlazadores o adaptadores. Los promotores para uso en sistemas bacterianos también contendrán una secuencia Shine-Dalgarno (S.D.) operativamente enlazada a la secuencia codificante. En ciertos casos, también, la célula hospedante se puede modificar genéticamente para ajustar las concentraciones de metabolito o las proteínas transportadoras inductoras de modo que todas las células del cultivo sean inducidas de forma equivalente.

Los promotores adecuados para células eucariotas, p. ej., células de levadura, se conocen en la técnica. Prácticamente todos los genes eucariotas poseen una región rica en AT localizada aproximadamente 25 a 30 bases en dirección 5' del sitio en donde se inicia la transcripción. Otra secuencia hallada 70 a 80 bases en dirección 5' del inicio de la transcripción de muchos genes es una región CXCAAT en 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 del extremo poli-A al extremo 3' de la secuencia codificante. Todas estas secuencias se insertan adecuadamente en los vectores de expresión eucariotas. Los ejemplos de secuencias promotoras adecuadas para uso con hospedantes de levadura incluyen los promotores para 3-fosfoglicerato cinasa u otras enzimas glicolíticas, como enolasa, gliceraldehído-3-fosfato deshidrogenasa, hexocinasa, piruvato descarboxilasa, fosfofructocinasa, glucosa-6-fosfato isomerasa, 3-fosfoglicerato mutasa, piruvato cinasa, triosefosfato isomerasa, fosfoglucosa isomerasa y glucocinasa.

Otros promotores de levadura, que son promotores inducibles que tienen la ventaja adicional de transcripción controlada por condiciones de crecimiento, son las regiones promotoras para alcohol deshidrogenasa 2, isocitocromo C, fosfatasa ácida, enzimas degradativas asociadas con el metabolismo de nitrógeno, metalotioneína, gliceraldehído-3-fosfato deshidrogenasa, y enzimas responsables de la utilización de maltosa y galactosa. Los potenciadores de levadura se emplean ventajosamente con los promotores de levadura.

## Ejemplos

Proteasa específica del sitio segregadora. Los pasos en el proceso de segregación de proteasas e ingeniería de enzimas diana incluyen (i) analizar la vía para identificar una enzima o enzimas que se han de dirigir para inactivación, (ii) manipular un sitio de escisión de proteasas modificando la secuencia de ADN que codifica la enzima(s) diana, (iii) verificar la actividad de la enzima diana, (iv) manipular a nivel periplásmico las proteasas segregadas, (v) verificar la actividad de las proteasas expresadas a nivel periplásmico, (vi) demostrar el crecimiento celular metabólicamente saludable, y (vii) demostrar la inactivación de la enzima(s) diana y el aumento del flujo hacia el producto de interés post-lisis en una reacción activa, libre de células.

### Ejemplo 1. Segregación de proteasas al periplasma

Se construye una biblioteca de plásmidos que contienen varios genes de proteasa con diversas secuencias de señalización periplásmicas que dirigen las enzimas proteasa hacia el periplasma. Por ejemplo, podrían combinarse cinco proteasas específicas y seis secuencias de señalización periplásmicas para conformar una biblioteca de treinta miembros.

Se añaden secuencias de nucleótidos que codifican secuencias de señalización periplásmicas a cada gen de proteasa a través de ampliación de la reacción en cadena de la polimerasa (PCR). Alternativamente, los genes de proteasa se sintetizan para contener la secuencia de señalización líder periplásmica. La síntesis completa permite la optimización de codones para expresión eficiente en la cepa hospedante. Las proteasas adecuadas y secuencias de direccionamiento periplásmico/líderes se describieron anteriormente en las Tablas 2 y 3, respectivamente.

La biblioteca de genes de proteasa dirigidos a nivel periplásmico subclonados en un vector pDuet, u otro vector adecuado, para expresión inducible de la cepa de *E. coli* BL21(DE3), o cepa similar que expresa T7 polimerasa, se transforma con la biblioteca de proteasa dirigida a nivel periplásmico. La modificación de la expresión se logra con el uso de niveles variables de isopropil  $\beta$ -D-1-tiogalactopiranosido (IPTG) (cada vector expresa *LacI*, y el promotor T7 contiene un sitio operador *lac*), a través del uso de varios promotores y sitios de unión al ribosoma, además del uso

de una variación del número de copias entre diferentes vectores pDuet. También se pueden emplear otros plásmidos, sistemas de expresión o cepas familiares para los expertos en la técnica.

Se desarrollan cultivos separados que alojan miembros de bibliotecas individuales para intermediar la densidad óptica (OD) en medio definido rico antes de la inducción de expresión con 0,02 - 1 mM IPTG. La expresión se induce durante varias horas para permitir la acumulación de proteasa en el periplasma. El crecimiento post-inducción se monitorea para determinar qué miembros de la biblioteca, si es que alguno, obstaculizan el crecimiento celular supuestamente debido a una escisión fuera de la diana de las proteínas periplásmicas de *E. coli* nativas. La inhibición del crecimiento severo puede indicar que la proteasa activa expresada dirige una proteína(s) importante o esencial. Esto se puede mitigar identificando la proteína(s) fuera de la diana y modificando su secuencia(s) de aminoácidos para eliminar el sitio(s) de escisión específico. La proteasa dirigida a nivel periplásmico se extrae usando choque osmótico u otros métodos conocidos por el experto en la técnica. La verificación de la expresión de la proteína de longitud total se determina desnaturalizando electroforesis en gel de la proteína con estándares de tamaño apropiado. La exportación periplásmica más eficiente se puede alcanzar optimizando la expresión a través del uso de niveles variables de IPTG, diferentes orígenes de replicación de plásmidos en el plásmido que se esté utilizando y/o modificación del sitio de unión al ribosoma (RBS).

La actividad específica de la proteasa dirigida a nivel periplásmico se determina en extracto periplásmico diseñando una proteína diana que contiene un sitio de escisión y que, tras la escisión por la proteasa específica del sitio, ofrecerá una señal mensurable. Un ejemplo es señalado por Sarath y Schwartzbach en Current Protocols in Protein Science (2001) 21.9.1 - 21.9.10. Este método incluye el uso de una proteína indicadora fluorescente que contiene una etiqueta de purificación o bien en el término N o en el término C. Un sitio de escisión se inserta entre la etiqueta de purificación y el marco abierto de lectura del indicador fluorescente, creando una diana para una proteasa específica. La diana indicadora purificada se puede añadir a extracto periplásmico que contiene la proteasa específica del sitio, y esta mezcla se puede separar usando una resina de afinidad que une la etiqueta de purificación. El indicador fluorescente escindido, no unido, se separa, y se mide la fluorescencia para determinar el grado de escisión de proteasa. Alternativamente el indicador fluorescente manipulado se puede expresar en el citoplasma de la cepa que expresa una proteasa dirigida a nivel periplásmico. El indicador no escindido se puede separar del lisado, y el nivel de fluorescencia en el lisado se puede usar para determinar el grado de escisión diana de proteasa.

Un ejemplo no limitativo específico de una vía es la vía para la síntesis de ácido shikímico. En esta vía, por ejemplo, una reacción entre los compuestos commodity celulares fosfoenolpiruvato ( $S_{1A}$ ) y eritrosa-4-fosfato ( $S_{1B}$ ) se cataliza por las isoenzimas de DAHP sintasa, AroF, G o H ( $E_1$ ) para formar 3-desoxi-D-arabinoseheptulosa-7-fosfato (DAHP). DAHP ( $S_2$ ) es transformado a 3-deshidroquinato (3-DHQ) por la segunda enzima en la vía, DHQ sintasa, AroB ( $E_2$ ). 3-DHQ ( $S_3$ ) es deshidratado a 3-deshidroshikimato por la tercera enzima en la vía, 3-DHQ deshidratasa, AroD ( $E_3$ ). 3-deshidroshikimato ( $S_4$ ) se reduce a ácido shikímico ( $P_F$ ) por la cuarta enzima en la vía, shikimato deshidrogenasa, AroE ( $E_4$ ), usando NADPH como cofactor. Las enzimas de la vía se conocen en la técnica y se han caracterizado en un número de organismos, incluido, por ejemplo, *E. coli*, en donde las enzimas son codificadas por los locus genéticos de la siguiente manera: DAHP sintasa (*aroG*, *F* o *H*); DHQ sintasa (*aroB*); 3-DHQ deshidratasa (*aroD*); shikimato deshidrogenasa (*aroE*).

En la vía de ácido shikímico, la sobreexpresión de *aroE*, *aroF* y *tktA* ha demostrado incrementar la producción de ácido shikímico o bien directamente con *aroE* y *aroF*, o indirectamente con *tktA*, que incrementa el suministro de la molécula de sustrato eritrosa-4-fosfato.

Existen también enzimas competitivas relevantes a esta vía, que utilizan el producto final deseado como sustrato, en donde las enzimas son shikimato cinasa (AroK) y shikimato cinasa II (AroL). La presencia de enzimas competitivas puede ser indeseable, ya que reducen la cantidad de producto deseado en la mezcla de reacción.

La secuencia que codifica la enzima competitiva, shikimato cinasa II (número de acceso en Genbank NP\_414922.1) se torna susceptible a proteólisis por una proteasa específica del sitio, introduciendo un sitio de escisión de proteasa en la proteína diana a través de la introducción de mutaciones específicas, por recombinación, o ingeniería de genomas múltiple. El sitio para inactivación de proteasa se introduce en un sitio adecuado en la enzima. Alternativamente, se inactiva shikimato cinasa II por los métodos expuestos en Datsenko and Wanner (2000), y shikimato cinasa I se torna susceptible a proteólisis por una proteasa específica del sitio.

Incluida en la producción de ácido shikímico puede estar la expresión de DAHP sintasa y/o enzimas adicionales en la vía bioquímica a ácido shikímico:

Tabla 4. Ejemplos de enzimas en la vía bioquímica de ácido shikímico.		
EC#	Enzima	Gen
2.2.1.1	transcetolasa	<i>tktA</i>
4.2.3.4	deshidroquinato sintasa	<i>aroB</i>
4.2.1.10	deshidroquinato deshidratasa	<i>aroD</i>
1.1.1.25	shikimato deshidrogenasa	<i>aroE</i>

5 Se analiza la secuencia de la enzima para identificar la secuencia de aminoácidos más próxima al motivo de escisión de proteasas, de modo tal que se necesite el número más bajo de modificaciones a la secuencia de aminoácidos para resultar en un sitio de escisión de proteasas. El sitio ideal está en la superficie de la proteína, para minimizar los efectos sobre la función enzimática. Un sitio cercano a la superficie también asegura que sea accesible a la proteasa cuando se pliega. Alternativamente, una enzima de un organismo diferente que desempeña la misma función y contiene un sitio de escisión de proteasas se puede usar en lugar de una enzima nativa que no es lábil a proteasa.

10 Un sitio adecuado para escisión (una barra indica aminoácidos alternativos) es la secuencia ENLYFQ<sup>G/S</sup> (escindida por la proteasa del virus del grabado del tabaco). Un sitio alternativo para escisión es GARR<sup>G/S</sup> (escindido por la proteasa del virus de la fiebre amarilla). Un sitio alternativo para escisión es LVPRGS (escindido por trombina). Un sitio alternativo para escisión es I<sup>E/D</sup>GR (escindido por trombina). La secuencia genéticamente modificada, que se puede denominar forma lábil de proteasa de la secuencia, se expresa a partir de un plásmido con inactivación concomitante de la copia cromosómica, o se usa para alterar la secuencia cromosómica del organismo hospedante.

15 En una realización alternativa, el gen que codifica AroL se inactiva, y AroK se torna susceptible a proteólisis.

20 Se llevan a cabo ensayos enzimáticos, después de una breve exploración de las secuencias, para asegurar que la proteasa elegida escinda la proteína diana, pero no ninguna otra proteína necesaria para la biosíntesis de ácido shikímico. Para lograr esto, las células modificadas con enzima competitiva lábil de proteasa se modifican adicionalmente para expresar la proteasa análoga bajo el control de un promotor inducible. Las células se desarrollan en medio que contiene glucosa como fuente de carbono. La producción de ácido shikímico en las células se mide usando espectroscopia de resonancia magnética nuclear (<sup>1</sup>HNMR) o cromatografía de líquidos de alto rendimiento (HPLC), como lo describen Knop et al. (2001) J. Am. Chem. Soc. 123:10173-10182 o van Hess et al. (1999) Talanta, 5 de enero de 1999, páginas 173-17], o cualquier otra técnica de química analítica tal como GC-MS.

25 Alternativamente, se realizan los ensayos enzimáticos individuales, p. ej., en los cuales el agotamiento de un sustrato enzimático o la acumulación de su producto se mide en una disolución con la proteína enzimática purificada.

30 Si se halla que alguna proteína en la vía de ácido shikímico es susceptible a proteólisis, distinta de la enzima(s) competitiva, los sitios de escisión se eliminan efectuando cambios de aminoácidos conservadores por mutagénesis dirigida al sitio. Alternativamente, una enzima de un organismo diferente que lleva a cabo la misma función, pero que no contiene un sitio de escisión de proteasas, se puede usar en lugar de una enzima nativa que es lábil a proteasa. Dicho análisis se repite si es necesario hasta encontrar una proteasa apropiada.

### Ejemplo 2. Expresión y actividad de enterocinasa dirigida a nivel periplásmico

Plásmidos, crecimiento y expresión celular. El gen que codifica enterocinasa bovina (número de acceso en Genbank L19663.1) se sintetizó químicamente con contenido de codones optimizado para expresión en *E. coli* y con las siguientes secuencias de señalización periplásmicas:

Tabla 5. Secuencias de señalización periplásmicas	
DsbA (DA)	atgaaaaagatttggctggcgctggctggttagtttagcgttagcgcacatcggcg (SEQ ID NO:17)
OmpA (OA)	atgaaaaaacggcaattgcgatagcggttgctagctggtttgccacgggtggcgcaggct (SEQ ID NO:18)
MalE (ME)	atgaaaataaaacagggtgcacgcacatcctcgattatccgcattaacgacgatgatgtttccgctcggctctcgcc (SEQ ID NO:19)

Estos constructos se subclonaron en pACYC-Duet (EMD Chemicals, Gibbstown, NJ), y se transformó la cepa de *E. coli* BL21(DE3) con los plásmidos resultantes. Se desarrollaron cultivos de 250 ml de stocks de glicerol durante 6 h a 37 °C y 300 rpm en medio rico EZ con 34 µg/ml de cloranfenicol (cam) hasta una densidad óptica (OD600) de 0,6 – 0,8. Una cepa que porta el vector pACYC-Duet vacío se usó como control. La expresión se indujo con isopropil β-D-1-tiogalactopiranosido (IPTG) 0,1 mM, y los cultivos se desarrollaron durante 16 h a 25 °C. Después de la inducción, las células se cosecharon y se extrajo la fracción periplásmica.

Ensayo de extracción y actividad periplásmica. Las células se cosecharon a partir de 12 ml de cultivo por centrifugación a 3000 xg durante 30 min a 4 °C. Se desechó el sobrenadante, y el sedimento celular se resuspendió suavemente en 2 ml de Tris-HCl 1 mM enfriado con hielo. Esta suspensión luego se centrifugó a 3000 xg durante 30 min a 4 °C. Se recogió el sobrenadante como la fracción periplásmica. Los ensayos de actividad se llevaron a cabo incubando 3,6 µg de la fracción periplásmica con 4 µg de proteína control (una proteína escindible de enterocinasa provista por EMD) en tampón de reacción que contenía NaCl 50 mM, TrisHCl 20 mM, CaCl<sub>2</sub> 2 mM, pH 7,4. Las reacciones se incubaron a 30 °C durante 24 h. Las reacciones se inactivaron añadiendo 19,5 µl de tampón de corrida dodecil sulfato sódico (SDS)-PAGE que contenía ditioneitol (DTT) 0,04 M, y llevando a ebullición a 99 °C durante 5 min. Las reacciones se llevaron a cabo en un gel NuPAGE® 12% Bis-Tris Gel a 200 V durante 50 minutos, y los geles se tiñeron con tinte Coomassie. La escisión de la proteína control de 48 kDa por enterocinasa (EK) proporciona dos fragmentos proteolíticos de 32 y 16 kDa. Los geles que se muestran en la Figura 1 se emplearon para visualizar los productos de escisión de proteasa e indican que la enterocinasa localizada a nivel periplásmico es activa en escindir la proteína diana.

### Ejemplo 3. Expresión de enterocinasa dirigida a nivel periplásmico

La secuencia génica para enterocinasa bovina se optimizó con codones para expresión en *E. coli* y se subclonó en el vector de expresión pACYCDuet-1 (Novagen) con la adición de la secuencia de señalización de expresión periplásmica OmpA. El constructo se transformó en la cepa de *E. coli* BL21(DE3). Las células se desarrollaron hasta una densidad óptica (OD600) de 0,6, se indujo la expresión por adición de IPTG 0,8 mM y continuó el crecimiento a 37 °C durante 3 horas. El cultivo se sometió a 2,4 duplicaciones durante la inducción, indicando que la expresión de enterocinasa no inhibió el crecimiento. Las células se cosecharon por centrifugación, se resuspendieron en tampón de ensayo (Tris HCl 20, pH 7,4; NaCl 50 mM; CaCl<sub>2</sub> 2 mM) y se lisaron usando un homogeneizador.

### Ejemplo 4. Actividad de enterocinasa en lisado de células enteras

Se ensayó un lisado de células enteras de células que expresan enterocinasas para actividad de proteasa por incubación con una proteína control de escisión que contenía un sitio de escisión de enterocinasa (Asp-Asp-Asp-Asp-Lys) y se comparó con la actividad de enterocinasa recombinante. La enterocinasa recombinante y la proteína control de escisión se obtuvieron de Novagen. Se incubaron 40 µl de lisado o 2 unidades de enterocinasa recombinante en 40 µl de tampón de ensayo (Tris HCl 20 mM, pH 7,4; NaCl 50 mM; CaCl<sub>2</sub> 2 mM) con 4 µg de proteína control de escisión a temperatura ambiente. Se usó un lisado preparado de forma similar a partir de una cepa que no expresa enterocinasa como control negativo. Se analizaron muestras a las cero y 2 horas por inmunotransferencia con conjugado de peroxidasa de rábano (HRP) picante y proteína S que reconoce la etiqueta aminoterminal S en la proteína control de escisión. La Figura 2 muestra la detección del fragmento de escisión de proteasa de 16 kDa. La proteína diana se escindió específicamente por enterocinasa recombinante (senda 2) y por lisado de una cepa que expresa enterocinasa dirigida a nivel periplásmico (senda 4), pero no por un lisado que carece de expresión de enterocinasa (senda 6).

### Ejemplo 5. Inactivación de enzimas diana por liberación de proteasa del periplasma

La secuencia génica para enterocinasa bovina se optimiza pro codones para expresión en células de *E. coli* y se subclonó en un vector de expresión de plásmidos inducible con la adición de una secuencia de señalización de expresión periplásmica. Un gen que codifica la enzima AroL se modifica para incluir un sitio de escisión específico de enterocinasa. Se modifican células de *E. coli* para expresar la proteasa específica del sitio en el periplasma y la enzima diana modificada en el citoplasma. Las células se desarrollan hasta una densidad óptica (OD600) de 0,6, se induce la expresión de enterocinasa y se permite que las células sigan creciendo a 37 °C durante 3 horas.

Las células se lisan usando un homogeneizador, poniendo así en contacto la proteasa específica del sitio y la enzima modificada. El lisado celular se combina con uno o más sustratos, enzimas, nutrientes, co-factores, tampones, agentes reductores y/o sistemas generadores de ATP, según lo requerido para actividad de las enzimas de la vía, formando de este modo un sistema libre de células. Se optimiza el potencial de redox, pH, y fuerza iónica del sistema libre de células. La producción de ácido shikímico en el sistema libre de células se mide usando espectroscopia de <sup>1</sup>HNMR o HPLC.

### Ejemplo 6. Inactivación de PurF por liberación de proteasa del periplasma

Se genera una biblioteca, cada miembro de la cual codifica una enzima amidofosforibosil transferasa (*E. coli* PurF) que contiene un sitio de escisión específico de enterocinasa en diferentes ubicaciones dentro de la secuencia de

5 polipéptidos de la enzima. La elección de la ubicación del sitio de escisión es auxiliada por la estructura terciaria PurF publicada como se indica en Muchmore et al., Protein Science (1998) 7:39-51. Por ejemplo, Muchmore et al. describen una región que comprende un bucle expuesto a un disolvente flexible que es adecuado para incorporación del sitio de escisión. La biblioteca se genera por PCR, síntesis de genes u otros métodos de biología molecular estándar y se construye en vectores para expresión controlada en células de *E. coli*. Las células de *E. coli* se transforman con los vectores que expresan la biblioteca de enzimas modificadas. Las células se desarrollan hasta una densidad óptica (OD600) de 0,6, se induce la expresión de la biblioteca, y se permite que las células sigan creciendo a 37 °C durante 3 horas.

10 Las células se lisan usando un homogeneizador, y el lisado celular se trata con enterocinasa recombinante purificada (Novagen) para verificar la escisión mediada por enterocinasa de miembros de la biblioteca. La escisión de PurF se vigila por SDS-PAGE. Los miembros de la biblioteca que se escinden por enterocinasa recombinante se ensayan luego para actividad. Concretamente, el lisado de células que sobreexpresan estos miembros de la biblioteca se combina con PRPP y glutamina (10 mM cada uno), según se requiere para actividad enzimática, formando así un sistema libre de células. Las muestras de la reacción se analizan en una serie de puntos de tiempo. 15 La actividad de PurF modificada se vigila por el aspecto del glutamato y la desaparición concomitante de glutamina en la reacción libre de células según lo medido por HPLC usando derivación pre-columna con ortoaldehído. La inactivación de PurF modificada se ensaya por adición de enterocinasa recombinante al ensayo anteriormente descrito.

20 Los miembros más activos de la biblioteca de PurF que exhiben actividad en ausencia de enterocinasa, y se escinden e inactivan por enterocinasa en el curso de la reacción, se escogen para integración en una célula que contiene una vía de interés que se inhibe por actividad de PurF (p. ej., una vía que requiere PRPP, como una vía que produce un compuesto derivado de pirimidina). La célula adicionalmente contiene enterocinasa expresada a nivel periplásmico y cualquier enzima requerida para la vía de interés. El miembro de la biblioteca se inactiva en el locus cromosómico de PurF, reemplazando de este modo el gen endógeno. Se verifica que el crecimiento no esté adversamente afectado por el reemplazo del gen. 25

Las células se lisan usando un homogeneizador, poniendo en contacto de esta forma la proteasa específica del sitio con la enzima modificada. El lisado celular se combina con uno o más sustratos, enzimas, nutrientes, co-factores, tampones, agentes reductores y/o sistemas generadores de ATP, según se requiera para actividad de las enzimas de la vía, formando así un sistema libre de células. Se optimizan el potencial de rédox, el pH y la fuerza iónica del sistema libre de células. La producción del producto de interés en el sistema libre de células se mide usando métodos adecuados. 30

#### Otras realizaciones

A menos que se defina algo distinto, todos los términos técnico-científicos utilizados en este documento tienen el mismo significado que entiende comúnmente el experto en la técnica.

35 La cita de publicaciones o documentos no está destinada a admitir que ninguna de dichas publicaciones o documentos es técnica anterior pertinente, ni constituye ninguna admisión en cuanto a los contenidos o la fecha de estas publicaciones o documentos.

Tal como se emplea en este documento, "un" o "una" significa "por lo menos uno/a" o "uno/a o más" a menos que se indique otra cosa.

40 Tal como se emplea en este documento, los términos "aproximadamente" y "alrededor de" se usan de manera intercambiable, y el significado de los términos se puede obtener considerando el contexto de los términos tal como se usan en la memoria y en las reivindicaciones de la solicitud. En algunas realizaciones, "aproximadamente" un valor numérico particular abarca un intervalo de 30% del valor. En otras realizaciones, "aproximadamente" un valor numérico particular abarca un intervalo de 20% del valor. Incluso en otras realizaciones, "aproximadamente" un valor 45 numérico particular abarca un intervalo de 10% del valor.

#### LISTA DE SECUENCIAS

<110> GREENLIGHT BIOSCIENCES INC.

EL CONSEJO DIRECTIVO DE LA UNIVERSIDAD DE LELAND STANFORD JUNIOR UNIVERSITY

50 <120> MÉTODOS PARA CONTROL DE FLUJO EN VÍAS METABÓLICAS MEDIANTE MANIPULACIÓN DE PROTEASAS

<130> P57445EP1/KVC

<150> EP11752452.0

<151> 2001-08-31

<150> US 61/378.828

<151> 2010-08-31

<160> 19

<170> PatentIn versión 3.5

5 <210> 1

<211> 26

<212> PRT

<213> E. coli

<400> 1

Met Lys Ile Lys Thr Gly Ala Arg Ile Leu Ala Leu Ser Ala Leu Thr  
 1 5 10 15

10 Thr Met Met Phe Ser Ala Ser Ala Leu Ala  
 20 25

<210> 2

<211> 21

<212> PRT

<213> E. coli

15 <400> 2

Met Lys Lys Thr Ala Ile Ala Ile Ala Val Ala Leu Ala Gly Phe Ala  
 1 5 10 15

Thr Val Ala Gln Ala  
 20

<210> 3

<211> 21

<212> PRT

20 <213> E. coli

<400> 3

Met Lys Gln Ser Thr Ile Ala Leu Ala Leu Leu Pro Leu Leu Phe Thr  
 1 5 10 15

Pro Val Thr Lys Ala

20

<210> 4

25 <211> 25

<212> PRT

<213> E. coli

<400> 4

ES 2 839 424 T3

Met Met Ile Thr Leu Arg Lys Leu Pro Leu Ala Val Ala Val Ala Ala  
 1 5 10 15

Gly Val Met Ser Ala Gln Ala Met Ala  
 20 25

<210> 5

<211> 23

<212> PRT

5 <213> E. coli

<400> 5

Met Asn Lys Lys Val Leu Thr Leu Ser Ala Val Met Ala Ser Met Leu  
 1 5 10 15

Phe Gly Ala Ala Ala His Ala  
 20

<210> 6

<211> 22

10 <212> PRT

<213> E. coli

<400> 6

Met Lys Tyr Leu Leu Pro Thr Ala Ala Ala Gly Leu Leu Leu Leu Ala  
 1 5 10 15

Ala Gln Pro Ala Met Ala  
 20

<210> 7

15 <211> 23

<212> PRT

<213> E. coli

<400> 7

Met Lys Lys Asn Ile Ala Phe Leu Leu Ala Ser Met Phe Val Phe Ser  
 1 5 10 15

Ile Ala Thr Asn Ala Tyr Ala  
 20

20 <210> 8

<211> 19

<212> PRT

<213> E. coli

<400> 8

Met Lys Lys Ile Trp Leu Ala Leu Ala Gly Leu Val Leu Ala Phe Ser  
 1 5 10 15

25 Ala Ser Ala

ES 2 839 424 T3

<210> 9

<211> 23

<212> PRT

<213> E. coli

5 <400> 9

Met Met Thr Lys Ile Lys Leu Leu Met Leu Ile Ile Phe Tyr Leu Ile  
1 5 10 15

Ile Ser Ala Ser Ala His Ala  
20

<210> 10

<211> 21

<212> PRT

10 <213> E. coli

<400> 10

Met Lys Gln Ala Leu Arg Val Ala Phe Gly Phe Leu Ile Leu Trp Ala  
1 5 10 15

Ser Val Leu His Ala  
20

<210> 11

<211> 18

15 <212> PRT

<213> E. coli

<400> 11

Met Arg Val Leu Leu Phe Leu Leu Leu Ser Leu Phe Met Leu Pro Ala  
1 5 10 15

Phe Ser

<210> 12

20 <211> 7

<212> PRT

<213> Secuencia artificial

<220>

<223> sitio de escisión de proteasas dentro de la proteína TEV Nla

25 <400> 12

Glu Asn Leu Tyr Phe Gln Gly  
1 5

<210> 13

<211> 8

<212> PRT

<213> Secuencia artificial  
 <220>  
 <223> sitio de escisión de proteasas dentro de la proteína HRV 3C  
 <400> 13  
**Leu Glu Val Leu Phe Gln Gly Pro**  
 5 1 5  
 <210> 14  
 <211> 5  
 <212> PRT  
 <213> Secuencia artificial  
 10 <220>  
 <223> sitio de escisión de proteasas dentro de la proteína Enterocinasa  
 <400> 14  
**Asp Asp Asp Asp Lys**  
 1 5  
 <210> 15  
 15 <211> 4  
 <212> PRT  
 <213> Secuencia artificial  
 <220>  
 <223> sitio de escisión de proteasas dentro de la proteína del Factor Xa  
 20 <400> 15  
**Ile Glu Gly Arg**  
 1  
 <210> 16  
 <211> 5  
 <212> PRT  
 25 <213> Secuencia artificial  
 <220>  
 <223> sitio de escisión de proteasas dentro de la proteína Trombina (fIIa)  
 <400> 16  
**Leu Val Pro Arg Gly**  
 1 5  
 30 <210> 17  
 <211> 57  
 <212> ADN  
 <213> E. coli  
 <400> 17  
 35 atgaaaaaga ttggctggc gctggctggt ttagtttag cgtttagcgc atcggcg 57

# ES 2 839 424 T3

<210> 18

<211> 63

<212> ADN

<213> E. coli

5 <400> 18

atgaaaaaaa cggcaattgc gatagcggtt gcgctagctg gttttgccac ggtggcgcag 60

gct 63

<210> 19

<211> 78

<212> ADN

10 <213> E. coli

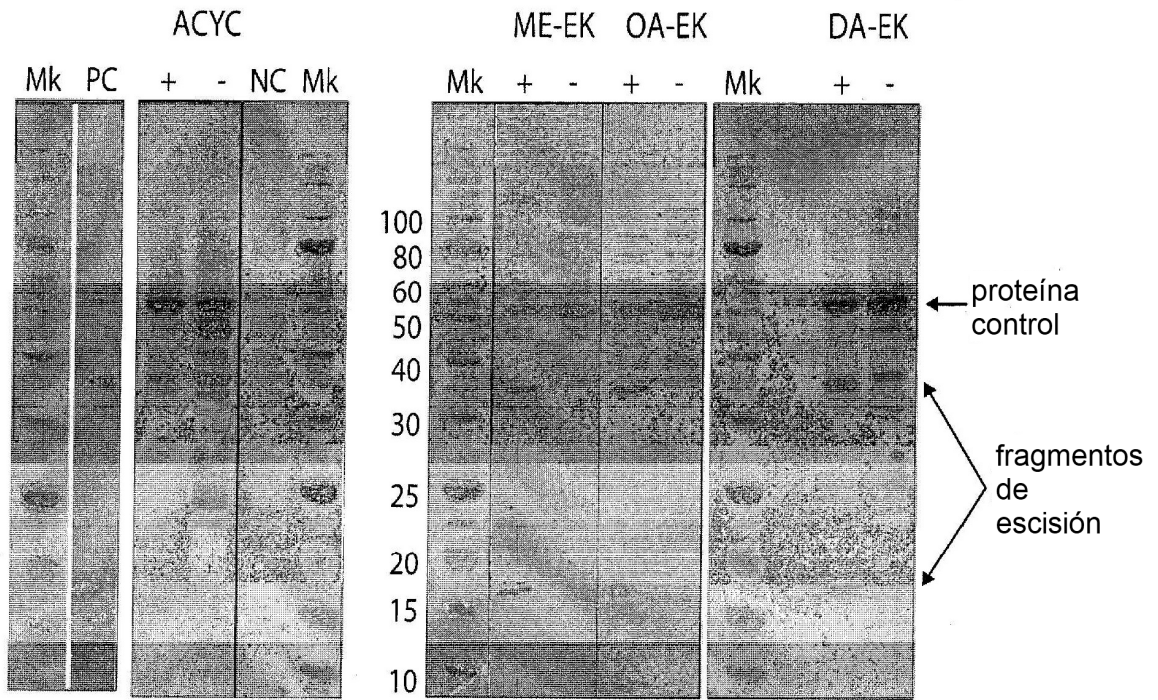
<400> 19

atgaaaataa aaacaggtgc acgcacctc gcattatccg cattaacgac gatgatgttt 60

tccgcctagg ctctagcc 78

**REIVINDICACIONES**

1. Un método para producir un producto de una vía de interés, en donde el método comprende:
- 5 (a) cultivar en medio de crecimiento hasta una densidad deseada, células genéticamente modificadas para comprender: (i) una secuencia codificante de HRV 3C proteasa que comprende una secuencia de reubicación, y (ii) una secuencia codificante para al menos una enzima diana que compite con una enzima implicada en una vía de interés, en donde la secuencia codificante para al menos una enzima diana está genéticamente modificada para comprender un sitio de reconocimiento de HRV 3C proteasa, en donde al menos una enzima diana desvía/desacelera el flujo fuera del producto de interés, y/o en donde la eliminación de las actividades enzimáticas de al menos una enzima diana afecta adversamente el crecimiento celular;
- 10 (b) lisar por lo menos una porción de las células para producir un lisado celular; y
- (c) combinar el lisado celular con uno o más sustratos, enzimas, nutrientes, co-factores, tampones, agentes reductores, y/o sistemas de generación de ATP,
- mediante los cual las enzimas en la vía de interés catalizan la producción del producto, y la HRV 3C proteasa escinde al menos una enzima diana; y
- 15 en donde el producto se selecciona del grupo que consiste en antibióticos, biotensioactivos, combustibles biológicos, aminoácidos, ácidos orgánicos, alcoholes, polioles, saporíferos, fragancias, nucleótidos, vitaminas, pigmentos, azúcares, polisacáridos, biopolímeros, plásticos, isoprenoides y terpenos.
2. El método según la reivindicación 1, en donde el método comprende además combinar el lisado celular con uno o más lisados celulares adicionales.
- 20 3. El método según la reivindicación 1, en donde la secuencia que codifica HRV 3C proteasa está operativamente enlazada a un promotor inducible.
4. El método según la reivindicación 1, en donde por lo menos una enzima diana compite con una enzima que aumenta la tasa de precursor suministrado a la vía de interés.
- 25 5. El método según la reivindicación 1, en donde por lo menos una enzima diana compite con una enzima clave de entrada a la vía.
6. El método según la reivindicación 1, en donde por lo menos una enzima diana es esencial para el crecimiento de las células.
7. El método según una cualquiera de las reivindicaciones 1-6, en donde el gen que codifica la contrapartida nativa de por lo menos una enzima diana se reemplaza con la secuencia codificante para al menos una enzima diana genéticamente modificada para comprender un sitio de reconocimiento de HRV 3C proteasa.
- 30 8. El método según la reivindicación 7, en donde el gen que codifica la enzima de contrapartida nativa se inactiva.
9. El método según la reivindicación 1, en donde por lo menos una enzima diana se sobreexpresa en las células.
10. El método según la reivindicación 1, en donde por lo menos una enzima diana está presente o bien en un vector episomal o en un cromosoma.
- 35 11. El método según la reivindicación 1, en donde el medio de crecimiento se ha modificado por la adición o potenciación de un factor que aumenta o preserva la actividad de por lo menos una enzima diana.
12. El método según la reivindicación 1, en donde las células son células bacterianas.
13. El método según la reivindicación 12, en donde las células bacterianas son células de *Escherichia coli*.
- 40 14. El método según la reivindicación 1, en donde la secuencia de reubicación es una secuencia diana periplásmica.
15. El método según la reivindicación 1, en donde la actividad de por lo menos una enzima diana se reduce en por lo menos 50% tras la escisión por HRV 3C proteasa.



+ = IPTG 0,1 mM  
 - = sin IPTG  
 NC = control neg. (sin proteína control)  
 PC = control pos. (EK purificado c/proteína control)  
 Mk = marcador de peso molecular (kDa)

Fig. 1

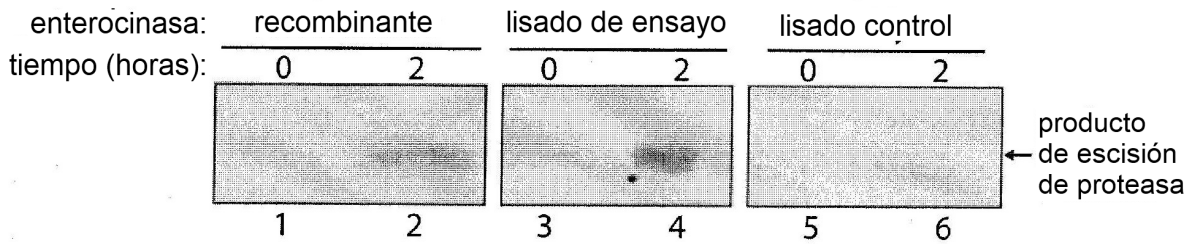


Fig. 2