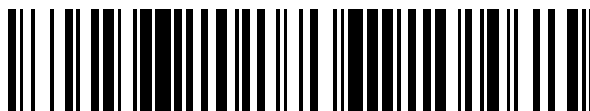


19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 540 728**

51 Int. Cl.:

C12N 9/18 (2006.01)

A23L 1/27 (2006.01)

C11B 3/00 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **14.06.2005** **E 12150400 (5)**

97 Fecha y número de publicación de la concesión europea: **01.04.2015** **EP 2468853**

54 Título: **Método para la decoloración enzimática de feofitina**

30 Prioridad:

16.06.2004 US 580447 P

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

13.07.2015

73 Titular/es:

DSM IP ASSETS B.V. (100.0%)
Het Overloon 1
6411 TE Heerlen, NL

72 Inventor/es:

LAM, DAVID;
WEINER, DAVID;
HITCHMAN, TIMOTHY;
BARTON, NELSON ROBERT y
BURK, MARK J.

74 Agente/Representante:

LEHMANN NOVO, María Isabel

Observaciones :

Véase nota informativa (Remarks) en el folleto original publicado por la Oficina Europea de Patentes

ES 2 540 728 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

Método para la decoloración enzimática de feofitina

CAMPO TÉCNICO

5 Esta invención se refiere a los campos del procesamiento industrial de alimentos, piensos o aceites vegetales, productos vegetales y animales, y a enzimología. En particular, la invención proporciona composiciones y métodos para el tratamiento enzimático ("blanqueo" o "decoloración") de composiciones que contienen clorofila o que están contaminadas con clorofila, por ejemplo preparaciones de algas, animales o vegetales, alimentos, piensos o aceites, por ejemplo aceites vegetales, incluyendo aceites procesados a partir de oleaginosas, tales como aceite de cáñola (colza) o aceite de haba de soja, o frutos oleosos, tales como aceite de palma. En un aspecto, la invención
10 proporciona métodos usando enzimas procedentes del catabolismo de la feofitina (por ejemplo, una clorofilasa) para la modificación enzimática de una feofitina, por ejemplo en una preparación de algas, animal o vegetal, o un alimento, un pienso o un aceite.

ANTECEDENTES

15 Los aceites vegetales que proceden de oleaginosas tales como cáñola o haba de soja, o frutos oleaginosos tales como palma, contienen clorofila. La clorofila se elimina durante muchas etapas del procedimiento de producción de aceite, incluyendo etapas de trituración de las semillas, extracción del aceite, desgomado, tratamiento cáustico y blanqueo. En la última de éstas, la clorofila residual del procedimiento de blanqueo se elimina hasta lograr niveles aceptables. Esta clorofila se elimina típicamente del aceite en una etapa del procedimiento de blanqueo que implica calentar el aceite y hacerlo pasar a través de un adsorbente para eliminar la clorofila y otros compuestos que poseen
20 color que afecta al aspecto y/o estabilidad del aceite acabado. Esta tecnología también se usa para tratar otros aceites o preparaciones vegetales o de algas que contienen clorofila, tales como aceites que contienen ácidos grasos poliinsaturados (PUFA) (por ejemplo, ácido eicosapentenoico (EPA) y ácido docosahexenoico (DHA)).

25 Un nivel elevado de pigmentos de clorofila da un color indeseable e induce la oxidación del aceite durante el almacenamiento, conduciendo al deterioro del aceite. En la industria del procesamiento de aceites comestibles, se emplea una etapa de blanqueo para reducir los niveles de clorofila tan bajos como 0,1 ppm para garantizar la calidad del aceite en términos de color y cualidad organoléptica. Los niveles de clorofila acabados deseados típicos están entre 0,02 y 0,05 ppm. Esta etapa de blanqueo incrementa el coste de procesamiento y reduce el rendimiento del aceite debido al arrastre en la arcilla de blanqueo.

30 Samaha y Kermasha, (1997), J. Chem. Techn. Biotechn. 68, 315-323, describen que la clorofilasa procedente de *Phaeodactylum triconutum* también es capaz de hidrolizar feofitina en feoforbida y fitol.

En plantas, la clorofilasa (clasa) es la primera enzima implicada en la degradación de la clorofila; cataliza la hidrólisis de un enlace de éster en la clorofila para producir clorofilida y fitol.

SUMARIO

35 La invención proporciona composiciones y métodos para el tratamiento enzimático ("blanqueo" o "decoloración") de composiciones que contienen feofitina o que están contaminadas con feofitina, tales como preparaciones vegetales, animales (por ejemplo, preparaciones de pescado, de carne) o de algas, alimentos, piensos o aceites, tales como aceites que contienen ácidos grasos poliinsaturados (PUFA) o que contienen ácido docosahexenoico (DHA), o composiciones que comprenden mezclas de los mismos. En un aspecto, el "blanqueo enzimático" de las composiciones y métodos de la descripción comprende el uso de una enzima modificadora de la clorofila, por
40 ejemplo un polipéptido que tiene actividad de clorofilasa, incluyendo clasas y clorofila clorofilido-hidrolasas, y péptidos relacionados, o cualquier enzima catabólica de la clorofila. De este modo, como se usa aquí, la expresión "blanqueo enzimático" incluye cualquier modificación de una molécula de clorofila o equivalente, incluyendo la decoloración parcial o completa. En un aspecto, las composiciones y métodos de la invención pueden reducir la pérdida de rendimiento debido al arrastre y división de grasas atribuidos a la catálisis por condiciones de arcilla/blanqueo.
45

En aspectos alternativos de los métodos y procedimientos de la descripción, se añade en cualquier momento o en cualquier lugar en el método o procedimiento, por ejemplo como se discute aquí, una clorofilasa, que puede ser una nueva clorofilasa de la descripción, o una enzima conocida, incluyendo clasas y clorofila clorofilido-hidrolasas y péptidos relacionados, o una combinación de los mismos, o cualquier enzima catabólica de la clorofila. Por ejemplo,
50 la clorofilasa (que puede ser una nueva clorofilasa de la descripción, o una enzima conocida, o una combinación de las mismas) y/o cualquier enzima catabólica de la clorofila se pueden añadir a una composición, tal como un aceite bruto, con o sin otra enzima, por ejemplo una fosfolipasa (por ejemplo, fosfolipasa C), en una etapa de mezclado o en una etapa de desgomado, en una etapa de tanque cáustico, en una mezcladora estática, en un tanque de día o en una mezcladora de retención. Como alternativa, en un aspecto de un método o procedimiento de la descripción,
55 la clorofilasa (de la descripción, o conocida) y/o cualquier enzima catabólica de la clorofila se pueden añadir en cualquier combinación de estas etapas, o en todas estas etapas.

La descripción proporciona métodos o procedimientos para la modificación enzimática de clorofila para facilitar su eliminación de una composición, por ejemplo a través de un procedimiento de separación acuoso, como se ilustra en la página 1, Apéndice A, o un procedimiento de separación hidrófobo, o un procedimiento de separación por afinidad, y similar.

- 5 La descripción proporciona métodos y procedimientos que comprenden la modificación enzimática (por ejemplo, catabolismo) de la clorofila, o compuestos equivalentes, en una composición (por ejemplo, un alimento, pienso, planta, animal, alga, etc.), que comprende además eliminar componentes de esa composición (por ejemplo, compuestos no deseables en un producto acabado), tal como clorofila residual (por ejemplo clorofila o compuestos equivalentes no modificados mediante una clorofilasa), un plaguicida, un hidrocarburo aromático policíclico, etc. Los
- 10 componentes indeseables, por ejemplo clorofila residual, plaguicidas, hidrocarburos aromáticos policíclicos, y similares, se pueden eliminar con cantidades significativamente menores de arcilla de blanqueo u otro adsorbente, tal como sílice o compuestos equivalentes.

- Estos componentes de la composición se eliminan usando una arcilla de blanqueo, por ejemplo en una pluralidad de etapas usando arcilla de blanqueo, en el que, en un aspecto, los componentes de la composición se eliminan con
- 15 cantidades significativamente menores de arcilla de blanqueo y/o al menos algún otro adsorbente (por ejemplo, una sílice). En un aspecto, los niveles de clorofila finales están entre alrededor de 0,02 ppm y 0,05 ppm. En este procedimiento ejemplar, la etapa de blanqueo puede incrementar los costes del procesamiento y reducir los rendimientos del aceite debido al arrastre en la arcilla de blanqueo. Las composiciones y procedimientos de la descripción pueden reducir la pérdida de rendimiento por arrastre y división de grasas atribuida a la catálisis por
- 20 condiciones de arcilla/blanqueo.

- En un método (reacción) ilustrado ejemplar de la descripción, una clorofilasa cataliza la hidrólisis de clorofila para generar clorofilida, que, en un aspecto, se extrae de forma acuosa, y fitol, que permanece en la fase oleosa. En otro método ejemplar, la feoforbida se puede eliminar de manera similar a la clorofilida. En un aspecto, poniendo en
- 25 práctica las composiciones y métodos de la descripción, un procedimiento de separación acuoso puede eliminar parcial o completamente la necesidad de agentes adsorbentes. Sin embargo, en otro aspecto, los métodos comprenden la extracción parcial o completa de la clorofilida o feoforbida soluble acuosa usando un procedimiento de extracción a base de sílice (por ejemplo, refinado con sílice libre de adsorbente o con poco adsorbente). En un aspecto, la clorofilasa se inmoviliza sobre una sílice (que entonces adsorbe la clorofilida), por ejemplo un gel de sílice. En un aspecto, la sílice comprende una sílice TriSyl o una sílice SORBSIL R™.

- La invención proporciona métodos, incluyendo procedimientos industriales, para el tratamiento enzimático de composiciones que contienen feofitina o contaminadas por feofitina, que comprenden las siguientes etapas: (a) proporcionar una composición que contiene feofitina o contaminada con feofitina; (b) proporcionar un polipéptido que
- 30 tiene una actividad de clorofilasa o feofitinasa (que puede ser una nueva clorofilasa de la descripción, o una enzima conocida, o una combinación de los mismos); y (c) hacer reaccionar la composición de la etapa (a) con el polipéptido de la etapa (b) en condiciones en las que el polipéptido puede catalizar una reacción que modifica a la feofitina. El derivado de clorofila con menor contenido de magnesio se denomina feofitina. La feofitina está coloreada y a menudo está presente en el aceite, especialmente si se ha usado tratamiento con ácido. En algunas aplicaciones, es deseable eliminar la feofitina. El producto de tratamiento de la feofitina con clorofilasa es feoforbida, que se puede eliminar de manera similar a la clorofilida.
- 35

- En un aspecto, las composiciones y métodos de la invención se ponen en práctica como o con procedimientos industriales, por ejemplo procedimientos de blanqueo o de neutralización cáustica o de desgomado del aceite. En un aspecto, el uso de las composiciones y métodos de la invención facilita la reducción de la cantidad de o la
- 40 eliminación de la necesidad de adsorbentes en el procesamiento de blanqueo actual, que implica típicamente calentar el aceite u otra composición que contiene clorofila y hacerla pasar a través de un adsorbente para eliminar clorofila u otros compuestos que poseen color que alteran el aspecto y/o estabilidad del aceite acabado. De este modo, en la práctica, este aspecto de la invención, eliminando parcial o completamente la necesidad de
- 45 adsorbentes, puede reducir los costes del procesamiento, por ejemplo se pueden reducir los costes de los adsorbentes (por ejemplo, arcilla), costes de eliminación de desechos, costes de agua, costes energéticos, costes de vapor. Otros beneficios al practicar diversos aspectos de la invención incluyen mejoras del rendimiento, por ejemplo menor cantidad de aceites arrastrados en sustratos adsorbentes, mayor valor del producto final, incluyendo retención de micronutrientes valiosos tales como beta-caroteno, eficiencias del procedimiento, incluyendo menores etapas del procesamiento, ahorros de capital y un beneficio medioambiental, por ejemplo reduciendo o eliminando el
- 50 vertido de adsorbentes de blanqueo.

- En un aspecto, en la práctica de las composiciones y métodos de la invención, los polipéptidos modificadores de la clorofila (que pueden ser una nueva clorofilasa de la descripción, o una enzima conocida, o una combinación de las mismas) se pueden emplear en cualquier punto en un procedimiento de desgomado (por ejemplo desgomado
- 55 enzimático). Por ejemplo, los polipéptidos modificadores de la clorofila se pueden añadir antes o después de cualquier etapa en un procedimiento, o antes o después de cualquier combinación de etapas, o antes o después de todas las etapas, en un procedimiento, por ejemplo antes de, durante o tras la extracción mecánica y/o química, y/o el desgomado y/o neutralización cáustica y/o blanqueo, y similar.
- 60

En aspectos alternativos de cualquiera de los métodos de la invención, al menos una etapa se lleva a cabo en una vasija de reacción, por ejemplo un aparato de desgomado de aceite. Como alternativa, en los métodos de la descripción, al menos una etapa se lleva a cabo en un extracto celular. En aspectos alternativos de cualquiera de los métodos de la invención, al menos una etapa se lleva a cabo en una célula completa. La célula puede ser de cualquier fuente, por ejemplo una célula vegetal, una célula bacteriana, una célula fúngica, una célula animal (por ejemplo, una célula de mamífero, una célula de pescado) o una célula de levadura.

La descripción proporciona métodos para el tratamiento enzimático de composiciones que contienen clorofila o contaminadas con clorofila, que comprenden las siguientes etapas: (a) proporcionar una composición que contiene clorofila o contaminada con clorofila (que puede ser una nueva clorofilasa) de la descripción, o una enzima conocida, o una combinación de las mismas; (b) proporcionar un polipéptido que tiene una actividad de clorofilasa; y (c) hacer reaccionar la composición de la etapa (a) con el polipéptido de la etapa (b) en condiciones en las que el polipéptido puede catalizar una reacción modificadora de la clorofila.

La descripción proporciona procedimientos industriales para el tratamiento enzimático ("blanqueo") de composiciones que contienen clorofila o contaminadas con clorofila, que comprenden las siguientes etapas: (a) proporcionar una composición que contiene clorofila o contaminada con clorofila (que puede ser una nueva clorofilasa de la descripción, o una enzima conocida, o una combinación de las mismas); (b) proporcionar un polipéptido que tiene actividad de clorofilasa; y (c) hacer reaccionar la composición de la etapa (a) con el polipéptido de la etapa (b) en condiciones en las que el polipéptido puede catalizar una reacción modificadora de la clorofila.

La descripción proporciona procedimientos de desgomado que comprenden una etapa para el blanqueado enzimático de composiciones que contienen clorofila o contaminadas con clorofila, que comprenden las siguientes etapas: (a) proporcionar una composición que contiene clorofila o contaminada con clorofila (que puede ser una nueva clorofilasa de la descripción, o una enzima conocida, o una combinación de las mismas); (b) proporcionar un polipéptido que tiene actividad de clorofilasa; y (c) hacer reaccionar la composición de la etapa (a) con el polipéptido de la etapa (b) en condiciones en las que el polipéptido puede catalizar una reacción modificadora de la clorofila.

Existe un segundo éster en las clorofilas y feofitinas – un éster metílico. Los métodos de la descripción pueden comprender además la hidrólisis de este éster metílico mediante una esterasa. Esto puede incrementar la tendencia del derivado de la reacción (ahora un diácido) a repartirse en una capa acuosa.

En un método ejemplar, una fosfolipasa, por ejemplo una fosfolipasa C, u otra hidrolasa (por ejemplo, una celulasa, una hemicelulasa, una esterasa, una proteasa y/o una fosfatasa) se usa, por ejemplo, para mejorar la extracción del aceite y el desgomado del aceite.

En aspectos alternativos, los métodos y procedimientos de la descripción pueden comprender además la hidrólisis del éster metílico en una clorofila o en una feofitina mediante una esterasa (que puede ser una nueva enzima de la descripción, o una enzima conocida, o una combinación de las mismas). En aspectos alternativos, los métodos de la descripción pueden comprender además eliminar la clorofila modificada en una extracción acuosa. Los métodos pueden comprender además modificar el pH (por ejemplo, incrementando el pH) para promover la separación acuosa de clorofilida. Las enzimas usadas en los métodos, por ejemplo una clorofilasa, se pueden añadir durante este pH incrementado o fase "cáustica" en el procedimiento de separación. Los métodos pueden comprender además una etapa de neutralización cáustica. Los métodos pueden comprender además una etapa de refinado con sílice libre de adsorbente o con poco adsorbente, para eliminar una clorofilida generada por la degradación enzimática de la clorofila. Los métodos pueden comprender además el uso de una hidrolasa, por ejemplo una fosfolipasa C.

En un aspecto de los métodos y procedimientos, el polipéptido es una esterasa (por ejemplo, una enzima de la descripción), por ejemplo una clorofilasa, o tiene actividad semejante a clorofilasa, o tiene actividad catabólica de clorofila. En un aspecto de los métodos, el polipéptido se inmoviliza. El polipéptido se puede inmovilizar sobre un soporte inorgánico o un soporte orgánico. El soporte inorgánico puede comprender alúmina, celite, Dowex-1-cloruro, perlas de vidrio o gel de sílice o equivalente. El polipéptido se puede inmovilizar sobre un hidrogel de alginato o perla de alginato o equivalente. En un aspecto de los métodos, el polipéptido comprende además un liposoma, un hidrogel o un gel.

En un aspecto de los métodos, al menos una etapa se lleva a cabo en una vasija de reacción, por ejemplo una vasija que comprende un dispositivo de separación de goma gravitacional o un tanque de retención o similar. En un aspecto de los métodos, al menos una etapa se lleva a cabo en un extracto celular, o una célula completa. La célula puede ser una célula vegetal, una célula bacteriana, una célula fúngica, una célula de levadura, una célula de mamífero, una célula de insecto, y similar.

En un aspecto de los métodos, la composición que contiene clorofila o contaminada con clorofila comprende un material vegetal, aceite vegetal o extracto vegetal. El material vegetal, aceite vegetal o extracto vegetal puede comprender un aceite vegetal o un aceite de semillas. El aceite vegetal puede comprender un aceite de palma o un aceite de cáñola. Como alternativa, el material vegetal, aceite vegetal o extracto vegetal puede comprender una preparación de algas. En un aspecto de los métodos, las composiciones que contienen clorofila o contaminadas con

clorofila comprenden un producto maderero o no maderero. En un aspecto de los métodos, las composiciones que contienen clorofila o contaminadas con clorofila comprenden un tejido o tela. En un aspecto de los métodos, las composiciones que contienen clorofila o contaminadas con clorofila comprenden una formulación farmacéutica, un alimento, un aceite, un pienso, o un suplemento dietético.

- 5 Las composiciones y métodos de la invención se pueden usar para tratar aceites brutos o refinados, por ejemplo aceites derivados de plantas (por ejemplo vegetales), algas, animales o peces, o de fuentes sintéticas. Las composiciones y métodos de la invención se pueden usar para tratar aceites brutos o refinados a concentraciones de aceite superiores, o, en un aspecto, se pueden usar para tratar aceites brutos sin refinar o no diluidos.

- 10 En un aspecto, los métodos descritos comprenden además eliminar una clorofilida generada mediante degradación enzimática de una clorofila mediante adsorción sobre un gel de sílice o equivalente. Las composiciones que contienen clorofila o contaminadas con clorofila pueden comprender un textil, tela, hilo o tejido o composición relacionada, un producto maderero o papelerero o un subproducto, tal como pasta maderera, una pasta de papel, una pasta Kraft, o un producto de papel no maderero o subproducto, tal como papel de arroz.

- 15 La descripción proporciona productos de fabricación que comprenden un sistema de desgomado para el tratamiento enzimático de composiciones que contienen clorofila o contaminadas con clorofila, que comprenden: (a) un aparato de refinado de aceite vegetal; y (b) un polipéptido que tiene actividad de clorofilasa (por ejemplo, una enzima de la descripción), en el que la actividad del polipéptido comprende la catálisis de una reacción modificadora de la clorofila, y el aparato de refinado de aceite vegetal puede hacer reaccionar una composición que contiene clorofila o contaminada con clorofila con el polipéptido en condiciones en las que el polipéptido puede catalizar una reacción modificadora de la clorofila. En un aspecto del producto de fabricación, el aparato de refinado de aceite vegetal comprende un expulsor que deja un aceite, un tanque de retención o un dispositivo de separación de goma gravitacional. Las reacciones modificadoras de la clorofila pueden comprender la generación de clorofilida y fitol.

- 20 La descripción proporciona detergentes que comprenden un tratamiento enzimático para tejidos que contienen clorofila o contaminados con clorofila, que comprenden: (a) una composición detergente; y (b) un polipéptido que tiene actividad de clorofilasa (por ejemplo, una enzima de la descripción), en el que la actividad comprende la catálisis de una reacción modificadora de la clorofila. En un aspecto, la reacción modificadora de la clorofila comprende la generación de clorofilida y fitol.

- 25 La descripción proporciona métodos para tratar enzimáticamente tejidos que contienen clorofila o contaminados con clorofila, que comprenden: (a) proporcionar una composición detergente que comprende un polipéptido que tiene actividad de clorofilasa (por ejemplo, una enzima de la descripción), en el que la actividad comprende catálisis de una reacción modificadora de la clorofila; y (b) poner en contacto la composición detergente con el tejido que contiene clorofila o contaminado con clorofila en condiciones en las que el polipéptido puede catalizar una reacción modificadora de la clorofila. En un aspecto, la reacción modificadora de la clorofila comprende la generación de clorofilida y fitol.

- 30 La descripción proporciona ácidos nucleicos aislados, sintéticos o recombinantes que comprende una secuencia de ácido nucleico que tiene al menos alrededor de 50%, 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, 60%, 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, o más, o identidad de secuencia completa (100%) con un ácido nucleico ejemplar de la descripción, por ejemplo SEC ID NO:1, SEC ID NO:3, SEC ID NO:5, SEC ID NO:7, SEC ID NO:9, SEC ID NO:11, SEC ID NO:13, SEC ID NO:15, SEC ID NO:17 o SEC ID NO:19, a lo largo de una región de al menos alrededor de 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 75, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, 1000, 1050, 1100, 1150, 1200, 1250, 1300, 1350, 1400, 1450, 1500, 1550, 1600, 1650, 1700, 1750, 1800, 1850, 1900, 1950, 2000, 2050, 2100, 2200, 2250, 2300, 2350, 2400, 2450, 2500, o más restos, secuencias sustancialmente idénticas a ella, y las secuencias complementarias a ella, y codifica al menos un polipéptido que tiene una actividad enzimática como se describe aquí, por ejemplo, una actividad de enzima esterasa.

- 35 En aspectos alternativos, las identidades de secuencia se determinan mediante análisis con un algoritmo de comparación de secuencias o mediante inspección visual. En un aspecto, el algoritmo de comparación de secuencias es el algoritmo BLAST, por ejemplo un algoritmo BLAST versión 2.2.2, en el que el ajuste de filtrado se ajusta a blastall -p blastp -d "nr pataa" -F F, y todas las otras opciones se establecen por defecto.

- 40 Los ácidos nucleicos ejemplares de la descripción también incluyen ácidos nucleicos aislados, sintéticos o recombinantes que codifican un polipéptido de la descripción, por ejemplo un polipéptido que tiene una secuencia como se expone en SEC ID NO:2, SEC ID NO:4, SEC ID NO:6, SEC ID NO:8, SEC ID NO:10, SEC ID NO:12, SEC ID NO:14, SEC ID NO:16, SEC ID NO:18 o SEC ID NO:20, y sus subsecuencias y sus variantes.

- 45 En aspectos alternativos, el polipéptido tiene una actividad de enzima esterasa, incluyendo actividad de clorofilasa (una clase), o actividad enzimática que comprende modificación enzimática de una molécula de clorofila, por ejemplo, en el que la modificación enzimática comprende el catabolismo de la molécula de clorofila. En un aspecto, la actividad de esterasa comprende una actividad de clorofila clorofilido-hidrolasa.

En un aspecto, el ácido nucleico aislado, sintético o recombinante de la descripción codifica un polipéptido que tiene una actividad enzimática que es termoestable. El polipéptido puede retener la actividad enzimática en condiciones que comprenden un intervalo de temperatura de entre alrededor de 37°C y alrededor de 95°C; entre alrededor de 55°C y alrededor de 85°C, entre alrededor de 70°C y alrededor de 95°C, o entre alrededor de 90°C y alrededor de 95°C.

En otro aspecto, un ácido nucleico aislado, sintético o recombinante de la descripción codifica un polipéptido que tiene una enzima que es termotolerante. El polipéptido puede retener actividad enzimática después de la exposición a una temperatura en el intervalo de más de 37°C a alrededor de 95°C, o en cualquier intervalo desde más de 55°C hasta alrededor de 85°C. El polipéptido puede retener la actividad enzimática tras la exposición a una temperatura en el intervalo entre alrededor de 1°C y alrededor de 5°C, entre alrededor de 5°C y alrededor de 15°C, entre alrededor de 15°C y alrededor de 25°C, entre alrededor de 25°C y alrededor de 37°C, entre alrededor de 37°C y alrededor de 95°C, entre alrededor de 55°C y alrededor de 85°C, entre alrededor de 70°C y alrededor de 75°C, o entre alrededor de 90°C y alrededor de 95°C, o más. En un aspecto, el polipéptido retiene la actividad enzimática tras la exposición a una temperatura en el intervalo de más de 90°C a alrededor de 95°C a pH 4,5.

La descripción proporciona ácidos nucleicos aislados, sintéticos o recombinantes que comprenden una secuencia que se hibrida en condiciones restrictivas a un ácido nucleico que comprende una secuencia de la descripción, por ejemplo una secuencia como se expone en SEC ID NO:1, SEC ID NO:3, SEC ID NO:5, SEC ID NO:7, SEC ID NO:9, SEC ID NO:11, SEC ID NO:13, SEC ID NO:15, SEC ID NO:17 o SEC ID NO:19, o sus fragmentos o subsecuencias (o sus complementos). En un aspecto, el ácido nucleico de la descripción codifica un polipéptido que tiene una actividad de enzima esterasa, incluyendo una actividad de clorofilasa (una clase), o actividad de enzima que comprende la modificación enzimática de una molécula de clorofila, por ejemplo, en el que la modificación enzimática comprende el catabolismo de la molécula de clorofila. En un aspecto, la actividad de esterasa comprende una actividad de clorofila clorofilido-hidrolasa. El ácido nucleico puede tener una longitud de al menos alrededor de 10, 15, 20, 25, 30, 35, 40, 45, 50, 75, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, 1000, 1050, 1100, 1150, 1200 o más restos de longitud, o toda la longitud del gen o transcrito. En un aspecto, las condiciones restrictivas incluyen una etapa de lavado que comprende un lavado en 0,2X SSC a una temperatura de alrededor de 65°C durante alrededor de 15 minutos.

La descripción proporciona una sonda de ácido nucleico para identificar un ácido nucleico que codifica un polipéptido que tiene una actividad enzimática como se describe aquí (por ejemplo, actividad de enzima esterasa, incluyendo actividad de clorofilasa (una clase)), en el que la sonda comprende al menos alrededor de 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, 1000 o más bases consecutivas de una secuencia que comprende una secuencia de la descripción, o sus fragmentos o subsecuencias, en el que la sonda identifica el ácido nucleico mediante unión o hibridación. La sonda puede comprender un oligonucleótido que comprende al menos alrededor de 10 a 50, alrededor de 20 a 60, alrededor de 30 a 70, alrededor de 40 a 80, o alrededor de 60 a 100 bases consecutivas de una secuencia que comprende una secuencia de la descripción, o sus fragmentos o subsecuencias.

La descripción proporciona una sonda de ácido nucleico para identificar un ácido nucleico que codifica un polipéptido que tiene al menos una actividad enzimática como se describe aquí (por ejemplo, actividad de enzima esterasa, incluyendo actividad de clorofilasa (una clase)), en el que la sonda comprende un ácido nucleico que comprende una secuencia de al menos alrededor de 10, 15, 20, 30, 40, 50, 60, 70, 80, 90, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, 1000 o más restos que tienen al menos alrededor de 50%, 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, 60%, 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, o más, o identidad de secuencia completa (100%) con un ácido nucleico de la descripción, en el que las identidades de secuencia se determinan mediante análisis con un algoritmo de comparación de secuencias o mediante inspección visual.

La sonda puede comprender un oligonucleótido que comprende al menos alrededor de 10 a 50, alrededor de 20 a 60, alrededor de 30 a 70, alrededor de 40 a 80, o alrededor de 60 a 100 bases consecutivas de una secuencia de ácido nucleico de la descripción, o una subsecuencia de la misma.

La descripción proporciona un par de cebadores de amplificación para amplificar un ácido nucleico que codifica un polipéptido que tiene al menos una actividad enzimática como se describe aquí (por ejemplo, actividad enzimática de esterasa, incluyendo actividad de clorofilasa (una clase)), en el que el par de cebadores es capaz de amplificar un ácido nucleico que comprende una secuencia de la descripción, o sus fragmentos o subsecuencias. Uno o cada miembro del par de secuencias de cebadores de amplificación puede comprender un oligonucleótido que comprende al menos alrededor de 10 a 50 bases consecutivas de la secuencia, o alrededor de 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30 o más bases consecutivas de la secuencia.

La descripción proporciona pares de cebadores de amplificación, en los que el par de cebadores comprende un primer miembro que tiene una secuencia como se expone en alrededor de los primeros (los 5') 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30 o más restos de un ácido nucleico de la descripción, y un

segundo miembro que tiene una secuencia como se expone en alrededor de los primeros (los 5') 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30 o más restos de la hebra complementaria del primer miembro.

La descripción proporciona ácidos nucleicos que codifican proteínas (por ejemplo, enzimas), incluyendo los polipéptidos de la descripción, generadas mediante amplificación, por ejemplo reacción en cadena de la polimerasa (PCR), usando un par de cebadores de amplificación de la descripción. La descripción proporciona ácidos nucleicos que codifican polipéptidos que tienen al menos una actividad enzimática como se describe aquí (por ejemplo, actividad enzimática de esterasa, incluyendo una actividad de clorofilasa (una clasa)) usando un par de cebadores de amplificación de la descripción. La descripción proporciona métodos para obtener y/o identificar enzimas mediante amplificación, por ejemplo reacción en cadena de la polimerasa (PCR), usando un par de cebadores de amplificación de la descripción. En un aspecto, el par de cebadores de amplificación amplifica un ácido nucleico de una biblioteca, por ejemplo una biblioteca génica, tal como una biblioteca medioambiental.

La descripción proporciona métodos para amplificar un ácido nucleico que codifica un polipéptido que tiene actividad enzimática, que comprenden la amplificación de un ácido nucleico molde con un par de secuencias cebadoras de amplificación capaz de amplificar una secuencia de ácido nucleico de la descripción, o sus fragmentos o subsecuencias.

La descripción proporciona casetes de expresión que comprenden un ácido nucleico de la descripción o una subsecuencia del mismo. En un aspecto, el casete de expresión puede comprender el ácido nucleico que está enlazado operablemente a un promotor. El promotor puede ser un promotor vírico, bacteriano, de mamífero o vegetal. En un aspecto, el promotor vegetal puede ser un promotor de patata, de arroz, de maíz, de trigo, de tabaco o de cebada. El promotor puede ser un promotor constitutivo. El promotor constitutivo puede comprender CaMV35S. En otro aspecto, el promotor puede ser un promotor inducible. En un aspecto, el promotor puede ser un promotor específico de tejido, o un promotor regulado medioambientalmente o un promotor regulado mediante el desarrollo. De este modo, el promotor puede ser, por ejemplo, un promotor específico de semillas, un promotor específico de hojas, un promotor específico de raíces, un promotor específico de tallos o un promotor inducido por abscisión. En un aspecto, el casete de expresión puede comprender además un vector de expresión vegetal o de virus vegetal.

La descripción proporciona vehículos de clonación que comprenden un casete de expresión (por ejemplo, un vector) de la descripción o un ácido nucleico de la descripción. El vehículo de clonación puede ser un vector vírico, un plásmido, un fago, un fagómido, un cósmido, un fósido, un bacteriófago o un cromosoma artificial. El vector vírico puede comprender un vector adenovírico, un vector retrovírico o un vector vírico adenoasociado. El vehículo de clonación puede comprender un cromosoma artificial bacteriano (BAC), un plásmido, un vector derivado del bacteriófago P1 (PAC), un cromosoma artificial de levadura (YAC), o un cromosoma artificial de mamífero (MAC).

La descripción proporciona célula transformada que comprende un ácido nucleico de la descripción o un casete de expresión (por ejemplo, un vector) de la descripción, o un vehículo de clonación de la descripción. En un aspecto, la célula transformada puede ser una célula bacteriana, una célula de mamífero, una célula fúngica, una célula de levadura, una célula de insecto o una célula vegetal. En un aspecto, la célula vegetal puede ser una célula de cereal, de patata, de trigo, de arroz, de maíz, de tabaco o de cebada.

La descripción proporciona animales transgénicos no humanos que comprenden un ácido nucleico de la descripción o un casete de expresión (por ejemplo, un vector) de la descripción. En un aspecto, el animal es un ratón.

La descripción proporciona plantas transgénicas que comprenden un ácido nucleico de la descripción o un casete de expresión (por ejemplo, un vector) de la descripción. La planta transgénica puede ser una planta de cereal, una planta de maíz, una planta de patata, una planta de tomate, una planta de trigo, una planta de oleaginosa, una planta de colza, una planta de haba de soja, una planta de arroz, una planta de cebada, o una planta de tabaco.

La descripción proporciona semillas transgénicas que comprenden un ácido nucleico de la descripción o un casete de expresión (por ejemplo, un vector) de la descripción. La semilla transgénica puede ser una planta de cereal, una semilla de maíz, una semilla de trigo, una oleaginosa, una semilla de colza, una semilla de haba de soja, una semilla de palma, una semilla de girasol, una semilla de sésamo, una semilla de cacahuete, o una semilla de planta de tabaco.

La descripción proporciona un oligonucleótido antisentido que comprende una secuencia de ácido nucleico complementaria a o capaz de hibridarse en condiciones restrictivas a un ácido nucleico de la descripción. La descripción proporciona métodos para inhibir la traducción de un mensaje enzimático (de una enzima de la descripción) en una célula, que comprenden administrar a la célula o expresar en la célula un oligonucleótido antisentido que comprende una secuencia de ácido nucleico complementaria a o capaz de hibridarse en condiciones restrictivas a un ácido nucleico de la descripción. En un aspecto, el oligonucleótido antisentido tiene una longitud entre alrededor de 10 y 50, alrededor de 20 y 60, alrededor de 30 y 70, alrededor de 40 y 80, o alrededor de 60 y 100 bases.

La descripción proporciona métodos para inhibir la traducción de un mensaje enzimático en una célula que comprenden administrar a la célula o expresar en la célula un oligonucleótido antisentido que comprende una secuencia de ácido nucleico complementaria a o capaz de hibridarse en condiciones restrictivas a un ácido nucleico

de la descripción. La descripción proporciona moléculas de ARN inhibidor (ARNi) bicatenario que comprenden una subsecuencia de una secuencia de la descripción. En un aspecto, el ARNi tiene una longitud de alrededor de 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25 o más nucleótidos dúplex. La descripción proporciona métodos para inhibir la expresión de un polipéptido (por ejemplo, una enzima de la descripción) en una célula, que comprenden administrar a la célula o expresar en la célula un ARN inhibidor bicatenario (ARNi), en el que el ARN comprende una subsecuencia de una secuencia de la descripción.

La descripción proporciona un polipéptido aislado, sintético o recombinante que comprende una secuencia de aminoácidos que tiene al menos alrededor de 50%, 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, 60%, 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, o más, o una identidad de secuencia completa (100%) con un polipéptido o péptido ejemplar de la descripción a lo largo de una región de al menos alrededor de 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 100, 125, 150, 175, 200, 225, 250, 275, 300, 325, 350 o más restos, o a lo largo de toda la longitud del polipéptido. En un aspecto, las identidades de secuencia se determinan mediante análisis con un algoritmo de comparación de secuencias o mediante una inspección visual. Las secuencias polipeptídicas o peptídicas ejemplares de la descripción incluyen SEC ID NO:2, SEC ID NO:4, SEC ID NO:6, SEC ID NO:8, SEC ID NO:10, SEC ID NO:12, SEC ID NO:14, SEC ID NO:16, SEC ID NO:18 o SEC ID NO:20, y sus subsecuencias y sus variantes. Los polipéptidos ejemplares también incluyen fragmentos de al menos alrededor de 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600 o más restos en longitud, o a lo largo de toda la longitud de una enzima. Las secuencias polipeptídicas o peptídicas ejemplares de la descripción incluyen secuencia codificada por un ácido nucleico de la descripción. Las secuencias polipeptídicas o peptídicas ejemplares de la descripción incluyen polipéptidos o péptidos específicamente unidos mediante un anticuerpo de la descripción. El péptido puede ser, por ejemplo, un fragmento inmunógeno, un motivo (por ejemplo, un sitio de unión), una secuencia señal, una secuencia prepro, un dominio catalítico (CDs) o un sitio activo.

En un aspecto, un polipéptido de la descripción tiene una actividad de esterasa, tal como una actividad de clorofilasa (una clase), o tiene una actividad enzimática que comprende modificación enzimática de una molécula de clorofila, por ejemplo, en el que la modificación enzimática comprende el catabolismo de la molécula de clorofila. En un aspecto, la actividad de esterasa comprende una actividad de clorofila clorofilido-hidrolasa.

Otro aspecto de la presente descripción proporciona un polipéptido o péptido aislado, sintético o recombinante, que incluye al menos 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95 ó 100 o más bases consecutivas de una secuencia polipeptídica o peptídica de la descripción, secuencias sustancialmente idénticas a ella, y las secuencias complementarias a ella. El péptido puede ser, por ejemplo, un fragmento inmunógeno, un motivo (por ejemplo, un sitio de unión), una secuencia señal, una secuencia prepro o un dominio catalítico (CDs) o un sitio activo.

La descripción proporciona sistemas biosintéticos que comprenden ácidos nucleicos y/o plásmidos de la descripción en una célula, por ejemplo una célula de levadura, una célula vegetal, una célula fúngica, o una célula microbiana (por ejemplo, bacteriana). En un aspecto, los sistemas biosintéticos de la descripción comprenden secuencias codificantes para todas las enzimas necesarias, o un subconjunto de las mismas, para el catabolismo de una molécula de clorofila. En un aspecto, las secuencias codificantes pueden estar en un plásmido, un vector recombinante o virus y similar.

En un aspecto, la actividad enzimática de un polipéptido de la descripción, es termoestable. El polipéptido de la descripción, puede retener actividad en condiciones que comprenden un intervalo de temperatura de entre alrededor de 1°C y alrededor de 5°C, entre alrededor de 5°C y alrededor de 15°C, entre alrededor de 15°C y alrededor de 25°C, o entre alrededor de 25°C y alrededor de 37°C, entre alrededor de 37°C y alrededor de 95°C, entre alrededor de 55°C y alrededor de 85°C, entre alrededor de 70°C y alrededor de 75°C, o entre alrededor de 90°C y alrededor de 95°C, o más. En otro aspecto, la actividad enzimática de un polipéptido de la descripción es termotolerante. El polipéptido puede retener actividad tras la exposición a una temperatura en el intervalo de más de 37°C a alrededor de 95°C, o en el intervalo de más de 55°C a alrededor de 85°C. En un aspecto, el polipéptido puede retener actividad tras la exposición a una temperatura en el intervalo de más de 90°C a alrededor de 95°C a pH 4,5.

En un aspecto, el polipéptido aislado, sintético o recombinante puede comprender el polipéptido de la descripción que carece de una secuencia señal. El polipéptido aislado, sintético o recombinante puede comprender el polipéptido de la descripción que comprende una secuencia señal heteróloga.

En un aspecto, la descripción proporciona proteínas quiméricas que comprenden un primer dominio que comprende una secuencia señal de la descripción y al menos un segundo dominio. La proteína puede ser una proteína de fusión. El segundo dominio puede comprender una enzima. La enzima quimérica puede comprender toda o una subsecuencia de al menos un polipéptido que tiene actividad como se describe aquí (por ejemplo, actividad enzimática de esterasa, incluyendo actividad de clorofilasa (una clase)).

La descripción proporciona polipéptidos quiméricos que comprenden al menos un primer dominio que comprende un péptido señal (SP), una secuencia prepro y/o un dominio catalítico (CD) de la descripción, y al menos un segundo

dominio que comprende un polipéptido o péptido heterólogo, en el que el polipéptido o péptido heterólogo no está asociado naturalmente con el péptido señal (SP), una secuencia prepro y/o un dominio catalítico (CD). En un aspecto, el polipéptido o péptido heterólogo no es un polipéptido que tiene actividad que comprende actividad enzimática de esterasa o actividad de catabolismo de la clorofila. El polipéptido o péptido heterólogo puede ser aminoterminal a, carboxiterminal a, o puede estar en ambos extremos del péptido señal (SP), secuencia prepro y/o dominio catalítico (CD).

La descripción proporciona ácidos nucleicos aislados, sintéticos o recombinantes que codifican un polipéptido quimérico, en el que el polipéptido quimérico comprende al menos un primer dominio que comprende un péptido señal (SP), un dominio prepro y/o un dominio catalítico (CD) de la descripción, y al menos un segundo dominio que comprende un polipéptido o péptido heterólogo, en el que el polipéptido o péptido heterólogo no está asociado naturalmente con el péptido señal (SP), secuencia prepro y/o dominio catalítico (CD).

La descripción proporciona secuencias señal aisladas, sintéticas o recombinantes (por ejemplo, péptidos señal) que consisten en o que comprenden una secuencia como se expone en los restos 1 a 14, 1 a 15, 1 a 16, 1 a 17, 1 a 18, 1 a 19, 1 a 20, 1 a 21, 1 a 22, 1 a 23, 1 a 24, 1 a 25, 1 a 26, 1 a 27, 1 a 28, 1 a 28, 1 a 30, 1 a 31, 1 a 32, 1 a 33, 1 a 34, 1 a 35, 1 a 36, 1 a 37, 1 a 38, 1 a 40; 1 a 41, 1 a 42, 1 a 43 o 1 a 44, de un polipéptido de la descripción, por ejemplo, SEC ID NO:2, SEC ID NO:4, SEC ID NO:6, SEC ID NO:8, SEC ID NO:10, SEC ID NO:12, SEC ID NO:14, SEC ID NO:16, SEC ID NO:18 o SEC ID NO:20. La descripción proporciona secuencias señal aisladas, sintéticas o recombinantes (por ejemplo, péptidos señal) que consisten en o que comprenden una secuencia como se expone en la Tabla 1, más abajo.

En un aspecto, una enzima de la descripción tiene una actividad específica a alrededor de 37°C en el intervalo de alrededor de 1 a alrededor de 1200 unidades por miligramo de proteína, o alrededor de 100 a alrededor de 1000 unidades por miligramo de proteína. En otro aspecto, una enzima de la descripción tiene una actividad específica de alrededor de 100 a alrededor de 1000 unidades por miligramo de proteína, o de alrededor de 500 a alrededor de 750 unidades por miligramo de proteína. Como alternativa, una enzima de la descripción tiene una actividad específica a 37°C en el intervalo de alrededor de 1 a alrededor de 750 unidades por miligramo de proteína, o de alrededor de 500 a alrededor de 1200 unidades por miligramo de proteína. En un aspecto, una enzima de la descripción tiene una actividad específica a 37°C en el intervalo de alrededor de 1 a alrededor de 500 unidades por miligramo de proteína, o de alrededor de 750 a alrededor de 1000 unidades por miligramo de proteína. En otro aspecto, una enzima de la descripción tiene una actividad específica a 37°C en el intervalo de alrededor de 1 a alrededor de 250 unidades por miligramo de proteína. Como alternativa, una enzima de la descripción tiene una actividad específica a 37°C en el intervalo de alrededor de 1 a alrededor de 100 unidades por miligramo de proteína. En otro aspecto, la termotolerancia comprende retención de al menos la mitad de la actividad específica de la enzima a 37°C después de calentarla hasta la temperatura elevada. Como alternativa, la termotolerancia puede comprender retención de la actividad específica a 37°C en el intervalo de alrededor de 1 a alrededor de 1200 unidades por miligramo de proteína, o de alrededor de 500 a alrededor de 1000 unidades por miligramo de proteína, después de calentarla hasta la temperatura elevada. En otro aspecto, la termotolerancia puede comprender retención de la actividad específica a 37°C en el intervalo de alrededor de 1 a alrededor de 500 unidades por miligramo de proteína después de calentarla hasta la temperatura elevada.

La descripción proporciona un polipéptido aislado, sintético o recombinante de la descripción, en el que el polipéptido comprende al menos un sitio de glucosilación. En un aspecto, la glucosilación puede ser una glucosilación enlazada mediante N. En un aspecto, el polipéptido se puede glucosilar después de ser expresado en una *P. pastoris* o una *S. pombe*.

En un aspecto, un polipéptido de la presente descripción puede retener actividad enzimática en condiciones que comprenden alrededor de pH 6,5, pH 6, pH 5,5, pH 5, pH 4,5 o pH 4 o más ácidas. En otro aspecto, un polipéptido de la descripción retiene actividad en condiciones que comprenden alrededor de pH 7, pH 7,5, pH 8,0, pH 8,5, pH 9, pH 9,5, pH 10, pH 10,5 o pH 11 o más básicas. En un aspecto, un polipéptido de la descripción retiene actividad enzimática en condiciones que comprenden alrededor de pH 6,5, pH 6, pH 5,5, pH 5, pH 4,5 o pH 4 o más ácidas. En otro aspecto, un polipéptido de la descripción retiene actividad en condiciones que comprenden alrededor de pH 7, pH 7,5, pH 8,0, pH 8,5, pH 9, pH 9,5, pH 10, pH 10,5 o pH 11 o más básicas.

La descripción proporciona preparaciones proteicas que comprenden un polipéptido de la descripción, en las que la preparación proteica comprende un líquido, un sólido o un gel.

La descripción proporciona heterodímeros que comprenden un polipéptido de la descripción y una segunda proteína o dominio. En un aspecto, el segundo miembro del heterodímero no es un polipéptido de la descripción, sino más bien es una enzima diferente u otra proteína. En un aspecto, el segundo dominio puede ser un polipéptido, y el heterodímero puede ser una proteína de fusión. En un aspecto, el segundo dominio puede ser un epítipo o una etiqueta. En un aspecto, la invención proporciona homodímeros que comprenden un polipéptido de la descripción.

La descripción proporciona polipéptidos inmovilizados de la descripción, un polipéptido codificado por un ácido nucleico de la descripción, o un polipéptido que comprende un polipéptido de la descripción y un segundo dominio. En un aspecto, el polipéptido se puede inmovilizar en una célula, un metal, una resina, un polímero, un material

cerámico, un vidrio, un microelectrodo, una partícula gráfica, una perla, un gel, una placa, una matriz o un tubo capilar.

La descripción proporciona matrices que comprenden un ácido nucleico inmovilizado de la descripción. La descripción proporciona matrices que comprenden un anticuerpo de la descripción.

- 5 La invención proporciona anticuerpos aislados, sintéticos o recombinantes que se unen específicamente a un polipéptido de la descripción o a un polipéptido codificado por un ácido nucleico de la descripción. El anticuerpo puede ser un anticuerpo monoclonal o policlonal. La descripción proporciona hibridomas que comprenden un anticuerpo de la descripción, por ejemplo un anticuerpo que se une específicamente a un polipéptido de la descripción o a un polipéptido codificado por un ácido nucleico de la descripción.
- 10 La descripción proporciona un método para aislar o identificar un polipéptido implicado en el catabolismo de clorofila o que tiene actividad de esterasa (por ejemplo, clorofilasa), en el que el método comprende las etapas de: (a) proporcionar un anticuerpo de la descripción; (b) proporcionar una muestra que comprende polipéptidos; y (c) poner en contacto la muestra de la etapa (b) con el anticuerpo de la etapa (a) en condiciones en las que el anticuerpo se puede unir específicamente al polipéptido, aislando o identificando de ese modo al polipéptido.
- 15 La descripción proporciona métodos para obtener un anticuerpo que se une específicamente a un polipéptido de la descripción (por ejemplo, una enzima u otro anticuerpo de la descripción), que comprenden administrar a un animal no humano un ácido nucleico de la descripción o un polipéptido de la descripción o sus subsecuencias, en una cantidad suficiente para generar una respuesta inmunitaria humoral, generando de ese modo una respuesta de anticuerpo. La descripción proporciona métodos para obtener una respuesta inmunitaria humoral o celular, que comprenden administrar a un animal no humano un ácido nucleico de la descripción o un polipéptido de la descripción o sus subsecuencias, en una cantidad suficiente para generar una respuesta inmunitaria.

- 25 La descripción proporciona métodos para producir un polipéptido recombinante, que comprenden las etapas de: (a) proporcionar un ácido nucleico de la descripción enlazado operablemente a un promotor; y (b) expresar el ácido nucleico de la etapa (a) en condiciones que permiten la expresión del polipéptido, produciendo de ese modo un polipéptido recombinante. En un aspecto, el método puede comprender además transformar una célula hospedante con el ácido nucleico de la etapa (a), seguido de la expresión del ácido nucleico de la etapa (a), produciendo de ese modo un polipéptido recombinante en una célula transformada.

- 30 La descripción proporciona métodos para identificar un polipéptido implicado en el catabolismo de la clorofila o que tiene una actividad de esterasa (por ejemplo, clorofilasa), que comprenden las siguientes etapas: (a) proporcionar un polipéptido de la descripción; o un polipéptido codificado por un ácido nucleico de la descripción; (b) proporcionar un sustrato apropiado (por ejemplo, sustrato del polipéptido); y, (c) poner en contacto el polipéptido o un fragmento o variante del mismo de la etapa (a) con el sustrato de la etapa (b) y detectar una disminución en la cantidad de sustrato o un incremento en la cantidad de un producto de reacción, en el que una disminución en la cantidad del sustrato o un incremento en la cantidad del producto de reacción detecta un polipéptido implicado en el catabolismo de la clorofila o que tiene una actividad de esterasa (por ejemplo, clorofilasa).

- 40 La descripción proporciona métodos para identificar un sustrato de un polipéptido implicado en un catabolismo de la clorofila o que tiene una actividad de esterasa (por ejemplo, clorofilasa), en el que el método comprende las siguientes etapas: (a) proporcionar un polipéptido de la descripción; o un polipéptido codificado por un ácido nucleico de la descripción; (b) proporcionar un sustrato de ensayo; y, (c) poner en contacto el polipéptido de la etapa (a) con el sustrato de ensayo de la etapa (b) y detectar una disminución en la cantidad de sustrato o un incremento en la cantidad de producto de reacción, en el que una disminución en la cantidad del sustrato o un incremento en la cantidad de un producto de reacción identifica el sustrato de ensayo, un sustrato de un polipéptido implicado en el catabolismo de la clorofila o que tiene una actividad de esterasa (por ejemplo, clorofilasa).

- 45 La descripción proporciona métodos para determinar si un compuesto de ensayo se une específicamente a un polipéptido, que comprenden las siguientes etapas: (a) expresar un ácido nucleico o un vector que comprende el ácido nucleico en condiciones que permitan la traducción del ácido nucleico a un polipéptido, en el que el ácido nucleico comprende un ácido nucleico de la descripción; o, proporcionar un polipéptido de la descripción; (b) proporcionar un compuesto de ensayo; (c) poner en contacto el polipéptido con el compuesto de ensayo; y (d) determinar si el compuesto de ensayo de la etapa (b) se une específicamente al polipéptido.

- 50 La descripción proporciona métodos para identificar un modulador de un polipéptido implicado en el catabolismo de la clorofila o que tiene una actividad de esterasa (por ejemplo, clorofilasa), que comprenden las siguientes etapas: (a) proporcionar un polipéptido de la descripción o un polipéptido codificado por un ácido nucleico de la descripción; (b) proporcionar un compuesto de ensayo; (c) poner en contacto el polipéptido de la etapa (a) con el compuesto de ensayo de la etapa (b) y medir una actividad del polipéptido, en el que un cambio en la actividad medido en presencia del compuesto de ensayo, comparado con la actividad en ausencia del compuesto de ensayo, proporciona una determinación de que el compuesto de ensayo modula la actividad del polipéptido. En un aspecto, la actividad del polipéptido se puede medir proporcionando un sustrato apropiado (por ejemplo, sustrato del polipéptido implicado en un catabolismo de la clorofila o que tiene una actividad de esterasa (por ejemplo, clorofilasa)) y
- 55

detectando una disminución en la cantidad del sustrato o un incremento en la cantidad de un producto de reacción, o un incremento en la cantidad del sustrato o una disminución en la cantidad de un producto de reacción. Una disminución en la cantidad del sustrato o un incremento en la cantidad del producto de reacción con el compuesto de ensayo, en comparación con la cantidad de sustrato o producto de reacción sin el compuesto de ensayo, identifica al compuesto de ensayo como un activador de la actividad. Un incremento en la cantidad del sustrato o una disminución en la cantidad del producto de reacción con el compuesto de ensayo, en comparación con la cantidad de sustrato o producto de reacción sin el compuesto de ensayo, identifica al compuesto de ensayo como un inhibidor de la actividad.

La descripción proporciona sistemas de ordenador que comprenden un procesador y un dispositivo de almacenamiento de datos, en los que dicho dispositivo de almacenamiento de datos tiene almacenada en él una secuencia polipeptídica o una secuencia de ácido nucleico de la descripción (por ejemplo, un polipéptido codificado por un ácido nucleico de la descripción). En un aspecto, el sistema de ordenador puede comprender además un algoritmo de comparación de secuencias y un dispositivo de almacenamiento de datos que tiene al menos una secuencia de referencia almacenada en él. En otro aspecto, el algoritmo de comparación de secuencias comprende un programa de ordenador que indica polimorfismos. En un aspecto, el sistema de ordenador puede comprender además un identificador que identifica una o más características en dicha secuencia. La descripción proporciona medios legibles por ordenador que tienen almacenados en ellos una secuencia polipeptídica o una secuencia de ácido nucleico de la descripción. La descripción proporciona métodos para identificar una característica en una secuencia, que comprenden las etapas de: (a) leer la secuencia usando un programa de ordenador que identifica una o más características en una secuencia, en el que la secuencia comprende una secuencia polipeptídica o una secuencia de ácido nucleico de la descripción; y, (b) identificar una o más características en la secuencia con el programa de ordenador. La descripción proporciona métodos para comparar una primera secuencia con una segunda secuencia, que comprenden las etapas de: (a) leer la primera secuencia y la segunda secuencia mediante uso de un programa de ordenador que compara secuencias, en el que la primera secuencia comprende una secuencia polipeptídica o una secuencia de ácido nucleico de la descripción; y, (b) determinar las diferencias entre la primera secuencia y la segunda secuencia con el programa de ordenador. La etapa de determinar diferencias entre la primera secuencia y la segunda secuencia puede comprender además la etapa de identificar polimorfismos. En un aspecto, el método puede comprender además un identificador que identifica una o más características en una secuencia. En otro aspecto, el método puede comprender leer la primera secuencia usando un programa de ordenador, e identificar una o más características en la secuencia.

La descripción proporciona métodos para aislar o recuperar un ácido nucleico que codifica un polipéptido que tiene actividad enzimática implicado en el catabolismo de la clorofila o que tiene una actividad de esterasa (por ejemplo, clorofilasa) a partir de una muestra medioambiental, que comprenden las etapas de: (a) proporcionar un par de secuencias de cebadores de amplificación para amplificar un ácido nucleico que codifica un polipéptido implicado en el catabolismo de la clorofila o que tiene una actividad de esterasa (por ejemplo, clorofilasa), en el que el par de cebadores es capaz de amplificar un ácido nucleico de la descripción; (b) aislar un ácido nucleico de la muestra medioambiental o tratar que la muestra medioambiental, tal como un ácido nucleico en la muestra, sea accesible para hibridación al par de cebadores de amplificación; y (c) combinar el ácido nucleico de la etapa (b) con el par de cebadores de amplificación de la etapa (a) y amplificar el ácido nucleico procedente de la muestra medioambiental, aislando o recuperando de ese modo un ácido nucleico que codifica un polipéptido implicado en el catabolismo de la clorofila o que tiene una actividad de esterasa (por ejemplo, clorofilasa) de una muestra medioambiental. En un aspecto, uno o cada miembro del par de cebadores de amplificación puede comprender un oligonucleótido que comprende al menos alrededor de 10 a 50 o más bases consecutivas de una secuencia de la descripción. En un aspecto, el par de cebadores de amplificación es un par de amplificación de la descripción.

La descripción proporciona métodos para aislar o recuperar un ácido nucleico que codifica un polipéptido implicado en el catabolismo de la clorofila o que tiene una actividad de esterasa (por ejemplo, clorofilasa) a partir de una muestra medioambiental, que comprenden las etapas de: (a) proporcionar una sonda polinucleotídica que comprende un ácido nucleico de la descripción, o una subsecuencia del mismo; (b) aislar un ácido nucleico a partir de la muestra medioambiental o tratar que la muestra medioambiental, tal como el ácido nucleico en la muestra, sea accesible para la hibridación a una sonda polinucleotídica de la etapa (a); (c) combinar el ácido nucleico aislado sintético o la muestra medioambiental tratada de la etapa (b) con la sonda polinucleotídica de la etapa (a); y, (d) aislar un ácido nucleico que se hibrida específicamente con la sonda polinucleotídica de la etapa (a), aislando o recuperando de ese modo un ácido nucleico que codifica un polipéptido implicado en el catabolismo de la clorofila o que tiene una actividad de esterasa (por ejemplo, clorofilasa) a partir de una muestra medioambiental. La muestra medioambiental puede comprender una muestra de agua, una muestra de líquido, una muestra de suelo, una muestra de aire o una muestra biológica. En un aspecto, la muestra biológica puede derivar de una célula bacteriana, una célula de protozoo, una célula de insecto, una célula de levadura, una célula vegetal, una célula fúngica o una célula de mamífero.

La descripción proporciona métodos para generar una variante de un ácido nucleico que codifica un polipéptido implicado en el catabolismo de la clorofila o que tiene una actividad de esterasa (por ejemplo, clorofilasa), que comprende las etapas de: (a) proporcionar un ácido nucleico molde que comprende un ácido nucleico de la descripción; y (b) modificar, suprimir o añadir uno o más nucleótidos en la secuencia molde, o una combinación de los mismos, para generar una variante del ácido nucleico molde. En un aspecto, el método puede comprender

además expresar el ácido nucleico variante para generar un polipéptido variante implicado en el catabolismo de la clorofila o que tiene una actividad de esterasa (por ejemplo, clorofilasa). Las modificaciones, adiciones o supresiones se pueden introducir mediante un método que comprende PCR propensa a errores, transposición de fragmentos, mutagénesis dirigida por oligonucleótidos, PCR de ensamblaje, mutagénesis de PCR sexual, mutagénesis *in vivo*, mutagénesis por inserción de un casete, mutagénesis de conjunto recursiva, mutagénesis de conjunto exponencial, mutagénesis específica del sitio, reensamblaje génico, mutagénesis por saturación de sitios génicos (GSSM), reensamblaje por ligación sintética (SLR) o una combinación de los mismos. En otro aspecto, las modificaciones, adiciones o supresiones se introducen mediante un método que comprende recombinación, recombinación de secuencia recursiva, mutagénesis de ADN modificado con fosfotioato, mutagénesis de molde que contiene uracilo, mutagénesis de dúplex con saltos, mutagénesis de reparación de desemparejamiento de punto, mutagénesis de hebra de hospedante deficiente en la reparación, mutagénesis química, mutagénesis radiogénica, mutagénesis de supresión, mutagénesis de restricción-selección, mutagénesis de restricción-purificación, síntesis génica artificial, mutagénesis de conjunto, creación de multímeros de ácidos nucleicos quiméricos y una combinación de los mismos.

En un aspecto, el método se puede repetir iterativamente hasta que se produce un polipéptido implicado en el catabolismo de la clorofila o que tiene una actividad de esterasa (por ejemplo, clorofilasa), que tiene una actividad alterada o diferente o una estabilidad alterada o diferente de la de un polipéptido codificado por el ácido nucleico molde. En un aspecto, el polipéptido variante es termotolerante, y retiene cierta actividad después de exponerlo a una temperatura elevada. En otro aspecto, el polipéptido variante tiene mayor glucosilación en comparación con el polipéptido codificado por un ácido nucleico molde. Como alternativa, el polipéptido variante tiene actividad a una temperatura elevada (o mayor), en el que la enzima codificada por el ácido nucleico molde no es activa a la temperatura elevada. En un aspecto, el método se puede repetir de forma iterativa hasta que se produce una secuencia codificante de enzima que tiene un uso de codones alterado con respecto al del ácido nucleico molde. En otro aspecto, el método se puede repetir de forma iterativa hasta que se produce un gen codificante de la enzima que tiene un nivel mayor o menor de expresión o estabilidad del mensaje con respecto al del ácido nucleico molde.

La descripción proporciona métodos para modificar codones en un ácido nucleico que codifica un polipéptido implicado en el catabolismo de la clorofila o que tiene actividad de esterasa (por ejemplo, clorofilasa) para incrementar su expresión en una célula hospedante, comprendiendo el método las siguientes etapas: (a) proporcionar un ácido nucleico de la descripción que codifica un polipéptido implicado en el catabolismo de la clorofila o que tiene actividad de esterasa (por ejemplo, clorofilasa); y, (b) identificar un codón no preferido o menos preferido en el ácido nucleico de la etapa (a) y sustituirlo por un codón preferido o usado de forma neutra que codifica el mismo aminoácido que el codón sustituido, en el que un codón preferido es un codón sobrerrepresentado en secuencias codificantes en genes en la célula hospedante, y un codón no preferido o menos preferido es un codón subrepresentado en secuencias codificantes en genes en la célula hospedante, modificando de ese modo el ácido nucleico para incrementar su expresión en una célula hospedante.

La descripción proporciona métodos para modificar codones en un ácido nucleico que codifica un polipéptido implicado en el catabolismo de la clorofila o que tiene actividad de esterasa (por ejemplo, clorofilasa), comprendiendo el método las siguientes etapas: (a) proporcionar un ácido nucleico de la descripción; y, (b) identificar un codón en el ácido nucleico de la etapa (a) y sustituirlo por un codón diferente que codifica el mismo aminoácido que el codón sustituido, modificando de ese modo codones en un ácido nucleico que codifica el polipéptido.

La descripción proporciona métodos para modificar codones en un ácido nucleico que codifica un polipéptido que tiene actividad enzimática implicado en el catabolismo de la clorofila o que tiene actividad de esterasa (por ejemplo, clorofilasa) para incrementar su expresión en una célula hospedante, comprendiendo el método las siguientes etapas: (a) proporcionar un ácido nucleico de la descripción que codifica un polipéptido que tiene actividad enzimática implicado en el catabolismo de la clorofila o que tiene actividad de esterasa (por ejemplo, clorofilasa); y, (b) identificar un codón no preferido o menos preferido en el ácido nucleico de la etapa (a) y sustituirlo por un codón preferido o usado de forma neutra que codifica el mismo aminoácido que el codón sustituido, en el que un codón preferido es un codón sobrerrepresentado en secuencias codificantes en genes en la célula hospedante, y un codón no preferido o menos preferido es un codón subrepresentado en secuencias codificantes en genes en la célula hospedante, modificando de ese modo el ácido nucleico para incrementar su expresión en una célula hospedante.

La descripción proporciona métodos para modificar un codón en un ácido nucleico que codifica un polipéptido que tiene actividad enzimática implicado en el catabolismo de la clorofila o que tiene actividad de esterasa (por ejemplo, clorofilasa) para disminuir su expresión en una célula hospedante, comprendiendo el método las siguientes etapas: (a) proporcionar un ácido nucleico de la descripción; y, (b) identificar al menos un codón preferido en el ácido nucleico de la etapa (a) y sustituirlo por un codón no preferido o menos preferido que codifica el mismo aminoácido que el codón sustituido, en el que un codón preferido es un codón sobrerrepresentado en secuencias codificantes en genes en una célula hospedante, y un codón no preferido o menos preferido es un codón subrepresentado en secuencias codificantes en genes en la célula hospedante, modificando de ese modo el ácido nucleico para disminuir su expresión en una célula hospedante. En un aspecto, la célula hospedante puede ser una célula bacteriana, una célula fúngica, una célula de insecto, una célula de levadura, una célula vegetal o una célula de mamífero.

La descripción proporciona métodos para producir una biblioteca de ácidos nucleicos que codifica una pluralidad de sitios activos modificados de polipéptidos (dominios catalíticos (CDs)) o sitios de unión a sustrato de polipéptidos que tienen actividad enzimática implicados en el catabolismo de la clorofila o que tienen actividad de esterasa (por ejemplo, clorofilasa), en el que los sitios activos modificados o sitios de unión a sustrato derivan de un primer ácido nucleico que comprende una secuencia que codifica un primer sitio activo o un primer sitio de unión a sustrato, comprendiendo el método las siguientes etapas: (a) proporcionar un primer ácido nucleico que codifica un primer sitio activo o un primer sitio de unión a sustrato, en el que la secuencia del primer ácido nucleico comprende una secuencia que se hibrida en condiciones restrictivas a un ácido nucleico de la descripción, y el ácido nucleico codifica un sitio activo o un sitio de unión a sustrato; (b) proporcionar un conjunto de oligonucleótidos mutagénicos que codifican variantes de aminoácidos de origen natural en una pluralidad de codones seleccionados como diana en el primer ácido nucleico; y, (c) usar el conjunto de oligonucleótidos mutagénicos para generar un conjunto de ácidos nucleicos variantes que codifican el sitio activo o que codifican el sitio de unión a sustrato que codifican un intervalo de variaciones de aminoácidos en cada codón de aminoácidos que se mutagenizó, produciendo de ese modo una biblioteca de ácidos nucleicos que codifica una pluralidad de sitios activos modificados o sitios de unión a sustrato de polipéptidos que tienen actividad enzimática implicados en el catabolismo de la clorofila o que tienen actividad de esterasa (por ejemplo, clorofilasa). En un aspecto, el método comprende mutagenizar el primer ácido nucleico de la etapa (a) mediante un método que comprende un sistema de evolución dirigida optimizada, mutagénesis de saturación de sitios génicos (GSSM), reensamblaje por ligación sintética (SLR), PCR propensa a errores, barajado, mutagénesis dirigida por oligonucleótidos, PCR de ensamblaje, mutagénesis de PCR sexual, mutagénesis in vivo, mutagénesis por inserción de casete, mutagénesis de conjunto recursiva, mutagénesis de conjunto exponencial, mutagénesis específica del sitio, reensamblaje génico, reensamblaje por ligación sintética (SLR) y/o una combinación de los mismos. En otro aspecto, el método comprende mutagenizar el primer ácido nucleico de la etapa (a) o variantes mediante un método que comprende recombinación, recombinación de secuencias recursiva, mutagénesis de ADN modificado con fosfotioato, mutagénesis de molde que contiene uracilo, mutagénesis de dúplex con saltos, mutagénesis de reparación de desemparejamiento de punto, mutagénesis de hebra hospedante deficiente en la reparación, mutagénesis química, mutagénesis radiogénica, mutagénesis de supresión, mutagénesis de restricción-selección, mutagénesis de restricción-purificación, síntesis génica artificial, mutagénesis de conjunto, creación de multímeros de ácidos nucleicos quiméricos, y/o una combinación de los mismos.

La descripción proporciona métodos para obtener una pequeña molécula, que comprenden las etapas de: (a) proporcionar una pluralidad de enzimas biosintéticas capaces de sintetizar o modificar una pequeña molécula, en el que una de las enzimas comprende un polipéptido de la descripción o es codificado por un ácido nucleico de la descripción; (b) proporcionar un sustrato para al menos una de las enzimas de la etapa (a); y, (c) hacer reaccionar el sustrato de la etapa (b) con las enzimas en condiciones que faciliten una pluralidad de reacciones biocatalíticas para generar una pequeña molécula mediante una serie de reacciones biocatalíticas. La descripción proporciona métodos para modificar una pequeña molécula, que comprenden las etapas de: (a) proporcionar una enzima en el que la enzima comprende un polipéptido de la descripción, o un polipéptido codificado por un ácido nucleico de la descripción, o una subsecuencia del mismo; (b) proporcionar una pequeña molécula; y, (c) hacer reaccionar la enzima de la etapa (a) con la pequeña molécula de la etapa (b) en condiciones que faciliten una reacción enzimática catalizada por el polipéptido de la descripción, o un polipéptido codificado por un ácido nucleico de la descripción, modificando de ese modo una pequeña molécula. En un aspecto, el método puede comprender una pluralidad de sustratos de pequeñas moléculas para la enzima de la etapa (a), generando de ese modo una biblioteca de pequeñas moléculas modificadas producidas mediante al menos una reacción enzimática catalizada por una enzima de la descripción. En un aspecto, el método puede comprender una pluralidad de enzimas adicionales en condiciones que facilitan una pluralidad de reacciones biocatalíticas mediante las enzimas para formar una biblioteca de pequeñas moléculas modificadas producidas mediante la pluralidad de reacciones enzimáticas. En otro aspecto, el método puede comprender además la etapa de ensayar la biblioteca para determinar si una pequeña molécula modificada particular que exhibe una actividad deseada está presente en la biblioteca. La etapa de ensayar la biblioteca puede comprender además las etapas de eliminar sistemáticamente todas menos una de las reacciones biocatalíticas usadas para producir una porción de la pluralidad de las pequeñas moléculas modificadas en la biblioteca ensayando la porción de la pequeña molécula modificada en busca de la presencia o ausencia de la pequeña molécula modificada particular con una actividad deseada, e identificar al menos una reacción biocatalítica específica que produce la pequeña molécula modificada particular de actividad deseada.

La descripción proporciona métodos para determinar un fragmento funcional de un polipéptido de la descripción, o un polipéptido codificado por un ácido nucleico de la descripción, que comprenden las etapas de: (a) proporcionar un polipéptido de la descripción, o un polipéptido codificado por un ácido nucleico de la descripción, o una subsecuencia del mismo; y, (b) suprimir una pluralidad de restos de aminoácidos de la secuencia de la etapa (a) y ensayar la subsecuencia que queda en busca de actividad enzimática o de unión, determinando de ese modo un fragmento funcional de la enzima. En un aspecto, la actividad se mide proporcionando un sustrato y detectando una disminución en la cantidad del sustrato o un incremento en la cantidad de un producto de reacción.

La descripción proporciona métodos para manipular células completas de fenotipos nuevos o modificados usando análisis de flujo metabólico en tiempo real, comprendiendo el método las siguientes etapas: (a) obtener una célula modificada modificando la composición genética de una célula, en la que la composición genética se modifica por

adición a la célula de un ácido nucleico de la descripción; (b) cultivar la célula modificada para generar una pluralidad de células modificadas; (c) medir al menos un parámetro metabólico de la célula monitorizando el cultivo celular de la etapa (b) en tiempo real; y, (d) analizar los datos de la etapa (c) para determinar si el parámetro medido difiere de una medida comparable en una célula no modificada en condiciones similares, identificando de ese modo un fenotipo manipulado en la célula usando análisis de flujo metabólico en tiempo real. En un aspecto, la composición genética de la célula se puede modificar por un método que comprende la supresión de una secuencia o modificación de una secuencia en la célula, o, genosuprimiendo la expresión de un gen. En un aspecto, el método puede comprender adicionalmente seleccionar una célula que comprende un fenotipo recientemente manipulado. En otro aspecto, el método puede comprender cultivar la célula seleccionada, generándose de ese modo una nueva cepa celular que comprende un fenotipo recientemente manipulado.

La descripción proporciona métodos para aumentar la termotolerancia o termoestabilidad de un polipéptido de la descripción, o un polipéptido codificado por un ácido nucleico de la descripción, comprendiendo el método la glucosilación de un polipéptido que comprende al menos treinta aminoácidos contiguos de un polipéptido de la descripción; o un polipéptido codificado por una secuencia de ácido nucleico de la descripción, incrementando de ese modo la termotolerancia o termoestabilidad del polipéptido. En un aspecto, la actividad específica puede ser termoestable o termotolerante a una temperatura en el intervalo de más de alrededor de 37°C a alrededor de 95°C.

La descripción proporciona métodos para sobreexpresar un polipéptido recombinante en una célula, que comprenden expresar un vector que comprende un ácido nucleico que comprende un ácido nucleico de la descripción, o una secuencia de ácido nucleico de la descripción, en el que las identidades de secuencia se determinan mediante análisis con un algoritmo de comparación de secuencias o mediante inspección visual, en el que la sobreexpresión se efectúa mediante el uso de un promotor de alta actividad, un vector dicistrónico o mediante amplificación génica del vector.

La descripción proporciona métodos para obtener una planta transgénica, que comprenden las siguientes etapas: (a) introducir una secuencia de ácido nucleico heteróloga en la célula, en el que la secuencia de ácido nucleico heteróloga comprende una secuencia de ácido nucleico de la descripción, produciendo de ese modo una célula vegetal transformada; y (b) producir una planta transgénica a partir de la célula transformada. En un aspecto, la etapa (a) puede comprender adicionalmente introducir la secuencia de ácido nucleico heteróloga mediante electroporación o microinyección de protoplastos de células vegetales. En otro aspecto, la etapa (a) puede comprender adicionalmente introducir la secuencia de ácido nucleico heteróloga directamente en tejido vegetal por medio de bombardeo de partículas de ADN. Alternativamente, la etapa (a) puede comprender adicionalmente introducir la secuencia de ácido nucleico heteróloga en el ADN de la célula vegetal usando un hospedante de *Agrobacterium tumefaciens*. En un aspecto, la célula vegetal puede ser una célula de patata, maíz, arroz, trigo, tabaco, o cebada.

La descripción proporciona métodos para expresar una secuencia de ácido nucleico heteróloga en una célula vegetal, que comprenden las siguientes etapas: (a) transformar la célula vegetal con una secuencia de ácido nucleico heteróloga enlazada operablemente a un promotor, en el que la secuencia de ácido nucleico heteróloga comprende un ácido nucleico de la descripción; (b) hacer crecer la planta en condiciones en las que la secuencia de ácido nucleico heteróloga se exprese en la célula vegetal. La descripción proporciona métodos para expresar una secuencia de ácido nucleico heteróloga en una célula vegetal, que comprenden las siguientes etapas: (a) transformar la célula vegetal con una secuencia de ácido nucleico heteróloga enlazada operablemente a un promotor, en el que la secuencia de ácido nucleico heteróloga comprende una secuencia de la descripción; (b) hacer crecer la planta en condiciones en las que la secuencia de ácido nucleico heteróloga se exprese en la célula vegetal. En otro aspecto, una enzima de la descripción se puede preparar mediante la expresión de un polinucleótido de la descripción en un organismo seleccionado tal como una bacteria, una levadura, una planta, un insecto, un hongo o un animal. Los organismos ejemplares para expresar polipéptidos de la descripción pueden ser *S. pombe*, *S. cerevisiae*, *Pichia* sp., por ejemplo *P. pastoris*, *E. coli*, *Streptomyces* sp., *Bacillus* sp. y *Lactobacillus* sp.

Otro aspecto de la descripción es un método para obtener un polipéptido de la descripción. El método incluye introducir un ácido nucleico que codifica el polipéptido en una célula hospedante, en el que el ácido nucleico está enlazado operablemente a un promotor, y cultivar la célula hospedante en condiciones que permitan la expresión del ácido nucleico. Otro aspecto de la descripción es un método para obtener un polipéptido o péptido de la descripción. El método incluye introducir un ácido nucleico que codifica el polipéptido en una célula hospedante, en el que el ácido nucleico está enlazado operablemente a un promotor, y cultivar la célula hospedante en condiciones que permitan la expresión del ácido nucleico, produciendo de ese modo el polipéptido.

Otro aspecto de la descripción es un método para generar una variante, que incluye obtener un ácido nucleico que tiene una secuencia de la descripción, secuencias esencialmente idénticas a ella, secuencias complementarias a una secuencia de la descripción, y fragmentos de las mismas, y cambiar uno o más nucleótidos en la secuencia por otro nucleótido, suprimir uno o más nucleótidos en la secuencia, o añadir uno o más nucleótidos a la secuencia.

La descripción proporciona sistemas biosintéticos para el catabolismo de la clorofila, que comprenden al menos una enzima de la descripción. La descripción proporciona sistemas biosintéticos para el catabolismo de la clorofila que comprenden al menos un ácido nucleico que codifica una enzima implicada en el catabolismo de la clorofila, en los

que el ácido nucleico comprende una secuencia de la descripción. En un aspecto, el sistema comprende una pluralidad de ácidos nucleicos que codifican enzimas, en el que las enzimas están implicadas en el catabolismo de la clorofila. En un aspecto, la pluralidad de ácidos nucleicos que codifican enzimas comprende todas las enzimas en una ruta de catabolismo de la clorofila. En un aspecto, la pluralidad de ácidos nucleicos que codifican enzimas está contenida en al menos un plásmido, casete de expresión o vector de expresión.

En un aspecto, el sistema biosintético de la descripción está contenido en (comprende) una célula. La célula puede ser una célula bacteriana, una célula de mamífero, una célula fúngica, una célula de levadura, una célula de insecto, o una célula vegetal. La célula de levadura puede ser una *Pichia* sp. o una *Saccharomyces* sp., tal como *Pichia pastoris*, *Saccharomyces cerevisiae* o *Schizosaccharomyces pombe*.

Los detalles de una o más realizaciones de la descripción se muestran en los dibujos adjuntos y la descripción a continuación. Otras características, objetos, y ventajas de la descripción resultarán evidentes a partir de la descripción y los dibujos, y a partir de las reivindicaciones.

DESCRIPCIÓN DE LOS DIBUJOS

La Figura 1 es una representación esquemática de clorofila (Figura 1A), fitol (Figura 1B) y clorofilida (Figura 1C).

La Figura 2 y la Figura 3 ilustran datos que muestran los resultados de un ensayo de actividad de esterasa (actividad de clorofilasa) usando enzimas ejemplares de la descripción, como se describe con detalle en el Ejemplo 1, más abajo.

La Figura 4 es un diagrama de bloques de un sistema de ordenador ejemplar de la descripción, como se describe con detalle más abajo.

La Figura 5 es un diagrama de flujo que ilustra un aspecto de un procedimiento de la descripción para comparar una nueva secuencia nucleotídica o proteica con una base de datos de secuencias a fin de determinar los niveles de homología entre la nueva secuencia y las secuencias en la base de datos, como se describe con detalle más abajo.

La Figura 6 es un diagrama de flujo que ilustra un aspecto de un procedimiento en un ordenador para determinar si dos secuencias son homólogas, como se describe con detalle más abajo.

La Figura 7 es un diagrama de flujo que ilustra un aspecto de un procedimiento identificador 300 para detectar la presencia de una característica en una secuencia, como se describe con detalle más abajo.

La Figura 8 ilustra la reacción de una esterasa ejemplar de la descripción en la degradación de la clorofila, como se describe con detalle más abajo.

La Figura 9 ilustra y compara la reacción de decoloración (blanqueo) enzimática tradicional frente a la ejemplar de la descripción, como se describe con detalle más abajo.

La Figura 10 ilustra una reacción de decoloración (blanqueo) enzimática ejemplar de la descripción, como se describe con detalle más abajo.

La Figura 11 ilustra un procedimiento de decoloración (blanqueo) enzimática ejemplar de la descripción, que combina etapas de desgomado, blanqueo ("decoloración") enzimático y neutralización cáustica, como se describe con detalle más abajo.

La Figura 12 ilustra un procedimiento de decoloración (blanqueo) enzimática ejemplar de la descripción, como se describe con detalle más abajo.

La Figura 13 ilustra un esquema de refinado de oleaginosa ejemplar, que comprende la extracción, el refinado y la modificación de una oleaginosa usando una esterasa de la descripción, como se describe con detalle más abajo.

La Figura 14 ilustra un procedimiento industrial ejemplar de la descripción – un procedimiento de biodesgomado –, que comprende usar al menos un polipéptido de la descripción, como se describe con detalle más abajo.

La Figura 15 ilustra otro procedimiento industrial ejemplar de la descripción, que comprende usar al menos un polipéptido de la descripción, como se describe con detalle más abajo.

La Figura 16 ilustra otro procedimiento industrial ejemplar de la descripción, que comprende usar al menos un polipéptido de la descripción que tiene actividad de enzima clorofilasa.

Símbolos de referencia similares en los diversos dibujos indican elementos similares.

DESCRIPCIÓN DETALLADA

La presente descripción proporciona nuevas composiciones y procedimientos para el tratamiento enzimático ("blanqueo" o "decoloración") de composiciones que contienen clorofila o contaminadas con clorofila, por ejemplo preparaciones vegetales, animales o de algas, alimentos, piensos o aceites. En un aspecto, el tratamiento (o

5 "blanqueo enzimático" o "decoloración") de clorofila usado en las composiciones y métodos de la descripción comprende el uso de una enzima clorofilasa, u otra enzima implicada en el catabolismo de la clorofila, para modificar la clorofila, por ejemplo para facilitar la eliminación del anillo de porfirina que posee color, mediante, por ejemplo, extracción acuosa. La clorofilasa cataliza la hidrólisis de la clorofila para generar clorofilida, que se puede extraer de forma acuosa, y fitol, que permanece en la fase oleosa.

10 Por ejemplo, en un aspecto, la descripción proporciona composiciones y procedimientos para el procesamiento enzimático (por ejemplo, hidrólisis) de clorofila en un pienso, alimento o aceite, por ejemplo aceites vegetales, incluyendo aceites procesados a partir de oleaginosas, tales como aceite de cáñola (colza) o aceite de haba de soja, o frutos oleosos, tales como aceite de palma. En un aspecto, la descripción proporciona métodos de blanqueo enzimático que usan una enzima clorofilasa para la hidrólisis enzimática de una clorofila o cualquier anillo de

15 porfirina que posea color en un aceite animal o vegetal, por ejemplo aceites vegetales.

La invención incluye métodos para tratar enzimáticamente (por ejemplo, "blanquear") alimentos o aceites que contienen clorofila vía técnicas *in vitro* o *in vivo*, por ejemplo protocolos de células completas, tales como fermentación u otros procedimientos biocatalíticos.

Generación y manipulación de ácidos nucleicos

20 La descripción proporciona ácidos nucleicos aislados, recombinantes y sintéticos (por ejemplo, un ácido nucleico ejemplar de la descripción, incluyendo SEC ID NO:1, SEC ID NO:3, SEC ID NO:5, SEC ID NO:7, SEC ID NO:9, SEC ID NO:11, SEC ID NO:13, SEC ID NO:15, SEC ID NO:17 o SEC ID NO:19), y secuencias que tienen una identidad de secuencia con un ácido nucleico ejemplar; ácidos nucleicos que codifican polipéptidos de la descripción, por ejemplo las secuencias de aminoácidos ejemplares como se exponen en SEC ID NO:2, SEC ID NO:4, SEC ID NO:6,

25 SEC ID NO:8, SEC ID NO:10, SEC ID NO:12, SEC ID NO:14, SEC ID NO:16, SEC ID NO:18 o SEC ID NO:20). La descripción también proporciona casetes de expresión tales como vectores de expresión, que comprenden ácidos nucleicos de la descripción, que incluyen polinucleótidos que codifican los polipéptidos de la descripción. La descripción también incluye métodos para descubrir nuevas secuencias polipeptídicas usando los ácidos nucleicos de la descripción. La descripción también incluye métodos para inhibir la expresión de genes, transcritos y

30 polipéptidos usando los ácidos nucleicos de la descripción. También se proporcionan métodos para modificar los ácidos nucleicos de la descripción, por ejemplo, mediante reensamblaje por ligación sintética, sistema de evolución dirigida optimizada y/o mutagénesis de saturación.

Los ácidos nucleicos usados de la descripción se pueden obtener, aislar y/o manipular, por ejemplo, mediante clonación y expresión de bibliotecas de ADNc, amplificación de ADN mensajero o genómico mediante PCR, y

35 similares.

Al poner en práctica los métodos de la descripción, los genes homólogos se pueden modificar manipulando un ácido nucleico molde, como se describe aquí. La descripción se puede poner en práctica conjuntamente con cualquier método o protocolo o dispositivo conocido en la técnica, que se describen bien en la bibliografía científica y de patentes.

40 Un aspecto de la descripción es un ácido nucleico aislado que comprende una de las secuencias de la descripción, o un fragmento que comprende al menos 10, 15, 20, 25, 30, 35, 40, 50, 75, 100, 150, 200, 300, 400, o 500 bases consecutivas de un ácido nucleico de la descripción. Los ácidos nucleicos aislados pueden comprender ADN, incluyendo ADNc, ADN genómico y ADN sintético. El ADN puede ser bicatenario o monocatenario, y si es monocatenario, puede ser la cadena codificante o la cadena no codificante (anti-sentido). Alternativamente, los

45 ácidos nucleicos aislados pueden comprender ARN.

Los ácidos nucleicos aislados de la descripción se pueden usar para preparar uno de los polipéptidos de la descripción, o fragmentos que comprenden al menos 5, 10, 15, 20, 25, 30, 35, 40, 50, 75, 100, o 150 aminoácidos consecutivos de uno de los polipéptidos de la descripción.

En consecuencia, otro aspecto de la descripción es un ácido nucleico aislado que codifica uno de los polipéptidos de la descripción, o fragmentos que comprenden al menos 5, 10, 15, 20, 25, 30, 35, 40, 50, 75, 100, o 150 aminoácidos consecutivos de uno de los polipéptidos de la descripción. Las secuencias codificantes de estos ácidos nucleicos pueden ser idénticas a una de las secuencias codificantes de uno de los ácidos nucleicos de la descripción, o pueden ser secuencias codificantes diferentes que codifican uno de los polipéptidos de la descripción que tienen al menos 5, 10, 15, 20, 25, 30, 35, 40, 50, 75, 100, o 150 aminoácidos consecutivos de uno de los polipéptidos de la invención, como resultado de la redundancia o degeneración del código genético. El código genético es bien conocido por los expertos en la técnica, y se puede obtener, por ejemplo, en la página 214 de B. Lewin, Genes VI, Oxford University Press, 1997.

55

El ácido nucleico aislado que codifica uno de los polipéptidos de la descripción puede incluir, pero no está limitado a: sólo la secuencia codificante de un ácido nucleico de la descripción y secuencias codificantes adicionales, tales como secuencias líder o secuencias de proproteínas, y secuencias no codificantes, tales como intrones o secuencias no codificantes 5' y/o 3' de la secuencia codificante. De este modo, como se usa aquí, la expresión

5 "polinucleótido que codifica un polipéptido" abarca un polinucleótido que incluye sólo la secuencia codificante para el polipéptido, así como un polinucleótido que incluye una secuencia codificante y/o secuencia no codificante adicional.

Alternativamente, las secuencias de ácido nucleico de la descripción se pueden mutagenizar usando técnicas convencionales, tales como mutagénesis dirigida al sitio, u otras técnicas familiares para los expertos en la técnica, para introducir cambios silenciosos en los polinucleótidos de la descripción. Como se usa en la presente memoria,

10 los "cambios silenciosos" incluyen, por ejemplo, cambios que no alteran la secuencia de aminoácidos codificada por el polinucleótido. Tales cambios pueden ser deseables a fin de incrementar el nivel del polipéptido producido por las células hospedantes que contienen un vector que codifica el polipéptido mediante la introducción de codones o pares de codones que aparecen frecuentemente en el organismo hospedante.

La descripción también se refiere a polinucleótidos que tienen cambios de nucleótidos que dan como resultado sustituciones, adiciones, supresiones, fusiones y truncamientos de aminoácidos en los polipéptidos de la descripción. Tales cambios de nucleótidos se pueden introducir usando técnicas tales como mutagénesis dirigida al sitio, mutagénesis química al azar, supresión mediante exonucleasa III y otras técnicas de ADN recombinante. Alternativamente, tales cambios de nucleótidos pueden ser variantes alélicas de origen natural que se aíslan identificando ácidos nucleicos que se hibridan específicamente a sondas que comprenden al menos 10, 15, 20, 25,

20 30, 35, 40, 50, 75, 100, 150, 200, 300, 400, o 500 bases consecutivas de una de las secuencias de la descripción (o las secuencias complementarias a ellas) en condiciones de restricción elevada, moderada, o baja como se proporciona aquí.

Como se usa aquí, el término "aislado" significa que el material es separado de su entorno original (por ejemplo, el entorno natural si es de origen natural). Por ejemplo, un polinucleótido o polipéptido de origen natural presente en un animal vivo no está aislado, pero el mismo polinucleótido o polipéptido, separado de alguno o de todos los materiales coexistentes en el sistema natural, está aislado. Tales polinucleótidos podían ser parte de un vector, y/o tales polinucleótidos o polipéptidos podían ser parte de una composición y todavía estar aislados por cuanto tal vector o composición no es parte de su entorno natural. Como se usa aquí, el término "purificado" no requiere pureza absoluta; más bien, se pretende que sea una definición relativa. Los ácidos nucleicos individuales obtenidos de una biblioteca se han purificado convencionalmente hasta la homogeneidad electroforética. Las secuencias obtenidas de estos clones no se pudieron obtener directamente a partir de la biblioteca o a partir del ADN humano total. Los ácidos nucleicos purificados de la descripción han sido purificados a partir del resto del ADN genómico en el organismo en al menos 10^4 - 10^6 veces. Sin embargo, el término "purificado" también incluye ácidos nucleicos que han sido purificados a partir del resto del ADN genómico o a partir de otras secuencias en una biblioteca u otro entorno en al menos un orden de magnitud, típicamente dos o tres órdenes, y más típicamente cuatro o cinco órdenes de magnitud.

Como se usa aquí, el término "recombinante" significa que el ácido nucleico está adyacente a un ácido nucleico "esqueleto" al que no está adyacente en su entorno natural. Adicionalmente, para estar "enriquecidos", los ácidos nucleicos representarán 5% o más del número de insertos de ácido nucleico en una población de moléculas esqueléticas de ácido nucleico. Las moléculas esqueléticas de acuerdo con la descripción incluyen ácidos nucleicos tales como vectores de expresión, ácidos nucleicos autorreplicantes, virus, ácidos nucleicos integrantes, y otros vectores o ácidos nucleicos usados para mantener o manipular un inserto de ácido nucleico de interés. Típicamente, los ácidos nucleicos enriquecidos representan 15% o más del número de insertos de ácido nucleico en la población de moléculas esqueléticas recombinantes. Más típicamente, los ácidos nucleicos enriquecidos representan 50% o más del número de insertos de ácido nucleico en la población de moléculas esqueléticas recombinantes. En un aspecto, los ácidos nucleicos enriquecidos representan 90% o más del número de insertos de ácido nucleico en la población de moléculas esqueléticas recombinantes.

Los polipéptidos o proteínas "recombinantes" hacen referencia a polipéptidos o proteínas producidos mediante técnicas de ADN recombinante; es decir, producidos a partir de células transformadas por un constructo de ADN exógeno que codifica el polipéptido o proteína deseados. Los polipéptidos o proteínas "sintéticos" son aquellos preparados por medio de síntesis química. Los métodos de síntesis peptídica química en fase sólida también se pueden usar para sintetizar el polipéptido o fragmentos de la descripción. Tal método se conoce en la técnica desde los comienzos de 1960 (Merrifield, R. B., J. Am. Chem. Soc., 85:2149-2154, 1963) (Véase también Stewart, J. M. y Young, J. D., Solid Phase Peptide Synthesis, 2a Ed., Pierce Chemical Co., Rockford, Ill., p. 11-12)) y ha sido empleado recientemente en kits de diseño y síntesis peptídica de laboratorio comercialmente disponibles (Cambridge Research Biochemicals). Tales kits de laboratorio disponibles comercialmente han utilizado generalmente las enseñanzas de H. M. Geysen et al, Proc. Natl. Acad. Sci., USA, 81:3998 (1984), y proporcionan la síntesis de péptidos sobre las puntas de una multitud de "varillas" o "clavijas", todas las cuales están conectadas a una sola placa. Cuando se usa tal sistema, una placa de varillas o clavijas se invierte y se inserta en una segunda placa de pocillos o reservorios correspondientes, que contienen disoluciones para la unión o el anclaje de un aminoácido apropiado a las puntas de las clavijas o varillas. Repitiendo tal etapa procedimental, es decir, invirtiendo e insertando las puntas de las varillas y clavijas en las disoluciones apropiadas, los aminoácidos se arman para dar

los péptidos deseados. Además, hay disponibles varios sistemas de síntesis peptídica FMOC asequibles. Por ejemplo, el ensamblaje de un polipéptido o fragmento se puede llevar a cabo sobre un soporte sólido utilizando un sintetizador peptídico automático Applied Biosystems, Inc. Modelo 431A. Tal equipo proporciona acceso fácil a los péptidos de la descripción, ya sea mediante síntesis directa o mediante síntesis de una serie de fragmentos que se pueden acoplar usando otras técnicas conocidas.

Una secuencia promotora está “enlazada operablemente a” una secuencia codificante cuando el ARN polimerasa que inicia la transcripción en el promotor transcribirá la secuencia codificante en ARNm.

“Plásmidos” se designan mediante una “p” en minúsculas, precedida y/o seguida de mayúsculas y/o números. Los plásmidos de partida aquí están comercialmente disponibles, están públicamente disponibles en una base no restringida, o se pueden construir a partir de plásmidos disponibles según procedimientos publicados. Además, los plásmidos equivalentes a los descritos aquí son conocidos en la técnica y serán manifiestos para un experto normal.

“Digestión” de ADN hace referencia a la escisión catalítica del ADN con una enzima de restricción que actúa solamente en ciertas secuencias en el ADN. Las diversas enzimas de restricción usadas aquí están comercialmente disponibles, y sus condiciones de reacción, cofactores y otros requisitos se usaron como conocen los expertos normales. Con fines analíticos, típicamente se usa 1 µg de plásmido o fragmento de ADN con alrededor de 2 unidades de enzima en alrededor de 20 µl de disolución tampón. Con el fin de aislar fragmentos de ADN para la construcción del plásmido, se digieren típicamente 5 a 50 µg de ADN con 20 a 250 unidades de enzima en un volumen más grande. Los tampones y las cantidades de sustrato apropiados para las enzimas de restricción particulares se especifican por el fabricante. Normalmente se usan tiempos de incubación de alrededor de 1 hora a 37°C, pero pueden variar de acuerdo con las instrucciones del fabricante. Después de la digestión, se puede llevar a cabo la electroforesis en gel para aislar el fragmento deseado.

“Hibridación” se refiere al proceso mediante el cual una hebra de ácido nucleico se une con una hebra complementaria a través de emparejamiento de bases. Las reacciones de hibridación pueden ser sensibles y selectivas, de manera que se puede identificar una secuencia particular de interés incluso en muestras en las que está presente a concentraciones bajas. Adecuadamente, las condiciones restrictivas se pueden definir, por ejemplo, mediante las concentraciones de sal o formamida en las disoluciones de prehibridación e hibridación, o mediante la temperatura de hibridación, y son bien conocidas en la técnica. En particular, la restricción se puede incrementar reduciendo la concentración de sal, incrementando la concentración de formamida, o elevando la temperatura de hibridación. En aspectos alternativos, los ácidos nucleicos de la descripción se definen por su capacidad para hibridarse en diversas condiciones de restricción (por ejemplo, elevada, media, y baja), como se expone aquí.

Por ejemplo, la hibridación en condiciones de restricción elevada se podría producir en alrededor de 50% de formamida a alrededor de 37°C a 42°C. La hibridación se podría producir en condiciones de restricción reducida en alrededor de 35% a 25% de formamida a alrededor de 30°C a 35°C. En particular, la hibridación se podría producir en condiciones de restricción elevada a 42°C en 50% de formamida, 5X SSPE, 0,3% de SDS, y 200 n/ml de ADN de esperma de salmón cizallado y desnaturalizado. La hibridación se podría producir en condiciones de restricción reducida como se describe anteriormente, pero en 35% de formamida a una temperatura reducida de 35°C. El intervalo de temperatura que corresponde a un nivel particular de restricción se puede estrechar adicionalmente calculando la relación de purina a pirimidina del ácido nucleico de interés y ajustando la temperatura en consecuencia. Las variaciones en los intervalos y condiciones anteriores son bien conocidas en la técnica.

El término “variante” hace referencia a polinucleótidos o polipéptidos de la descripción modificados en uno o más pares de bases, codones, intrones, exones, o restos de aminoácidos (respectivamente), conservando aún la actividad biológica de una enzima de la descripción. Las variantes se pueden producir por cualquier número de medios, incluyendo métodos tales como, por ejemplo, PCR propensa a error, transposición de fragmentos, mutagénesis dirigida por oligonucleótidos, PCR de ensamblaje, mutagénesis de PCR sexual, mutagénesis *in vivo*, mutagénesis por inserción de un casete, mutagénesis de conjunto recursiva, mutagénesis de conjunto exponencial, mutagénesis específica del sitio, reensamblaje génico, GSSM, y cualquier combinación de los mismos.

La expresión “mutagénesis por saturación”, “mutagénesis por saturación de sitios génicos” o “GSSM” incluye un método que usa cebadores oligonucleotídicos degenerados para introducir mutaciones puntuales en un polinucleótido, como se describe con detalle más abajo.

La expresión “sistema de evolución dirigida optimizada” o “evolución dirigida optimizada” incluye un método para reensamblar fragmentos de secuencias de ácidos nucleicos relacionadas, por ejemplo genes relacionados, y se explica con detalle, más abajo.

La expresión “reensamblaje por ligación sintética” o “SLR” incluye un método para ligar fragmentos oligonucleotídicos de una manera no estocástica, y se explica con detalle, más abajo.

Las expresiones “ácido nucleico” o “secuencia de ácido nucleico”, como se usan aquí, se refieren a un oligonucleótido, nucleótido, polinucleótido, o a un fragmento de cualquiera de estos, a ADN o ARN de origen genómico o sintético que puede ser monocatenario o bicatenario y puede representar una hebra sentido o

antisentido (complementaria), a ácido nucleico peptídico (PNA), o a cualquier material semejante a ADN o semejante a ARN, de origen natural o sintético. Las expresiones "ácido nucleico" o "secuencia de ácido nucleico" incluyen oligonucleótido, nucleótido, polinucleótido, o a un fragmento de cualquiera de estos, a ADN o ARN (por ejemplo, ARNm, ARNr, ARNt, ARNi) de origen genómico o sintético que puede ser monocatenario o bicatenario y puede representar una hebra sentido o antisentido, a ácido nucleico peptídico (PNA), o a cualquier material semejante a ADN o semejante a ARN, de origen natural o sintético, incluyendo, por ejemplo, ARNi, ribonucleoproteínas (por ejemplo, ARNi bicatenarios, por ejemplo RNPi). La expresión engloba ácidos nucleicos, es decir, oligonucleótidos, que contienen análogos conocidos de nucleótidos naturales. La expresión también engloba estructuras semejantes a ácidos nucleicos, con cadenas principales sintéticas; véanse, por ejemplo, Mata (1997) Toxicol. Appl. Pharmacol. 144:189-197; Strauss-Soukup (1997) Biochemistry 36:8692-8698; Samstag (1996) Antisense Nucleic Acid Drug Dev 6:153-156. "Oligonucleótido" incluye un polidesoxinucleótido monocatenario o dos hebras polidesoxinucleotídicas complementarias que se pueden sintetizar químicamente. Tales oligonucleótidos sintéticos no tienen 5' fosfato, y de este modo no se ligarán a otro oligonucleótido sin añadir un fosfato con un ATP en presencia de una cinasa. Un oligonucleótido sintético se puede ligar a un fragmento que no se ha desfosforilado.

Una "secuencia codificante de" o una "secuencia nucleotídica que codifica" un polipéptido o proteína particular es una secuencia de ácido nucleico que es transcrita y traducida en un polipéptido o proteína cuando se coloca bajo el control de secuencias reguladoras apropiadas.

El término "gen" significa el segmento de ADN implicado en producir una cadena polipeptídica; incluye regiones que preceden y que siguen a la región codificante (líder y tráiler), así como también, cuando sea aplicable, secuencias interventoras (intrones) entre segmentos codificantes individuales (exones). "Enlazado operablemente", como se usa aquí, hace referencia a una relación funcional entre dos o más segmentos de ácido nucleico (por ejemplo, ADN). En un aspecto, hace referencia a la relación funcional de secuencia reguladora transcripcional a una secuencia transcrita. Por ejemplo, un promotor está enlazado operablemente a una secuencia codificante, tal como un ácido nucleico de la descripción, si estimula o modula la transcripción de la secuencia codificante en una célula hospedante u otro sistema de expresión apropiado. En un aspecto, las secuencias reguladoras de la transcripción promotoras que están enlazadas operablemente a una secuencia transcrita son contiguas físicamente a la secuencia transcrita, es decir, actúan en cis. Sin embargo, algunas secuencias reguladoras de la transcripción, tales como los potenciadores, no necesitan estar contiguos físicamente o situados en estrecha proximidad a las secuencias codificantes cuya transcripción potencian.

La expresión "casete de expresión", como se usa aquí, se refiere a una secuencia nucleotídica que es capaz de efectuar la expresión de un gen estructural (es decir, una secuencia codificante proteica, tal como una enzima de la descripción) en un hospedante compatible con tales secuencias. Los casetes de expresión incluyen al menos un promotor enlazado operablemente con la secuencia codificante polipeptídica; y, opcionalmente, con otras secuencias, por ejemplo señales de terminación de la transcripción. También se pueden usar factores adicionales necesarios o útiles para efectuar la expresión, por ejemplo potenciadores. De este modo, los casetes de expresión también incluyen plásmidos, vectores de expresión, virus recombinantes, cualquier forma de vector de "ADN desnudo" recombinante, y similar. Un "vector" comprende un ácido nucleico que puede infectar, transfectar, transducir transitoria o permanentemente una célula. Se reconocerá que un vector puede ser un ácido nucleico desnudo, o un ácido nucleico complejado con proteína o lípido. El vector comprende opcionalmente ácidos nucleicos y/o proteínas víricas o bacterianas, y/o membranas (por ejemplo, una membrana celular, una cubierta lipídica vírica, etc.). Los vectores incluyen, pero no se limitan a, replicones (por ejemplo, replicones de ARN, bacteriófagos) a los que se pueden unir fragmentos de ADN y se pueden replicar. De este modo, los vectores incluyen, pero no se limitan a, ARN, ADN o ARN circular o lineal autorreplicante autónomo (por ejemplo, plásmidos, virus, y similares; véase, por ejemplo, la patente U.S. nº 5.217.879), e incluyen tanto los plásmidos de expresión como de no expresión. Cuando un microorganismo recombinante o cultivo celular se describe como hospedante de un "vector de expresión", esto incluye tanto ADN circular y lineal extracromosómico como también ADN que se ha incorporado en el cromosoma o cromosomas del hospedante. Cuando un vector se mantiene mediante una célula hospedante, el vector se puede replicar de forma estable mediante las células durante mitosis como una estructura autónoma, o se incorpora en el genoma del hospedante.

Como se usa aquí, el término "promotor" incluye todas las secuencias capaces de dirigir la transcripción de una secuencia codificante en una célula, por ejemplo una célula vegetal. De este modo, los promotores usados en los constructos de la descripción incluyen elementos de control de la transcripción que actúan en cis y secuencias reguladoras que están implicadas en la regulación o la modulación de la duración y/o la velocidad de la transcripción de un gen. Por ejemplo, un promotor puede ser un elemento de control de la transcripción que actúa en cis, incluyendo un potenciador, un promotor, un terminador de la transcripción, un origen de replicación, una secuencia de integración cromosómica, regiones no traducidas 5' y 3', o una secuencia intrónica, que están implicados en la regulación de la transcripción. Estas secuencias que actúan en cis interactúan típicamente con proteínas u otras biomoléculas para llevar a cabo (activar/desactivar, regular, modular, etc.) la transcripción. Los promotores "constitutivos" son aquellos que dirigen la expresión continuamente en la mayoría de las condiciones medioambientales y estados de desarrollo o diferenciación celular. Los promotores "inducibles" o "regulables" dirigen la expresión del ácido nucleico de la descripción bajo la influencia de las condiciones medioambientales o condiciones de desarrollo. Los ejemplos de las condiciones medioambientales que puede afectar a la transcripción por promotores inducibles incluyen condiciones anaerobias, temperatura elevada, sequía, o la presencia de luz.

Los promotores “específicos de tejido” son elementos de control de la transcripción que únicamente son activos en células o tejidos u órganos concretos, por ejemplo en plantas o animales. La regulación específica del tejido se puede lograr por medio de ciertos factores intrínsecos que aseguran que se expresen los genes que codifican proteínas específicas para un tejido dado. Se sabe que tales factores existen en mamíferos y plantas para permitir que se desarrollen tejidos específicos.

Técnicas generales

La presente descripción proporciona nuevas composiciones y procedimientos para tratar enzimáticamente (por ejemplo, “blanquear”) composiciones que contienen clorofila, tales como plantas, algas, alimentos o aceites. El experto reconocerá que los compuestos usados en los métodos de la descripción (por ejemplo, compuestos catalíticos, de partida o intermedios) se pueden sintetizar usando una variedad de procedimientos y metodologías, que están bien descritos en la bibliografía científica y de patentes, *por ejemplo*, Organic Syntheses Collective Volumes, Gilman et al. (Eds) John Wiley & Sons, Inc., NY; Venuti (1989) Pharm Res. 6:867-873. La invención se puede poner en práctica juntamente con cualquier método o protocolo conocido en la técnica, que están bien descritos en la bibliografía científica y de patente.

Los ácidos nucleicos usados para poner en práctica esta descripción, ya sean ARN, ARNi, ácido nucleico antisentido, ADNc, ADN genómico, vectores, virus, o híbridos de los mismos, se pueden aislar de una variedad de fuentes, construir mediante ingeniería genética, amplificar, y/o expresar/generar recombinantemente. Los polipéptidos recombinantes (por ejemplo, enzimas de la descripción) generados a partir de estos ácidos nucleicos se pueden aislar o clonar individualmente y someter a ensayo para determinar una actividad deseada. Se puede usar cualquier sistema de expresión recombinante, incluyendo sistemas de expresión bacterianos, de mamíferos, de levadura, de insecto o de células vegetales.

Alternativamente, estos ácidos nucleicos se pueden sintetizar *in vitro* por medio de técnicas de síntesis química bien conocidas, como se describe, por ejemplo, en Adams (1983) J. Am. Chem. Soc. 105:661; Belousov (1997) Nucleic Acids Res. 25:3440-3444; Frenkel (1995) Free Radic. Biol. Med. 19:373-380; Blommers (1994) Biochemistry 33:7886-7896; Narang (1979) Meth. Enzymol. 68:90; Brown (1979) Meth. Enzymol. 68:109; Beaucage (1981) Tetra. Lett. 22:1859; patente U.S. n° 4.458.066.

Las técnicas para la manipulación de ácidos nucleicos, tales como, por ejemplo, subclonación, sondas de marcaje (por ejemplo, marcaje del cebador al azar usando polimerasa de Klenow, traducción por muescas, amplificación), secuenciación, hibridación, y similares, están bien descritas en la bibliografía científica y de patentes, véanse, por ejemplo, Sambrook, ed., MOLECULAR CLONING: A LABORATORY MANUAL (2ª ED.), Vols. 1-3, Cold Spring Harbor Laboratory, (1989); CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, Ausubel, ed. John Wiley & Sons, Inc., Nueva York (1997); LABORATORY TECHNIQUES IN BIOCHEMISTRY AND MOLECULAR BIOLOGY: HYBRIDIZATION WITH NUCLEIC ACID PROBES, Part I. Theory and Nucleic Acid Preparation, Tijssen, ed. Elsevier, N.Y. (1993).

Otros medios útiles para obtener y manipular ácidos nucleicos usados para poner en práctica los métodos de la descripción es la clonación a partir de muestras genómicas, y, si se desea, escrutar y re-clonar insertos aislados o amplificados a partir de, por ejemplo, clones genómicos o clones de ADNc. Las fuentes de ácido nucleico usadas en los métodos de la descripción incluyen bibliotecas genómicas o de ADNc contenidas, por ejemplo, en cromosomas artificiales de mamífero (MACs), véanse, por ejemplo, las patentes U.S. n°s 5.721.118; 6.025.155; cromosomas artificiales humanos, véase, por ejemplo, Rosenfeld (1997) Nat. Genet. 15:333-335; cromosomas artificiales de levadura (YAC); cromosomas artificiales bacterianos (BAC); cromosomas artificiales P1, véase, por ejemplo, Woon (1998) Genomics 50:306-316; vectores derivados de P1 (PACs), véase, por ejemplo, Kern (1997) Biotechniques 23:120-124; cósmidos, virus recombinantes, fagos o plásmidos.

La presente descripción proporciona nuevas composiciones y procedimientos para tratar enzimáticamente (por ejemplo, “blanquear”) composiciones que contienen clorofila, tales como plantas, algas, alimentos o aceites. El experto reconocerá que los compuestos usados en los métodos de la descripción (por ejemplo, compuestos catalíticos, de partida o intermedios) se pueden sintetizar usando una variedad de procedimientos y metodologías, que están bien descritos en la bibliografía científica y de patentes, por ejemplo, Organic Syntheses Collective Volumes, Gilman et al. (Eds) John Wiley & Sons, Inc., NY; Venuti (1989) Pharm Res. 6:867-873. La descripción se puede poner en práctica juntamente con cualquier método o protocolo conocido en la técnica, que están bien descritos en la bibliografía científica y de patente.

En un aspecto, un ácido nucleico que codifica un polipéptido de la descripción se ensambla en una fase apropiada con una secuencia líder capaz de dirigir secreción del polipéptido traducido o un fragmento del mismo.

La descripción proporciona proteínas de fusión y ácidos nucleicos que las codifican. Un polipéptido de la descripción se puede fusionar a un péptido o polipéptido heterólogo, tal como péptidos de identificación N-terminales que proporcionan características deseadas, tales como estabilidad incrementada o purificación simplificada. Los péptidos y polipéptidos de la descripción también se pueden sintetizar y expresar como proteínas de fusión con uno o más dominios adicionales enlazados a ellas para, por ejemplo, producir un péptido más inmunógeno, para aislar más

fácilmente un péptido sintetizado recombinantemente, para identificar y aislar anticuerpos y linfocitos B que expresan anticuerpos, y similares. Los dominios que facilitan la detección y purificación incluyen, por ejemplo, péptidos quelantes de metales tales como tramos de polihistidina y módulos de histidina-triptófano que permiten la purificación sobre metales inmovilizados, dominios de proteína A que permiten la purificación sobre inmunoglobulina inmovilizada, y el dominio utilizado en el sistema de purificación de extensión/afinidad de FLAGS (Immunex Corp, Seattle WA). La inclusión de secuencias ligadoras escindibles, tales como el Factor Xa o enterocinasa (Invitrogen, San Diego CA) entre un dominio de purificación y el péptido o polipéptido que comprende un motivo para facilitar la purificación. Por ejemplo, un vector de expresión puede incluir una secuencia de ácido nucleico que codifica un epítipo enlazada a seis restos de histidina seguido de una tiorredoxina y un sitio de escisión de enterocinasa (véanse, por ejemplo, Williams (1995) *Biochemistry* 34:1787-1797; Dobeli (1998) *Protein Expr. Purif.* 12:404-414). Los restos de histidina facilitan la detección y purificación, mientras que el sitio de escisión de enterocinasa proporciona un medio para purificar el epítipo del resto de la proteína de fusión. La tecnología que pertenece a vectores que codifican proteínas de fusión, y la aplicación de las proteínas de fusión, está bien descrita en la bibliografía científica y de patentes; véase, por ejemplo, Kroll (1993) *DNA Cell. Biol.*, 12:441-53.

Secuencias de control transcripcionales y traduccionales

La descripción proporciona secuencias de ácido nucleico (por ejemplo, ADN) de la descripción operablemente enlazadas a secuencia o secuencias de control de la expresión (por ejemplo, transcripcionales o traduccionales), por ejemplo promotores o potenciadores, para dirigir o modular la síntesis/expresión del ARN. La secuencia de control de la expresión puede estar en un vector de expresión. Los promotores bacterianos ejemplares incluyen *lacI*, *lacZ*, T3, T7, *gpt*, *lambda PR*, *PL* y *trp*. Los promotores eucariotas ejemplares incluyen el inmediato temprano del CMV, HSV timidina cinasa, temprano y tardío del SV40, LTRs de retrovirus, y metalotioneína I de ratón.

Los promotores adecuados para expresar un polipéptido en bacterias incluyen los promotores *lac* o *trp* de *E. coli*, el promotor *lacI*, el promotor *lacZ*, el promotor T3, el promotor T7, el promotor *gpt*, el promotor *lambda PR*, el promotor *lambda PL*, promotores de operones que codifican enzimas glicolíticas tales como 3-fosfoglicerato cinasa (PGK), y el promotor de fosfatasa ácida. Los promotores eucariotas incluyen el promotor temprano inmediato de CMV, el promotor de HSV timidina cinasa, promotores de choque térmico, el promotor de SV40 temprano y tardío, LTRs de retrovirus, y el promotor de metalotioneína I de ratón. También se pueden usar otros promotores que se sabe que controlan la expresión de genes en células procariotas o eucariotas, o sus virus. Los promotores adecuados para expresar el polipéptido o su fragmento en bacterias incluyen los promotores *lac* o *trp* de *E. coli*, el promotor *lacI*, el promotor *lacZ*, el promotor T3, el promotor T7, el promotor *gpt*, el promotor *lambda PR*, el promotor *lambda PL*, promotores de operones que codifican enzimas glicolíticas tales como 3-fosfoglicerato cinasa (PGK), y el promotor de fosfatasa ácida. Los promotores fúngicos incluyen el promotor del factor ∇ . Los promotores eucariotas incluyen el promotor temprano inmediato de CMV, el promotor de HSV timidina cinasa, promotores de choque térmico, el promotor de SV40 temprano y tardío, LTRs de retrovirus, y el promotor de metalotioneína I de ratón. También se pueden usar otros promotores que se sabe que controlan la expresión de genes en células procariotas o eucariotas, o sus virus.

Promotores Vegetales Específicos de Tejidos

La descripción proporciona casetes de expresión que se pueden expresar de una manera específica del tejido, por ejemplo, que puede expresar una enzima de la descripción de una manera específica del tejido. La descripción también proporciona plantas o semillas que expresan una enzima de la descripción de una manera específica del tejido. La especificidad del tejido puede ser específica de la semilla, específica del tallo, específica de la hoja, específica de la raíz, específica del fruto, y similar.

En un aspecto, un promotor constitutivo tal como el promotor 35S de CaMV se puede usar para la expresión en partes específicas de la planta o semilla o en toda la planta. Por ejemplo, para la sobreexpresión, se puede emplear un fragmento de promotor vegetal que dirigirá la expresión de un ácido nucleico en algunos o en todos los tejidos de una planta, por ejemplo una planta regenerada. Tales promotores son denominados aquí como promotores "constitutivos", y son activos en la mayoría de las condiciones medioambientales y estados de desarrollo o diferenciación celular. Los ejemplos de los promotores constitutivos incluyen la región de iniciación de la transcripción del virus del mosaico de la coliflor (CaMV) 35S, el promotor 1' o 2' derivado de ADN-T de *Agrobacterium tumefaciens*, y otras regiones de iniciación de la transcripción procedentes de diversos genes vegetales conocidos por los expertos en la técnica. Tales genes incluyen, por ejemplo, *ACT11* de *Arabidopsis* (Huang (1996) *Plant Mol. Biol.* 33:125-139); *Cat3* de *Arabidopsis* (GenBank No. U43147, Zhong (1996) *Mol. Gen. Genet.* 251:196-203); el gen codificante de la desaturasa de la proteína portadora de estearoilo-acilo de *Brassica napus* (Genbank No. X74782, Solocombe (1994) *Plant Physiol.* 104:1167-1176); *Gpc1* del maíz (GenBank No. X15596; Martínez (1989) *J. Mol. Biol.* 208:551-565); *Gpc2* del maíz (GenBank No. U45855, Manjunath (1997) *Plant Mol. Biol.* 33:97-112); los promotores vegetales descritos en las patentes U.S. n^{os} 4.962.028 y 5.633.440.

La descripción usa promotores específicos de tejidos o promotores constitutivos derivados de virus que pueden incluir, por ejemplo, el promotor subgenómico de tobamovirus (Kumagai (1995) *Proc. Natl. Acad. Sci. USA* 92:1679-1683; el virus baciliforme tungro del arroz (RTBV), que se replica solamente en células del floema en plantas de arroz infectadas, con su promotor que dirige la expresión fuerte del gen informador específico del floema; el

promotor del virus del mosaico de la nervadura de la yuca (CVMV), con la actividad más alta en los elementos vasculares, en células del mesófilo de las hojas, y en las puntas de las raíces (Verdaguer (1996) Plant Mol. Biol. 31:1129-1139).

Alternativamente, el promotor vegetal puede dirigir la expresión del ácido nucleico que expresa la enzima en un tejido específico, órgano o tipo celular (es decir, promotores específicos de tejido), o puede estar de otro modo bajo control medioambiental o de desarrollo más preciso, o bajo el control de un promotor inducible. Los ejemplos de las condiciones medioambientales que pueden afectar a la transcripción incluyen condiciones anaerobias, temperatura elevada, la presencia de luz, o pulverización con productos químicos/hormonas. Por ejemplo, la descripción incorpora el promotor inducible por sequía del maíz (Busk (1997), más arriba); el promotor inducible por frío, sequía, y alta concentración de sal de la patata (Kirch (1997) Plant Mol. Biol. 33:897 909).

Los promotores específicos de tejido pueden promover la transcripción solamente en un cierto marco temporal del estado de desarrollo en ese tejido. Véase, por ejemplo, Blázquez (1998) Plant Cell 10:791-800, que caracteriza el promotor del gen LEAFY de *Arabidopsis*. Véase también Cardon (1997) Plant J 12:367-77, que describe el factor de transcripción SPL3, que reconoce un motivo de secuencia conservado en la región promotora del gen AP1 de identidad del meristemo floral de *A. thaliana*; y Mandel (1995) Plant Molecular Biology, Vol. 29, págs. 995-1004, que describen el promotor del meristemo eIF4. Se pueden usar promotores específicos de tejido que son activos a lo largo del ciclo de vida de un tejido concreto. En un aspecto, los ácidos nucleicos de la descripción están enlazados operablemente a un promotor activo principalmente durante las etapas de alargamiento de células de fibra de algodón, por ejemplo como se describe por Rinehart (1996), más arriba. Los ácidos nucleicos pueden estar enlazados operablemente al promotor del gen Fb12A para ser expresados preferentemente en células de fibra de algodón (Ibid). Véase también, John (1997) Proc. Natl. Acad. Sci. USA 89:5769-5773; John, et al., patentes U.S. n^{os} 5.608.148 y 5.602.321, que describen promotores específicos de fibra de algodón y métodos para la construcción de plantas de algodón transgénicas. También se pueden usar promotores específicos de raíz para expresar los ácidos nucleicos de la descripción. Los ejemplos de los promotores específicos de raíz incluyen el promotor del gen de la alcohol deshidrogenasa (DeLisle (1990) Int. Rev. Cytol. 123:39-60). Otros promotores que se pueden usar para expresar los ácidos nucleicos de la descripción incluyen, por ejemplo, promotores específicos de ovulo, específicos de embrión, específicos de endospermo, específicos de integumento, específicos del revestimiento del tegumento, o alguna de sus combinaciones; un promotor específico de hoja (véase, por ejemplo, Busk (1997) Plant J. 11:1285 1295, que describe un promotor específico de hoja en el maíz); el promotor ORF13 de *Agrobacterium rhizogenes* (que muestra actividad elevada en raíces; véase, por ejemplo, Hansen (1997), más arriba); un promotor específico de polen de maíz (véase, por ejemplo, Guerrero (1990) Mol. Gen. Genet. 224:161 168); se puede usar un promotor de tomate activo durante la maduración del fruto, la senescencia y la abscisión de las hojas y, en menor grado, de las flores (véase, por ejemplo, Blume (1997) Plant J. 12:731 746); un promotor específico del pistilo procedente del gen SK2 de la patata (véase, por ejemplo, Ficker (1997) Plant Mol. Biol. 35:425 431); el gen Blec4 del guisante, que es activo en tejido epidérmico de ápices de brotes vegetativos y florales de alfalfa transgénica, que lo hacen una herramienta útil para dirigir la expresión de genes foráneos a la capa epidérmica de brotes o fibras que crecen activamente; el gen BEL1 específico de óvulos (véase, por ejemplo, Reiser (1995) Cell 83:735-742, GenBank No. U39944); y/o, el promotor de Klee, patente U.S. n^o 5.589.583, que describe una región promotora vegetal que es capaz de conferir niveles elevados de transcripción en tejido meristémico y/o células que se dividen rápidamente.

Alternativamente, los promotores vegetales que son inducibles con la exposición a hormonas vegetales, tales como auxinas, se usan para expresar los ácidos nucleicos de la descripción. Por ejemplo, la descripción puede usar el fragmento de promotor E1 de los elementos de respuesta a auxinas (AuxREs) en la soja (*Glycine max* L.) (Liu (1997) Plant Physiol. 115:397-407); el promotor GST6 de *Arabidopsis* sensible a auxinas (también sensible a ácido salicílico y peróxido de hidrógeno) (Chen (1996) Plant J. 10: 955-966); el promotor parC inducible por auxinas del tabaco (Sakai (1996) 37:906-913); un elemento de respuesta a biotina vegetal (Streit (1997) Mol. Plant Microbe Interact. 10:933-937); y, el promotor sensible a la hormona del estrés ácido abscísico (Sheen (1996) Science 274:1900-1902).

Los ácidos nucleicos de la descripción se pueden enlazar también operablemente a promotores vegetales que son inducibles con la exposición a sustancias químicas que se pueden aplicar a la planta, tales como herbicidas o antibióticos. Por ejemplo, se puede usar el promotor In2-2 de maíz, activado por protectores contra herbicidas de bencenosulfonamida (De Veylder (1997) Plant Cell Physiol. 38:568-577); la aplicación de diferentes protectores contra herbicidas induce distintos patrones de expresión génica, incluyendo expresión en la raíz, hidatodos, y el meristemo apical de los brotes. La secuencia codificante puede estar bajo el control de, por ejemplo, un promotor inducible por tetraciclina, por ejemplo como se describe con las plantas de tabaco transgénicas que contienen el gen de la arginina descarboxilasa de *Avena sativa* L. (avena) (Masgrau (1997) Plant J. 11:465-473); o un elemento sensible al ácido salicílico (Stange (1997) Plant J. 11:1315-1324). Usando promotores inducidos químicamente (por ejemplo, por hormonas o plaguicidas), es decir, un promotor sensible a un agente químico que se puede aplicar a la planta transgénica en el campo, se puede inducir la expresión de un polipéptido de la descripción en una etapa concreta de desarrollo de la planta. De este modo, la descripción también proporciona plantas transgénicas que contienen un gen inducible que codifica polipéptidos de la descripción cuya gama de anfitriones está limitada a especies de plantas diana, tales como maíz, arroz, cebada, trigo, patata u otros cultivos, inducible en cualquier etapa de desarrollo del cultivo.

Un experto en la técnica advertirá que un promotor vegetal específico de tejido puede dirigir la expresión de secuencias enlazadas operablemente en tejidos distintos del tejido diana. De este modo, un promotor específico de tejido es el que dirige la expresión preferentemente en el tejido diana o tipo celular, pero puede además conducir también a cierta expresión en otros tejidos.

Los ácidos nucleicos de la invención también pueden estar enlazados operablemente a promotores vegetales que son inducibles con la exposición a reactivos químicos. Estos reactivos incluyen, por ejemplo, herbicidas, auxinas sintéticas, o antibióticos que se pueden aplicar, por ejemplo pulverizar, sobre las plantas transgénicas. La expresión inducible de los ácidos nucleicos que producen la enzima de la descripción permitirá al productor seleccionar plantas con expresión y/o actividad enzimática óptima. De este modo se puede controlar el desarrollo de partes de las plantas. De este modo, la descripción proporciona el medio para facilitar la cosecha de plantas y partes de plantas. Por ejemplo, en diversas realizaciones, se usa el promotor In2-2 de maíz, activado por protectores contra herbicidas de bencenosulfonamida (De Veylder (1997) *Plant Cell Physiol.* 38:568-577); la aplicación de diferentes protectores contra herbicidas induce distintos patrones de expresión génica, incluyendo expresión en la raíz, hidatodos, y el meristemo apical del brote. Las secuencias codificantes de la descripción también están bajo el control de un promotor de tetraciclina inducible, por ejemplo como se describe con plantas de tabaco transgénicas que contienen el gen de la arginina descarboxilasa de *Avena sativa* L. (avena) (Masgrau (1997) *Plant J.* 11:465-473); o un elemento sensible al ácido salicílico (Stange (1997) *Plant J.* 11:1315-1324).

En algunos aspectos, la expresión apropiada del polipéptido puede requerir una región de poliadenilación en el extremo 3' de la región codificante. La región de poliadenilación puede derivar del gen natural, de una variedad de genes de otra planta (o animal u otro), o de genes en el ADN-T agrobacteriano.

El término "planta" incluye plantas completas, partes de plantas (por ejemplo, hojas, tallos, flores, raíces, etc.), protoplastos vegetales, semillas y células vegetales y sus progenies. La clase de plantas que se pueden usar en el método de la descripción es generalmente tan amplio como la clase de plantas superiores susceptibles de técnicas de transformación, incluyendo angiospermas (plantas monocotiledóneas y dicotiledóneas), así como gimnospermas. Incluye plantas de una variedad de niveles de ploidía, incluyendo los estados poliploide, diploide, haploide y hemizigoto. Como se usa aquí, el término "planta transgénica" incluye plantas o células vegetales en las que se ha insertado una secuencia de ácido nucleico heteróloga, por ejemplo, los ácidos nucleicos y diversos constructos recombinantes (por ejemplo, casetes de expresión) de la descripción.

Vectores de expresión y vehículos de clonación

La descripción proporciona vectores de expresión y vehículos de clonación que comprenden ácidos nucleicos de la descripción, por ejemplo secuencias que codifican las enzimas de la descripción. Los vectores de expresión y los vehículos de clonación de la descripción pueden comprender partículas víricas, baculovirus, fago, plásmidos, fagómidos, cósmidos, fósidos, cromosomas artificiales bacterianos, ADN vírico (por ejemplo de la vacuna, adenovirus, virus de la viruela, pseudorrabia y derivados de SV40), cromosomas artificiales a base de P1, plásmidos de levadura, cromosomas artificiales de levadura, y cualesquiera otros vectores específicos para hospedantes específicos de interés (tales como *Bacillus*, *Aspergillus* y levadura). Los vectores de la descripción pueden incluir secuencias de ADN sintéticas, cromosómicas y no cromosómicas. Un gran número de vectores adecuados son conocidos por los expertos en la técnica, y están comercialmente disponibles. Los vectores ejemplares incluyen: vectores bacterianos: pQE (Qiagen), plásmidos pBluescript, vectores pNH, (vectores lambda-ZAP (Stratagene); ptrc99a, pKK223-3, pDR540, pRIT2T (Pharmacia); eucariotas: pXT1, pSG5 (Stratagene), pSVK3, pBPV, pMSG, pSVLSV40 (Pharmacia). Sin embargo, se puede usar cualquier otro plásmido u otro vector con tal de que sean replicables y viables en el hospedante. Con la presente descripción, se pueden emplear vectores con bajo número de copias o alto número de copias.

Los "plásmidos" pueden estar comercialmente disponibles, públicamente disponibles en una base no restringida, o se pueden construir a partir de plásmidos disponibles según procedimientos publicados. Los plásmidos equivalentes a los descritos aquí son conocidos en la técnica y serán manifiestos para un experto normal.

El vector de expresión puede comprender un promotor, un sitio de unión al ribosoma para el inicio de la traducción, y un terminador de la transcripción. El vector también puede incluir secuencias apropiadas para amplificar la expresión. El vector puede incluir también las secuencias apropiadas para amplificar la expresión. Los vectores de expresión de mamíferos pueden comprender un origen de replicación, cualesquiera sitios de unión al ribosoma necesarios, un sitio de poliadenilación, sitios dadores y aceptores de ajuste, secuencias de terminación de la transcripción, secuencias no transcritas flanqueantes 5'. En algunos aspectos, se pueden usar secuencias de ADN derivadas de los sitios de ajuste y poliadenilación de SV40 para proporcionar los elementos genéticos no transcritos requeridos.

En un aspecto, los vectores de expresión contienen uno o más genes marcadores seleccionables, para permitir la selección de células hospedantes que contienen el vector. Tales marcadores seleccionables incluyen genes que codifican dihidrofolato reductasa, o genes que confieren resistencia a neomicina para cultivo de células eucariotas, genes que confieren resistencia a tetraciclina o a ampicilina en *E. coli*, y el gen TRP1 de *S. cerevisiae*. Las regiones

promotoras se pueden seleccionar de cualquier gen deseado usando vectores de cloranfenicol transferasa (CAT) u otros vectores con marcadores seleccionables.

Los vectores para expresar el polipéptido o su fragmento en células eucariotas también pueden contener potenciadores para incrementar los niveles de expresión. Los potenciadores son elementos de ADN que actúan en cis, habitualmente de alrededor de 10 a alrededor de 300 pb de longitud, que actúan sobre un promotor para incrementar su transcripción. Los ejemplos incluyen el potenciador de SV40 en el lado tardío del origen de replicación pb 100 a 270, el potenciador del promotor temprano de citomegalovirus, el potenciador de polioma en el lado tardío del origen de replicación, y los potenciadores adenovíricos.

Una secuencia de ácido nucleico se puede insertar en un vector mediante una variedad de procedimientos. En general, la secuencia se liga a la posición deseada en el vector tras la digestión del inserto y el vector con endonucleasas de restricción apropiadas. Como alternativa, se pueden ligar extremos romos tanto en el inserto como en el vector. En la técnica se conoce una variedad de técnicas de clonación, por ejemplo como se describe en Ausubel y Sambrook. Se considera que tales procedimientos y otros están dentro del alcance de los expertos en la técnica.

El vector puede estar en forma de un plásmido, una partícula vírica, o un fago. Otros vectores incluyen secuencias de ADN cromosómicas, no cromosómicas y sintéticas, derivados de SV40, plásmidos bacterianos, ADN fágico, baculovirus, plásmidos de levadura, vectores derivados de combinaciones de plásmidos y ADN fágico, ADN vírico tal como el virus de la vacuna, adenovirus, virus de la viruela, y pseudorrabia. Una variedad de vectores de clonación y de expresión para uso con hospedantes procariotas y eucariotas se describen, por ejemplo, por Sambrook.

Los vectores bacterianos particulares que se pueden usar incluyen los plásmidos comercialmente disponibles que comprenden elementos genéticos del vector de clonación bien conocido pBR322 (ATCC 37017), pKK223-3 (Pharmacia Fine Chemicals, Uppsala, Suecia), GEM1 (Promega Biotec, Madison, WI, USA) pQE70, pQE60, pQE-9 (Qiagen), pD10, psiX174 pBluescript II KS, pNH8A, pNH16a, pNH18A, pNH46A (Stratagene), ptrc99a, pKK223-3, pKK233-3, DR540, pRIT5 (Pharmacia), pKK232-8 y pCM7. Los vectores eucariotas particulares incluyen pSV2CAT, pOG44, pXT1, pSG (Stratagene) pSVK3, pBPV, pMSG, y pSVL (Pharmacia). Sin embargo, se puede usar cualquier otro vector en tanto que sea replicable y viable en la célula hospedante.

Los ácidos nucleicos de la descripción se pueden expresar en casetes de expresión, vectores o virus, y se pueden expresar transitoria o establemente en células vegetales y semillas. Un sistema de expresión transitorio ejemplar usa sistemas de expresión episómicos, por ejemplo, ARN vírico del virus del mosaico de la coliflor (CaMV) generado en el núcleo mediante transcripción de un mini-cromosoma episómico que contiene ADN superenrollado, véase, por ejemplo, Covey (1990) Proc. Natl. Acad. Sci. USA 87:1633-1637. Como alternativa, secuencias codificantes, es decir, todas o sub-fragmentos de secuencias de la descripción se pueden insertar en un genoma de una célula hospedante vegetal convirtiéndose en una parte integrante del ADN cromosómico del hospedante. De esta manera, se pueden expresar transcritos sentido o antisentido. Un vector que comprende las secuencias (por ejemplo, promotores o regiones codificantes) de ácidos nucleicos de la descripción puede comprender un gen marcador que confiere un fenotipo seleccionable sobre una célula vegetal o una semilla. Por ejemplo, el marcador puede codificar resistencia a biocidas, particularmente resistencia a antibióticos, tal como resistencia a kanamicina, G418, bleomicina, higromicina, o resistencia a herbicidas, tal como resistencia a clorosulfurón o Basta.

Los vectores de expresión capaces de expresar ácidos nucleicos y proteínas en plantas son bien conocidos en la técnica, y pueden incluir, por ejemplo, vectores de *Agrobacterium* spp., virus X de la patata (véase, por ejemplo, Angell (1997) EMBO J. 16:3675-3684), virus del mosaico del tabaco (véase, *por ejemplo*, Casper (1996) Gene 173:69-73), virus del enanismo arbustivo del tomate (véase, por ejemplo, Hillman (1989) Virology 169:42-50), virus del grabado del tabaco (véase, por ejemplo, Dolja (1997) Virology 234:243-252), virus del mosaico amarillo de la judía (véase, *por ejemplo*, Morinaga (1993) Microbiol Immunol. 37:471-476), virus del mosaico de la coliflor (véase, *por ejemplo*, Cecchini (1997) Mol. Plant Microbe Interact. 10:1094-1101), elemento transponible del maíz Ac/Ds (véanse, *por ejemplo*, Rubin (1997) Mol. Cell. Biol. 17:6294-6302; Kunze (1996) Curr. Top. Microbiol. Immunol. 204:161-194), y elemento transponible supresor-mutador (Spm) del maíz (véase, *por ejemplo*, Schlappi (1996) Plant Mol. Biol. 32:717-725); y derivados de los mismos.

En un aspecto, el vector de expresión puede tener dos sistemas de replicación para permitirle mantenerse en dos organismos, por ejemplo en células de mamífero o insecto para su expresión y en un hospedante procariota para su clonación y amplificación. Además, para los vectores de expresión integrantes, el vector de expresión puede contener al menos una secuencia homóloga al genoma de la célula hospedante. Puede contener dos secuencias homólogas que flanquean el constructo de expresión. El vector integrante puede dirigirse a un locus específico en la célula hospedante seleccionando la secuencia homóloga apropiada para la inclusión en el vector. Los constructos para los vectores integrantes son bien conocidos en la técnica.

Los vectores de expresión de la descripción pueden incluir también un gen marcador seleccionable para permitir la selección de cepas bacterianas que se han transformado, por ejemplo genes que vuelven la bacteria resistente a fármacos tales como ampicilina, cloranfenicol, eritromicina, kanamicina, neomicina y tetraciclina. Los marcadores

seleccionables pueden incluir también genes biosintéticos, tales como los de las rutas biosintéticas de la histidina, el triptófano y la leucina.

La secuencia de ADN en el vector de expresión está enlazada operativamente a una secuencia o secuencias de control de la expresión (promotor) apropiadas, para dirigir la síntesis de ARN. Los promotores bacterianos designados concretos incluyen *lacI*, *lacZ*, *T3*, *T7*, *gpt*, P_R , P_L lambda y *trp*. Los promotores eucariotas incluyen el temprano inmediato de CMV, timidina cinasa del HSV, temprano y tardío de SV40, las LTR de retrovirus, y metalotioneína I de ratón. La selección del vector y promotor apropiados está al nivel de conocimiento normal en la técnica. El vector de expresión también contiene un sitio de unión al ribosoma para el comienzo de la traducción, y un terminador de la transcripción. El vector también puede incluir secuencias apropiadas para amplificar la expresión. Las regiones promotoras se pueden seleccionar entre cualquier gen deseado usando vectores de la cloranfenicol transferasa (CAT) u otros vectores con marcadores seleccionables. Además, en un aspecto, los vectores de expresión contienen uno o más genes marcadores seleccionables para proporcionar un rasgo fenotípico para la selección de células hospedantes transformadas, tal como dihidrofolato reductasa o resistencia a neomicina para un cultivo de células eucariotas, o tal como resistencia a tetraciclina o ampicilina en *E. coli*.

Los vectores de expresión de mamíferos puede comprender también un origen de replicación, cualesquiera sitios de unión al ribosoma necesarios, un sitio de poliadenilación, sitios dadores o aceptores de ajuste, secuencias de terminación de la transcripción y secuencias no transcritas flanqueantes 5'. En algunos aspectos, se pueden usar secuencias de ADN derivadas de los sitios de ajuste y poliadenilación de SV40 para proporcionar los elementos genéticos no transcritos requeridos.

Los vectores para expresar el polipéptido o su fragmento en células eucariotas también pueden contener potenciadores para incrementar los niveles de expresión. Los potenciadores son elementos de ADN que actúan en cis, habitualmente de alrededor de 10 a alrededor de 300 pb de longitud, que actúan sobre un promotor para incrementar su transcripción. Los ejemplos incluyen el potenciador de SV40 en el lado tardío del origen de replicación pb 100 a 270, el potenciador del promotor temprano de citomegalovirus, el potenciador de polioma en el lado tardío del origen de replicación, y los potenciadores adenovíricos.

Además, la vectores de expresión contienen típicamente uno o más genes marcadores seleccionables para permitir la selección de células hospedantes que contienen el vector. Tales marcadores seleccionables incluyen genes que codifican genes de la dihidrofolato reductasa o genes que confieren resistencia a la neomicina para el cultivo de células eucariotas, genes que confieren resistencia a tetraciclina o ampicilina en *E. coli* y el gen *TRP1* de *S. cerevisiae*.

En algunos aspectos, el ácido nucleico que codifica uno de los polipéptidos de la descripción, o fragmentos que comprenden al menos alrededor de 5, 10, 15, 20, 25, 30, 35, 40, 50, 75, 100, o 150 aminoácidos consecutivos del mismo, se ensambla en la fase apropiada con una secuencia líder capaz de dirigir la secreción del polipéptido traducido o un fragmento del mismo. Opcionalmente, el ácido nucleico puede codificar un polipéptido de fusión en el que uno de los polipéptidos de la descripción, o fragmentos que comprenden al menos 5, 10, 15, 20, 25, 30, 35, 40, 50, 75, 100, o 150 aminoácidos consecutivos del mismo, se fusiona a péptidos o polipéptidos heterólogos, tales como péptidos de identificación N-terminal que confieren las características deseadas, tales como mayor estabilidad o purificación simplificada.

La secuencia de ADN apropiada se puede insertar en el vector por medio de una variedad de procedimientos. En general, la secuencia de ADN se liga a la posición deseada en el vector después de la digestión del inserto y el vector con endonucleasas de restricción apropiadas. Alternativamente, se pueden ligar los extremos romos tanto en el inserto como en el vector. Una variedad de técnicas de clonación son descritas en Ausubel et al. Current Protocols in Molecular Biology, John Wiley 503 Sons, Inc. 1997 y Sambrook et al., Molecular Cloning: A Laboratory Manual 2a Ed., Cold Spring Harbor Laborator y Press (1989). Se considera que tales procedimientos y otros están al alcance de los expertos en la técnica.

El vector puede estar, por ejemplo, en forma de un plásmido, una partícula vírica, o un fago. Otros vectores incluyen secuencias de ADN cromosómicas, no cromosómicas y sintéticas, derivados de SV40, plásmidos bacterianos, ADN fágico, baculovirus, plásmidos de levadura, vectores derivados de combinaciones de plásmidos y ADN fágico, ADN vírico tal como el virus de la vacuna, adenovirus, virus de la viruela, y pseudorrabia. Una variedad de vectores de clonación y de expresión para uso con hospedantes procariotas y eucariotas son descritos por Sambrook, et al., Molecular Cloning: A Laboratory Manual, 2a Ed., Cold Spring Harbor, N.Y., (1989).

Células hospedantes y células transformadas

La descripción también proporciona una célula transformada que comprende una secuencia de ácido nucleico de la descripción, por ejemplo una secuencia que codifica una enzima de la descripción, o un vector de la descripción. La célula hospedante puede ser cualquiera de las células hospedantes familiares para los expertos en la técnica, incluyendo células procariotas, células eucariotas, tales como células bacterianas, células fúngicas, células de levadura, células de mamífero, células de insecto, o células vegetales. Las células bacterianas ejemplares incluyen *E. coli*, *Lactococcus lactis*, *Streptomyces*, *Bacillus subtilis*, *Bacillus cereus*, *Salmonella typhimurium* o cualquier

especie dentro de los géneros *Bacillus*, *Streptomyces*, y *Staphylococcus*. Las células de insecto ejemplares incluyen S2 de *Drosophila* y Sf9 de *Spodoptera*. Las células de levadura ejemplares incluyen *Pichia pastoris*, *Saccharomyces cerevisiae* o *Schizosaccharomyces pombe*. Las células animales ejemplares incluyen CHO, COS o melanoma de Bowes, o cualquier línea celular de ratón o humana. La selección de un hospedante apropiado se encuentra dentro de las capacidades de los expertos en la técnica. Las técnicas para la transformación de una amplia variedad de especies de plantas superiores son bien conocidas y se describen en la bibliografía técnica y científica. Véanse, por ejemplo, Weising (1988) Ann. Rev. Genet. 22:421-477; patente U.S. 5.750.870.

El vector se puede introducir en las células hospedantes usando cualquiera de una variedad de técnicas, incluyendo transformación, transfección, transducción, infección vírica, pistolas génicas, o transferencia génica mediada por plásmidos Ti. Los métodos particulares incluyen transfección con fosfato de calcio, transfección mediada por DEAE-Dextrano, lipofección, o electroporación (Davis, L., Dibner, M., Battey, I., Basic Methods in Molecular Biology, (1986)).

En un aspecto, los ácidos nucleicos o vectores de la descripción se introducen en las células para escrutar, de ese modo, los ácidos nucleicos que entran en las células de una manera adecuada para la posterior expresión del ácido nucleico. El método de introducción está dictado en gran parte por el tipo de célula seleccionada como diana. Los métodos ejemplares incluyen precipitación con CaPO_4 , fusión de liposomas, lipofección (por ejemplo, LIPOFECTIN™), electroporación, infección vírica, etc. Los ácidos nucleicos candidato se pueden integrar establemente en el genoma de la célula hospedante (por ejemplo, con introducción retrovírica) o pueden existir transitoria o establemente en el citoplasma (es decir, a través del uso de plásmidos tradicionales, utilizando secuencias reguladoras convencionales, marcadores de selección, etc.). Puesto que muchos escrutinios farmacéuticamente importantes requieren dianas celulares humanas o de mamíferos modelo, se pueden usar vectores retrovíricos capaces de transfectar tales dianas.

Cuando sea apropiado, las células hospedantes manipuladas mediante ingeniería se pueden cultivar en medios nutrientes convencionales modificados según sea apropiado para activar promotores, seleccionar transformantes o amplificar los genes de la descripción. Después de la transformación de una cepa hospedante adecuada y el crecimiento de la cepa hospedante hasta una densidad celular apropiada, se puede introducir el promotor seleccionado por medio de métodos apropiados (por ejemplo, desplazamiento de temperatura o inducción química) y las células se pueden cultivar durante un período adicional para permitirles producir el polipéptido deseado o fragmento del mismo.

Las células se pueden cosechar mediante centrifugación, se pueden destruir por medios físicos o químicos, y el extracto bruto resultante se retiene para purificación adicional. Las células microbianas empleadas para la expresión de proteínas se pueden destruir mediante cualquier método conveniente, incluyendo ciclos de congelación-descongelación, tratamiento con ultrasonidos, interrupción mecánica, o uso de agentes de lisado celular. Tales métodos son bien conocidos por los expertos en la técnica. El polipéptido expresado o su fragmento se puede recuperar y purificar a partir de cultivos de células recombinantes mediante métodos que incluyen precipitación con etanol o sulfato de amonio, extracción ácida, cromatografía de intercambio aniónico o catiónico, cromatografía de fosfocelulosa, cromatografía de interacción hidrófoba, cromatografía de afinidad, cromatografía con hidroxipatita, y cromatografía con lecitina. Las etapas de replegamiento proteico se pueden usar, según sea necesario, para completar la configuración del polipéptido. Si se desea, para las etapas de purificación final, se puede emplear cromatografía de líquidos de altas prestaciones (HPLC).

Los constructos en células hospedantes se pueden usar de manera convencional para producir el producto génico codificado por la secuencia recombinante. Dependiendo del hospedante empleado en un procedimiento de producción recombinante, los polipéptidos producidos por células hospedantes que contienen el vector pueden estar glicosilados o pueden no estar glicosilados. Los polipéptidos de la descripción también pueden incluir o no un resto de aminoácido metionina inicial.

Para producir un polipéptido de la descripción, también se pueden emplear sistemas de traducción libres de células. Los sistemas de traducción libres de células pueden usar ARNm transcritos a partir de un constructo de ADN que comprende un promotor enlazado operablemente a un ácido nucleico que codifica el polipéptido o fragmento del mismo. En algunos aspectos, el constructo de ADN se puede linealizar antes de llevar a cabo una reacción de transcripción in vitro. El ARNm transcrito se incuba entonces con un extracto de traducción libre de células apropiado, tal como un extracto de reticulocitos de conejo, para producir el polipéptido deseado o fragmento del mismo.

Los vectores de expresión pueden contener uno o más genes marcadores seleccionables para proporcionar un rasgo fenotípico para la selección de células hospedantes transformadas, tales como dihidrofolato reductasa o resistencia a neomicina para cultivo de células eucariotas, o tales como resistencia a tetraciclina o ampicilina en *E. coli*.

Las células hospedantes que contienen los polinucleótidos de interés, por ejemplo ácidos nucleicos de la descripción, se pueden cultivar en medios nutrientes convencionales modificados según sea apropiado para activar promotores, seleccionar transformantes o amplificar genes. Las condiciones de cultivo, tales como temperatura, pH y

similares, son las que se han usado previamente con la célula hospedante seleccionada para la expresión, y resultarán evidentes para los expertos normales en la técnica. Los clones que se identifican que tienen la actividad enzimática especificada se pueden secuenciar a continuación para identificar la secuencia polinucleotídica que codifica una enzima que tiene la actividad intensificada.

- 5 La descripción proporciona un método para sobreexpresar una enzima recombinante en una célula, que comprende expresar un vector que comprende un ácido nucleico de la descripción, por ejemplo un ácido nucleico que comprende una secuencia de ácido nucleico con al menos alrededor de 50%, 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, 60%, 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 10 97%, 98%, 99%, o más de identidad de secuencia con una secuencia ejemplar de la descripción, a lo largo de una región de al menos alrededor de 50, 75, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, 1000, 1050, 1100, 1150 o más restos, o toda la longitud de un gen o transcrito, en el que las identidades de secuencia se determinan mediante análisis con un algoritmo de comparación de secuencias o mediante inspección visual, o un ácido nucleico que se hibrida en condiciones restrictivas con una secuencia de 15 ácido nucleico de la descripción. La sobreexpresión se puede efectuar por cualquier medio, por ejemplo uso de un promotor de alta actividad, un vector dicistrónico o mediante amplificación génica del vector.

- Los ácidos nucleicos de la descripción se pueden expresar, o sobreexpresar, en cualquier sistema de expresión in vitro o in vivo. Se puede emplear cualquier sistema de cultivo celular para expresar, o sobreexpresar, proteína recombinante, incluyendo cultivos bacterianos, de insecto, de levadura, fúngicos o de mamíferos. La sobreexpresión 20 se puede efectuar por medio de la elección apropiada de promotores, potenciadores, vectores (por ejemplo, uso de vectores de replicón, vectores dicistrónicos (véase, por ejemplo, Gurtu (1996) Biochem. Biophys. Res. Commun. 229:295-8), medios, sistemas de cultivo y similares. En un aspecto, la amplificación génica que usa marcadores de selección, por ejemplo glutamina sintetasa (véase, por ejemplo, Sanders (1987) Dev. Biol. Stand. 66:55-63), en sistemas celulares se usa para sobreexpresar los polipéptidos de la descripción.

- 25 La célula hospedante puede ser cualquiera de las células hospedantes familiares para los expertos en la técnica, incluyendo células procariotas, células eucariotas, células de mamífero, células de insecto, o células vegetales. Como ejemplos representativos de los hospedantes apropiados, se pueden mencionar: células bacterianas, tales como *E. coli*, *Streptomyces*, *Bacillus subtilis*, *Salmonella typhimurium* y diversas especies de los géneros *Streptomyces* y *Staphylococcus*, células fúngicas, tales como levadura, células de insecto tales como S2 de 30 *Drosophila* y Sf9 de *Spodoptera*, células animales tales como CHO, COS o melanoma de Bowes, y adenovirus. La selección de un hospedante apropiado se encuentra dentro de las capacidades de los expertos en la técnica.

- El vector se puede introducir en las células hospedantes usando cualquiera de una variedad de técnicas, incluyendo transformación, transfección, transducción, infección vírica, pistolas génicas, o transferencia génica mediada por plásmidos Ti. Los métodos particulares incluyen transfección con fosfato de calcio, transfección mediada por DEAE-Dextrano, lipofección, o electroporación (Davis, L., Dibner, M., Battey, I., Basic Methods in Molecular Biology, 35 (1986)).

- Cuando sea apropiado, las células hospedantes manipuladas mediante ingeniería genética se pueden cultivar en medios nutrientes convencionales modificados según sea apropiado para activar promotores, seleccionar transformantes o amplificar los genes de la descripción. Tras la transformación de una cepa hospedante adecuada y 40 el crecimiento de la cepa hospedante hasta una densidad celular apropiada, el promotor seleccionado se puede inducir por medios apropiados (por ejemplo, desplazamiento de la temperatura o inducción química), y las células se pueden cultivar durante un período adicional para permitirles producir el polipéptido deseado o fragmento del mismo.

- Las células se pueden cosechar mediante centrifugación, se pueden destruir por medios físicos o químicos, y el extracto bruto resultante se retiene para purificación adicional. Las células microbianas empleadas para la expresión 45 de proteínas se pueden destruir mediante cualquier método conveniente, incluyendo ciclos de congelación-descongelación, tratamiento con ultrasonidos, interrupción mecánica, o uso de agentes de lisado celular. Tales métodos son bien conocidos por los expertos en la técnica. El polipéptido expresado o su fragmento se puede recuperar y purificar a partir de cultivos de células recombinantes mediante métodos que incluyen precipitación con etanol o sulfato de amonio, extracción ácida, cromatografía de intercambio aniónico o catiónico, cromatografía de fosfocelulosa, cromatografía de interacción hidrófoba, cromatografía de afinidad, cromatografía con hidroxipatita, y 50 cromatografía con lecitina. Las etapas de plegamiento proteico se pueden usar, según sea necesario, para completar la configuración del polipéptido. Si se desea, para las etapas de purificación final, se puede emplear cromatografía de líquidos de altas prestaciones (HPLC).

- También se pueden emplear diversos sistemas de cultivo de células de mamíferos para expresar proteína recombinante. Los ejemplos de sistemas de expresión de mamíferos incluyen las estirpes COS-7 de fibroblastos de riñón de mono (descritas por Gluzman, Cell, 23:175, 1981) y otras estirpes celulares capaces de expresar proteínas 55 a partir de un vector compatible, tales como las estirpes celulares C127, 3T3, CHO, HeLa y BHK.

Los constructos en las células hospedantes se pueden utilizar de una manera convencional para producir el producto génico codificado por la secuencia recombinante. Dependiendo del hospedante empleado en un

procedimiento de producción recombinante, los polipéptidos producidos por las células hospedantes que contienen el vector pueden estar glucosilados o pueden estar no glucosilados. Los polipéptidos de la descripción también pueden incluir o no un resto de aminoácido de metionina inicial.

- 5 Como alternativa, los polipéptidos de la descripción, o fragmentos que comprenden al menos 5, 10, 15, 20, 25, 30, 35, 40, 50, 75, 100, o 150 o más aminoácidos consecutivos de los mismos, se pueden producir sintéticamente por medio de sintetizadores peptídicos convencionales. En otros aspectos, se pueden emplear fragmentos o porciones de los polipéptidos para producir el polipéptido de longitud completa correspondiente mediante síntesis peptídica; por lo tanto, los fragmentos se pueden emplear como intermedios para la producción de polipéptidos de longitud completa.
- 10 Para producir uno de los polipéptidos de la descripción, o fragmentos que comprenden al menos 5, 10, 15, 20, 25, 30, 35, 40, 50, 75, 100, o 150 aminoácidos consecutivos de los mismos, también se pueden emplear sistemas de traducción libres de células, usando ARNm transcritos a partir de un constructo de ADN que comprende un promotor enlazado operablemente a un ácido nucleico que codifica el polipéptido o fragmento del mismo. En algunos aspectos, el constructo de ADN se puede linealizar antes de llevar a cabo una reacción de transcripción *in vitro*. El
- 15 ARNm transcrito se incuba entonces con un extracto de traducción libre de células apropiado, tal como un extracto de reticulocitos de conejo, para producir el polipéptido deseado o fragmento del mismo.

Amplificación de Ácidos Nucleicos

- 20 Al poner en práctica la descripción, los ácidos nucleicos de la descripción y los ácidos nucleicos que codifican las enzimas de la descripción, o los ácidos nucleicos modificados de la descripción, se pueden reproducir mediante amplificación. La amplificación también se puede usar para clonar o modificar los ácidos nucleicos de la descripción. De este modo, la descripción proporciona pares de secuencias de cebadores de amplificación para amplificar ácidos nucleicos de la descripción. Un experto en la técnica puede diseñar pares de secuencias de cebadores de amplificación para cualquier parte de o para toda la longitud de estas secuencias.

- 25 En un aspecto, la descripción proporciona un ácido nucleico amplificado por un par de cebadores de la descripción, por ejemplo un par de cebadores como se expone en alrededor de los primeros (los 5') 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, o 25 restos de un ácido nucleico de la descripción, y alrededor de los primeros (los 5') 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, o 25 restos de la hebra complementaria.

- 30 La descripción proporciona un par de secuencias cebadoras de amplificación para amplificar un ácido nucleico que codifica un polipéptido que tiene una actividad enzimática, en el que el par de cebadores es capaz de amplificar un ácido nucleico que comprende una secuencia de la descripción, o fragmentos o subsecuencias del mismo. Uno o cada miembro del par de secuencias cebadoras de amplificación pueden comprender un oligonucleótido que comprende al menos alrededor de 10 a 50 bases consecutivas de la secuencia, o alrededor de 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, o 25 bases consecutivas de la secuencia. La descripción proporciona pares de cebadores de amplificación, en el que el par de cebadores comprende un primer miembro que tiene una secuencia como la mostrada mediante alrededor de los primeros (los 5') 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, o 25 restos de un ácido nucleico de la descripción, y a segundo miembro que tiene una secuencia como la mostrada mediante alrededor de los primeros (los 5') 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, o 25 restos de la hebra complementaria del primer miembro. La descripción proporciona ácidos nucleicos que codifican enzimas generados mediante amplificación, por ejemplo reacción en cadena de la polimerasa (PCR), usando un par de cebadores de
- 35 amplificación de la descripción. La descripción proporciona métodos para obtener ácidos nucleicos que codifican enzimas mediante amplificación, por ejemplo reacción en cadena de la polimerasa (PCR), usando un par de cebadores de amplificación de la descripción. En un aspecto, el par de cebadores de amplificación amplifica un ácido nucleico de una biblioteca, por ejemplo una genoteca, tal como una biblioteca medioambiental.

- 40 Las reacciones de amplificación también se pueden usar para cuantificar la cantidad de ácido nucleico en una muestra (tal como la cantidad de mensaje en una muestra celular), para marcar el ácido nucleico (por ejemplo, para aplicarlo a una matriz o a una transferencia), para detectar el ácido nucleico, o para cuantificar la cantidad de un ácido nucleico específico en una muestra. En un aspecto de la descripción, se amplifica el mensaje aislado de una célula o una biblioteca de ADNc.

- 45 El experto en la técnica puede seleccionar y diseñar cebadores de amplificación oligonucleotídicos adecuados. Los métodos de amplificación son también bien conocidos en la técnica, e incluyen, por ejemplo, reacción en cadena de la polimerasa, PCR (véanse, por ejemplo, PCR PROTOCOLS, A GUIDE TO METHODS AND APPLICATIONS, ed. Innis, Academic Press, N.Y. (1990) y PCR STRATEGIES (1995), ed. Innis, Academic Press, Inc., N.Y., reacción en cadena de la ligasa (LCR) (véanse, por ejemplo, Wu (1989) Genomics 4:560; Landegren (1988) Science 241:1077; Barringer (1990) Gene 89:117); amplificación basada en la transcripción (véase, por ejemplo, Kwok (1989) Proc. Natl. Acad. Sci. USA 86:1173); y replicación de secuencias autosostenida (véase, por ejemplo, Guatelli (1990) Proc. Natl. Acad. Sci. USA 87:1874); amplificación de la replicasa Q Beta (véase, por ejemplo, Smith (1997) J. Clin. Microbiol. 35:1477-1491), ensayo de amplificación de la replicasa Q Beta automático (véase, por ejemplo, Burg (1996) Mol. Cell. Probes 10:257-271) y otras técnicas mediadas por ARN polimerasa (por ejemplo, NASBA,
- 50
- 55

Cangene, Mississauga, Ontario); véanse también Berger (1987) *Methods Enzymol.* 152:307-316; Sambrook; Ausubel; patentes U.S. n^{os} 4.683.195 y 4.683.202; Sooknanan (1995) *Biotechnology* 13:563-564.

Determinación del grado de identidad de secuencia

La descripción proporciona ácidos nucleicos que comprenden secuencias que tienen al menos alrededor de 50%, 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, 60%, 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% de identidad o más, o identidad de secuencia completa (100%) con un ácido nucleico ejemplar de la descripción (por ejemplo, SEC ID NO:1, SEC ID NO:3, SEC ID NO:5, SEC ID NO:7, SEC ID NO:9, SEC ID NO:11, SEC ID NO:13, SEC ID NO:15, SEC ID NO:17 o SEC ID NO:19) a lo largo de una región de al menos alrededor de 10, 20, 30, 40, 50, 60, 70, 75, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, 1000, 1050, 1100, 1150, 1200, 1250, 1300, 1350, 1400, 1450, 1500, 1550 o más restos. La descripción proporciona polipéptidos que comprenden secuencias que tienen al menos alrededor de 50%, 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, 60%, 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% de identidad o más, o identidad de secuencia completa (100%) con un polipéptido ejemplar de la descripción. El grado de identidad de secuencia (homología) se puede determinar usando cualquier programa informático y parámetros asociados, incluyendo los descritos aquí, tales como BLAST 2.2.2. o FASTA versión 3.0t78, con los parámetros por defecto.

Las secuencias de ácidos nucleicos de la descripción pueden comprender al menos 10, 15, 20, 25, 30, 35, 40, 50, 75, 100, 150, 200, 300, 400, o 500 nucleótidos consecutivos de una secuencia ejemplar de la descripción, y secuencias sustancialmente idénticas a éstas. Las secuencias homólogas y los fragmentos de las secuencias de ácido nucleico de la descripción hacen referencia a una secuencia que tiene una homología de al menos 99%, 98%, 97%, 96%, 95%, 90%, 85%, 80%, 75%, 70%, 65%, 60%, 55%, o 50% con estas secuencias. La homología se puede determinar utilizando cualquiera de los programas informáticos y los parámetros descritos aquí, incluyendo FASTA versión 3.0t78 con los parámetros por defecto. Las secuencias homólogas también incluyen secuencias de ARN en las que las uridinas sustituyen a las timinas en las secuencias de ácido nucleico de la invención. Las secuencias homólogas se pueden obtener usando cualquiera de los procedimientos descritos en la presente memoria, o pueden resultar de la corrección de un error de secuenciación. Se apreciará que las secuencias de ácido nucleico de la descripción se pueden representar en el formato de un solo carácter tradicional (Véase la contraportada de Stryer, Lubert. *Biochemistry*, 3^a Ed., W. H Freeman & Co., Nueva York.) o en cualquier otro formato que registre la identidad de los nucleótidos en una secuencia.

Se contemplan particularmente para uso en este aspecto de la descripción diversos programas de comparación de secuencias identificados en otra parte en esta memoria descriptiva de patente. Las homologías de las secuencias de proteínas y/o ácidos nucleicos se pueden evaluar usando cualquiera de la variedad de algoritmos de comparación de secuencias y programas conocidos en la técnica. Tales algoritmos y programas incluyen, pero no están limitados de ningún modo a, TBLASTN, BLASTP, FASTA, TFASTA y CLUSTALW (Pearson y Lipman, *Proc. Natl. Acad. Sci. USA* 85 (8):2444-2448, 1988; Altschul et al., *J. Mol. Biol.* 215(3):403-410, 1990; Thompson et al., *Nucleic Acids Res.* 22(2):4673-4680, 1994; Higgins et al., *Methods Enzymol.* 266:383-402, 1996; Altschul et al., *J. Mol. Biol.* 215(3):403-410, 1990; Altschul et al., *Nature Genetics* 3:266-272, 1993).

La homología o identidad se miden con frecuencia utilizando programas de análisis de secuencias (por ejemplo, Sequence Analysis Software Package del Genetics Computer Group, University of Wisconsin Biotechnology Center, 1710 University Avenue, Madison, WI 53705). Tales programas emparejan secuencias similares mediante la asignación de grados de homología a diversas supresiones, sustituciones u otras modificaciones. Los términos "homología" e "identidad", en el contexto de dos o más ácidos nucleicos o secuencias polipeptídicas, hacen referencia a dos o más secuencias o subsecuencias que son la misma o tienen un porcentaje especificado de restos de aminoácidos o nucleótidos que son los mismos cuando se comparan y alinean para una correspondencia máxima a lo largo de una ventana de comparación o región designada según se mide usando cualquier número de algoritmos de comparación de secuencias o mediante alineamiento manual e inspección visual.

Para la comparación de secuencias, típicamente una secuencia actúa como una secuencia de referencia, con la que se comparan las secuencias de ensayo. Cuando se usa un algoritmo de comparación de secuencias, las secuencias de ensayo y de referencia se introducen en un ordenador, se diseñan las coordenadas de la subsecuencia, si fuera necesario, y se diseñan los parámetros del programa de algoritmo de secuencias. Se pueden usar parámetros del programa por defecto, o se pueden diseñar parámetros alternativos. El algoritmo de comparación de secuencias calcula a continuación el porcentaje de identidades de secuencia para las secuencias de ensayo con respecto a la secuencia de referencia, basándose en los parámetros del programa.

Una "ventana de comparación", según se usa aquí, incluye la referencia a un segmento de una cualquiera de las varias posiciones contiguas seleccionadas del grupo que consiste en 20 a 600, habitualmente alrededor de 50 a alrededor de 200, más habitualmente alrededor de 100 a alrededor de 150 en las que una secuencia se puede comparar con una secuencia de referencia del mismo número de posiciones contiguas después de haber alineado óptimamente las dos secuencias. Los métodos para el alineamiento de secuencias para comparación son bien

conocidos en la técnica. El alineamiento óptimo de secuencias para comparación se puede llevar a cabo, *por ejemplo*, por medio del algoritmo de homología local de Smith y Waterman, Adv. Appl. Math. 2:482, 1981, mediante el algoritmo de alineamiento de homología de Needleman y Wunsch, J. Mol. Biol. 48:443, 1970, por medio del método de búsqueda de similitud de Pearson y Lipman, Proc. Nat'l. Acad. Sci. USA 85:2444, 1988, por medio de implementaciones computarizadas de estos algoritmos (GAP, BESTFIT, FASTA y TFASTA en el Wisconsin Genetics Software Package, Genetics Computer Group, 575 Science Dr., Madison, WI), o por medio de alineamiento manual e inspección visual. Otros algoritmos para determinar la homología o identidad incluyen, por ejemplo, además del programa BLAST (Basic Local Alignment Search Tool en el National Center for Biological Information, "Herramienta de Búsqueda de Alineamiento Local Básico"), ALIGN, AMAS (Analysis of Multiply Aligned Sequences, "Análisis de Secuencias Alineadas Multiplicadas"), AMPS (Protein Multiple Sequence Alignment, "Alineamiento de Secuencias Múltiples de Proteínas"), ASSET (Aligned Segment Statistical Evaluation Tool, "Herramienta de Evaluación Estadística de Segmentos Alineados"), BANDS, BESTSCOR, BIOSCAN (Biological Sequence Comparative Analysis Node, "Nodo de Análisis Comparativo de Secuencias Biológicas"), BLIMPS (BLOCKS IMPROVED Searcher, "Búsqueda de Similitud Contra una Base de Datos de Bloques"), FASTA, Intervals & Points, BMB, CLUSTAL V, CLUSTAL W, CONSENSUS, LCONSENSUS, WCONSENSUS, algoritmo de Smith-Waterman, DARWIN, algoritmo Las Vegas, FNAT (Forced Nucleotide Alignment Tool, "Herramienta de Alineamiento de Nucleótidos Forzada"), Framealign, Framesearch, DYNAMIC, FILTER, FSAP (Fristensky Sequence Analysis Package, "Paquete de Análisis de Secuencia de Fristensky"), GAP (Global Alignment Program, "Programa de Alineamiento Global"), GENAL, GIBBS, GenQuest, ISSC (Sensitive Sequence Comparison, "Comparación de Secuencias Sensibles"), LALIGN (Local Sequence Alignment, "Alineamiento de Secuencia Local"), LCP (Local Content Program, "Programa de Contenido Local"), MACAW (Multiple Alignment Construction & Analysis Workbench, "Construcción de Alineamiento Múltiple y Análisis Workbench"), MAP (Multiple Alignment Program, "Programa de Alineamiento Múltiple"), MBLKP, MBLKN, PIMA (Pattern-Induced Multi-secuencia Alignment, "Alineamiento Multisequencia Inducido por Patrón"), SAGA (Sequence Alignment by Genetic Algorithm, "Alineamiento de Secuencia mediante Algoritmo Genético") y WHAT-IF. Tales programas de alineamiento también se pueden utilizar para escrutar bases de datos genómicas para identificar secuencias polinucleotídicas que tienen secuencias sustancialmente idénticas. Existen varias bases de datos genómicas, por ejemplo una porción sustancial del genoma humano está disponible como parte del Proyecto de Secuenciación del Genoma Humano (J. Roach, http://weber.u.Washington.edu/~roach/human_genome_progress_2.html) (Gibbs, 1995). Ya se han secuenciado al menos veintiún genomas distintos, incluyendo, por ejemplo, *M. genitalium* (Fraser et al., 1995), *M. jannaschii* (Bult et al., 1996), *H. influenzae* (Fleischmann et al., 1995), *E. coli* (Blattner et al., 1997) y levadura (*S. cerevisiae*) (Mewes et al., 1997) y *D. melanogaster* (Adams et al., 2000). También se han hecho progresos significativos en la secuenciación de los genomas de organismos modelo, tales como ratón, *C. elegans* y *Arabidopsis sp.* Varias bases de datos que contienen información genómica anotada con cierta información funcional son mantenidas por diferentes organizaciones y son accesibles a través de Internet.

Un ejemplo de un algoritmo útil son los algoritmos BLAST y BLAST 2.0, que se describen por Altschul et al., en Nuc. Acids Res. 25:3389-3402, 1997 y Altschul et al., J. Mol. Biol. 215:403-410, 1990, respectivamente. Los programas para llevar a cabo los análisis BLAST están disponibles al público a través del National Center for Biotechnology Information. Este algoritmo implica identificar en primer lugar pares de secuencias de alta puntuación (HSPs) identificando palabras cortas de longitud W en la secuencia problema, que coinciden o satisfacen cierta puntuación T umbral con valor positivo cuando se alinean con una palabra de la misma longitud en una secuencia de la base de datos. T es referido como el umbral de puntuación de la palabra vecina (Altschul et al., más arriba). Estos éxitos con las palabras vecinas iniciales actúan como semillas para iniciar búsquedas para encontrar HSPs más largos que las contienen. Los éxitos con las palabras se prolongan en ambas direcciones a lo largo de cada secuencia tanto como se pueda aumentar la puntuación de alineamiento acumulativa. Las puntuaciones acumulativas se calculan usando, para las secuencias nucleotídicas, los parámetros M (la puntuación de recompensa para un par de restos que se emparejan; siempre >0). Para las secuencias de aminoácidos, se usa una matriz de puntuación para calcular la puntuación acumulativa. La prolongación de los éxitos con las palabras en cada dirección se detiene cuando: la puntuación de alineamiento acumulativa disminuye en la cantidad X de su valor máximo alcanzado; la puntuación acumulativa tiende a cero o menos, debido a la acumulación de uno o más alineamientos de restos de puntuación negativa; o se alcanza el final de cualquier secuencia. Los parámetros del algoritmo BLAST W , T y X determinan la sensibilidad y la velocidad del alineamiento. El programa BLASTN (para secuencias nucleotídicas) usa como parámetros por defecto una longitud de palabra (W) de 11, una expectativa (E) de 10, $M=5$, $N=-4$ y una comparación de ambas hebras. Para las secuencias de aminoácidos, el programa BLASTP usa como parámetros por defecto una longitud de palabra de 3 y expectativas (E) de 10 y la matriz de puntuación BLOSUM62 (véase Henikoff y Henikoff, Proc. Natl. Acad. Sci. USA 89:10915, 1989) alineamientos (B) de 50, expectativa (E) de 10, $M=5$, $N=-4$ y una comparación de ambas hebras.

El algoritmo BLAST también realiza un análisis estadístico de la similitud entre dos secuencias (véase, *por ejemplo*, Karlin y Altschul, Proc. Natl. Acad. Sci. USA 90:5873, 1993). Una medida de similitud proporcionada por el algoritmo BLAST es la probabilidad de la suma más pequeña ($P(N)$), que proporciona una indicación de la probabilidad por la cual ocurriría por casualidad un emparejamiento entre dos secuencias nucleotídicas o de aminoácidos. Por ejemplo, un ácido nucleico se considera similar a una secuencia de referencia si la probabilidad de la suma más pequeña en una comparación del ácido nucleico de ensayo con respecto al ácido nucleico de referencia es menor que alrededor de 0,2, más en un aspecto menor que alrededor de 0,01, y mucho más en un aspecto menor que alrededor de 0,001.

En un aspecto, las homologías de las secuencias de proteínas y de ácidos nucleicos se evalúan usando la Herramienta de Búsqueda de Alineamiento Local Básico ("BLAST"). En particular, se usan cinco programas BLAST específicos para realizar la siguiente tarea:

- 5 (1) BLASTP y BLAST3 comparan una secuencia problema de aminoácidos frente a una base de datos de secuencias proteicas;
- (2) BLASTN compara una secuencia problema de nucleótidos frente a una base de datos de secuencias nucleotídicas;
- (3) BLASTX compara los productos de traducción conceptual de seis marcos de una secuencia nucleotídica problema (ambas hebras) frente a una base de datos de secuencias proteicas;
- 10 (4) TBLASTN compara una secuencia problema de proteínas frente a una base de datos de secuencias nucleotídicas traducidas en los seis marcos de lectura (ambas hebras); y
- (5) TBLASTX compara las traducciones de seis marcos de una secuencia problema de nucleótidos frente a las traducciones de seis marcos de una base de datos de secuencias nucleotídicas.

15 Los programas BLAST identifican secuencias homólogas identificando segmentos similares, que son referidos aquí como "pares de segmentos de alta puntuación", entre una secuencia de aminoácidos o de ácido nucleico y una secuencia de ensayo que, en un aspecto, se obtiene a partir de una base de datos de secuencias de proteínas o de ácidos nucleicos. En un aspecto, los pares de segmentos de alta puntuación se identifican (es decir, se alinean) por medio de una matriz de puntuación, muchas de las cuales son conocidas en la técnica. En un aspecto, la matriz de puntuación usada es la matriz BLOSUM62 (Gonnet et al., Science 256:1443-1445, 1992; Henikoff y Henikoff, Proteins 17:49-61, 1993). En un aspecto menos preferible, también se pueden usar las matrices PAM o PAM250 (véase, por ejemplo, Schwartz y Dayhoff, eds., 1978, Matrices for Detecting Distance Relationships: Atlas of Protein Sequence and Structure, Washington: National Biomedical Research Foundation). Los programas BLAST son accesibles a través de la U.S. National Library of Medicine.

25 Los parámetros usados con los algoritmos anteriores se pueden adaptar dependiendo de la longitud de la secuencia y el grado de homología estudiado. En algunos aspectos, los parámetros pueden ser los parámetros por defecto usados por los algoritmos en ausencia de instrucciones desde el usuario.

30 En un aspecto, la frase "sustancialmente idéntico", en el contexto de dos ácidos nucleicos o polipéptidos, hace referencia a dos o más secuencias que tienen, por ejemplo, al menos alrededor de 50%, 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, 60%, 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% o más de identidad de nucleótidos o restos de aminoácidos (secuencia), cuando se comparan y alinean para una correspondencia máxima, según se mide usando uno de los algoritmos de comparación de secuencias conocidos o mediante inspección visual. En aspectos alternativos, existe identidad sustancial a lo largo de una región de al menos alrededor de 50, 75, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, 1000, 1050, 1100, 1150 o más restos, o toda la longitud de un gen o transcrito. En algunos aspectos, las secuencias son sustancialmente idénticas a lo largo de toda la longitud de las regiones codificantes.

Sistemas informáticos y productos de programas informáticos

40 Para determinar e identificar identidades de secuencia, homologías estructurales, motivos y similares vía simulación computacional, una secuencia de ácido nucleico o polipeptídica de la descripción se puede almacenar, registrar, y manipular en cualquier medio que pueda ser leído y al que se pueda acceder mediante un ordenador.

45 En consecuencia, la descripción proporciona ordenadores, sistemas informáticos, medios legibles por ordenador, productos de programas informáticos y similares que tienen registradas o almacenadas en ellos las secuencias de ácidos nucleicos y polipeptídicas de la descripción. Como se usa aquí, las palabras "registrado" y "almacenado" hacen referencia a un procedimiento para almacenar información en un medio informático. Un experto en la técnica puede adoptar fácilmente cualquier método conocido para registrar información en un medio legible por ordenador para generar construcciones que comprenden una o más de las secuencias de ácidos nucleicos y/o polipeptídicas de la descripción.

50 Los polipéptidos de la descripción incluyen las secuencias polipeptídicas de la descripción, por ejemplo las secuencias ejemplares de la descripción, y secuencias sustancialmente idénticas a ellas, y fragmentos de cualquiera de las secuencias anteriores. Las secuencias polipeptídicas sustancialmente idénticas, u homologas, hacen referencia a una secuencia polipeptídica que tiene una identidad de secuencia de al menos 50%, 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, 60%, 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, o más, o identidad completa (100%) con una secuencia ejemplar de la descripción.

La homología se puede determinar usando cualquiera de los programas informáticos y los parámetros descritos aquí, incluyendo FASTA versión 3.0t78 con los parámetros por defecto o con cualesquiera parámetros modificados. Las secuencias homólogas se pueden obtener usando cualquiera de los procedimientos descritos en la presente memoria, o pueden resultar de la corrección de un error de secuenciación. Los fragmentos polipeptídicos comprenden al menos alrededor de 10, 15, 20, 25, 30, 35, 40, 45, 50, 75, 100, 150, 200, 250, 300, 350, 400, 450, 500 o más aminoácidos consecutivos de los polipéptidos de la descripción. Se apreciará que los códigos del polipéptido como se muestra en las secuencias de aminoácidos de la descripción se pueden representar en el formato de un solo carácter o en el formato de tres letras tradicional (véase la contraportada de Stryer, Lubert. Biochemistry, 3ª Ed., W. H Freeman & Co., Nueva York.) o en cualquier otro formato que refiera la identidad de los polipéptidos en una secuencia.

Una secuencia de ácido nucleico o polipeptídica de la descripción se puede almacenar, registrar y manipular en cualquier medio que pueda ser leído y al que se pueda acceder mediante un ordenador. Como se usa en la presente memoria, las palabras “registrar” y “almacenar” hacen referencia a un procedimiento para almacenar información en un medio informático. Un experto en la técnica puede adoptar fácilmente cualquiera de los métodos conocidos en la actualidad para registrar información en un medio legible por ordenador para generar construcciones que comprenden una o más de las secuencias de ácidos nucleicos de la descripción, una o más de las secuencias polipeptídicas de la descripción. Otro aspecto de la descripción es un medio legible por ordenador que tiene registradas en él al menos 2, 5, 10, 15, o 20 o más secuencias de ácido nucleico de la descripción.

Otro aspecto de la descripción es un medio legible por ordenador que tiene registradas en él una o más de las secuencias de ácidos nucleicos de la descripción. Otro aspecto de la descripción es un medio legible por ordenador que tiene registradas en él una o más de las secuencias polipeptídicas de la descripción. Otro aspecto de la descripción es un medio legible por ordenador que tiene registradas en él al menos 2, 5, 10, 15, o 20 o más de las secuencias como se exponen anteriormente.

Como se usa aquí, los términos “ordenador”, “programa de ordenador” y “procesador” se usan en sus contextos generales más amplios, e incorporan todos los citados dispositivos, como se describe con detalle más abajo. Una “secuencia codificante de” o una “secuencia codifica” un polipéptido o proteína particular es una secuencia de ácido nucleico que es transcrita y traducida en un polipéptido o proteína cuando se coloca bajo el control de secuencias reguladoras apropiadas.

Los medios legibles por ordenador incluyen medios legibles magnéticamente, medios legibles ópticamente, medios legibles electrónicamente y medios magnéticos/ópticos. Por ejemplo, los medios legibles por ordenador pueden ser un disco duro, un disco flexible, una cinta magnética, un CD-ROM, un Disco Versátil Digital (DVD), una Memoria de Acceso Aleatorio (RAM), o una Memoria de Solo Lectura (ROM), así como otros tipos de otros medios conocidos por los expertos en la técnica.

Aspectos de la descripción incluyen sistemas (*por ejemplo*, sistemas basados en Internet), particularmente sistemas informáticos, que almacenan y manipulan la información de la secuencia descrita aquí. Un ejemplo de un sistema informático 100 se ilustra en forma de diagrama de bloques en la Figura 7. Como se usa aquí, “un sistema informático” hace referencia a los componentes del equipo físico, a los componentes de soporte lógico y a los componentes de almacenamiento de datos usados para analizar una secuencia nucleotídica de una secuencia de ácido nucleico de la descripción, o una secuencia polipeptídica de la descripción. El sistema informático 100 incluye típicamente un procesador para procesar, acceder y manipular los datos de la secuencia. El procesador 105 puede ser cualquier tipo bien conocido de unidad central de procesamiento, tal como, por ejemplo, Pentium III de Intel Corporación, o un procesador similar de Sun, Motorola, Compaq, AMD o International Business Machines.

Típicamente, el sistema informático 100 es un sistema de finalidad general que comprende el procesador 105 y uno o más componentes de almacenamiento de datos interno 110 para el almacenamiento de datos, y uno o más dispositivos de recuperación de datos para recuperar los datos almacenados en los componentes de almacenamiento de datos. Un experto en la técnica apreciará fácilmente que es adecuado uno cualquiera de los sistemas informáticos disponibles en la actualidad.

En un aspecto particular, el sistema informático 100 incluye un procesador 105 conectado a un bus que está conectado a una memoria principal 115 (en un aspecto, implementada como RAM) y uno o más dispositivos de almacenamiento de datos internos 110, tal como un disco duro y/u otros medios legibles por ordenador que tienen datos registrados en el mismo. En algunos aspectos, el sistema informático 100 incluye adicionalmente uno o más dispositivos de recuperación de datos 118 para leer los datos almacenados en los dispositivos de almacenamiento de datos internos 110.

El dispositivo de recuperación de datos 118 puede representar, por ejemplo, una unidad para disco flexible, una unidad para disco compacto, una unidad para cinta magnética, o un módem susceptible de conexión a un sistema de almacenamiento de datos remoto (por ejemplo, vía Internet) etc. En algunos aspectos, el dispositivo de almacenamiento de datos interno 110 es un medio legible por ordenador extraíble, tal como un disco flexible, un disco compacto, una cinta magnética, etc., que contiene registrados lógica de control y/o datos en el mismo. El sistema informático 100 puede incluir ventajosamente o puede ser programado por el soporte lógico apropiado para

leer la lógica de control y/o los datos del componente de almacenamiento de datos una vez insertados en el dispositivo de recuperación de datos.

El sistema informático 100 incluye una pantalla 120 que se utiliza para la visualización por un usuario del ordenador. También se debe observar que el sistema informático 100 puede estar conectado a otros sistemas informáticos 125a-c en una red o red de área extendida para proporcionar acceso centralizado al sistema informático 100.

El soporte lógico para acceder a y procesar las secuencias nucleotídicas de una secuencia de ácido nucleico de la descripción, o una secuencia polipeptídica de la descripción, (tales como herramientas de búsqueda, herramientas de comparación y herramientas de formación de modelos, etc.) puede residir en la memoria principal 115 durante la ejecución.

En algunos aspectos, el sistema informático 100 puede comprender adicionalmente un algoritmo de comparación de secuencias para comparar una secuencia de ácido nucleico de la descripción, o una secuencia polipeptídica de la descripción, almacenada en un medio legible por ordenador, con una secuencia o secuencias nucleotídicas o polipeptídicas de referencia almacenadas en un medio legible por ordenador. Un "algoritmo de comparación de secuencias" hace referencia a uno o más programas que son implementados (local o remotamente) en el sistema informático 100 para comparar una secuencia nucleotídica con otras secuencias nucleotídicas y/o compuestos almacenados en los medios de almacenamiento de datos. Por ejemplo, el algoritmo de comparación de secuencias puede comparar las secuencias nucleotídicas de una secuencia de ácido nucleico de la descripción, o una secuencia polipeptídica de la descripción, almacenada en un medio legible por ordenador, con secuencias de referencia almacenadas en un medio legible por ordenador, para identificar homologías o motivos estructurales.

La Figura 5 es un diagrama de flujo que ilustra un proceso 200 para comparar una nueva secuencia nucleotídica o proteica con una base de datos de secuencias a fin de determinar los niveles de homología entre la nueva secuencia y las secuencias en la base de datos. La base de datos de secuencias puede ser una base de datos privada, almacenada en el sistema 100 de ordenador, o una base de datos pública, tal como GENBANK, que está disponible a través de Internet.

El proceso 200 comienza en el estado 201 de partida, y después continúa a un estado 202, en el que la nueva secuencia a comparar se almacena en una memoria en un sistema informático 100. Como se explica anteriormente, la memoria podría ser cualquier tipo de memoria, incluyendo RAM o un dispositivo de almacenamiento interno.

El proceso 200 continúa entonces a un estado 204, en el que se abre una base de datos de secuencias para el análisis y comparación. El proceso 200 continúa entonces a un estado 206, en el que la primera secuencia almacenada en la base de datos se lee en una memoria en el ordenador. Después se lleva a cabo una comparación en un estado 210 para determinar si la primera secuencia es la misma que la segunda secuencia. Es importante observar que esta etapa no está limitada a llevar a cabo una comparación exacta entre la nueva secuencia y la primera secuencia en la base de datos. Aquellos expertos en la técnica conocen métodos bien conocidos para comparar dos secuencias nucleotídicas o proteicas, incluso si no son idénticas. Por ejemplo, se pueden introducir saltos en una secuencia a fin de elevar el nivel de homología entre las dos secuencias ensayadas. Los parámetros que controlan si se introducen saltos u otras características en una secuencia durante la comparación son introducidos normalmente por el usuario del sistema de ordenador.

Una vez que se ha llevado a cabo una comparación de las dos secuencias en el estado 210, se realiza una determinación en un estado 210 de decisión sobre si las dos secuencias son iguales. Por supuesto, el término "iguales" no está limitado a secuencias que son absolutamente idénticas. Las secuencias que están dentro de los parámetros de homología introducidos por el usuario se marcarán como "iguales" en el proceso 200.

Si se realiza una determinación de que las dos secuencias son iguales, el proceso 200 continúa hasta un estado 214, en el que se presenta al usuario el nombre de la secuencia de la base de datos. Este estado notifica al usuario que la secuencia con el nombre presentado cumple las restricciones de homología que se introdujeron. Una vez que se presenta al usuario el nombre de la secuencia almacenada, el proceso 200 continúa a un estado 218 de decisión, en el que se realiza una determinación sobre si existen más secuencias en la base de datos. Si no existen más secuencias en la base de datos, entonces el proceso 200 termina en el estado 220 final. Sin embargo, si existen más secuencias en la base de datos, entonces el proceso 200 continúa a un estado 224, en el que un puntero se mueve a la siguiente secuencia en la base de datos de manera que se puede comparar con la nueva secuencia. De esta manera, la nueva secuencia se alinea y se compara con cada secuencia en la base de datos.

Se debería señalar que si se ha hecho una determinación en el estado 212 de decisión de que las secuencias no fueron homólogas, entonces el proceso 200 continuaría inmediatamente al estado 218 de decisión, a fin de determinar si cualesquiera otras secuencias estaban disponibles en la base de datos para comparación.

En consecuencia, un aspecto de la descripción es un sistema informático que comprende un procesador, un dispositivo de almacenamiento de datos que tiene almacenado en él una secuencia de ácido nucleico de la descripción, o una secuencia polipeptídica de la descripción, un dispositivo de almacenamiento de datos recuperables que tiene almacenado en él secuencias nucleotídicas o secuencias polipeptídicas de referencia para compararlas con una secuencia de ácido nucleico de la descripción o con una secuencia polipeptídica de la

descripción, y un comparador de secuencias para llevar a cabo la comparación. El comparador de secuencias puede indicar un nivel de homología entre las secuencias comparadas, o puede identificar motivos estructurales en el código de ácido nucleico descrito anteriormente de una secuencia de ácido nucleico de la descripción, o una secuencia polipeptídica de la descripción, o puede identificar motivos estructurales en secuencias que se comparan con estos códigos de ácido nucleico y códigos polipeptídicos. En algunos aspectos, el dispositivo de almacenamiento de datos puede tener almacenadas en él las secuencias de al menos 2, 5, 10, 15, 20, 25, 30 ó 40 o más de las secuencias de ácidos nucleicos de la descripción, o las secuencias polipeptídicas de la descripción.

Otro aspecto de la descripción es un método para determinar el nivel de homología entre una secuencia de ácido nucleico de la descripción, o una secuencia polipeptídica de la descripción, y una secuencia nucleotídica de referencia. El método incluye leer el código de ácido nucleico o el código polipeptídico y la secuencia nucleotídica o polipeptídica de referencia a través del uso de un programa informático que determina los niveles de homología, y determinar la homología entre el código de ácido nucleico o el código polipeptídico y la secuencia nucleotídica o polipeptídica de referencia con el programa informático. El programa informático puede ser cualquiera de un número de programas informáticos para determinar niveles de homología, incluyendo los enumerados específicamente aquí (*por ejemplo*, BLAST2N con los parámetros por defecto o con parámetros modificados cualesquiera). El método se puede implementar usando los sistemas informáticos descritos anteriormente. El método también se puede llevar a cabo leyendo al menos 2, 5, 10, 15, 20, 25, 30 o 40 o más de las secuencias de ácidos nucleicos descritas anteriormente de la descripción, o las secuencias polipeptídicas de la descripción, a través del uso del programa informático, y determinando la homología entre los códigos de ácido nucleico o los códigos polipeptídicos y las secuencias nucleotídicas o las secuencias polipeptídicas de referencia.

La Figura 6 es un diagrama de flujo que ilustra un aspecto de un proceso 250 en un ordenador para determinar si dos secuencias son homólogas. El proceso 250 comienza en un estado 252 de partida, y continúa entonces hacia un estado 254, en el que una primera secuencia a comparar se almacena en una memoria. La segunda secuencia a comparar se almacena entonces en una memoria en un estado 256. El proceso 250 continúa entonces hacia un estado 260, en el que se lee el primer carácter en la primera secuencia, y después hacia un estado 262, en el que se lee el primer carácter de la segunda secuencia. Se debería entender que si la secuencia es una secuencia nucleotídica, entonces el carácter sería normalmente A, T, C, G o U. Si la secuencia es una secuencia proteica, entonces es, en un aspecto, un código de aminoácido de una sola letra, de manera que las secuencias primera y segunda se pueden comparar fácilmente.

Entonces se realiza una determinación en un estado 264 de decisión sobre si los dos caracteres son iguales. Si son iguales, entonces el proceso 250 continúa hacia un estado 268, en el que se leen los siguientes caracteres en las secuencias primera y segunda. Entonces se realiza una determinación sobre si los siguientes caracteres son iguales. Si lo son, entonces el proceso 250 continúa este bucle hasta que dos caracteres no son iguales. Si se realiza una determinación de que los siguientes dos caracteres no son iguales, el proceso 250 continúa hacia un estado 274 de decisión para determinar si hay más caracteres en cualquier secuencia a leer.

Si no hay más caracteres a leer, entonces el proceso 250 continúa hacia un estado 276, en el que se presenta al usuario el nivel de homología entre las secuencias primera y segunda. El nivel de homología se determina calculando la proporción de caracteres entre las secuencias que fueron las mismas del número total de secuencias en la primera secuencia. De este modo, si cada carácter en una primera secuencia de 100 nucleótidos se alinea con cada carácter en una segunda secuencia, el nivel de homología sería 100%.

Como alternativa, el programa informático puede ser un programa informático que compara las secuencias nucleotídicas de una secuencia de ácido nucleico como se expone en la descripción, con una o más secuencias nucleotídicas de referencia con el fin de determinar si el código de ácido nucleico de la descripción difiere de una secuencia de ácido nucleico de referencia en una o más posiciones. Opcionalmente, tal programa registra la longitud e identidad de los nucleótidos insertados, suprimidos o sustituidos, con respecto al polinucleótido de referencia o a una secuencia de ácido nucleico de la descripción. En un aspecto, el programa informático puede ser un programa que determina si una secuencia de ácido nucleico de la descripción contiene un polimorfismo de un solo nucleótido (SNP) con respecto a una secuencia nucleotídica de referencia.

En consecuencia, otro aspecto de la descripción es un método para determinar si una secuencia de ácido nucleico de la descripción difiere en uno o más nucleótidos de una secuencia nucleotídica de referencia, que comprende las etapas de leer el código de ácido nucleico y la secuencia nucleotídica de referencia mediante el uso de un programa informático que identifica diferencias entre secuencias de ácidos nucleicos, e identificar diferencias entre el código de ácido nucleico y la secuencia nucleotídica de referencia con el programa informático. En algunos aspectos, el programa informático es un programa que identifica polimorfismos de un solo nucleótido. El método se puede implementar mediante los sistemas informáticos descritos anteriormente y el método ilustrado en la Figura 6. El método también se puede llevar a cabo leyendo al menos 2, 5, 10, 15, 20, 25, 30, o 40 o más de las secuencias de ácidos nucleicos de la descripción y las secuencias nucleotídicas de referencia mediante el uso del programa informático, e identificar diferencias entre el código de ácido nucleico y las secuencias nucleotídicas de referencia con el programa informático.

En otros aspectos, el sistema a base de ordenador puede comprender además un identificador para identificar características dentro de una secuencia de ácido nucleico de la descripción o una secuencia polipeptídica de la descripción.

- 5 Un “identificador” se refiere a uno o más programas que identifica ciertas características dentro de una secuencia de ácido nucleico de la descripción, o de una secuencia polipeptídica de la descripción. En un aspecto, el identificador puede comprender un programa que identifica un marco de lectura abierto en una secuencia de ácido nucleico de la descripción.

10 La Figura 7 es un diagrama de flujo que ilustra un aspecto de un proceso 300 identificador para detectar la presencia de una característica en una secuencia. El proceso 300 comienza en un estado 302 de partida y continúa entonces hacia un estado 304, en el que una primera secuencia que se va a comprobar en busca de características se almacena en una memoria 115 en el sistema informático 100. El proceso 300 continúa entonces hacia un estado 306, en el que se abre una base de datos de características de secuencias. Tal base de datos incluiría una lista de cada atributo de la característica, junto con el nombre de la característica. Por ejemplo, un nombre de característica podría ser “Codón de Iniciación”, y el atributo sería “ATG”. Otro ejemplo sería el nombre de característica “Caja TAATAA”, y el atributo de la característica sería “TAATAA”. Un ejemplo de tal base de datos se produce por el University of Wisconsin Genetics Computer Group. Como alternativa, las características pueden ser motivos polipeptídicos estructurales, tales como hélices alfa, láminas beta, o motivos polipeptídicos funcionales tales como dominios catalíticos enzimáticos (CDs), o sitios activos, motivos de hélice-vuelta-hélice, u otros motivos conocidos por los expertos en la técnica.

20 Una vez que la base de datos de las características se abre en el estado 306, el proceso 300 continúa hacia un estado 308, en el que la primera característica se lee de la base de datos. Entonces se lleva a cabo en un estado 310 una comparación del atributo de la primera característica con la primera secuencia. Entonces se realiza una determinación en un estado 316 de decisión sobre si el atributo de la característica se encontró en la primera secuencia. Si se encontró el atributo, entonces el proceso 300 continúa hacia un estado 318, en el que se presenta al usuario el nombre de la característica encontrada.

25 El proceso 300 continúa entonces hacia un estado 320 de decisión, en el que se realiza una determinación sobre si existen más características en la base de datos. Si no existen más características, entonces el proceso 300 termina en un estado final 324. Sin embargo, si existen más características en la base de datos, entonces el proceso 300 lee la siguiente característica de secuencia en una etapa 326 y vuelve nuevamente al estado 310, en el que el atributo de la siguiente característica se compara frente a la primera secuencia. Se debería señalar que si no se encuentra el atributo de la característica en la primera secuencia en el estado 316 de decisión, el proceso 300 continúa directamente al estado 320 de decisión, a fin de determinar si existen más características en la base de datos.

30 En consecuencia, otro aspecto de la descripción, es un método para identificar una característica dentro de una secuencia de ácido nucleico de la descripción, o una secuencia polipeptídica de la descripción, que comprende leer el código o códigos del ácido nucleico o el código o códigos del polipéptido mediante el uso de un programa informático que identifica características en ellos, e identificar características en el código o códigos del ácido nucleico con el programa informático. En un aspecto, un programa informático comprende un programa informático que identifica marcos de lectura abiertos. El método se puede llevar a cabo leyendo una sola secuencia o al menos 2, 5, 10, 15, 20, 25, 30, o 40 de las secuencias de ácido nucleico de la descripción, o las secuencias polipeptídicas de la descripción, mediante el uso del programa informático, e identificar características dentro de los códigos de los ácidos nucleicos o de los códigos de los polipéptidos con el programa informático.

35 Una secuencia de ácido nucleico de la descripción, o una secuencia polipeptídica de la descripción, se puede almacenar y manipular en una variedad de programas procesadores de datos, en una variedad de formatos. Por ejemplo, una secuencia de ácido nucleico de la descripción, o una secuencia polipeptídica de la descripción, se puede almacenar como texto en un archivo de procesamiento de palabras, tal como MicrosoftWORD™ o WORDPERFECT™, o como un archivo ASCII en una variedad de programas de bases de datos familiares para los expertos en la técnica, tales como DB2™, SYBASE™, u ORACLE™. Además, muchos programas informáticos y bases de datos se pueden usar como algoritmos de comparación de secuencias, identificadores, o fuentes de secuencias nucleotídicas o secuencias polipeptídicas de referencia a comparar con una secuencia de ácido nucleico de la descripción, o una secuencia polipeptídica de la descripción. Se pretende que la siguiente lista no limite la descripción sino que proporcione apoyo a los programas y bases de datos que son útiles con las secuencias de ácido nucleico de la descripción, o las secuencias polipeptídicas de la descripción.

40 Los programas y bases de datos que se pueden usar incluyen, pero no se limitan a: MacPattern (EMBL), DiscoveryBase (Molecular Applications Group), GeneMine (Molecular Applications Group), Look (Molecular Applications Group), MacLook (Molecular Applications Group), BLAST y BLAST2 (NCBI), BLASTN y BLASTX (Altschul et al, J. Mol. Biol. 215: 403, 1990), FASTA (Pearson y Lipman, Proc. Natl. Acad. Sci. USA, 85: 2444, 1988), FASTDB (Brutlag et al. Comp. App. Biosci. 6:237-245, 1990), Catalyst (Molecular Simulations Inc.), Catalyst/SHAPE (Molecular Simulations Inc.), Cerius².DBAccess (Molecular Simulations Inc.), HypoGen (Molecular Simulations Inc.), Insight II, (Molecular Simulations Inc.), Discover (Molecular Simulations Inc.), CHARMM (Molecular Simulations Inc.), Felix (Molecular Simulations Inc.), DelPhi, (Molecular Simulations Inc.), QuanteMM, (Molecular Simulations Inc.),

Homology (Molecular Simulations Inc.), Modeler (Molecular Simulations Inc.), ISIS (Molecular Simulations Inc.), Quanta/Protein Design (Molecular Simulations Inc.), WebLab (Molecular Simulations Inc.), WebLab Diversity Explorer (Molecular Simulations Inc.), Gene Explorer (Molecular Simulations Inc.), SeqFold (Molecular Simulations Inc.), la base de datos del MDL Available Chemicals Directory, la base de datos del MDL Drug Data Report, la base de datos de Comprehensive Medicinal Chemistry, la base de datos de Derwent's World Drug Index, la base de datos de BioByteMasterFile, la base de datos Genbank, y la base de datos Genseq. Muchos otros programas y bases de datos serán manifiestos para un experto en la técnica dada la presente descripción.

Los motivos que se pueden detectar usando los programas anteriores incluyen secuencias que codifican cremalleras de leucina, motivos de hélice-vuelta-hélice, sitios de glucosilación, sitios de ubiquitinación, hélices alfa y láminas beta, secuencias señal que codifican péptidos señal que dirigen la secreción de las proteínas codificadas, secuencias implicadas en la regulación de la transcripción, tales como homeocajas, tramos ácidos, sitios activos enzimáticos (dominios catalíticos (CDs)), sitios de unión a sustrato, y sitios de escisión enzimática.

Hibridación de ácidos nucleicos

La descripción proporciona ácidos nucleicos aislados, sintéticos o recombinantes que se hibridan en condiciones restrictivas a una secuencia ejemplar de la descripción (por ejemplo, SEC ID NO:1, SEC ID NO:3, SEC ID NO:5, SEC ID NO:7, SEC ID NO:9, SEC ID NO:11, SEC ID NO:13, SEC ID NO:15, SEC ID NO:17 o SEC ID NO:19). Las condiciones restrictivas pueden ser condiciones muy restrictivas, condiciones restrictivas medias y/o condiciones restrictivas bajas, incluyendo las condiciones de elevada restricción y de restricción reducida descritas aquí. En un aspecto, es la restricción de las condiciones de lavado la que establece las condiciones que determinan si un ácido nucleico está dentro del alcance de la descripción, como se explica más abajo.

En aspectos alternativos, los ácidos nucleicos de la descripción como se definen por su capacidad para hibridarse en condiciones restrictivas pueden tener entre alrededor de cinco restos y la longitud completa del ácido nucleico de la descripción; por ejemplo, pueden tener al menos 5, 10, 15, 20, 25, 30, 35, 40, 50, 55, 60, 65, 70, 75, 80, 90, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, 1000, o más, restos de longitud. También están incluidos los ácidos nucleicos más cortos que la longitud completa. Estos ácidos nucleicos pueden ser útiles como, por ejemplo, sondas de hibridación, sondas de marcaje, sondas oligonucleotídicas de PCR, ARNi (mono- o bicatenario), antisentido o secuencias que codifican péptidos que se unen a anticuerpos (epítomos), motivos, sitios activos (dominios catalíticos (CDs)) y similares.

En un aspecto, los ácidos nucleicos de la descripción se definen por su capacidad para hibridarse en condiciones de restricción elevada que comprenden alrededor de 50% de formamida a alrededor de 37°C a 42°C. En un aspecto, los ácidos nucleicos se definen por su capacidad para hibridarse en condiciones de restricción reducida que comprenden de alrededor de 35% a 25% de formamida a alrededor de 30°C a 35°C.

Como alternativa, los ácidos nucleicos de la descripción se definen por su capacidad para hibridarse en condiciones de restricción elevada que comprenden a 42°C en 50% de formamida, 5X SSPE, 0,3% de SDS, y un ácido nucleico de bloqueo de secuencia repetitiva, tal como cot-1 o ADN de esperma de salmón (por ejemplo, 200 n/ml de ADN de esperma de salmón cizallado y desnaturalizado). En un aspecto, los ácidos nucleicos de la descripción se definen por su capacidad para hibridarse en condiciones de restricción reducida que comprenden 35% de formamida a una temperatura reducida de 35°C.

En reacciones de hibridación de ácidos nucleicos, las condiciones usadas para lograr un nivel particular de restricción variarán, dependiendo de la naturaleza de los ácidos nucleicos a hibridar. Por ejemplo, se pueden considerar la longitud, el grado de complementariedad, la composición de la secuencia nucleotídica (por ejemplo, contenido de GC frente a AT), y el tipo de ácidos nucleicos (por ejemplo, ARN frente a ADN) de las regiones que se hibridan de los ácidos nucleicos a la hora de seleccionar las condiciones de hibridación. Una consideración adicional es si uno de los ácidos nucleicos está inmovilizado, por ejemplo, sobre un filtro.

La hibridación se puede llevar a cabo en condiciones de baja restricción, restricción moderada o restricción elevada. Como ejemplo de hibridación de ácidos nucleicos, una membrana polimérica que contiene los ácidos nucleicos desnaturalizados inmovilizados se prehibrida primero durante 30 minutos a 45°C en una disolución que consiste en 0,9 M de NaCl, 50 mM de NaH₂PO₄, pH 7,0, 5,0 mM de Na₂EDTA, 0,5% de SDS, 10X de disolución de Denhardt, y 0,5 mg/ml de ácido polirriboadenílico. Entonces se añaden a la disolución aproximadamente 2×10^7 cpm (actividad específica $4-9 \times 10^8$ cpm/ug) de sonda oligonucleotídica marcada en el extremo con ³²P. Después de 12-16 horas de incubación, la membrana se lava durante 30 minutos a temperatura ambiente en 1X SET (150 mM de NaCl, 20 mM de Tris hidrocloreuro, pH 7,8, 1 mM de Na₂EDTA) que contiene 0,5% de SDS, seguido de un lavado durante 30 minutos en 1X SET reciente a T_m -10°C para la sonda oligonucleotídica. La membrana se expone entonces a una película autorradiográfica para la detección de las señales de hibridación.

Todas las hibridaciones anteriores se considerarían en condiciones de restricción elevada.

Tras la hibridación, el filtro se puede lavar para eliminar cualquier sonda detectable no unida específicamente. La restricción usada para lavar los filtros también se puede variar dependiendo de la naturaleza de los ácidos nucleicos que se hibridan, la longitud de los ácidos nucleicos que se hibridan, el grado de complementariedad, la composición

de la secuencia nucleotídica (por ejemplo, contenido de GC frente a AT), y el tipo de ácido nucleico (por ejemplo, ARN frente a ADN). Los ejemplos de lavados en condiciones de restricción progresivamente mayores son los siguientes: 2X SSC, 0,1% de SDS a temperatura ambiente durante 15 minutos (baja restricción); 0,1X SSC, 0,5% de SDS a temperatura ambiente durante 30 minutos hasta 1 hora (restricción moderada); 0,1X SSC, 0,5% de SDS durante 15 a 30 minutos a una temperatura entre la temperatura de hibridación y 68°C (restricción elevada); y 0,15M de NaCl durante 15 minutos a 72°C (restricción muy elevada). Se puede llevar a cabo un lavado de baja restricción final en 0,1X SSC a temperatura ambiente. Los ejemplos anteriores son meramente ilustrativos de un conjunto de condiciones que se puede usar para lavar filtros. El experto en la técnica sabrá que hay numerosas recetas para lavados con diferentes restricciones.

- 10 Los ácidos nucleicos que se han hibridado a la sonda se pueden identificar mediante autorradiografía u otras técnicas convencionales.

El procedimiento anterior se puede modificar para identificar ácidos nucleicos que tienen grados decrecientes de homología con la secuencia de la sonda. Por ejemplo, para obtener ácidos nucleicos de homología decreciente con la sonda detectable, se pueden usar condiciones menos restrictivas. Por ejemplo, la temperatura de hibridación se puede disminuir en incrementos de 5°C desde 68°C hasta 42°C en un tampón de hibridación que tiene una concentración de Na⁺ de aproximadamente 1M. Tras la hibridación, el filtro se puede lavar con 2X SSC, 0,5% de SDS a la temperatura de hibridación. Estas condiciones se consideran condiciones “moderadas” por encima de 50°C, y condiciones “bajas” por debajo de 50°C. Un ejemplo específico de condiciones de hibridación “moderadas” es cuando la hibridación anterior se lleva a cabo a 55°C. Un ejemplo específico de condiciones de hibridación de “baja restricción” es cuando la hibridación anterior se lleva a cabo a 45°C.

Como alternativa, la hibridación se puede llevar a cabo en tampones, tales como 6X SSC, que contienen formamida a una temperatura de 42°C. En este caso, la concentración de formamida en el tampón de hibridación se puede reducir en incrementos de 5% desde 50% hasta 0% para identificar clones que tienen niveles decrecientes de homología con la sonda. Tras la hibridación, el filtro se puede lavar con 6X SSC, 0,5% de SDS a 50°C. Estas condiciones se consideran condiciones “moderadas” por encima de 25% de formamida, y condiciones “bajas” por debajo de 25% de formamida. Un ejemplo específico de condiciones de hibridación “moderadas” es cuando la hibridación anterior se lleva a cabo a 30% de formamida. Un ejemplo específico de condiciones de hibridación de “baja restricción” es cuando la hibridación anterior se lleva a cabo a 10% de formamida.

Sin embargo, la selección de un formato de hibridación no es crítico – es el rigor de las condiciones de lavado lo que establece las condiciones que determinan si un ácido nucleico está dentro del alcance de la descripción. Las condiciones de lavado usadas para identificar ácidos nucleicos dentro del alcance de la descripción incluyen, por ejemplo: una concentración de sal de alrededor de 0,02 molar a pH 7 y una temperatura de al menos alrededor de 50°C o alrededor de 55°C a alrededor de 60°C; o, una concentración de sal de alrededor de NaCl 0,15 M a 72°C durante alrededor de 15 minutos; o, una concentración de sal de alrededor de 0,2X SSC a una temperatura de al menos alrededor de 50°C o alrededor de 55°C a alrededor de 60°C durante alrededor de 15 a alrededor de 20 minutos; o, el complejo de hibridación se lava dos veces con una disolución con una concentración de sal de alrededor de 2X SSC que contiene SDS al 0,1% a temperatura ambiente durante 15 minutos y a continuación se lava dos veces con 0,1X SSC que contiene SDS al 0,1% a 68°C durante 15 minutos; o, condiciones equivalentes. Véase Sambrook, Tijssen y Ausubel para una descripción de un tampón SSC y condiciones equivalentes.

Estos métodos se pueden usar para aislar ácidos nucleicos de la descripción. Por ejemplo, los métodos anteriores se pueden usar para aislar ácidos nucleicos que tienen una secuencia con una homología de al menos alrededor de 97%, al menos 95%, al menos 90%, al menos 85%, al menos 80%, al menos 75%, al menos 70%, al menos 65%, al menos 60%, al menos 55%, o al menos 50% con una secuencia de ácido nucleico seleccionada del grupo que consiste en una de las secuencias de la descripción, o fragmentos que comprenden al menos alrededor de 10, 15, 20, 25, 30, 35, 40, 50, 75, 100, 150, 200, 300, 400, o 500 bases consecutivas de las mismas y las secuencias complementarias a ellas. La homología se puede medir usando el algoritmo de alineamiento. Por ejemplo, los polinucleótidos homólogos pueden tener una secuencia codificante que es una variante alélica de origen natural de una de las secuencias codificantes descritas aquí. Tales variantes alélicas pueden tener una sustitución, supresión o adición de uno o más nucleótidos cuando se comparan con los ácidos nucleicos de la descripción.

Adicionalmente, los procedimientos anteriores se pueden usar para aislar ácidos nucleicos que codifican polipéptidos que tienen una homología de al menos alrededor de 99%, 95%, al menos 90%, al menos 85%, al menos 80%, al menos 75%, al menos 70%, al menos 65%, al menos 60%, al menos 55%, o al menos 50% con un polipéptido de la descripción, o fragmentos que comprenden al menos 5, 10, 15, 20, 25, 30, 35, 40, 50, 75, 100, o 150 aminoácidos consecutivos de los mismos como se determina usando un algoritmo de alineamiento de secuencias (*por ejemplo*, tal como el algoritmo FASTA versión 3.0t78 con los parámetros por defecto).

Sondas oligonucleotídicas y métodos para usarlas

La descripción también proporciona sondas de ácidos nucleicos que se pueden usar, por ejemplo, para identificar ácidos nucleicos que codifican un polipéptido con una actividad enzimática o sus fragmentos, o para identificar genes u otros ácidos nucleicos que codifican polipéptidos que tienen una actividad de enzima clorofilasa o enzimas

implicadas en el catabolismo de la clorofila. En un aspecto, la sonda comprende al menos 10 bases consecutivas de un ácido nucleico de la descripción. Alternativamente, una sonda puede tener al menos alrededor de 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 110, 120, 130, 150 o alrededor de 10 a 50, alrededor de 20 a 60, alrededor de 30 a 70, bases consecutivas de una secuencia como la mostrada en un ácido nucleico de la descripción. Las sondas identifican un ácido nucleico mediante unión y/o hibridación. Las sondas se pueden usar en matrices de la descripción, véase la discusión más abajo, incluyendo, por ejemplo, matrices capilares. Las sondas de la descripción también se pueden usar para aislar otros ácidos nucleicos o polipéptidos.

Los ácidos nucleicos aislados de la descripción, las secuencias complementarias a éstas, o un fragmento que comprende al menos 10, 15, 20, 25, 30, 35, 40, 50, 75, 100, 150, 200, 300, 400, o 500 bases consecutivas de una de las secuencias de las secuencias de la descripción también se pueden usar como sondas para determinar si una muestra biológica, tal como una muestra de suelo, contiene un organismo que tiene una secuencia de ácido nucleico de la descripción o un organismo a partir del cual se obtuvo el ácido nucleico. En tales procedimientos, se obtiene una muestra biológica que posee potencialmente el organismo a partir del cual se aisló el ácido nucleico, y los ácidos nucleicos se obtienen de la muestra. Los ácidos nucleicos se ponen en contacto con la sonda en condiciones que permiten que la sonda se hibride específicamente a cualesquiera secuencias complementarias presentes en allí.

Cuando sea necesario, las condiciones que permiten que la sonda se hibride específicamente a secuencias complementarias se pueden determinar colocando la sonda en contacto con secuencias complementarias procedentes de muestras que se sabe que contienen la secuencia complementaria, así como secuencias de control que no contienen la secuencia complementaria. Las condiciones de hibridación, tales como la concentración salina del tampón de hibridación, la concentración de formamida del tampón de hibridación, o la temperatura de hibridación, se pueden variar para identificar condiciones que permitan que la sonda se hibride específicamente a ácidos nucleicos complementarios.

Si la muestra contiene el organismo a partir del cual se aisló el ácido nucleico, entonces se detecta hibridación específica de la sonda. La hibridación se puede detectar marcando la sonda con un agente detectable tal como un isótopo radioactivo, un colorante fluorescente o una enzima capaz de catalizar la formación de un producto detectable.

Muchos métodos para usar sondas marcadas para detectar la presencia de ácidos nucleicos complementarios en una muestra son familiares para los expertos en la técnica. Estos incluyen transferencias Southern, transferencias Northern, procedimientos de hibridación de colonias y transferencias de puntos. Los protocolos para cada uno de estos procedimientos son proporcionados por Ausubel et al. *Current Protocols in Molecular Biology*, John Wiley 503 Sons, Inc. (1997) y Sambrook et al., *Molecular Cloning: A Laboratory Manual* 2^a Ed., Cold Spring Harbor Laboratory Press (1989).

Como alternativa, se puede usar más de una sonda (al menos una de las cuales es capaz de hibridarse específicamente a cualesquiera secuencias complementarias que estén presentes en la muestra de ácido nucleico) en una reacción de amplificación para determinar si la muestra contiene un organismo que contiene una secuencia de ácido nucleico de la descripción (por ejemplo, un organismo a partir del cual se aisló el ácido nucleico). Típicamente, las sondas comprenden oligonucleótidos. En un aspecto, la reacción de amplificación puede comprender una reacción de PCR. Los protocolos de PCR se describen en Ausubel y Sambrook, más arriba. Como alternativa, la amplificación puede comprender una reacción en cadena de la ligasa, 3SR, o una reacción de desplazamiento de la hebra. (Véanse Barany, F., "The Ligase Chain Reaction in a PCR World", *PCR Methods and Applications* 1:5-16, 1991; E. Fahy et al., "Self-sustained Sequence Replication (3SR): An Isothermal Transcription-based Amplification System Alternative to PCR", *PCR Methods and Applications* 1:25-33, 1991; y Walker G.T. et al., "Strand Displacement Amplification-an Isothermal in vitro ADN Amplification Technique", *Nucleic Acid Research* 20:1691-1696, 1992). En tales procedimientos, los ácidos nucleicos en la muestra se ponen en contacto con las sondas, se lleva a cabo la reacción de amplificación, y se detecta cualquier producto de amplificación resultante. El producto de la amplificación se puede detectar llevando a cabo la electroforesis en gel sobre los productos de reacción y tiñendo el gel con un intercalador tal como bromuro de etidio. Como alternativa, una o más de las sondas se puede marcar con un isótopo radioactivo, y la presencia de un producto de amplificación radioactivo se puede detectar mediante autorradiografía tras la electroforesis en gel.

Las sondas derivadas de secuencias próximas a los extremos de las secuencias de la descripción también se pueden usar en procedimientos de paseo cromosómico, para identificar clones que contienen secuencias genómicas situadas adyacentes a las secuencias de la descripción. Tales métodos permiten el aislamiento de genes que codifican proteínas adicionales a partir del organismo hospedante.

Los ácidos nucleicos aislados de la descripción, las secuencias complementarias a éstas, o un fragmento que comprende al menos 10, 15, 20, 25, 30, 35, 40, 50, 75, 100, 150, 200, 300, 400, o 500 bases consecutivas de una de las secuencias de la descripción, o las secuencias complementarias a éstas, se pueden usar como sondas para identificar y aislar ácidos nucleicos relacionados. En algunos aspectos, los ácidos nucleicos relacionados pueden ser ADNc o ADN genómicos procedentes de organismos distintos de aquel a partir del cual se aisló el ácido nucleico. Por ejemplo, los otros organismos pueden ser organismos relacionados. En tales procedimientos, una muestra de

ácido nucleico se pone en contacto con la sonda en condiciones que permiten que la sonda se hibride específicamente a secuencias relacionadas. La hibridación de la sonda a ácidos nucleicos procedentes del organismo relacionado se detecta entonces usando cualquiera de los métodos descritos anteriormente.

Variando la restricción de las condiciones de hibridación usadas para identificar ácidos nucleicos, tales como ADNc o ADN genómicos, que se hibridan a la sonda detectable, se pueden identificar y aislar ácidos nucleicos que tienen niveles diferentes de homología con la sonda. La restricción se puede variar llevando a cabo la hibridación a temperaturas variables por debajo de las temperaturas de fusión de las sondas. La temperatura de fusión, T_m , es la temperatura (a fuerza iónica y pH definidos) a la que el 50% de la secuencia diana se hibrida a una sonda perfectamente complementaria. Las condiciones muy restrictivas se seleccionan para que sean iguales a o alrededor de 5°C menores que la T_m para una sonda particular. La temperatura de fusión de la sonda se puede calcular usando las siguientes fórmulas:

Para sondas entre 14 y 70 nucleótidos de longitud, la temperatura de fusión (T_m) se calcula usando la fórmula: $T_m = 81,5 + 16,6 (\log [Na^+]) + 0,41 (\text{fracción G} + \text{C}) - (600/N)$, en la que N es la longitud de la sonda.

Si la hibridación se lleva a cabo en una disolución que contiene formamida, la temperatura de fusión se puede calcular usando la ecuación: $T_m = 81,5 + 16,6 (\log [Na^+]) + 0,41 (\text{fracción G} + \text{C}) - (0,63\% \text{ de formamida}) - (600/N)$, en la que N es la longitud de la sonda.

La prehibridación se puede llevar a cabo en 6X SSC, 5X de reactivo de Denhardt, 0,5% de SDS, 100 µg de ADN de esperma de salmón fragmentado desnaturalizado, o 6X SSC, 5X de reactivo de Denhardt, 0,5% de SDS, 100 µg de ADN de esperma de salmón fragmentado desnaturalizado, 50% de formamida. Las fórmulas para SSC y las disoluciones de Denhardt se dan, por ejemplo, en Sambrook et al., más arriba.

En un aspecto, la hibridación se lleva a cabo añadiendo la sonda detectable a las disoluciones de prehibridación enumeradas anteriormente. Cuando la sonda comprende ADN bicatenario, se desnaturaliza antes de la adición a la disolución de hibridación. El filtro se pone en contacto con la disolución de hibridación durante un período de tiempo suficiente para permitir que la sonda se hibride a ADNc o ADN genómicos que contienen secuencias complementarias a ella u homólogas a ella. Para sondas de alrededor de 200 nucleótidos de longitud, la hibridación se puede llevar a cabo a 15-25°C por debajo de la T_m . Para sondas más cortas, tal como sondas oligonucleotídicas, la hibridación se puede realizar a 5-10°C por debajo de la T_m . En un aspecto, para hibridaciones en 6X SSC, las hibridaciones se llevan a cabo a aproximadamente 68°C. En un aspecto, para hibridaciones en disoluciones que contienen 50% de formamida, la hibridación se lleva a cabo a aproximadamente 42°C.

Inhibición de la expresión de enzimas

La descripción proporciona ácidos nucleicos complementarios a (por ejemplo, secuencias antisentido a) los ácidos nucleicos de la descripción, por ejemplo ácidos nucleicos que codifican polipéptidos que tienen una actividad enzimática implicada en el catabolismo de la clorofila o que tienen actividad de esterasa (por ejemplo, clorofilasa). Las secuencias antisentido son capaces de inhibir el transporte, ajuste o transcripción de genes que codifican la enzima. La inhibición se puede efectuar a través de la selección de ADN genómico o ARN mensajero. La transcripción o función del ácido nucleico seleccionado como diana se puede inhibir, por ejemplo, mediante hibridación y/o escisión. Un conjunto particularmente útil de inhibidores proporcionados por la presente invención incluye oligonucleótidos que son capaces de unirse a un gen o mensaje, evitando o inhibiendo en cualquier caso la producción o función de la enzima deseada. La asociación puede ser a través de hibridación específica de secuencias. Otra clase útil de inhibidores incluye oligonucleótidos que provocan la inactivación o escisión de ácidos nucleicos que tienen una actividad enzimática implicados en el catabolismo de la clorofila o que tienen actividad de esterasa (por ejemplo, clorofilasa). El oligonucleótido puede tener actividad enzimática, que provoca tal escisión, tal como ribozimas. El oligonucleótido puede estar químicamente modificado o conjugado a una enzima o composición capaz de escindir el ácido nucleico complementario. Se puede cribar un conjunto de muchos de tales oligonucleótidos diferentes en busca de aquellos con la actividad deseada. De este modo, la descripción proporciona diversas composiciones para la inhibición de la expresión enzimática a nivel de ácido nucleico y/o proteína, por ejemplo, secuencias antisentido, ARNi y ribozimas que comprenden secuencias de ácidos nucleicos de la descripción y anticuerpos de la descripción.

La inhibición de la expresión de ácidos nucleicos que codifican polipéptidos que tienen una actividad enzimática implicados en el catabolismo de clorofila o que tienen una actividad de esterasa (por ejemplo, clorofilasa) puede tener una variedad de aplicaciones industriales. Por ejemplo, las composiciones de la descripción para la inhibición de la expresión enzimática (por ejemplo, secuencias antisentido, ARNi, ribozimas, anticuerpos) se pueden usar como composiciones farmacéuticas, por ejemplo como agentes antipatógenos o en otras terapias, por ejemplo en las que la enzima inhibida tiene un efecto indeseado, perjudicial o tóxico.

Oligonucleótidos antisentido

La descripción proporciona oligonucleótidos antisentido capaces de unirse a un mensaje de enzima o un gen que pueden inhibir un gen diana o mensaje para, por ejemplo, inhibir un polipéptido implicado en el catabolismo de la clorofila o que tiene actividad de esterasa (por ejemplo, clorofilasa) al seleccionar como diana a ARNm. Las

estrategias para diseñar oligonucleótidos antisentido están bien descritas en la bibliografía científica y de patentes, y el experto puede diseñar tales oligonucleótidos usando los nuevos reactivos de la descripción. Por ejemplo, los protocolos de paseo génico/cartografiado de ARN para cribar oligonucleótidos antisentido eficaces son bien conocidos en la técnica, véase, por ejemplo, Ho (2000) *Methods Enzymol.* 314:168-183, que describe un ensayo de cartografiado de ARN, que se basa en técnicas moleculares estándar para proporcionar un método fácil y fiable para una selección potente de las secuencias antisentido. Véase también Smith (2000) *Eur. J. Pharm. Sci.* 11:191-198.

Los ácidos nucleicos de origen natural se usan como oligonucleótidos antisentido. Los oligonucleótidos antisentido pueden tener cualquier longitud; por ejemplo, en aspectos alternativos, los oligonucleótidos antisentido tienen entre alrededor de 5 y 100, alrededor de 10 y 80, alrededor de 15 y 60, alrededor de 18 y 40. La longitud óptima se puede determinar mediante cribado habitual. Los oligonucleótidos antisentido pueden estar presentes en cualquier concentración. La concentración óptima se puede determinar mediante cribado habitual. Se conoce una amplia variedad de análogos nucleotídicos y de ácidos nucleicos sintéticos, de origen no natural, que pueden resolver este problema potencial. Por ejemplo, se pueden usar ácidos nucleicos peptídicos (PNAs) que contienen cadenas principales no iónicas, tales como unidades de N-(2-aminoetil)glicina. También se pueden usar oligonucleótidos antisentido que tienen enlaces de fosforotioato, como se describe en los documentos WO 97/03211; WO 96/39154; Mata (1997) *Toxicol Appl Pharmacol* 144:189-197; *Antisense Therapeutics*, ed. Agrawal (Humana Press, Totowa, N.J., 1996). Los oligonucleótidos antisentido que tienen análogos de cadena principal de ADN sintéticos proporcionados por la descripción también pueden incluir ácidos nucleicos con fosforoditioato, metilfosfonato, fosforamidato, fosfotriéster de alquilo, sulfamato, 3'-tioacetal, metileno(metilimino), 3'-N-carbamato, y morfolino carbamato, como se describe anteriormente.

La metodología de química combinatoria se puede usar para crear grandes números de oligonucleótidos que se pueden cribar rápidamente en busca de oligonucleótidos específicos que tienen afinidades y especificidades de unión apropiadas con respecto a cualquier diana, tal como las secuencias sentido y antisentido de la descripción (véase, por ejemplo, Gold (1995) *J. of Biol. Chem.* 270:13581-13584).

Ribozimas inhibitoras

La descripción proporciona ribozimas capaces de unirse a un mensaje o genes que codifican polipéptidos de la descripción, o que codifican polipéptidos implicados en el catabolismo de clorofila o que tienen una actividad de esterasa (por ejemplo, clorofilasa). Estas ribozimas pueden inhibir la actividad, por ejemplo, seleccionando como dianas a ARNm. Las estrategias para diseñar ribozimas y seleccionar la secuencia antisentido específica de la enzima para la selección como diana están bien descritas en la bibliografía científica y de patentes, y el experto puede diseñar tales ribozimas usando los nuevos reactivos de la descripción. Las ribozimas actúan mediante la unión a un ARN diana a través de la porción de unión a ARN diana de una ribozima que se mantiene en estrecha proximidad a una porción enzimática del ARN que escinde el ARN diana. De este modo, la ribozima reconoce y se une a un ARN diana a través de un emparejamiento de bases complementarias, y una vez unida al sitio correcto, actúa enzimáticamente para escindir e inactivar el ARN diana. La escisión de un ARN diana de tal manera destruirá su capacidad para dirigir la síntesis de una proteína codificada si la escisión se produce en la secuencia codificante. Después de que una ribozima se ha unido y ha escindido su diana de ARN, puede ser liberada de ese ARN para unirse y escindir a nuevas dianas repetidamente.

En algunas circunstancias, la naturaleza enzimática de una ribozima puede ser ventajosa con respecto a otras tecnologías, tales como la tecnología antisentido (en la que una molécula de ácido nucleico se une simplemente a una diana de ácido nucleico para bloquear su transcripción, traducción o asociación con otra molécula), ya que la concentración eficaz de ribozima necesaria para efectuar un tratamiento terapéutico puede ser menor que la de un oligonucleótido antisentido. Esta ventaja potencial refleja la capacidad de la ribozima para actuar enzimáticamente. De este modo, una única molécula de ribozima es capaz de escindir muchas moléculas de ARN diana. Además, una ribozima es típicamente un inhibidor muy específico, dependiendo la especificidad de la inhibición no sólo del mecanismo de unión del emparejamiento de bases, sino también del mecanismo mediante el cual la molécula inhibe la expresión del ARN al que se une. Esto es, la inhibición es provocada por la escisión de la diana de ARN, y de este modo la especificidad se define como la relación de la velocidad de escisión del ARN seleccionado como diana con respecto a la velocidad de escisión de ARN no seleccionado como diana. Este mecanismo de escisión depende de factores adicionales a los implicados en el emparejamiento de bases. De este modo, la especificidad de acción de una ribozima puede ser mayor que la del oligonucleótido antisentido que se une al mismo sitio de ARN.

La ribozima de la descripción, por ejemplo una molécula de ARN de ribozima enzimática, se puede conformar en un motivo de cabeza de martillo, un motivo de horquilla de pelo, como un motivo del virus de la hepatitis delta, un motivo de intrón del grupo I, y/o ARN similar a ARNasaP en asociación con una secuencia guía de ARN. Los ejemplos de motivos de cabeza de martillo se describen, por ejemplo, por Rossi (1992) *Aids Research and Human Retroviruses* 8:183; los motivos de horquilla de pelo se describen por Hampel (1989) *Biochemistry* 28:4929, y Hampel (1990) *Nuc. Acids Res.* 18:299; el motivo de virus de la hepatitis delta se describe por Perrotta (1992) *Biochemistry* 31:16; el motivo de ARNasaP por Guerrier-Takada (1983) *Cell* 35:849; y el intrón del grupo I por la patente U.S. nº 4.987.071 de Cech. Las citas de estos motivos específicos no pretenden ser limitantes. Los expertos en la técnica reconocerán que una ribozima de la descripción, por ejemplo una molécula de ARN enzimática de esta descripción, puede tener un sitio de unión a sustrato específico, complementario a una o más de las regiones de

ARN del gen diana. Una ribozima de la descripción puede tener una secuencia nucleotídica en o rodeando a ese sitio de unión a sustrato que proporciona una actividad de escisión de ARN a la molécula.

ARN de interferencia (ARNi)

En un aspecto, la descripción proporciona una molécula inhibidora de ARN, una molécula denominada "ARNi", que comprende una secuencia de la descripción. La molécula de ARNi comprende una molécula de ARN bicatenario (ARNbc). El ARNi puede inhibir la expresión de un ácido nucleico que codifica un polipéptido implicado en el catabolismo de la clorofila o que tiene una actividad de esterasa (por ejemplo, clorofilasa), por ejemplo, como se describe aquí. En un aspecto, el ARNi tiene alrededor de 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25 o más nucleótidos de dúplex de longitud. Aunque la invención no está limitada por ningún mecanismo particular de acción, el ARNi puede entrar en una célula y provocar la degradación de un ARN monocatenario (ARNmc) de secuencias similares o idénticas, incluyendo ARNm endógenos. Cuando una célula se expone a ARN bicatenario (ARNbc), el ARNm del gen homólogo se degrada selectivamente mediante un proceso denominado interferencia del ARN (ARNi). Un mecanismo básico posible detrás del ARNi es la ruptura de un ARN bicatenario (ARNbc) que coincide con una secuencia génica específica en trozos cortos denominados ARN de interferencia corto, que dispara la degradación de ARNm que coincide con su secuencia. En un aspecto, los ARNi de la descripción se usan en terapéutica de silenciamiento de genes; véase, por ejemplo, Shuey (2002) *Drug Discov. Today* 7:1040-1046. En un aspecto, la descripción proporciona métodos para degradar selectivamente ARN usando los ARNi de la invención. El proceso se puede poner en práctica *in vitro*, *ex vivo* o *in vivo*. En un aspecto, las moléculas de ARNi de la descripción se pueden usar para generar una mutación de pérdida de función en una célula, un órgano o un animal. Los métodos para obtener y usar moléculas de ARNi para degradar selectivamente ARN son bien conocidos en la técnica, véanse, por ejemplo, las patentes U.S. n^{os} 6.506.559; 6.511.824; 6.515.109; 6.489.127.

Modificación de ácidos nucleicos

La descripción proporciona métodos para generar variantes de los ácidos nucleicos de la descripción, por ejemplo aquellos que codifican un polipéptido implicado en el catabolismo de la clorofila o que tiene actividad de esterasa (por ejemplo, clorofilasa), por ejemplo enzimas descritas aquí. Estos métodos se pueden repetir o usar en diversas combinaciones para generar polipéptidos implicados en el catabolismo de la clorofila o que tienen actividad de esterasa (por ejemplo, clorofilasa), que tienen una actividad alterada o diferente o una estabilidad alterada o diferente con respecto a la de una enzima codificada por el ácido nucleico molde. Estos métodos también se pueden repetir o usar en diversas combinaciones, por ejemplo, para generar variaciones en la expresión génica/expresión del mensaje, traducción del mensaje o estabilidad del mensaje. En otro aspecto, la composición genética de una célula se altera, por ejemplo, mediante modificación de un gen homólogo *ex vivo*, seguido de su reinserción en la célula.

Un ácido nucleico de la descripción se puede alterar por cualquier medio. Por ejemplo, métodos al azar o estocásticos, o métodos no estocásticos o de "evolución dirigida"; véase, por ejemplo, la patente U.S. n^o 6.361.974. Los métodos para la mutación al azar de genes son bien conocidos en la técnica; véase, por ejemplo, la patente U.S. n^o 5.830.696. Por ejemplo, se pueden usar mutágenos para mutar de forma aleatoria un gen. Los mutágenos incluyen, por ejemplo, irradiación con luz ultravioleta o con rayos gamma, o un mutágeno químico, por ejemplo mitomicina, ácido nitroso, psoralenos fotoactivados, solos o en combinación, para inducir rupturas del ADN susceptibles de reparación mediante recombinación. Otros mutágenos químicos incluyen, por ejemplo, bisulfito de sodio, ácido nitroso, hidroxilamina, hidrazina o ácido fórmico. Otros mutágenos son análogos de precursores nucleotídicos, por ejemplo nitrosoguanidina, 5-bromouracilo, 2-aminopurina, o acridina. Estos agentes se pueden añadir a una reacción de PCR en lugar del precursor nucleotídico, mutando de ese modo la secuencia. También se pueden usar agentes intercalantes, tales como proflavina, acriflavina, quinacrina y similares.

Se puede usar cualquier técnica en biología molecular, por ejemplo mutagénesis por PCR aleatoria, véase, por ejemplo, Rice (1992) *Proc. Natl. Acad. Sci. USA* 89:5467-5471; o mutagénesis de múltiples casetes combinatoria, por ejemplo véase Crameri (1995) *Biotechniques* 18:194-196. Como alternativa, los ácidos nucleicos, por ejemplo genes, se pueden reensamblar tras la fragmentación al azar, o "estocástica", véanse, por ejemplo, las patentes U.S. n^{os} 6.291.242; 6.287.862; 6.287.861; 5.955.358; 5.830.721; 5.824.514; 5.811.238; 5.605.793. En aspectos alternativos, se introducen modificaciones, adiciones o supresiones mediante PCR propensa a error, barajado, mutagénesis dirigida por oligonucleótidos, PCR de ensamblaje, mutagénesis de PCR sexual, mutagénesis *in vivo*, mutagénesis de inserción de casete, mutagénesis de conjunto recursiva, mutagénesis de conjunto exponencial, mutagénesis específica del sitio, reensamblaje génico, mutagénesis de saturación de sitios génicos (GSSMTM), reensamblaje por ligación sintética (SLR), recombinación, recombinación de secuencia recursiva, mutagénesis de ADN modificado con fosfotioato, mutagénesis de molde que contiene uracilo, mutagénesis de dúplex con saltos, mutagénesis de reparación de desemparejamientos de punto, mutagénesis de cepa hospedante deficiente en la reparación, mutagénesis química, mutagénesis radiogénica, mutagénesis de supresión, mutagénesis de restricción-selección, mutagénesis de restricción-purificación, síntesis de genes artificiales, mutagénesis de conjunto, creación de multímeros de ácidos nucleicos quiméricos, y/o una combinación de estos y otros métodos.

Las siguientes publicaciones describen una variedad de procedimientos y/o métodos de recombinación recursivos que se pueden incorporar en los métodos de la descripción: Stemmer (1999) "Molecular breeding of viruses for

targeting and other clinical properties" *Tumor Targeting* 4:1-4; Ness (1999) *Nature Biotechnology* 17:893-896; Chang (1999) "Evolution of a cytokine using DNA family shuffling" *Nature Biotechnology* 17:793-797; Minshull (1999) "Protein evolution by molecular breeding" *Current Opinion in Chemical Biology* 3:284-290; Christians (1999) "Directed evolution of thymidine kinase for AZT phosphorylation using DNA family shuffling" *Nature Biotechnology* 17:259-264; Cramer (1998) "DNA shuffling of a family of genes from diverse species accelerates directed evolution" *Nature* 391:288-291; Cramer (1997) "Molecular evolution of an arsenate detoxification pathway by DNA shuffling", *Nature Biotechnology* 15:436-438; Zhang (1997) "Directed evolution of an effective fucosidase from a galactosidase by DNA shuffling and screening" *Proc. Natl. Acad. Sci. USA* 94:4504-4509; Patten et al. (1997) "Applications of DNA Shuffling to Pharmaceuticals and Vaccines" *Current Opinion in Biotechnology* 8:724-733; Cramer et al. (1996) "Construction and evolution of antibody-phage libraries by DNA shuffling" *Nature Medicine* 2:100-103; Gates et al. (1996) "Affinity selective isolation of ligands from peptide libraries through display on a lac repressor 'headpiece dimer'" *Journal of Molecular Biology* 255:373-386; Stemmer (1996) "Sexual PCR and Assembly PCR" En: *The Encyclopedia of Molecular Biology*. VCH Publishers, Nueva York. p. 447-457; Cramer y Stemmer (1995) "Combinatorial multiple cassette mutagenesis creates all the permutations of mutant and wildtype cassettes" *BioTechniques* 18:194-195; Stemmer et al. (1995) "Single-step assembly of a gene and entire plasmid form large numbers of oligodeoxyribonucleotides" *Gene*, 164:49-53; Stemmer (1995) "The Evolution of Molecular Computation" *Science* 270: 1510; Stemmer (1995) "Searching Sequence Space" *Bio/Technology* 13:549-553; Stemmer (1994) "Rapid evolution of a protein in vitro by DNA shuffling" *Nature* 370:389-391; y Stemmer (1994) "DNA shuffling by random fragmentation and reassembly: In vitro recombination for molecular evolution" *Proc. Natl. Acad. Sci. USA* 91:10747-10751.

Los métodos mutacionales para generar diversidad incluyen, por ejemplo, mutagénesis dirigida al sitio (Ling et al. (1997) "Approaches to DNA mutagenesis: an overview" *Anal Biochem.* 254(2): 157-178; Dale et al. (1996) "Oligonucleotide-directed random mutagenesis using the phosphorothioate method" *Methods Mol. Biol.* 57:369-374; Smith (1985) "In vitro mutagenesis" *Ann. Rev. Genet.* 19:423-462; Botstein y Shortle (1985) "Strategies and applications of in vitro mutagenesis" *Science* 229:1193-1201; Carter (1986) "Site-directed mutagenesis" *Biochem. J.* 237:1-7; y Kunkel (1987) "The efficiency of oligonucleotide directed mutagenesis" in *Nucleic Acids & Molecular Biology* (Eckstein, F. y Lilley, D. M. J. eds., Springer Verlag, Berlín); mutagénesis que usa moldes que contienen uracilo (Kunkel (1985) "Rapid and efficient site-specific mutagenesis without phenotypic selection" *Proc. Natl. Acad. Sci. USA* 82:488-492; Kunkel et al. (1987) "Rapid and efficient site-specific mutagenesis without phenotypic selection" *Methods in Enzymol.* 154, 367-382; y Bass et al. (1988) "Mutant Trp repressors with new DNA-binding specificities" *Science* 242:240-245); mutagénesis dirigida por oligonucleótidos (Methods in Enzymol. 100: 468-500 (1983); Methods in Enzymol. 154: 329-350 (1987); Zoller (1982) "Oligonucleotide-directed mutagenesis using M13-derived vectors: an efficient and general procedure for the production of point mutations in any DNA fragment" *Nucleic Acids Res.* 10:6487-6500; Zoller y Smith (1983) "Oligonucleotide-directed mutagenesis of DNA fragments cloned into M13 vectors" *Methods in Enzymol.* 100:468-500; y Zoller (1987) "Oligonucleotide-directed mutagenesis: a simple method using two oligonucleotide primers and a single-stranded DNA template" *Methods in Enzymol.* 154:329-350); mutagénesis de ADN modificado con fosforotioato (Taylor (1985) "The use of phosphorothioate-modified DNA in restriction enzyme reactions to prepare nicked DNA" *Nucl. Acids Res.* 13: 8749-8764; Taylor (1985) "The rapid generation of oligonucleotide-directed mutations at high frequency using phosphorothioate-modified DNA" *Nucl. Acids Res.* 13: 8765-8787 (1985); Nakamaye (1986) "Inhibition of restriction endonuclease Nci I cleavage by phosphorothioate groups and its application to oligonucleotide-directed mutagenesis" *Nucl. Acids Res.* 14: 9679-9698; Sayers (1988) "Y-T Exonucleases in phosphorothioate-based oligonucleotide-directed mutagenesis" *Nucl. Acids Res.* 16:791-802; y Sayers et al. (1988) "Strand specific cleavage of phosphorothioate-containing DNA by reaction with restriction endonucleases in the presence of ethidium bromide" *Nucl. Acids Res.* 16: 803-814); mutagénesis que usa ADN dúplex consaltos (Kramer et al. (1984) "The gapped duplex DNA approach to oligonucleotide-directed mutation construction" *Nucl. Acids Res.* 12: 9441-9456; Kramer y Fritz (1987) *Methods in Enzymol.* "Oligonucleotide-directed construction of mutations via gapped duplex DNA" 154:350-367; Kramer (1988) "Improved enzymatic in vitro reactions in the gapped duplex DNA approach to oligonucleotide-directed construction of mutations" *Nucl. Acids Res.* 16: 7207; y Fritz (1988) "Oligonucleotide-directed construction of mutations: a gapped duplex DNA procedure without enzymatic reactions in vitro" *Nucl. Acids Res.* 16: 6987-6999).

Los protocolos adicionales que se pueden usar para la práctica de la descripción incluyen reparación de desemparejamientos de punto (Kramer (1984) "Point Mismatch Repair" *Cell* 38:879-887), mutagénesis que usa cepas hospedantes deficientes en la reparación (Carter et al. (1985) "Improved oligonucleotide site-directed mutagenesis using M13 vectors" *Nucl. Acids Res.* 13: 4431-4443; y Carter (1987) "Improved oligonucleotide-directed mutagenesis using M13 vectors" *Methods in Enzymol.* 154: 382-403), mutagénesis de supresión (Eghtedarzadeh (1986) "Use of oligonucleotides to generate large deletions" *Nucl. Acids Res.* 14: 5115), mutagénesis de restricción-selección y restricción-selección y restricción-purificación (Wells et al. (1986) "Importance of hydrogen-bond formation in stabilizing the transition state of subtilisin" *Phil. Trans. R Soc. Lond. A* 317: 415-423), mutagénesis mediante síntesis génica total (Nambiar et al. (1984) "Total synthesis and cloning of a gene coding for the ribonuclease S protein" *Science* 223: 1299-1301; Sakamar (1988) "Total synthesis and expression of a gene for the a-subunit of bovine rod outer segment guanine nucleotide-binding protein (transducin)" *Nucl. Acids Res.* 14: 6361-6372; Wells et al. (1985) "Cassette mutagenesis: an efficient method for generation of multiple mutations at defined sites" *Gene* 34:315-323; y Grundstrom et al. (1985) "Oligonucleotide-directed mutagenesis by microscale 'shot-gun' gene synthesis" *Nucl. Acids Res.* 13: 3305-3316), reparación de ruptura bicatenaria (Mandecki (1986); Arnold (1993)

“Protein engineering for unusual environments” Current Opinion in Biotechnology 4:450-455. “Oligonucleotide-directed double-strand break repair in plasmids of Escherichia coli: a method for site-specific mutagenesis” Proc. Natl. Acad. Sci. USA, 83:7177-7181). En Methods in Enzymology Volumen 154, que también describe controles útiles para resolver problemas con diversos métodos de mutagénesis, se pueden encontrar detalles adicionales sobre muchos de los métodos anteriores.

Los protocolos que se pueden usar para la práctica de la descripción se describen, por ejemplo, en las patentes U.S. n° 5.605.793 de Stemmer (25 de febrero de 1997), “Methods for In Vitro Recombination”; Patente U.S. n° 5.811.238 de Stemmer et al. (22 de septiembre de 1998) “Methods for Generating Polynucleotides having Desired Characteristics by Iterative Selection and Recombination”; Patente U.S. n° 5.830.721 de Stemmer et al. (3 de noviembre de 1998), “DNA Mutagenesis by Random Fragmentation and Reassembly”; Patente U.S. n° 5.834.252 de Stemmer, et al. (10 de noviembre de 1998) “End-Complementary Polymerase Reaction”; Patente U.S. n° 5.837.458 de Minshull, et al. (17 de noviembre de 1998), “Methods and Compositions for Cellular and Metabolic Engineering”; documento WO 95/22625, Stemmer y Cramer, “Mutagenesis by Random Fragmentation and Reassembly”; documento WO 96/33207 por Stemmer y Lipschutz “End Complementary Polymerase Chain Reaction”; documento WO 97/20078 por Stemmer y Cramer “Methods for Generating Polynucleotides having Desired Characteristics by Iterative Selection and Recombination”; documento WO 97/35966 por Minshull y Stemmer, “Methods and Compositions for Cellular and Metabolic Engineering”; documento WO 99/41402 por Punnonen et al. “Targeting of Genetic Vaccine Vectors”; documento WO 99/41383 por Punnonen et al. “Antigen Library Immunization”; documento WO 99/41369 por Punnonen et al. “Genetic Vaccine Vector Engineering”; documento WO 99/41368 por Punnonen et al. “Optimization of Immunomodulatory Properties of Genetic Vaccines”; documento EP 752008 por Stemmer y Cramer, “DNA Mutagenesis by Random Fragmentation and Reassembly”; documento EP 0932670 por Stemmer “Evolving Cellular DNA Uptake by Recursive Sequence Recombination”; documento WO 99/23107 por Stemmer et al., “Modification of Virus Tropism and Host Range by Viral Genome Shuffling”; documento WO 99/21979 por Apt et al., “Human Papillomavirus Vectors”; documento WO 98/31837 por del Cardayre et al. “Evolution of Whole Cells and Organisms by Recursive Sequence Recombination”; documento WO 98/27230 por Patten y Stemmer, “Methods and Compositions for Polypeptide Engineering”; documento WO 98/27230 por Stemmer et al., “Methods for Optimization of Gene Therapy by Recursive Sequence Shuffling and Selection”, documento WO 00/00632, “Methods for Generating Highly Diverse Libraries”, documento WO 00/09679, “Methods for Obtaining in Vitro Recombined Polynucleotide Sequence Banks and Resulting Sequences”, documento WO 98/42832 por Arnold et al., “Recombination of Polynucleotide Sequences Using Random or Defined Primers”, documento WO 99/29902 por Arnold et al., “Method for Creating Polynucleotide and Polypeptide Sequences”, documento WO 98/41653 por Vind, “An in Vitro Method for Construction of a DNA Library”, documento WO 98/41622 por Borchert et al., “Method for Constructing a Library Using DNA Shuffling”, y documento WO 98/42727 por Pati y Zarlign, “Sequence Alterations using Homologous Recombination”.

Los protocolos que se pueden usar para la práctica de la descripción (que proporcionan detalles con respecto a diversos métodos que generan diversidad) se describen, por ejemplo, en la solicitud de patente U.S. serie n° (USSN) 09/407.800, “SHUFFLING OF CODON ALTERED GENES” por Patten et al. presentada el 28 de septiembre de 1999; “EVOLUTION OF WHOLE CELLS AND ORGANISMS BY RECURSIVE SEQUENCE RECOMBINATION” por del Cardayre et al., patente de Estados Unidos de América número 6.379.964; “OLIGONUCLEOTIDE MEDIATED NUCLEIC ACID RECOMBINATION” por Cramer et al., patentes de Estados Unidos de América números 6.319.714; 6.368.861; 6.376.246; 6.423.542; 6.426.224 y PCT/US00/01203; “USE OF CODON-VARIED OLIGONUCLEOTIDE SYNTHESIS FOR SYNTHETIC SHUFFLING” por Welch et al., patente de Estados Unidos de América número 6.436.675; “METHODS FOR MAKING CHARACTER STRINGS, POLYNUCLEOTIDES & POLYPEPTIDES HAVING DESIRED CHARACTERISTICS” por Selifonov et al., presentada el 18 de enero, 2000, (PCT/US00/01202) y, por ejemplo “METHODS FOR MAKING CHARACTER STRINGS, POLYNUCLEOTIDES & POLYPEPTIDES HAVING DESIRED CHARACTERISTICS” por Selifonov et al., presentada el 18 de julio, 2000 (U.S. Ser. n° 09/618.579); “METHODS OF POPULATING DATA STRUCTURES FOR USE IN EVOLUTIONARY SIMULATIONS” por Selifonov y Stemmer, presentada el 18 de enero, 2000 (PCT/US00/01138); y “SINGLE-STRANDED NUCLEIC ACID TEMPLATE-MEDIATED RECOMBINATION AND NUCLEIC ACID FRAGMENT ISOLATION” por Affholter, presentada el 6 de septiembre, 2000 (U.S. Ser. n° 09/656.549); y patentes de Estados Unidos de América números 6.177.263; 6.153.410.

Los métodos no estocásticos, o “de evolución dirigida”, incluyen, por ejemplo, mutagénesis de saturación del sitio génico (GSSM), reensamblaje por ligación sintética (SLR), o una combinación de los mismos, que se usan para modificar los ácidos nucleicos de la descripción para generar polipéptidos implicados en el catabolismo de la clorofila o que tienen una actividad de esterasa (por ejemplo, clorofilasa) con propiedades nuevas o alteradas (por ejemplo, actividad en condiciones muy ácidas o alcalinas, temperaturas elevadas o bajas, y similares). Los polipéptidos codificados por los ácidos nucleicos modificados se pueden cribar en busca de una actividad antes del ensayo para determinar una hidrólisis de glucano u otro polisacárido, u otra actividad. Se puede usar cualquier modalidad o protocolo de ensayo, por ejemplo usando una plataforma de matriz capilar. Véanse, por ejemplo, las patentes U.S. n°s 6.361.974; 6.280.926; 5.939.250.

Mutagénesis de saturación de sitios génicos (GSSM)

En un aspecto, los cebadores de los codones que contienen una secuencia degenerada N,N,G/T se usan para introducir mutaciones de punto en un polinucleótido, por ejemplo un ácido nucleico de la descripción, para generar un conjunto de polipéptidos progenie en los que un intervalo completo de sustituciones de un solo aminoácido se representa en cada posición de aminoácido, por ejemplo un resto de aminoácido en un sitio activo enzimático (dominios catalíticos (CDs)) o un sitio de unión a ligando seleccionado para ser modificado. Estos oligonucleótidos pueden comprender una primera secuencia homóloga contigua, una secuencia degenerada N,N,G/T, y, opcionalmente, una segunda secuencia homóloga. Los productos traduccionales de progenie aguas abajo procedentes del uso de tales oligonucleótidos incluyen todos los posibles cambios de aminoácidos en cada sitio de aminoácido a lo largo del polipéptido, debido a que la degeneración de la secuencia N,N,G/T incluye codones para los 20 aminoácidos. En un aspecto, uno de tales oligonucleótidos degenerados (compuesto de, por ejemplo, un casete degenerado N,N,G/T) se usa para someter a cada codón original en un molde polinucleotídico progenitor a un intervalo completo de sustituciones de codones. En otro aspecto, se usan al menos dos casetes degenerados – ya sea en el mismo oligonucleótido o no, para someter al menos dos codones originales en un molde polinucleotídico progenitor a un intervalo completo de sustituciones de codones. Por ejemplo, más de una secuencia N,N,G/T puede estar contenida en un oligonucleótido para introducir mutaciones de aminoácidos en más de un sitio. Esta pluralidad de secuencias N,N,G/T puede estar directamente contigua, o separada por una o más secuencias nucleotídicas adicionales. En otro aspecto, los oligonucleótidos aprovechables para introducir adiciones y supresiones se pueden usar solos o en combinación con los codones que contienen una secuencia N,N,G/T, para introducir cualquier combinación o permutación de adiciones, supresiones y/o sustituciones de aminoácidos.

En un aspecto, la mutagénesis simultánea de dos o más posiciones de aminoácidos contiguas se realiza usando un oligonucleótido que contiene tripletes N,N,G/T contiguos, es decir, una secuencia degenerada (N,N,G/T)_n. En otro aspecto, se usan casetes degenerados que tienen una degeneración menor que la secuencia N,N,G/T. Por ejemplo, puede ser deseable en algunos casos usar (por ejemplo, en un oligonucleótido) una secuencia triplete degenerada compuesta solamente de un N, en la que dicho N puede estar en la primera, segunda o tercera posición del triplete. Cualquier otra base, incluyendo cualesquiera combinaciones y permutaciones de las mismas, se puede usar en las dos posiciones restantes del triplete. Como alternativa, puede ser deseable en algunos casos usar (por ejemplo, en un oligo) una secuencia triplete N,N,N degenerada.

En un aspecto, el uso de tripletes degenerados (por ejemplo, tripletes N,N,G/T) permite la generación sistemática y fácil de un intervalo completo de aminoácidos naturales posibles (para un total de 20 aminoácidos) en todas y cada una de las posiciones de aminoácidos en un polipéptido (en aspectos alternativos, los métodos también incluyen la generación de menos de todas las sustituciones posibles por resto de aminoácido, o codón, posición). Por ejemplo, para un polipéptido de 100 aminoácidos, se pueden generar 2000 especies diferentes (es decir, 20 aminoácidos posibles por posición X 100 posiciones de aminoácidos). Mediante el uso de un oligonucleótido o conjunto de oligonucleótidos que contienen un triplete N,N,G/T degenerado, 32 secuencias individuales pueden codificar los 20 aminoácidos naturales posibles. De este modo, en una vasija de reacción en la que una secuencia polinucleotídica progenitora se somete a mutagénesis de saturación usando al menos uno de tales oligonucleótidos, se generan 32 polinucleótidos progenie distintos que codifican 20 polipéptidos distintos. Por el contrario, el uso de un oligonucleótido no degenerado en mutagénesis dirigida al sitio conduce a sólo un producto polipeptídico progenie por vasija de reacción. Los oligonucleótidos no degenerados se pueden usar opcionalmente en combinación con cebadores degenerados descritos; por ejemplo, los oligonucleótidos no degenerados se pueden usar para generar mutaciones de punto específicas en un polinucleótido de trabajo. Esto proporciona un medio para generar mutaciones de punto silenciosas específicas, mutaciones de punto que conducen a cambios de aminoácidos correspondientes, y mutaciones de punto que provocan la generación de codones de parada y la expresión correspondiente de fragmentos polipeptídicos.

En un aspecto, cada vasija de reacción de mutagénesis de saturación contiene polinucleótidos que codifican al menos 20 moléculas polipeptídicas progenie (por ejemplo, polipéptidos de la descripción implicados en el catabolismo de la clorofila o que tienen actividad de esterasa (por ejemplo, clorofilasa)) de manera que los 20 aminoácidos naturales están representados en la una posición de aminoácido específica correspondiente a la posición de codón mutagenizada en el polinucleótido progenitor (otros aspectos usan menos de 20 combinaciones naturales). Los polipéptidos progenie degenerados 32 veces, generados a partir de cada vasija de reacción de mutagénesis de saturación, se pueden someter a amplificación clonal (por ejemplo, clonados en un hospedante adecuado, por ejemplo hospedante de *E. coli*, usando, por ejemplo, un vector de expresión) y se pueden someter a cribado de expresión. Cuando un polipéptido progenie individual se identifica mediante cribado para presentar un cambio favorable en la propiedad (cuando se compara con el polipéptido progenitor, tal como actividad de hidrólisis de glucano incrementada en condiciones alcalinas o ácidas), se puede secuenciar para identificar la sustitución de aminoácido correspondientemente favorable contenida en él.

En un aspecto, al mutagenizar todas y cada una de las posiciones de aminoácidos en un polipéptido progenitor usando mutagénesis de saturación como se describe aquí, se pueden identificar cambios de aminoácidos favorables en más de una posición de aminoácido. Se pueden generar una o más nuevas moléculas progenie que contienen una combinación de todas o partes de estas sustituciones de aminoácidos favorables. Por ejemplo, si se identifican 2 cambios de aminoácidos favorables específicos en cada una de 3 posiciones de aminoácidos en un polipéptido, las permutaciones incluyen 3 posibilidades en cada posición (ningún cambio del aminoácido original, y cada uno de dos cambios favorables) y 3 posiciones. De este modo, hay 3 x 3 x 3 ó 27 posibilidades totales, incluyendo 7 que se

examinaron previamente – 6 mutaciones de un solo punto (es decir, 2 en cada una de tres posiciones) y ningún cambio en ninguna posición.

En todavía otro aspecto, se puede usar mutagénesis de saturación de sitio junto con barajado, quimerización, recombinación y otros procedimientos mutagenizantes, junto con cribado. Esta descripción proporciona el uso de procedimientos de mutagenización cualesquiera, incluyendo mutagénesis de saturación, de una manera iterativa. En una ejemplificación, se usa el uso iterativo de procedimientos de mutagenización cualesquiera combinado con cribado.

La descripción también proporciona el uso de cebadores codónicos patentados (que contienen una secuencia N,N,N degenerada) para introducir mutaciones puntuales en un polinucleótido, para generar un conjunto de polipéptidos progenie en el que está representada en cada posición de aminoácido una gama completa de sustituciones individuales de aminoácidos (mutagénesis de saturación de sitios génicos™ (GSSM™)). Los oligos usados están compuestos contiguamente de una primera secuencia homóloga, una secuencia N,N,N degenerada y, en un aspecto pero no necesariamente, una segunda secuencia homóloga. Los productos traduccionales de la progenie aguas abajo a partir del uso de tales oligos incluyen todos los cambios de aminoácidos posibles en cada sitio de aminoácido a lo largo del polipéptido, puesto que la degeneración de la secuencia N,N,N incluye codones para los 20 aminoácidos.

En un aspecto, uno de tales oligos degenerados (compuestos de un casete N,N,N degenerado) se usa para someter cada codón original en un molde polinucleotídico progenitor a una gama completa de sustituciones codónicas. En otro aspecto, se usan al menos dos casetes N,N,N degenerados – en el mismo oligo o no, para someter al menos dos codones originales en un molde polinucleotídico progenitor a una gama completa de sustituciones codónicas. De este modo, más de una secuencia N,N,N puede estar contenida en un oligo para introducir mutaciones de aminoácidos en más de un sitio. Esta pluralidad de secuencias N,N,N puede estar directamente contigua, o separada por una o más secuencias nucleotídicas adicionales. En otro aspecto, los oligos utilizables para introducir adiciones y supresiones se pueden usar solos o combinados con los codones que contienen una secuencia N,N,N, para introducir cualquier combinación o permutación de adiciones, supresiones y/o sustituciones de aminoácidos.

En una ejemplificación particular, es posible mutagenizar simultáneamente dos o más posiciones de aminoácidos contiguas usando un oligo que contienen tripletes N,N,N contiguos, es decir, una secuencia (N,N,N)_n degenerada.

En otro aspecto, la presente descripción proporciona el uso de casetes degenerados que tienen menos degeneración que la secuencia N,N,N. Por ejemplo, puede ser deseable en algunos casos usar (por ejemplo, en un oligo) una secuencia triplete degenerada que está compuesta de solo un N, en la que el N puede estar en la primera, segunda o tercera posición del triplete. Se pueden usar otras bases cualesquiera, incluyendo combinaciones y permutaciones cualesquiera de las mismas, en las dos posiciones restantes del triplete. Alternativamente, puede ser deseable en algunos casos usar (por ejemplo, en un oligo) una secuencia triplete N,N,N degenerada, N,N,G/T, o una secuencia triplete N,N,G/C.

Se aprecia, sin embargo, que el uso de un triplete degenerado (tal como N,N,G/T o una secuencia triplete N,N,G/C) como se describe en la presente descripción es ventajoso por varias razones. En un aspecto, esta descripción proporciona un medio para generar sistemáticamente y de forma bastante fácil la sustitución de la gama completa de aminoácidos posibles (para un total de 20 aminoácidos) en todas y cada una de las posiciones de aminoácidos en un polipéptido. De este modo, para un polipéptido de 100 aminoácidos, la descripción proporciona una manera de generar sistemáticamente y de forma bastante fácil 2000 especies distintas (es decir, 20 aminoácidos posibles por posición en 100 posiciones de aminoácidos). Se aprecia que se proporcionan, a través del uso de un oligo que contiene una secuencia triplete N,N,G/T o N,N,G/C degenerada, 32 secuencias individuales que codifican 20 aminoácidos posibles. De este modo, en una vasija de reacción en la que una secuencia polinucleotídica progenitora se somete a mutagénesis de saturación usando uno de tales oligos, se generan 32 polinucleótidos progenie distintos que codifican 20 polipéptidos distintos. En contraste, el uso de un oligo no degenerado en la mutagénesis dirigida al sitio conduce a solamente un producto polipeptídico progenie por vasija de reacción.

Esta descripción también proporciona el uso de oligos neurodegenerados, que se pueden usar opcionalmente en combinación con cebadores degenerados descritos. Se aprecia que, en algunas situaciones, es ventajoso usar oligos no degenerados para generar mutaciones de punto específicas en un polinucleótido de trabajo. Esto proporciona un medio para generar mutaciones de punto silenciosas específicas, mutaciones de punto que conducen a cambios de aminoácidos correspondientes, y mutaciones de punto que causan la generación de codones de parada y la expresión correspondiente de fragmentos peptídicos.

De este modo, en un aspecto de esta descripción, cada vasija de reacción de mutagénesis de saturación contiene polinucleótidos que codifican al menos 20 moléculas polipeptídicas progenie de manera que los 20 aminoácidos están representados en la una posición de aminoácido específica correspondiente a la posición de codón mutagenizada en el polinucleótido progenitor. Los polipéptidos progenie degenerados 32 veces, generados a partir de cada vasija de reacción de mutagénesis de saturación, se pueden someter a amplificación clonal (por ejemplo, clonados en un hospedante de *E. coli* adecuado, usando un vector de expresión) y se pueden someter a cribado de expresión. Cuando un polipéptido progenie individual se identifica mediante cribado para presentar un cambio

favorable en la propiedad (cuando se compara con el polipéptido progenitor), se puede secuenciar para identificar la sustitución de aminoácido correspondientemente favorable contenida en él.

Sea precia que al mutagenizar todas y cada una de las posiciones de aminoácidos en un polipéptido progenitor usando mutagénesis de saturación como se describe aquí, se pueden identificar cambios de aminoácidos favorables en más de una posición de aminoácido. Se pueden generar una o más nuevas moléculas proge que contienen una combinación de todas o partes de estas sustituciones de aminoácidos favorables. Por ejemplo, si se identifican 2 cambios de aminoácidos favorables específicos en cada una de 3 posiciones de aminoácidos en un polipéptido, las permutaciones incluyen 3 posibilidades en cada posición (ningún cambio del aminoácido original, y cada uno de dos cambios favorables) y 3 posiciones. De este modo, hay $3 \times 3 \times 3$ ó 27 posibilidades totales, incluyendo 7 que se examinaron previamente – 6 mutaciones de un solo punto (es decir, 2 en cada una de tres posiciones) y ningún cambio en ninguna posición.

De este modo, en una ejemplificación no limitante, esta descripción proporciona el uso de mutagénesis de saturación en combinación con procedimientos de mutagenización adicionales, tal como un procedimiento en el que se introducen dos o más polinucleótidos relacionados en una célula hospedante adecuada de manera que se genera una polinucleótido híbrido mediante recombinación y redistribución reductiva.

Además de llevar a cabo la mutagénesis a lo largo de la secuencia completa de un gen, la presente descripción proporciona que la mutagénesis se puede usar para sustituir cada una de cualquier número de bases en una secuencia polinucleotídica, en la que el número de bases que se va a mutagenizar es, en un aspecto, cada número entero de 15 a 100.000. De este modo, en lugar de mutagenizar cada posición a lo largo de una molécula, se puede someter cada base o un número discreto de bases (en un aspecto, totalizando un subconjunto de 15 a 100.000) a mutagénesis. En un aspecto, se usa un nucleótido distinto para mutagenizar cada posición o grupo de posiciones a lo largo de una secuencia polinucleotídica. Un grupo de 3 posiciones a mutagenizar puede ser un codón. Las mutaciones se pueden introducir preferiblemente usando un cebador mutagénico, que contiene un casete heterólogo, también denominado casete mutagénico. Los casetes ilustrativos pueden tener de 1 a 500 bases. Cada posición nucleotídica en tales casetes heterólogos pueden ser N, A, C, G, T, A/C, A/G, A/T, C/G, C/T, G/T, C/G/T, A/G/T, A/C/T, A/C/G, o E, en la que E es cualquier base que no sea A, C, G, o T (E se puede referir como oligo diseñador).

En un sentido general, la mutagénesis de saturación comprende mutagenizar un conjunto completo de casetes mutagénicos (en el que cada casete tiene en un aspecto alrededor de 1-500 bases de longitud) en secuencias polinucleotídicas definidas a mutagenizar (en el que la secuencia a mutagenizar tiene en un aspecto de alrededor de 15 a 100.000 bases de longitud). De este modo, se introduce un grupo de mutaciones (que oscila de 1 a 100 mutaciones) en cada casete que se va a mutagenizar. Un agrupamiento de mutaciones que se va a introducir en un casete puede ser diferente o igual que un segundo agrupamiento de mutaciones que se va a introducir en un segundo casete durante la aplicación de una ronda de mutagénesis de saturación. Tales agrupamientos se ejemplifican mediante supresiones, adiciones, agrupamientos de codones particulares, y agrupamientos de casetes nucleotídicos particulares.

Las secuencias definidas a mutagenizar incluyen un gen completo, una ruta, un ADNc, un marco de lectura abierto (ORF) completo y un promotor completo, un potenciador, un represor/transactivador, un origen de replicación, un intrón, un operador, o cualquier grupo funcional polinucleotídico. Generalmente, las “secuencias definidas” para este fin pueden ser polinucleótidos cualesquiera que tienen una secuencia polinucleotídica de 15 bases y secuencias polinucleotídicas de entre 15 bases y 15.000 bases de longitud (esta descripción nombra específicamente cada número entero intermedio). Las consideraciones en la elección de los agrupamientos de codones incluyen los tipos de aminoácidos codificados por un casete mutagénico degenerado.

En una ejemplificación de un agrupamiento de mutaciones que se pueden introducir en un casete mutagénico, esta descripción proporciona específicamente sustituciones codónicas degeneradas (usando oligos degenerados) que codifican 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 y 20 aminoácidos en cada posición y una biblioteca de polipéptidos codificada de ese modo.

Reensamblaje por ligación sintética (SLR)

La descripción proporciona un sistema de modificación génica no estocástico denominado “reensamblaje por ligación sintética”, o simplemente “SLR”, un “proceso de evolución dirigida”, para generar polipéptidos, por ejemplo enzimas de la descripción, con propiedades nuevas o alteradas.

SLR es un método de ligar juntos fragmentos oligonucleotídicos de forma no estocástica. Este método difiere del barajado oligonucleotídico estocástico por cuanto los bloques constructores de ácido nucleico no se barajan, concatenan o quimerizan aleatoriamente, sino más bien se ensamblan de forma no estocástica. Véase, por ejemplo, la Solicitud de Patente U.S. Serie nº (USSN) 09/332.835 titulada “Synthetic Ligation Reassembly in Directed Evolution” y presentada el 14 de junio de 1999 (“USSN 09/332.835”). En un aspecto, SLR comprende las siguientes etapas: (a) proporcionar un polinucleótido molde, en el que el polinucleótido molde comprende una secuencia que codifica un gen homólogo; (b) proporcionar una pluralidad de polinucleótidos de bloques constructores, en el que los

polinucleótidos de bloques constructores se diseñan para el reensamblaje cruzado con el polinucleótido molde a una secuencia predeterminada, y un polinucleótido de bloque constructor comprende una secuencia que es una variante del gen homólogo y una secuencia homóloga al polinucleótido molde que flanquea la secuencia variante; (c) combinar un polinucleótido de bloque constructor con un polinucleótido molde de manera que el polinucleótido de bloque constructor se reensambla de forma cruzada con el polinucleótido molde para generar polinucleótidos que comprenden variaciones de secuencias génicas homólogas.

SLR no depende de la presencia de niveles elevados de homología entre los polinucleótidos a reordenar. De este modo, este método se puede usar para generar no estocásticamente bibliotecas (o conjuntos) de moléculas progenie que comprenden alrededor de 10^{100} quimeras diferentes. SLR se puede usar para generar bibliotecas compuestas de alrededor de 10^{1000} quimeras progenie diferentes. De este modo, los aspectos de la presente descripción incluyen métodos no estocásticos para producir un conjunto de moléculas de ácidos nucleicos quiméricas finalizadas que tienen un orden de ensamblaje global que se escoge mediante diseño. Este método incluye las etapas de generar mediante diseño una pluralidad de bloques constructores de ácidos nucleicos específicos que tienen extremos ligables compatibles mutuamente utilizables, y de ensamblar estos bloques constructores de ácidos nucleicos, de manera que se logre un orden de ensamblaje global diseñado.

Los extremos ligables mutuamente compatibles de los bloques constructores de ácidos nucleicos a ensamblar se consideran "utilizables" para este tipo de ensamblaje ordenado si permiten que los bloques constructores se acoplen en órdenes predeterminados. De este modo, el orden de ensamblaje global en el que se pueden acoplar los bloques constructores de ácidos nucleicos se especifica mediante el diseño de los extremos ligables. Si se usa más de una etapa de ensamblaje, entonces el orden de ensamblaje global en el que se pueden acoplar los bloques constructores de ácidos nucleicos también se especifica mediante el orden secuencial de la etapa o etapas de ensamblaje. En un aspecto, los trozos constructores hibridados se tratan con una enzima, tal como una ligasa (por ejemplo, T4 ADN ligasa), para lograr un enlace covalente de los trozos constructores.

En un aspecto, el diseño de los bloques constructores oligonucleotídicos se obtiene analizando un conjunto de moldes de secuencias de ácidos nucleicos progenitoras que sirven como base para producir un conjunto progenie de polinucleótidos quiméricos finalizados. Estos moldes oligonucleotídicos progenitores sirven así como una fuente de información de secuencia que ayuda en el diseño de los bloques constructores de ácidos nucleicos que se van a mutagenizar, por ejemplo quimerizar o barajar. En un aspecto de este método, las secuencias de una pluralidad de moldes de ácidos nucleicos progenitores se alinean a fin de seleccionar uno o más puntos de demarcación. Los puntos de demarcación se pueden localizar en un área de homología, y comprenden uno o más nucleótidos. Estos puntos de demarcación son compartidos en un aspecto por al menos dos de los moldes progenitores. Los puntos de demarcación se pueden usar de ese modo para delinear los límites de los bloques constructores oligonucleotídicos a generar a fin de reordenar los polinucleótidos progenitores. Los puntos de demarcación identificados y seleccionados en las moléculas progenitoras sirven como puntos de quimerización potenciales en el ensamblaje de las moléculas progenie quiméricas finales. Un punto de demarcación puede ser un área de homología (compuesta de al menos una base nucleotídica homóloga) compartida por al menos dos secuencias polinucleotídicas progenitoras. Como alternativa, un punto de demarcación puede ser un área de homología que es compartida por al menos la mitad de las secuencias polinucleotídicas progenitoras, o puede ser un área de homología que es compartida por al menos dos tercios de las secuencias polinucleotídicas progenitoras. Incluso más en un aspecto, un punto de demarcación usable es un área de homología que es compartida por al menos tres cuartos de las secuencias polinucleotídicas progenitoras, o puede ser compartida por casi todas las secuencias polinucleotídicas progenitoras. En un aspecto, un punto de demarcación es un área de homología que es compartida por todas las secuencias polinucleotídicas progenitoras.

En un aspecto, un proceso de reensamblaje por ligación se lleva a cabo exhaustivamente a fin de generar una biblioteca exhaustiva de polinucleótidos quiméricos progenie. En otras palabras, todas las posibles combinaciones ordenadas de los bloques constructores de ácidos nucleicos se representan en el conjunto de moléculas de ácidos nucleicos quiméricos finalizados. Al mismo tiempo, en otro aspecto, el orden de ensamblaje (es decir, el orden de ensamblaje de cada bloque constructor en la secuencia 5' a 3' de cada ácido nucleico quimérico finalizado) en cada combinación es mediante diseño (o no estocástico) como se describe anteriormente. Debido a la naturaleza no estocástica de esta descripción, se reduce enormemente la posibilidad de productos secundarios indeseados.

En otro aspecto, el método de reensamblaje por ligación se lleva a cabo sistemáticamente. Por ejemplo, el método se lleva a cabo a fin de generar una biblioteca sistemáticamente compartimentalizada de moléculas progenie, con compartimientos que se pueden cribar sistemáticamente, por ejemplo uno a uno. En otras palabras, esta descripción proporciona que, mediante el uso selectivo y juicioso de bloques constructores de ácidos nucleicos específicos, acoplado con el uso selectivo y juicioso de reacciones de ensamblaje secuencialmente por etapas, se puede lograr un diseño en el que se obtienen conjuntos específicos de productos progenie en cada una de las varias vasijas de reacción. Esto permite que se lleve a cabo un procedimiento de examen y cribado sistemático. De este modo, estos métodos permiten que se examine sistemáticamente en grupos más pequeños un número potencialmente muy grande de moléculas progenie. Debido a su capacidad para llevar a cabo quimerizaciones de una manera que es muy flexible aunque igualmente exhaustiva y sistemática, particularmente cuando hay un bajo nivel de homología entre las moléculas progenitoras, estos métodos proporcionan la generación de una biblioteca (o conjunto) compuesta de un gran número de moléculas progenie. Debido a la naturaleza no estocástica del presente

reensamblaje por ligación, las moléculas progenie generadas comprenden en un aspecto una biblioteca de moléculas de ácidos nucleicos quiméricos finalizados que tienen un orden de ensamblaje global que se escoge por diseño. La mutagénesis de saturación y los métodos de evolución dirigida optimizada también se pueden usar para generar diferentes especies moleculares progenie. Se aprecia que la descripción proporciona libertad de elección y control con respecto a la selección de los puntos de demarcación, el tamaño y número de los bloques constructores de ácidos nucleicos, y el tamaño y diseño de los acoplamientos. Además, se aprecia que el requisito de homología intermolecular está muy relajado para la operabilidad de esta descripción. De hecho, los puntos de demarcación se pueden escoger incluso en áreas de poca o ninguna homología intermolecular. Por ejemplo, debido al bamboleo de los codones, es decir, la degeneración de los codones, se pueden introducir sustituciones nucleotídicas en los bloques constructores de ácidos nucleicos sin alterar el aminoácido codificado originalmente en el molde progenitor correspondiente. Como alternativa, un codón se puede alterar de manera que se altere la codificación de un aminoácido original. Esta descripción proporciona que tales sustituciones se puedan introducir en el bloque constructor de ácido nucleico a fin de incrementar la incidencia de puntos de demarcación homólogos intramoleculares, y de este modo permite que se logre un número creciente de acoplamientos entre los bloques constructores, lo que a su vez permite que se genere un mayor número de moléculas quiméricas progenie.

En un aspecto, la presente descripción proporciona un método no estocástico denominado reensamblaje génico sintético, que está relacionado en cierto modo con el barajado estocástico, salvo que los bloques constructores de los ácidos nucleicos no están barajados o concatenados o quimerizados aleatoriamente, sino que en vez de eso se ensamblan no estocásticamente.

El método de reensamblaje de genes sintéticos no depende de la presencia de un nivel elevado de homología entre los polinucleótidos que se van a barajar. El método se puede usar para generar no estocásticamente bibliotecas (o conjuntos) de moléculas progenie que comprenden alrededor de 10^{100} quimeras diferentes. Posiblemente, el reensamblaje de genes sintéticos se puede usar incluso para generar bibliotecas que comprenden alrededor de 10^{1000} quimeras progenie diferentes.

De este modo, en un aspecto, la descripción proporciona un método no estocástico para producir un conjunto de moléculas de ácido nucleico quiméricas finalizadas que tienen un orden de ensamblaje global que se elige mediante diseño, método el cual comprende las etapas de generar mediante diseño una pluralidad de bloques constructores de ácidos nucleicos específicos que tienen extremos ligables compatibles mutuamente utilizables, y de ensamblar estos bloques constructores de ácidos nucleicos, de manera que se logre un orden de ensamblaje global diseñado.

Los extremos ligables mutuamente compatibles de los bloques constructores de ácidos nucleicos a ensamblar se consideran "utilizables" para este tipo de ensamblaje ordenado si permiten que los bloques constructores se acoplen en órdenes predeterminados. De este modo, en un aspecto, el orden de ensamblaje global en el que se pueden acoplar los bloques constructores de ácidos nucleicos se especifica mediante el diseño de los extremos ligables y, si se va a usar más de una etapa de ensamblaje, entonces el orden de ensamblaje global en el que se pueden acoplar los bloques constructores de ácidos nucleicos también se especifica mediante el orden secuencial de la etapa o etapas de ensamblaje. En un aspecto de la descripción, los trozos constructores hibridados se tratan con una enzima, tal como una ligasa (por ejemplo, T4 ADN ligasa), para lograr un enlace covalente de los trozos constructores.

En otro aspecto, el diseño de los bloques constructores de ácidos nucleicos se obtiene analizando las secuencias de un conjunto de moldes de ácidos nucleicos progenitores que sirven como base para producir un conjunto progenie de moléculas de ácidos nucleicos quiméricos finalizados. Estos moldes de ácidos nucleicos progenitores sirven así como una fuente de información de secuencia que ayuda en el diseño de los bloques constructores de ácidos nucleicos que se van a mutagenizar, es decir, quimerizar o barajar.

En una ejemplificación, la descripción proporciona la quimerización de una familia de genes relacionados y su familia codificada de productos relacionados. En una ejemplificación particular, los productos codificados son enzimas. Los polipéptidos de la presente descripción se pueden mutagenizar de acuerdo con los métodos descritos aquí.

De este método según un aspecto de la descripción, las secuencias de una pluralidad de moldes de ácidos nucleicos progenitores (por ejemplo, polinucleótidos de la descripción) se alinean a fin de seleccionar uno o más puntos de demarcación, puntos de demarcación los cuales se pueden localizar en un área de homología. Los puntos de demarcación se pueden usar de ese modo para delinear los límites de los bloques constructores de ácidos nucleicos a generar. De este modo, los puntos de demarcación identificados y seleccionados en las moléculas progenitoras sirven como puntos de quimerización potenciales en el ensamblaje de las moléculas progenie.

Típicamente, un punto de demarcación utilizable es un área de homología (compuesta de al menos una base nucleotídica homóloga) compartida por al menos dos moldes progenitores, pero el punto de demarcación puede ser un área de homología que es compartida por al menos la mitad de los moldes progenitores, al menos dos tercios de los moldes progenitores, al menos tres cuartos de los moldes progenitores, o en un aspecto casi todos los moldes progenitores. Incluso más en un aspecto todavía, un punto de demarcación utilizable es un área de homología que es compartida por todos los moldes progenitores.

- En un aspecto, un proceso de reensamblaje génico se lleva a cabo exhaustivamente a fin de generar una biblioteca exhaustiva. En otras palabras, todas las posibles combinaciones ordenadas de los bloques constructores de ácidos nucleicos se representan en el conjunto de moléculas de ácidos nucleicos quiméricos finalizados. Al mismo tiempo, el orden de ensamblaje (es decir, el orden de ensamblaje de cada bloque constructor en la secuencia 5' a 3' de cada ácido nucleico quimérico finalizado) en cada combinación es mediante diseño (o no estocástico). Debido a la naturaleza no estocástica del método, se reduce enormemente la posibilidad de productos secundarios indeseados.
- En otro aspecto, el método proporciona que el proceso de reensamblaje génico se lleva a cabo sistemáticamente, por ejemplo a fin de generar una biblioteca sistemáticamente compartimentalizada, con compartimientos que se pueden cribar sistemáticamente, por ejemplo uno a uno. En otras palabras, la descripción proporciona que, mediante el uso selectivo y juicioso de bloques constructores de ácidos nucleicos específicos, acoplado con el uso selectivo y juicioso de reacciones de ensamblaje secuencialmente por etapas, se puede lograr un diseño experimental en el que se obtienen conjuntos específicos de productos progenie en cada una de las varias vasijas de reacción. Esto permite que se lleve a cabo un procedimiento de examen y cribado sistemático. De este modo, permiten que se examine sistemáticamente en grupos más pequeños un número potencialmente muy grande de moléculas progenie.
- Debido a su capacidad para llevar a cabo quimerizaciones de una manera que es muy flexible aunque igualmente exhaustiva y sistemática, particularmente cuando hay un bajo nivel de homología entre las moléculas progenitoras, la actual descripción proporciona la generación de una biblioteca (o conjunto) compuesta de un gran número de moléculas progenie. Debido a la naturaleza no estocástica de la presente descripción de reensamblaje génico, las moléculas progenie generadas comprenden en un aspecto una biblioteca de moléculas de ácidos nucleicos quiméricos finalizados que tienen un orden de ensamblaje global que se escoge por diseño. En un aspecto particular, tal biblioteca generada comprende más de 10^5 a más de 10^{1000} especies moleculares progenie diferentes.
- En un aspecto, un conjunto de moléculas de ácido nucleico quiméricas finalizadas, producidas como se ha descrito, comprende un polinucleótido que codifica un polipéptido. De acuerdo con un aspecto, este polinucleótido es un gen, que puede ser un gen elaborado por el hombre. De acuerdo con otro aspecto, este polinucleótido forma parte de una ruta génica, que puede ser una ruta génica elaborada por el hombre. La descripción proporciona que uno o más genes elaborados por el hombre, generados por la descripción, se pueden incorporar a una ruta génica elaborada por el hombre, tal como una ruta operable en un organismo eucariota (incluyendo una planta).
- En otra ejemplificación, la naturaleza sintética de la etapa en la que se generan los bloques constructores permite el diseño e introducción de nucleótidos (por ejemplo, uno o más nucleótidos, que pueden ser, por ejemplo, codones o intrones, o secuencias reguladoras) que más tarde se pueden eliminar en un proceso *in vitro* (por ejemplo mediante mutagénesis) con un proceso *in vivo* (por ejemplo utilizando la capacidad de ajuste génico de un organismo hospedante). Se aprecia que en muchos casos la introducción de estos nucleótidos también puede ser deseable por muchas otras razones, además del beneficio potencial de crear un punto de demarcación utilizable.
- De este modo, según otro aspecto, la descripción proporciona que un bloque constructor de ácido nucleico se puede usar para introducir un intrón. De este modo, la descripción proporciona que los intrones funcionales se pueden introducir en un gen de la descripción elaborado por el hombre. En consecuencia, la invención proporciona la generación de un polinucleótido quimérico que es un gen elaborado por el hombre que contiene uno (o más) intrones introducidos artificialmente.
- En consecuencia, la invención también proporciona la generación de un polinucleótido quimérico que es una ruta génica elaborada por el hombre que contiene uno (o más) intrones introducidos artificialmente. En un aspecto, el intrón o intrones introducidos artificialmente son funcionales en una o más células hospedantes para el ajuste génico de la misma manera que los intrones de origen natural sirven funcionalmente en el ajuste génico. La descripción proporciona un procedimiento para producir polinucleótidos que contienen intrones elaborados por el hombre para introducirlos en organismos hospedantes para la recombinación y/o el ajuste.
- Un gen elaborado por el hombre producido usando la descripción puede servir también como sustrato para la recombinación con otro ácido nucleico. Igualmente, una ruta génica elaborada por el hombre producida usando la descripción puede servir también como sustrato para la recombinación con otro ácido nucleico. En un aspecto, la recombinación se facilita mediante, o se produce en, áreas de homología entre el gene que contiene intrones elaborado por el hombre y un ácido nucleico, que sirve como una pareja de recombinación. En un aspecto, la pareja de recombinación puede ser también un ácido nucleico generado mediante la descripción, incluyendo un gen elaborado por el hombre o una ruta génica elaborada por el hombre. La recombinación puede ser facilitada por, o se puede producir en áreas de homología que existen en uno (o más) intrones introducidos artificialmente en el gen elaborado por el hombre.
- El método de reensamblaje sintético de genes de la descripción utiliza una pluralidad de bloques constructores de ácidos nucleicos, cada uno de los cuales tiene en un aspecto dos extremos ligables. Los dos extremos ligables en cada bloque constructor de ácidos nucleicos puede ser dos extremos romos (es decir, teniendo cada uno un saliente de cero nucleótidos), o en un aspecto un extremo romo y un saliente, o todavía más en un aspecto dos salientes.

Un saliente útil para este fin puede ser un saliente 3' o un saliente 5'. De este modo, un bloque constructor de ácidos nucleicos puede tener un saliente 3' o alternativamente un saliente 5', o alternativamente dos salientes 3' o alternativamente dos salientes 5'. El orden global en el que se ensamblan los bloques constructores de ácidos nucleicos para formar una molécula de ácido nucleico quimérica finalizada se determina mediante diseño experimental intencionado, y no es aleatorio.

En un aspecto, un bloque constructor de ácidos nucleicos se genera mediante síntesis química de dos ácidos nucleicos monocatenarios (también denominados oligos monocatenarios) y poniéndolos en contacto para permitirles que se hibriden para formar un bloque constructor de ácidos nucleicos bicatenario.

Un bloque constructor de ácidos nucleicos bicatenario puede tener un tamaño variable. Los tamaños de estos bloques constructores pueden ser pequeños o grandes. Los tamaños ilustrativos para el bloque constructor oscilan de 1 par de bases (sin incluir ningún saliente) a 100.000 pares de bases (sin incluir ningún saliente). Se proporcionan también otros intervalos de tamaños ilustrativos, que tienen límites inferiores de 1 pb a 10.000 pb (incluyendo cada valor entero intermedio) y límites superiores de 2 pb a 100.000 pb (incluyendo cada valor entero intermedio).

Existen muchos métodos por medio de los cuales se puede generar un bloque constructor de ácidos nucleicos bicatenario que sea utilizable por la descripción; y éstos son conocidos en la técnica y pueden ser realizados fácilmente por el experto en la técnica.

Según un aspecto, se genera un bloque constructor de ácidos nucleicos bicatenario generando en primer lugar dos ácidos nucleicos bicatenarios y permitiéndoles que se hibriden para formar un bloque constructor de ácidos nucleicos bicatenario. Las dos hebras de un bloque constructor de ácidos nucleicos bicatenario pueden ser complementarias a cada nucleótido, aparte de cualquiera de los que forman un saliente; no conteniendo de este modo emparejamientos erróneos, aparte de cualquier saliente o salientes. Según otro aspecto, las dos hebras de un bloque constructor de ácidos nucleicos bicatenario son complementarias en no todos los nucleótidos, aparte de cualquiera de los que forman un saliente. De este modo, según este aspecto, se puede usar un bloque constructor de ácidos nucleicos bicatenario para introducir degeneración codónica. En un aspecto, la degeneración codónica se introduce usando la mutagénesis de saturación de sitio descrita aquí, usando uno o más casetes N,N,G/T o alternativamente usando uno o más casetes N,N,N.

El método de recombinación *in vivo* de la descripción se puede llevar a cabo a ciegas en un conjunto de híbridos o alelos desconocidos de un polinucleótido o secuencia específicos. Sin embargo, no es necesario conocer la secuencia de ADN o ARN real del polinucleótido específico.

El enfoque en el que se usa la recombinación en una población mixta de genes puede ser útil para la generación de cualquier proteína útil, por ejemplo interleucina I, anticuerpos, tPA y hormona del crecimiento. Este enfoque se puede usar para generar proteínas que tienen especificidad o actividad alteradas. El enfoque también puede ser útil para la generación de secuencias híbridas de ácidos nucleicos, por ejemplo regiones promotoras, intrones, exones, secuencias potenciadoras, regiones no traducidas 3' o regiones no traducidas 5' de genes. De este modo, este enfoque se puede usar para generar genes que tienen tasas de expresión incrementadas. Este enfoque también puede ser útil en el estudio de secuencias de ADN repetitivas. Finalmente, este enfoque puede ser útil para mutar ribozimas o aptámeros.

En un aspecto, la descripción descrita aquí se refiere al uso de ciclos repetidos de redistribución reductiva, recombinación y selección, que permiten la evolución molecular dirigida de secuencias lineales altamente complejas, tales como ADN, ARN o proteínas, a través de recombinación.

Sistema de evolución dirigida optimizada

La descripción proporciona un sistema de modificación génica no estocástico denominado "sistema de evolución dirigida optimizada" para generar polipéptidos, por ejemplo enzimas o anticuerpos de la descripción, con propiedades nuevas o alteradas. La evolución dirigida optimizada se refiere al uso de ciclos repetidos de redistribución reductiva, recombinación y selección que permite la evolución molecular dirigida de ácidos nucleicos mediante recombinación. La evolución dirigida optimizada permite la generación de una gran población de secuencias quiméricas evolucionadas, en el que la población generada está significativamente enriquecida en secuencias que tienen un número predeterminado de sucesos de cruzamiento.

Un suceso de cruzamiento es un punto en una secuencia quimérica en el que se produce un desplazamiento en la secuencia desde una variante parental a otra variante parental. Tal punto está normalmente en la unión donde se ligan juntos oligonucleótidos de dos progenitores para formar una única secuencia. Este método permite el cálculo de las concentraciones correctas de secuencias oligonucleotídicas de manera que la población quimérica final de secuencias está enriquecida para el número escogido de sucesos de cruzamiento. Esto proporciona más control sobre la elección de variantes quiméricas que tienen un número predeterminado de sucesos de cruzamiento.

Además, este método proporciona un medio conveniente para explorar una tremenda cantidad del espacio de variantes proteicas posibles en comparación con otros sistemas. Previamente, si se generaron, por ejemplo, 10^{13} moléculas quiméricas durante una reacción, sería extremadamente difícil ensayar tal número elevado de variantes

quiméricas en busca de una actividad particular. Además, una porción significativa de la población progeñie tendría un número muy elevado de sucesos de cruzamiento, lo que daría como resultado proteínas que tendrían menos probabilidad de tener niveles incrementados de una actividad particular. Usando estos métodos, la población de moléculas quiméricas se puede enriquecer en aquellas variantes que tienen un número particular de sucesos de cruzamiento. De este modo, aunque todavía se pueden generar 10^{13} moléculas quiméricas durante una reacción, cada una de las moléculas escogidas para el análisis posterior tendrá muy probablemente, por ejemplo, sólo tres sucesos de cruzamiento. Debido a que la población progeñie resultante puede estar sesgada para que tenga un número predeterminado de sucesos de cruzamiento, los límites en la variedad funcional entre las moléculas quiméricas se reducen. Esto proporciona un número más manejable de variables cuando se calcula qué oligonucleótido de los polinucleótidos parentales originales puede ser responsable de afectar un rasgo particular.

Un método para crear una secuencia polinucleotídica progeñie quimérica es crear oligonucleótidos que corresponden a fragmentos o porciones de cada secuencia parental. Cada oligonucleótido incluye en un aspecto una región única de solapamiento de manera que el mezclamiento de todos los oligonucleótidos da como resultado una nueva variante que tiene cada fragmento oligonucleotídico ensamblado en el orden correcto. En USSN 09/332.835, patente U.S. nº 6.361.974, se puede encontrar también información adicional.

El número de oligonucleótidos generados para cada variante parental posee una relación con el número total de cruzamientos resultantes en la molécula quimérica que se crea finalmente. Por ejemplo, se pueden proporcionar tres variantes de secuencias nucleotídicas parentales para sufrir una reacción de ligación a fin de encontrar una variante quimérica que tiene, por ejemplo, mayor actividad a temperatura elevada. Como ejemplo, se puede generar un conjunto de 50 secuencias oligonucleotídicas que corresponden a cada una de las porciones de cada variante parental. En consecuencia, durante el proceso de reensamblaje por ligación, podría haber hasta 50 sucesos de cruzamiento en cada una de las secuencias quiméricas. La probabilidad de que cada uno de los polinucleótidos quiméricos generados contendrá oligonucleótidos procedentes de cada variante parental en orden alterno es muy baja. Si cada fragmento oligonucleotídico está presente en la reacción de ligación en la misma cantidad molar, es probable que en algunas posiciones los oligonucleótidos procedentes del mismo polinucleótido parental se ligarán próximos entre sí, y de este modo no darán como resultado un suceso de cruzamiento. Si la concentración de cada oligonucleótido procedente de cada progenitor se mantiene constante durante cualquier etapa de ligación en este ejemplo, hay una posibilidad de 1/3 (suponiendo 3 progenitores) de que un oligonucleótido procedente de la misma variante parental se ligue en la secuencia quimérica y no produzca cruzamiento.

En consecuencia, se puede determinar una función de densidad de probabilidad (PDF) para predecir la población de sucesos de cruzamiento que es probable que se produzcan durante cada etapa en la reacción de ligación dado un número fijo de variantes parentales, un número de oligonucleótidos que corresponden a cada variante, y las concentraciones de cada variante durante cada etapa en la reacción de ligación. Más abajo se describen las estadísticas y matemáticas detrás de la determinación de la PDF. Utilizando estos métodos, se puede calcular tal función de densidad de probabilidad, y de este modo enriquecer la población progeñie quimérica para un número predeterminado de sucesos de cruzamiento que resultan de una reacción de ligación particular. Además, se puede determinar un número diana de sucesos de cruzamiento, y el sistema se puede programar entonces para calcular las cantidades de partida de cada oligonucleótido parental durante cada etapa en la reacción de ligación para dar como resultado una función de densidad de probabilidad que se centra en el número predeterminado de sucesos de cruzamiento. Estos métodos se refieren al uso de ciclos repetidos de redistribución reductiva, recombinación y selección, que permiten la evolución molecular dirigida de un ácido nucleico que codifica un polipéptido a través de recombinación. Este sistema permite la generación de una gran población de secuencias quiméricas evolucionadas, en el que la población generada está significativamente enriquecida en secuencias que tienen un número predeterminado de sucesos de cruzamiento. Un suceso de cruzamiento es un punto en una secuencia quimérica en el que se produce un desplazamiento en la secuencia desde una variante parental a otra variante parental. Tal punto está normalmente en la unión donde se ligan juntos oligonucleótidos de dos progenitores para formar una única secuencia. El método permite el cálculo de las concentraciones correctas de secuencias oligonucleotídicas de manera que la población quimérica final de secuencias está enriquecida para el número escogido de sucesos de cruzamiento. Esto proporciona más control sobre la elección de variantes quiméricas que tienen un número predeterminado de sucesos de cruzamiento.

Además, estos métodos proporcionan un medio conveniente para explorar una tremenda cantidad del espacio de variantes proteicas posible en comparación con otros sistemas. Usando los métodos descritos aquí, la población de moléculas quiméricas se puede enriquecer para aquellas variantes que tienen un número particular de sucesos de cruzamiento. De este modo, aunque todavía se pueden generar 10^{13} moléculas quiméricas durante una reacción, cada una de las moléculas escogidas para análisis posterior tiene muy probablemente, por ejemplo, sólo tres sucesos de cruzamiento. Debido a que la población progeñie resultante se puede sesgar para que tenga un número predeterminado de sucesos de cruzamiento, los límites en la variedad funcional entre las moléculas quiméricas se reducen. Esto proporciona un número más manejable de variables cuando se calcula qué oligonucleótido de los polinucleótidos parentales originales puede ser responsable de afectar a un rasgo particular.

En un aspecto, el método crea una secuencia polinucleotídica progeñie quimérica creando oligonucleótidos que corresponden a fragmentos o porciones de cada secuencia parental. Cada oligonucleótido incluye en un aspecto una región única de solapamiento de manera que el mezclamiento de los oligonucleótidos juntos da como resultado

una nueva variante que tiene cada fragmento oligonucleotídico ensamblado en el orden correcto. Véase también USSN 09/332.835.

Determinación de los sucesos de cruzamiento

5 Aspectos de la descripción incluyen un sistema y software que recibe una función de densidad de probabilidad (PDF) de cruzamiento deseado, el número de genes progenitores a reensamblar, y el número de fragmentos en el reensamblaje, como entradas. La salida de este programa es una "PDF de fragmentos" que se puede usar para determinar una receta para producir genes reensamblados, y la PDF de cruzamiento estimada de esos genes. En un aspecto, el procesamiento descrito aquí se lleva a cabo en MATLABTM (The Mathworks, Natick, Massachusetts), un lenguaje de programación y entorno de desarrollo para computación técnica.

10 Procesos iterativos

En la práctica de la descripción, estos procesos se pueden repetir iterativamente. Por ejemplo, un ácido nucleico (o el ácido nucleico) responsable de un fenotipo alterado o nuevo se identifica, se vuelve a aislar (por ejemplo, usando un ácido nucleico de la descripción), se modifica nuevamente, se vuelve a ensayar en busca de su actividad. Este proceso se puede repetir iterativamente hasta que se manipule un fenotipo deseado. Por ejemplo, se puede manipular en una célula una ruta anabólica o catabólica bioquímica completa, incluyendo, por ejemplo, una ruta biosintética o degradativa (por ejemplo, clorofila) nueva o alterada.

De forma similar, si se determina que un oligonucleótido particular no tiene efecto en absoluto sobre el rasgo deseado (por ejemplo, un fenotipo de ruta biosintética o degradativa (por ejemplo, clorofila) nueva o alterada), se puede eliminar como una variable sintetizando oligonucleótidos parentales más grandes que incluyen la secuencia a eliminar. Puesto que la incorporación de la secuencia en una secuencia más grande evita cualesquiera sucesos de cruzamiento, ya no habrá ninguna variación de esta secuencia en los polinucleótidos de la progenie. Esta práctica iterativa de determinar qué oligonucleótidos están más relacionados con el rasgo deseado, y cuáles no están relacionados, permite una exploración más eficiente de todas las variantes proteicas posibles que pueden proporcionar un rasgo o actividad particular.

25 Barajado *in vivo*

El barajado *in vivo* de las moléculas se usa en métodos de la descripción que proporcionan variantes de polipéptidos de la descripción, por ejemplo anticuerpos, enzimas, y similares. El barajado *in vivo* se puede llevar a cabo utilizando la propiedad natural de las células para recombinar multímeros. Aunque la recombinación *in vivo* ha proporcionado la ruta natural principal para la diversidad molecular, la recombinación genética sigue siendo un proceso relativamente complejo que implica 1) el reconocimiento de homologías; 2) escisión de las hebras, invasión de las hebras, y etapas metabólicas que conducen a la producción de quiasma recombinante; y finalmente, 3) la resolución de quiasma en moléculas recombinadas discretas. La formación del quiasma requiere el reconocimiento de secuencias homólogas.

En un aspecto, la descripción incluye un método para producir un polinucleótido híbrido a partir de al menos un primer polinucleótido y un segundo polinucleótido. La descripción se puede usar para producir un polinucleótido híbrido introduciendo al menos un primer polinucleótido y un segundo polinucleótido que comparten al menos una región de homología de secuencia parcial (por ejemplo, SEC ID NOS: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19 y sus combinaciones) en una célula hospedante adecuada. Las regiones de homología de secuencia parcial promueven procesos que dan como resultado la reorganización de secuencias que produce un polinucleótido híbrido. La expresión "polinucleótido híbrido", como se usa aquí, es cualquier secuencia nucleotídica que resulta del método de la presente descripción y contiene secuencia de al menos dos secuencias polinucleotídicas originales. Tales polinucleótidos híbridos pueden resultar de sucesos de recombinación intermolecular que promueven la integración de las secuencias entre moléculas de ADN. Además, tales polinucleótidos híbridos pueden resultar de procesos de redistribución reductiva intramolecular que utilizan secuencias repetidas para alterar una secuencia nucleotídica dentro de una molécula de ADN.

La reclasificación *in vivo* se centra en procesos "intermoleculares", denominados colectivamente como "recombinación", lo cuales, en bacterias, generalmente se denomina como un fenómeno "dependiente de RecA". La descripción se puede basar en procesos de recombinación de una célula hospedante para recombinar y reclasificar secuencias, o la capacidad de las células para mediar procesos reductivos para disminuir la complejidad de secuencias cuasi repetidas en la célula mediante supresión. Este proceso de "redistribución reductiva" se produce mediante un proceso "intramolecular", dependiente de RecA.

Por lo tanto, en otro aspecto de la descripción, se pueden generar nuevos polinucleótidos mediante el proceso de reclasificación reductiva. El método implica la generación de constructos que contienen secuencias consecutivas (secuencias codificantes originales), su inserción en un vector apropiado, y su introducción subsiguiente en una célula hospedante apropiada. La reclasificación de las identidades moleculares individuales se produce mediante procesos combinatorios entre las secuencias consecutivas en el constructo que posee regiones de homología, o entre unidades cuasi repetidas. El proceso de reclasificación recombina y/o reduce la complejidad y extensión de las secuencias repetidas, y da como resultado la producción de nuevas especies moleculares. Diversos tratamientos se

pueden aplicar para potenciar la tasa de reclasificación. Estos podrían incluir el tratamiento con luz ultravioleta, o sustancias químicas que dañan el ADN, y/o el uso de estirpes celulares hospedantes que presentan niveles potenciados de “inestabilidad genética”. De este modo, el proceso de reclasificación puede implicar recombinación homóloga o la propiedad natural de secuencias cuasi repetidas para dirigir su propia evolución.

- 5 Las secuencias repetidas o “cuasi repetidas” desempeñan un papel en la inestabilidad genética. En la presente descripción, “cuasi repetidas” son repeticiones que no están restringidas a su estructura de unidad original. Las unidades cuasi repetidas se pueden presentar como un conjunto de secuencias en un constructo; unidades consecutivas de secuencias similares. Una vez ligadas, las uniones entre las secuencias consecutivas se hacen esencialmente invisibles, y la naturaleza cuasi repetitiva del constructo resultante es ahora continua a nivel
- 10 molecular. El proceso de supresión que lleva a cabo la célula para reducir la complejidad del constructo resultante funciona entre las secuencias cuasi repetidas. Las unidades cuasi repetidas proporcionan un repertorio prácticamente ilimitado de moldes sobre los que se pueden producir sucesos de deslizamiento. Los constructos que contienen las cuasi repeticiones proporcionan así eficazmente elasticidad molecular suficiente de que se pueden producir sucesos de supresión (y potencialmente inserción) virtualmente en cualquier parte en las unidades cuasi
- 15 repetitivas.

- Cuando las secuencias cuasi repetidas están todas ligadas en la misma orientación, por ejemplo cabeza a cola o viceversa, la célula no puede distinguir unidades individuales. En consecuencia, el proceso reductivo se puede producir a través de las secuencias. Por el contrario, cuando, por ejemplo, las unidades se presentan cabeza a cabeza, en lugar de cabeza a cola, la inversión delinea los puntos finales de la unidad adyacente, de manera que la
- 20 formación de supresión favorecerá la pérdida de unidades discretas. De este modo, es preferible con el presente método que las secuencias estén en la misma orientación. La orientación al azar de secuencias cuasi repetidas dará como resultado la pérdida de eficiencia de la reclasificación, mientras que la orientación consistente de las secuencias ofrecerá la eficiencia más elevada. Sin embargo, aunque el tener un menor número de las secuencias contiguas en la misma orientación disminuye la eficiencia, todavía puede proporcionar suficiente elasticidad para la recuperación eficaz de nuevas moléculas. Se pueden obtener constructos con las secuencias cuasi repetidas en la
- 25 misma orientación para permitir una mayor eficiencia.

Las secuencias se pueden ensamblar en la orientación cabeza a cola usando cualquiera de una variedad de métodos, incluyendo los siguientes:

- 30 a) Se pueden utilizar cebadores que incluyen una cabeza poly-A y una cola poly-T que cuando son monocatenarios proporcionarían orientación. Esto se logra obteniendo las primeras pocas bases de los cebadores a partir de ARN y por tanto ARNasa H fácilmente eliminada.
- b) Se pueden utilizar cebadores que incluyen sitios de escisión de restricción únicos. Se requerirían múltiples sitios, una batería de secuencias únicas, y etapas repetidas de síntesis y ligación.
- 35 c) Las pocas bases internas del cebador se podrían tiolar y se podría usar una exonucleasa para producir moléculas con una cola apropiada.

La recuperación de las secuencias reclasificadas se basa en la identificación de vectores de clonación con un índice repetitivo (RI) reducido. Las secuencias codificantes reclasificadas se pueden recuperar entonces mediante amplificación. Los productos se vuelven a clonar y se expresan. La recuperación de vectores de clonación con RI reducido se puede ver afectada por:

- 40 1) El uso de vectores mantenidos sólo establemente cuando el constructo se reduce en complejidad.
- 2) La recuperación física de vectores acortados mediante procedimientos físicos. En este caso, el vector de clonación se recuperaría usando procedimientos de aislamiento de plásmidos estándar y se fraccionaría en tamaño en un gel de agarosa, o columna con un corte de bajo peso molecular utilizando procedimientos estándar.
- 45 3) La recuperación de vectores que contienen genes interrumpidos que se pueden seleccionar cuando disminuye el tamaño del inserto.
- 4) El uso de técnicas de selección directas con un vector de expresión y la selección apropiada.

- Las secuencias codificantes (por ejemplo, genes) de organismos relacionados pueden demostrar un grado elevado de homología y codificar productos proteicos bastante diversos. Estos tipos de secuencias son particularmente útiles
- 50 en la presente descripción como cuasi repeticiones. Sin embargo, aunque los ejemplos ilustrados más abajo demuestran la reclasificación de secuencias codificantes originales casi idénticas (cuasi repeticiones), este proceso no está limitado a tales repeticiones casi idénticas.

- El siguiente ejemplo demuestra un método de la descripción. Se describen secuencias de ácidos nucleicos codificantes (cuasi repeticiones) derivadas de tres (3) especies únicas. Cada secuencia codifica una proteína con un
- 55 conjunto distinto de propiedades. Cada una de las secuencias difiere en un único par o unos pocos pares de bases

en una posición única en la secuencia. Las secuencias cuasi repetidas se amplifican separada o colectivamente y se ligan en ensamblajes aleatorios de manera que todas las posibles permutaciones y combinaciones están disponibles en la población de moléculas ligadas. El número de unidades cuasi repetidas se puede controlar mediante las condiciones de ensamblaje. El número medio de unidades cuasi repetidas en un constructo se define como el índice repetitivo (RI).

Una vez formados, los constructos se pueden fraccionar o no por tamaños en un gel de agarosa según protocolos publicados, se pueden insertar en un vector de clonación, y se pueden transfectar en una célula hospedante apropiada. Las células se propagan entonces y se efectúa la "reclasificación reductiva". La velocidad del proceso de reclasificación reductiva se puede estimular mediante la introducción de daño de ADN si se desea. Si la reducción en RI está mediada por la formación de supresión entre secuencias repetidas mediante un mecanismo "intramolecular", o mediada por sucesos parecidos a recombinación a través de mecanismos "intermoleculares", es inmaterial. El resultado final es una reclasificación de las moléculas en todas las posibles combinaciones.

Opcionalmente, el método comprende la etapa adicional de cribar los miembros de la biblioteca del conjunto barajado para identificar miembros de la biblioteca barajados individuales que tienen la capacidad de unirse o de otro modo interactuar o catalizar una reacción particular (por ejemplo, tal como un dominio catalítico de una enzima) con una macromolécula predeterminada, tal como por ejemplo un receptor proteinoso, un oligosacárido, virión, u otro compuesto o estructura predeterminada.

Los polipéptidos que se identifican a partir de tales bibliotecas se pueden usar para fines terapéuticos, de diagnóstico, de investigación y fines relacionados (por ejemplo, catalizadores, solutos para incrementar la osmolaridad de una disolución acuosa, y similares), y/o se pueden someter a uno o más ciclos adicionales de barajado y/o selección.

En otro aspecto, se prevé que antes o durante la recombinación o reclasificación, los polinucleótidos generados mediante el método de la descripción se puedan someter a agentes o procesos que promuevan la introducción de mutaciones en los polinucleótidos originales. La introducción de tales mutaciones incrementaría la diversidad de polinucleótidos híbridos resultantes y polinucleótidos codificados a partir de ellos. Los agentes o procesos que promueven la mutagénesis pueden incluir, pero no se limitan a: (+)-CC-1065, o un análogo sintético tal como (+)-CC-1065-(N3-adenina (véase Sun y Hurley, (1992); un aducto 4'-fluoro-4-aminobifenílico N-acetilado o desacetilado capaz de inhibir la síntesis de ADN (véase, por ejemplo, van de Poll et al. (1992)); o un aducto 4-aminobifenílico N-acetilado o desacetilado capaz de inhibir la síntesis de ADN (véase también van de Poll et al. (1992), p. 751-758); cromo trivalente, una sal de cromo trivalente, un aducto de ADN con hidrocarburo aromático policíclico (PAH) capaz de inhibir la replicación del ADN, tal como 7-bromometilbenz[a]antraceno ("BMA"), tris(2,3-dibromopropil)fosfato ("Tris-BP"), 1,2-dibromo-3-cloropropano ("DBCP"), 2-bromoacroleína (2BA), benzo[a]pireno-7,8-dihidrodiol-9-10-epóxido ("BPDE"), una sal de halogenuro de platino(II), N-hidroxi-2-amino-3-metilimidazo[4,5-f]-quinolina ("N-hidroxi-IQ"), y N-hidroxi-2-amino-1-metil-6-fenilimidazo[4,5-f]-piridina ("N-hidroxi-PhIP"). Los medios ejemplares para ralentizar o detener la amplificación por PCR consiste en luz UV (+)-CC-1065 y (+)-CC-1065-(N3-adenina). Los medios particularmente englobados son aductos de ADN o polinucleótidos que comprenden los aductos de ADN de los polinucleótidos o conjunto de polinucleótidos, que se pueden liberar o eliminar mediante un proceso que incluye calentar la disolución que comprende los polinucleótidos antes del procesamiento posterior.

En otro aspecto, la descripción se refiere a un método para producir proteínas recombinantes que tienen actividad biológica tratando una muestra que comprende polinucleótidos molde bicatenarios que codifican una proteína de tipo salvaje en condiciones según la descripción que proporcionan la producción de polinucleótidos híbridos o reordenados.

Producción de variantes de secuencias

La descripción también proporciona métodos adicionales para obtener variantes de secuencias de las secuencias de ácido nucleico de la descripción. La descripción también proporciona métodos adicionales para aislar polipéptidos de la descripción. En un aspecto, la descripción proporciona variantes de secuencias codificantes (por ejemplo, un gen, ADNc o mensaje) de la descripción, que se pueden alterar por cualquier medio, incluyendo, por ejemplo, métodos aleatorios o estocásticos, o métodos no estocásticos o de "evolución dirigida", como se describe anteriormente.

Las variantes aisladas pueden ser de origen natural. La variante también se puede crear *in vitro*. Las variantes se pueden crear usando técnicas de ingeniería genética tales como mutagénesis dirigida al sitio, mutagénesis química al azar, procedimientos de supresión con exonucleasas III, y técnicas de clonación estándar. Como alternativa, tales variantes, fragmentos, análogos, o derivados se pueden crear usando procedimientos de síntesis o modificación química. Otros métodos para obtener variantes también son familiares para los expertos en la técnica. Estos incluyen procedimientos en los que las secuencias de ácidos nucleicos obtenidas de aislados naturales se modifican para generar ácidos nucleicos que codifican polipéptidos que tienen características que potencian su valor en aplicaciones industriales o de laboratorio. En tales procedimientos, se genera y se caracteriza un gran número de secuencias variantes que tienen una o más diferencias nucleotídicas con respecto a la secuencia obtenida del aislado natural. Estas diferencias nucleotídicas pueden dar como resultado cambios de aminoácidos con respecto a los polipéptidos codificados por los ácidos nucleicos de los aislados naturales.

Por ejemplo, se pueden crear variantes usando PCR propensa a errores. En PCR propensa a errores, la PCR se lleva a cabo en condiciones en las que la fidelidad de copiado de la ADN polimerasa es baja, de manera que se obtiene una tasa elevada de mutaciones de punto a lo largo de toda la longitud del producto de PCR. La PCR propensa a error se describe, por ejemplo, en Leung, D.W., et al., *Technique*, 1:11-15, 1989) y en Caldwell, R. C. y Joyce G.F., *PCR Methods Applic.*, 2:28-33, 1992. De forma breve, en tales procedimientos, los ácidos nucleicos a mutagenizar se mezclan con cebadores de PCR, tampón de reacción, $MgCl_2$, $MnCl_2$, Taq polimerasa y una concentración apropiada de dNTPs para lograr una tasa elevada de mutación de punto a lo largo de toda la longitud del producto de PCR. Por ejemplo, la reacción se puede llevar a cabo usando 20 fmoles de ácido nucleico a mutagenizar, 30 pmoles de cada cebador de PCR, un tampón de reacción que comprende 50 mM de KCl, 10 mM de Tris HCl (pH 8,3) y 0,01% de gelatina, 7 mM de $MgCl_2$, 0,5 mM de $MnCl_2$, 5 unidades de Taq polimerasa, 0,2 mM de dGTP, 0,2 mM de dATP, 1 mM de dCTP, y 1 mM de dTTP. La PCR se puede llevar a cabo durante 30 ciclos de 94°C durante 1 min., 45°C durante 1 min., y 72°C durante 1 min. Sin embargo, se apreciará que estos parámetros se pueden variar según sea apropiado. Los ácidos nucleicos mutagenizados se clonan en un vector apropiado, y se evalúan las actividades de los polipéptidos codificados por los ácidos nucleicos mutagenizados.

También se pueden crear variantes usando mutagénesis dirigida por oligonucleótidos, para generar mutaciones específicas del sitio en cualquier ADN clonado de interés. La mutagénesis oligonucleotídica se describe, por ejemplo, en Reidhaar-Olson (1988) *Science* 241:53-57. De forma breve, en tales procedimientos, se sintetiza una pluralidad de oligonucleótidos bicatenarios que poseen una o más mutaciones a introducir en el ADN clonado, y se inserta en el ADN clonado a mutagenizar. Los clones que contienen el ADN mutagenizado se recuperan, y se evalúan las actividades de los polipéptidos que codifican.

Otro método para generar variantes es PCR de ensamblaje. La PCR de ensamblaje implica el ensamblaje de un producto de PCR a partir de una mezcla de fragmentos pequeños de ADN. Se produce paralelamente en el mismo vial un gran número de reacciones de PCR diferentes, cebando los productos de una reacción a los productos de otra reacción. La PCR de ensamblaje se describe en, por ejemplo, la patente U.S. nº 5.965.408.

Todavía otro método para generar variantes es mutagénesis de PCR sexual. En la mutagénesis de PCR sexual, se produce *in vitro* la recombinación homóloga forzada entre moléculas de ADN de una secuencia de ADN diferente pero muy relacionada, como resultado de la fragmentación al azar de la molécula de ADN basado en la homología de secuencia, seguido de la fijación del cruzamiento mediante extensión del cebador en una reacción de PCR. La mutagénesis de PCR sexual se describe, por ejemplo, en Stemmer (1994) *Proc. Natl. Acad. Sci. USA* 91:10747-10751. De forma breve, en tales procedimientos, una pluralidad de ácidos nucleicos a recombinar se digiere con ADNasa para generar fragmentos que tienen un tamaño medio de 50-200 nucleótidos. Los fragmentos del tamaño medio deseado se purifican y resuspenden en una mezcla de PCR. La PCR se lleva a cabo en condiciones que facilitan la recombinación entre los fragmentos de ácidos nucleicos. Por ejemplo, la PCR se puede llevar a cabo resuspendiendo los fragmentos purificados a una concentración de 10-30 ng/μl en una disolución de 0,2 mM de cada dNTP, 2,2 mM de $MgCl_2$, 50 mM de KCl, 10 mM de Tris HCl, pH 9,0, y 0,1% de Triton X-100. Se añaden 2,5 unidades de Taq polimerasa por 100:1 de mezcla de reacción, y la PCR se lleva a cabo usando el siguiente régimen: 94°C durante 60 segundos, 94°C durante 30 segundos, 50-55°C durante 30 segundos, 72°C durante 30 segundos (30-45 veces) y 72°C durante 5 minutos. Sin embargo, se apreciará que estos parámetros se pueden variar según sea apropiado. En algunos aspectos, se pueden incluir oligonucleótidos en las reacciones de PCR. En otros aspectos, se puede usar el fragmento de Klenow de ADN polimerasa I en un primer conjunto de reacciones de PCR, y se puede usar Taq polimerasa en un conjunto subsiguiente de reacciones de PCR. Las secuencias recombinantes se aíslan, y se evalúan las actividades de los polipéptidos que codifican.

Las variantes también se pueden crear mediante mutagénesis *in vivo*. En algunos aspectos, las mutaciones al azar en una secuencia de interés se generan propagando la secuencia de interés en una cepa bacteriana, tal como una cepa de *E. coli*, que posee mutaciones en una o más de las rutas de reparación del ADN. Tales cepas "mutadoras" tienen una tasa de mutación al azar mayor que la de un progenitor de tipo salvaje. La propagación del ADN en una de estas cepas generará eventualmente mutaciones al azar dentro del ADN. Las cepas mutadoras adecuadas para uso para mutagénesis *in vivo* se describen en la Publicación PCT nº WO 91/16427, publicada el 31 de octubre de 1991, titulada "Methods for Phenotype Creation from Multiple Gene Population".

Las variantes también se pueden generar usando mutagénesis por inserción de casete. En la mutagénesis por inserción de casete, una pequeña región de una molécula de ADN bicatenaria se sustituye por un "casete" oligonucleotídico sintético que difiere de la secuencia nativa. El oligonucleótido contiene a menudo secuencia nativa completa y/o parcialmente aleatorizada.

También se puede usar la mutagénesis de conjunto recursiva para generar variantes. La mutagénesis de conjunto recursiva es un algoritmo para la manipulación de proteínas (mutagénesis proteica) desarrollado para producir diversas poblaciones de mutantes genotípicamente relacionados cuyos miembros difieren en la secuencia de aminoácidos. Este método usa un mecanismo de retroalimentación para controlar rondas sucesivas de mutagénesis por inserción de casete combinatoria. La mutagénesis de conjunto recursiva se describe, por ejemplo, en Arkin, A.P. y Youvan, D.C., *PNAS, USA*, 89:7811-7815, 1992.

En algunos aspectos, se crean variantes usando mutagénesis de conjunto exponencial. La mutagénesis de conjunto exponencial es un proceso para generar bibliotecas combinatorias con un porcentaje elevado de mutantes únicos y funcionales, en el que pequeños grupos de restos se distribuyen al azar en paralelo para identificar, en cada posición alterada, aminoácidos que conducen a proteínas funcionales. La mutagénesis de conjunto exponencial se describe, por ejemplo, en Delegrave, S. y Youvan, D.C., *Biotechnology Research*, 11:1548-1552, 1993. Las mutagénesis al azar y dirigida al sitio se describen, por ejemplo, en Arnold, F.H., *Current Opinion in Biotechnology* 4:450-455, 1993.

En algunos aspectos, las variantes se crean usando procedimientos de barajado en los que porciones de una pluralidad de ácidos nucleicos que codifican polipéptidos distintos se fusionan juntas para crear secuencias de ácidos nucleicos quiméricas que codifican polipéptidos quiméricos, como se describe en la patente U.S. nº 5.965.408, presentada el 9 de julio, 1996, titulada "Method of DNA Reassembly by Interrupting Synthesis", y en la patente U.S. nº 5.939.250, presentada el 22 de mayo, 1996, titulada "Production of Enzymes Having Desired Activities by Mutagenesis".

Las variantes de los polipéptidos de la descripción pueden ser variantes en las que uno o más de los restos de aminoácidos de los polipéptidos de las secuencias de la descripción se sustituyen por un resto de aminoácido conservado o no conservado (en un aspecto, un resto de aminoácido conservado), y tal resto de aminoácido sustituido puede ser o no uno codificado por el código genético.

La descripción proporciona realizaciones alternativas de los polipéptidos de la descripción (y los ácidos nucleicos que los codifican) que comprenden al menos una sustitución de aminoácido conservativa, como se explica aquí (por ejemplo, las sustituciones de aminoácido conservativas son aquellas que sustituyen un aminoácido dado en un polipéptido por otro aminoácido de características similares). La descripción proporciona polipéptidos (y los ácidos nucleicos que los codifican) en los que cualquiera, algunos o todos los restos de aminoácidos están sustituidos por otro aminoácido de características similares, por ejemplo una sustitución de aminoácido conservativa.

Las sustituciones conservativas son aquellas que sustituyen un aminoácido dado en un polipéptido por otro aminoácido de características similares. Típicamente observadas como sustituciones conservativas son las siguientes sustituciones: sustituciones de un aminoácido alifático tal como alanina, valina, leucina e isoleucina por otro aminoácido alifático; la sustitución de una serina por una treonina o viceversa; la sustitución de un resto ácido tal como ácido aspártico y ácido glutámico por otro resto ácido; la sustitución de un resto que posee un grupo amida, tal como asparagina y glutamina, por otro resto que posee un grupo amida; el intercambio de un resto básico tal como lisina y arginina por otro resto básico; y la sustitución de un resto aromático tal como fenilalanina, tirosina por otro resto aromático. Como alternativa, estas sustituciones conservativas pueden ser también equivalentes sintéticos de estos aminoácidos.

Otras variantes son aquellas en las que uno o más de los restos de aminoácidos de un polipéptido de la descripción incluyen un grupo sustituyente.

Todavía otras variantes son aquellas en las que el polipéptido está asociado con otro compuesto, tal como un compuesto para incrementar la semivida del polipéptido (por ejemplo polietilenglicol).

Variantes adicionales son aquellas en las que aminoácidos adicionales se fusionan al polipéptido, tal como una secuencia líder, una secuencia secretora, una secuencia proproteínica o una secuencia que facilita la purificación, enriquecimiento o estabilización del polipéptido.

En algunos aspectos, los fragmentos, derivados y análogos retienen la misma función o actividad biológica que los polipéptidos de la descripción. En otros aspectos, el fragmento, derivado, o análogo incluye una proproteína, de manera que el fragmento, derivado, o análogo se puede activar mediante escisión de la porción proproteínica para producir un polipéptido activo.

Optimización de codones para lograr niveles elevados de expresión proteica en células hospedantes

La descripción proporciona métodos para modificar ácidos nucleicos que codifican polipéptidos implicados en el catabolismo de la clorofila o que tienen actividad de esterasa (por ejemplo, clorofilasa) modificando el uso de codones. En un aspecto, la descripción proporciona métodos para modificar codones en un ácido nucleico que codifica un polipéptido para incrementar o disminuir su expresión en una célula hospedante. La descripción también proporciona ácidos nucleicos que codifican polipéptidos implicados en el catabolismo de la clorofila o que tienen actividad de esterasa (por ejemplo, clorofilasa) modificados para incrementar su expresión en una célula hospedante, enzimas así modificadas, y métodos para obtener los polipéptidos modificados implicados en el catabolismo de la clorofila o que tienen actividad de esterasa (por ejemplo, clorofilasa). El método comprende identificar un codón "no preferido" o un codón "menos preferido" en ácido nucleico que codifica la enzima, y sustituir uno o más de estos codones no preferidos o menos preferidos por un "codón preferido" que codifica el mismo aminoácido que el codón sustituido, y al menos un codón no preferido o menos preferido en el ácido nucleico se ha sustituido por un codón preferido que codifica el mismo aminoácido. Un codón preferido es un codón sobrerrepresentado en secuencias codificantes en genes en la célula hospedante, y un codón no preferido o menos preferido es un codón subrepresentado en secuencias codificantes en genes en la célula hospedante.

Las células hospedantes para expresar los ácidos nucleicos, casetes de expresión y vectores de la descripción incluyen bacterias, levaduras, hongos, células vegetales, células de insecto y células de mamífero. De este modo, la descripción proporciona métodos para optimizar el uso de codones en todas estas células, ácidos nucleicos alterados en los codones, y polipéptidos obtenidos mediante los ácidos nucleicos alterados en los codones. Las células hospedantes ejemplares incluyen bacterias gramnegativas, tales como *Escherichia coli*; bacterias grampositivas, tales como *Streptomyces*, *Lactobacillus gasseri*, *Lactococcus lactis*, *Lactococcus cremoris*, *Bacillus* sp., *Bacillus subtilis*, *Bacillus cereus*. Las células hospedantes ejemplares también incluyen organismos eucariotas, por ejemplo diversas levaduras, tales como *Saccharomyces* sp., que incluyen *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, *Pichia pastoris*, y *Kluyveromyces lactis*, *Hansenula polymorpha*, *Aspergillus niger*, y células y estirpes celulares de mamíferos y células y estirpes celulares de insectos. De este modo, la descripción también incluye ácidos nucleicos y polipéptidos optimizados para la expresión en estos organismos y especies, por ejemplo los ácidos nucleicos están optimizados en los codones para la expresión en una célula hospedante, por ejemplo, una *Pichia* sp., por ejemplo, *P. pastoris*, una *Saccharomyces* sp., o una *Bacillus* sp., una *Streptomyces* sp., y similares.

Por ejemplo, los codones de un ácido nucleico que codifica un polipéptido de la descripción o enzima similar aislada de una célula bacteriana se modifican de manera que el ácido nucleico (que codifica la enzima) se exprese óptimamente en una célula bacteriana diferente de la bacteria a partir de la cual derivó la enzima (por ejemplo, un polipéptido de la descripción), una levadura, un hongo, una célula vegetal, una célula de insecto o una célula de mamífero. Los métodos para optimizar codones son bien conocidos en la técnica; véanse, por ejemplo, la patente U.S. nº 5.795.737; Baca (2000) Int. J. Parasitol. 30:113-118; Hale (1998) Protein Expr. Purif. 12:185-188; Narum (2001) Infect. Immun. 69:7250-7253. Véase también Narum (2001) Infect. Immun. 69:7250-7253, que describe la optimización de codones en sistemas de ratón; Outchkourov (2002) Protein Expr. Purif. 24:18-24, que describe la optimización de codones en levadura; Feng (2000) Biochemistry 39:15399-15409, que describe la optimización de codones en *E. coli*; Humphreys (2000) Protein Expr. Purif. 20:252-264, que describe la optimización del uso de codones que afecta a la secreción en *E. coli*; Gao (2004) Biotechnol Prog. 20:443-448, que describe "UpGene", una aplicación de un algoritmo de optimización del uso de codones de ADN a base de web.

Animales no humanos transgénicos

La descripción proporciona animales no humanos transgénicos que comprenden un ácido nucleico, un polipéptido, un casete de expresión o vector o una célula transfectada o transformada de la descripción. La descripción también proporciona métodos para obtener y usar estos animales no humanos transgénicos.

Los animales no humanos transgénicos pueden ser, por ejemplo, cabras, conejos, ovejas, cerdos, vacas, ratas y ratones, que comprenden los ácidos nucleicos de la descripción. Estos animales se pueden usar, por ejemplo, como modelos *in vivo* para estudiar la actividad de la enzima, o como modelos para cribar agentes que cambian la actividad enzimática *in vivo*. Las secuencias codificantes para los polipéptidos a expresar en los animales no humanos transgénicos se pueden diseñar para que sean constitutivas, o estén bajo el control de factores reguladores transcripcionales específicos de tejido, específicos del desarrollo, o inducibles. Los animales no humanos transgénicos se pueden diseñar y generar usando cualquier método conocido en la técnica; véanse, por ejemplo, las patentes U.S. nºs 6.211.428; 6.187.992; 6.156.952; 6.118.044; 6.111.166; 6.107.541; 5.959.171; 5.922.854; 5.892.070; 5.880.327; 5.891.698; 5.639.940; 5.573.933; 5.387.742; 5.087.571, que describen la obtención y uso de células transformadas y huevos y ratones transgénicos, ratas, conejos, ovejas, cerdos y vacas. Véase también, por ejemplo, Pollock (1999) J. Immunol. Methods 231:147-157, que describe la producción de proteínas recombinantes en la leche de animales productores de leche transgénicos; Baguisi (1999) Nat. Biotechnol. 17:456-461, que demuestra la producción de cabras transgénicas. La patente U.S. nº 6.211.428 describe la obtención y uso de mamíferos no humanos transgénicos que expresan en sus cerebros un constructo de ácido nucleico que comprende una secuencia de ADN. La patente U.S. nº 5.387.742 describe la inyección de secuencias de ADN recombinantes o sintéticas clonadas en óvulos de ratón fertilizados, implantando los óvulos inyectados en hembras falsamente preñadas, y haciéndolos crecer hasta ratones transgénicos cuyas células expresan proteínas relacionadas con la patología de la enfermedad de Alzheimer. La patente U.S. nº 6.187.992 describe la obtención y uso de un ratón transgénico cuyo genoma comprende una interrupción del gen que codifica la proteína precursora de amiloide (APP).

"Animales a los que se les ha suprimido un gen" también se puede usar para la práctica de los métodos de la descripción. Por ejemplo, en un aspecto, los animales transgénicos o modificados de la descripción comprenden un "animal genosuprimido", por ejemplo un "ratón genosuprimido", manipulado para que no exprese un gen endógeno, que se sustituye por un gen que expresa una enzima de la descripción, o una proteína de fusión que comprende una enzima de la descripción.

Plantas y semillas transgénicas

La descripción proporciona plantas y semillas transgénicas que comprenden un ácido nucleico, un polipéptido (por ejemplo, un polipéptido implicado en el catabolismo de la clorofila o que tiene una actividad de esterasa (por ejemplo, clorofilasa)), un casete o vector de expresión, o una célula transfectada o transformada de la descripción. La descripción también proporciona productos vegetales, por ejemplo aceites, semillas, hojas, extractos y similares,

que comprenden un ácido nucleico y/o un polipéptido de la descripción. La planta transgénica puede ser dicotiledónea (una dicot) o monocotiledónea (una monocot). La descripción también proporciona métodos para obtener y usar estas plantas y semillas transgénicas. La planta o célula vegetal transgénica que expresa un polipéptido de la descripción se puede construir según cualquier método conocido en la técnica. Véase, por ejemplo, la patente U.S. nº 6.309.872.

Los ácidos nucleicos y constructos de expresión de la descripción se pueden introducir en una célula vegetal por cualquier medio. Por ejemplo, los ácidos nucleicos o constructos de expresión se pueden introducir en el genoma de un hospedante vegetal deseado, o los ácidos nucleicos o constructos de expresión pueden ser episomas. La introducción en el genoma de una planta deseada puede ser tal que los elementos de control transcripcionales y/o traduccionales endógenos del hospedante regulen la actividad del ácido nucleico introducido, ya sea que esté integrado o sea episómico. La descripción también proporciona "plantas genosuprimidas" en las que la inserción de una secuencia génica mediante, por ejemplo, recombinación homóloga ha interrumpido la expresión del gen endógeno. Los medios para generar plantas "genosuprimidas" son bien conocidos en la técnica; véanse, por ejemplo, Strepp (1998) *Proc Natl. Acad. Sci. USA* 95:4368-4373; Miao (1995) *Plant J* 7:359-365. Véase la discusión sobre plantas transgénicas, más abajo.

Los ácidos nucleicos de la descripción se pueden usar para conferir rasgos deseados en esencialmente cualquier planta, por ejemplo plantas que producen almidón, tales como patata, trigo, arroz, cebada, y similares. Los ácidos nucleicos de la descripción se pueden usar para manipular rutas metabólicas de una planta a fin de optimizar o alterar la expresión en el hospedante de un polipéptido de la descripción, o una enzima homóloga en el hospedante. Esto puede cambiar la actividad de la enzima (por ejemplo, clorofilasa) o el producto de la ruta biosintética (una ruta degradativa de la clorofila) en la planta. Como alternativa, una enzima o ácido nucleico de la descripción se puede usar en la producción de una planta transgénica para producir un compuesto no producido naturalmente por esa planta. Esto puede reducir los costes de producción o crear un nuevo producto.

En un aspecto, la primera etapa en la producción de una planta transgénica implica obtener un constructo de expresión para la expresión en una célula vegetal. Estas técnicas son bien conocidas en la técnica. Pueden incluir seleccionar y clonar un promotor, una secuencia codificante para facilitar la unión eficiente de ribosomas a ARNm, y seleccionar las secuencias terminadoras génicas apropiadas. Un promotor constitutivo ejemplar es CaMV35S, procedente del virus del mosaico de la coliflor, que generalmente da como resultado un grado elevado de expresión en plantas. Otros promotores son más específicos y responden a señales en el entorno interno o externo de la planta. Un promotor ejemplar inducible por la luz es el promotor del gen *cab*, que codifica la proteína de unión a clorofila *a/b* mayor.

En un aspecto, el ácido nucleico se modifica para lograr una mayor expresión en una célula vegetal. Por ejemplo, es probable que una secuencia de la descripción tenga un mayor porcentaje de pares nucleotídicos A-T en comparación con el observado en una planta, algunas de los cuales prefieren pares nucleotídicos G-C. Por lo tanto, los nucleótidos A-T en la secuencia codificante se pueden sustituir por nucleótidos G-C sin cambiar significativamente la secuencia de aminoácidos para potenciar la producción del producto génico en células vegetales.

Se puede añadir un gen marcador seleccionable al constructo génico a fin de identificar células o tejidos vegetales que han integrado con éxito el transgén. Esto puede ser necesario debido a que el logro de la incorporación y expresión de genes en células vegetales es un suceso raro, que se produce en sólo un pequeño porcentaje de los tejidos o células seleccionados como dianas. Los genes marcadores seleccionables codifican proteínas que proporcionan resistencia a agentes que son normalmente tóxicos para las plantas, tales como antibióticos o herbicidas. Sólo las células vegetales que tienen integrado el gen marcador seleccionable sobrevivirán cuando se hagan crecer en un medio que contiene el antibiótico o herbicida apropiado. En cuanto a otros genes insertados, los genes marcadores también requieren secuencias promotoras y de terminación para el funcionamiento apropiado.

En un aspecto, la obtención de plantas o semillas transgénicas comprende incorporar secuencias de la descripción y, opcionalmente, genes marcadores en un constructo de expresión diana (por ejemplo, un plásmido), junto con el posicionamiento de las secuencias promotoras y terminadoras. Esto puede implicar transferir el gen modificado a la planta a través de un método adecuado. Por ejemplo, se puede introducir directamente un constructo en el ADN genómico de la célula vegetal usando técnicas tales como electroporación y microinyección de protoplastos de células vegetales, o los constructos se pueden introducir directamente al tejido vegetal usando métodos balísticos, tal como bombardeo de partículas de ADN. Por ejemplo, véase, por ejemplo, Christou (1997) *Plant Mol. Biol.* 35:197-203; Pawlowski (1996) *Mol. Biotechnol.* 6:17-30; Klein (1987) *Nature* 327:70-73; Takumi (1997) *Genes Genet. Syst.* 72:63-69, que explican el uso del bombardeo de partículas para introducir transgenes en trigo; y Adam (1997) más arriba, para el uso de bombardeo de partículas para introducir YACs en células vegetales. Por ejemplo, Rinehart (1997), más arriba, usó bombardeo de partículas para generar plantas de algodón transgénicas. El aparato para acelerar partículas se describe en la patente U.S. nº 5.015.580; y el instrumento de aceleración de partículas comercialmente disponible de BioRad (Biolistics) PDS-2000; véase también John, patente U.S. nº 5.608.148; y Ellis, patente U.S. nº 5.681.730, que describen la transformación de gimnospermas mediada por partículas.

En un aspecto, los protoplastos se pueden inmovilizar e inyectar con ácidos nucleicos, por ejemplo un constructo de expresión. Aunque la regeneración vegetal a partir de protoplastos no es fácil con cereales, la regeneración vegetal es posible en legumbres usando embriogénesis somática a partir de callo derivado de protoplasto. Los tejidos organizados se pueden transformar con ADN desnudo usando técnica de pistola génica, en la que ADN se reviste sobre microproyectiles de wolframio, disparados 1/100 del tamaño de las células, que poseen el ADN profundamente en las células y orgánulos. El tejido transformado es inducido entonces a regenerarse, habitualmente mediante embriogénesis somática. Esta técnica ha tenido éxito en varias especies de cereales, incluyendo maíz y arroz.

Los ácidos nucleicos, por ejemplo constructos de expresión, también se pueden introducir en células vegetales usando virus recombinantes. Las células vegetales se pueden transformar usando vectores víricos, tales como, por ejemplo, vectores derivados del virus del mosaico del tabaco (Rouwendal (1997) *Plant Mol. Biol.* 33:989-999), véase Porta (1996) "Use of viral replicons for the expression of genes in plants", *Mol. Biotechnol.* 5:209-221.

Como alternativa, los ácidos nucleicos, por ejemplo un constructo de expresión, se puede combinar con regiones de flanco de T-DNA adecuadas y se pueden introducir en un vector hospedante de *Agrobacterium tumefaciens* convencional. Las funciones de virulencia del hospedante de *Agrobacterium tumefaciens* dirigirán la inserción del constructo y marcador adyacente en el ADN de la célula vegetal cuando la célula es infectada por la bacteria. Las técnicas de transformación mediadas por *Agrobacterium tumefaciens*, incluyendo el desarme y uso de vectores binarios, están bien descritas en la bibliografía científica. Véase, por ejemplo, Horsch (1984) *Science* 233:496-498; Fraley (1983) *Proc. Natl. Acad. Sci. USA* 80:4803 (1983); *Gene Transfer to Plants*, Potrykus, ed. (Springer-Verlag, Berlín 1995). El ADN en una célula de *A. tumefaciens* está contenido en el cromosoma bacteriano así como en otra estructura conocida como plásmido Ti (inductor de tumores). El plásmido Ti contiene un tramo de ADN denominado T-DNA (~20 kb de longitud) que es transferido a la célula vegetal en el proceso de infección y una serie de genes vir (virulencia) que dirigen el proceso de infección. *A. tumefaciens* puede infectar solamente a una planta a través de heridas: cuando una raíz o tallo vegetal es herido, da ciertas señales químicas, a cuya respuesta los genes vir de *A. tumefaciens* son activados y dirigen una serie de sucesos necesarios para la transferencia del T-DNA desde el plásmido Ti al cromosoma de la planta. El T-DNA entra entonces en la célula vegetal a través de la herida. Una suposición es que el T-DNA espera hasta que el ADN vegetal está siendo replicado o transcrito, y después se inserta él mismo en el ADN vegetal expuesto. A fin de usar *A. tumefaciens* como vector transgénico, se ha de eliminar la sección inductora de tumores de T-DNA, mientras que se retienen las regiones frontera de T-DNA y los genes vir. El transgén se inserta entonces entre las regiones frontera de T-DNA, donde es transferido a la célula vegetal y se integra en los cromosomas de la planta.

La descripción proporciona la transformación de plantas monocotiledóneas usando los ácidos nucleicos de la descripción, incluyendo cereales importantes; véase Hiei (1997) *Plant Mol. Biol.* 35:205-218. Véanse también, por ejemplo, Horsch, *Science* (1984) 233:496; Fraley (1983) *Proc. Natl. Acad. Sci. USA* 80:4803; Thykjaer (1997) más arriba; Park (1996) *Plant Mol. Biol.* 32:1135-1148, que explican la integración de T-DNA en ADN genómico. Véase también D'Halluin, patente U.S. nº 5.712.135, que describe un procedimiento para la integración estable de un ADN que comprende un gen que es funcional en una célula de un cereal, u otra planta monocotiledónea.

En un aspecto, la tercera etapa puede implicar la selección y regeneración de plantas completas capaces de transmitir a la siguiente generación el gen diana incorporado. Tales técnicas de regeneración se basan en la manipulación de ciertas fitohormonas en un medio de crecimiento de cultivo tisular, que se basa típicamente en un marcador biocida y/o herbicida que se ha introducido junto con las secuencias nucleotídicas deseadas. La regeneración vegetal a partir de los protoplastos cultivados se describe en Evans et al., *Protoplast Isolation and Culture*, Handbook of Plant Cell Culture, p. 124-176, MacMillan Publishing Company, Nueva York, 1983; y Binding, *Regeneration of Plants, Plant Protoplasts*, p. 21-73, CRC Press, Boca Raton, 1985. La regeneración también se puede obtener a partir de callo vegetal, explantes, órganos, o sus partes. Tales técnicas de regeneración se describen generalmente en Klee (1987) *Ann. Rev. of Plant Phys.* 38:467-486. Para obtener plantas completas a partir de tejidos transgénicos, tales como embriones inmaduros, se pueden hacer crecer en condiciones ambientales controladas en una serie de medios que contienen nutrientes y hormonas, un proceso conocido como cultivo de tejidos. Una vez que se generan plantas completas y se producen semillas, comienza la evaluación de la progenie.

Después de que el casete de expresión se incorpora de forma estable en plantas transgénicas, se puede introducir en otras plantas mediante cruzamiento sexual. Se puede usar cualquiera de un número de técnicas de reproducción estándar, dependiendo de la especie a cruzar. Puesto que la expresión transgénica de los ácidos nucleicos de la descripción conduce a cambios fenotípicos, las plantas que comprenden los ácidos nucleicos recombinantes de la descripción se pueden cruzar sexualmente con una segunda planta para obtener un producto final. De este modo, la semilla de la descripción se puede derivar de un cruce entre dos plantas transgénicas de la descripción, o un cruce entre una planta de la descripción y otra planta. Los efectos deseados (por ejemplo, expresión de los polipéptidos de la descripción para producir una planta en la que está alterado el comportamiento de la floración o de la siembra) se pueden potenciar cuando ambas plantas parentales expresan los polipéptidos de la descripción. Los efectos deseados se pueden pasar a las futuras generaciones de plantas por medios de propagación estándar.

Los ácidos nucleicos y polipéptidos de la descripción se expresan o insertan en cualquier planta o semilla. Las plantas transgénicas de la descripción pueden ser dicotiledóneas o monocotiledóneas. Los ejemplos de plantas transgénicas monocotiledóneas de la descripción son hierbas tales como poa común (hierba verde, *Poa*), hierba

forrajera tal como festuca, lolio, hierba templada, tal como *Agrostis*, y cereales, por ejemplo trigo, avena, centeno, cebada, arroz, sorgo, y maíz (maíz). Los ejemplos de plantas transgénicas dicotiledóneas de la descripción son tabaco, legumbres, tales como lupinos, patata, remolacha, guisante, haba y haba de soja, y plantas crucíferas (familia *Brassicaceae*), tal como coliflor, colza, y el organismo modelo estrechamente relacionado *Arabidopsis thaliana*. De este modo, las plantas y semillas transgénicas de la descripción incluyen un amplio intervalo de plantas, incluyendo, pero sin limitarse a, especies de los géneros *Anacardium*, *Arachis*, *Asparagus*, *Atropa*, *Avena*, *Brassica*, *Citrus*, *Citrullus*, *Capsicum*, *Carthamus*, *Cocos*, *Coffea*, *Cucumis*, *Cucurbita*, *Daucus*, *Elaeis*, *Fragaria*, *Glycine*, *Gossypium*, *Helianthus*, *Heterocallis*, *Hordeum*, *Hyoscyamus*, *Lactuca*, *Linum*, *Lolium*, *Lupinus*, *Lycopersicon*, *Malus*, *Manihot*, *Majorana*, *Medicago*, *Nicotiana*, *Olea*, *Oryza*, *Panicum*, *Pennisetum*, *Persea*, *Phaseolus*, *Pistachia*, *Pisum*, *Pyrus*, *Prunus*, *Raphanus*, *Ricinus*, *Secale*, *Senecio*, *Sinapis*, *Solanum*, *Sorghum*, *Theobromus*, *Trigonella*, *Triticum*, *Vicia*, *Vitis*, *Vigna*, y *Zea*.

En realizaciones alternativas, los ácidos nucleicos de la descripción se expresan en plantas que contienen células de fibras, incluyendo, por ejemplo, algodón, árbol de algodón de seda (Kapok, *Ceiba pentandra*), sauce del desierto, árbol de creosota, *Krascheninnikovia lanata*, balsa, ramio, kenaf, cáñamo, roda de Jamaica, yute, sisal, abacá y lino. En realizaciones alternativas, las plantas transgénicas de la descripción pueden ser miembros del género *Gossypium*, incluyendo miembros de cualquier especie de *Gossypium*, tales como *G. arboreum*, *G. herbaceum*, *G. barbadense*, y *G. hirsutum*.

La descripción también proporciona plantas transgénicas a usar para producir grandes cantidades de los polipéptidos (por ejemplo, enzimas o anticuerpo) de la descripción. Por ejemplo, véanse Palmgren (1997) Trends Genet. 13:348; Chong (1997) Transgenic Res. 6:289-296 (que producen la proteína de leche humana beta-caseína en plantas de patata transgénicas usando un promotor de manopina sintasa (*mas1'*,2') bidireccional, inducible con auxina, con métodos de transformación de discos de hoja mediada por *Agrobacterium tumefaciens*).

Usando procedimientos conocidos, el experto puede cribar plantas de la descripción para detectar el incremento o disminución de ARNm o proteína transgénica en plantas transgénicas. Los medios para detectar y cuantificar ARNm o proteínas son bien conocidos en la técnica.

Polipéptidos y péptidos

En un aspecto, la descripción proporciona polipéptidos aislados, sintéticos o recombinantes que tienen una identidad de secuencia (por ejemplo, al menos 50%, 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, 60%, 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99,5%, o más, o completa (100%) de identidad de secuencia) con una secuencia ejemplar de la descripción, por ejemplo proteínas que tiene una secuencia como se expone en SEC ID NO:2, SEC ID NO:4, SEC ID NO:6, SEC ID NO:8, SEC ID NO:10, SEC ID NO:12, SEC ID NO:14, SEC ID NO:16, SEC ID NO:18 o SEC ID NO:20.

En un aspecto, un polipéptido de la descripción tiene una actividad de esterasa, tal como una actividad de clorofilasa (una clase), o tiene una actividad enzimática que comprende modificación enzimática de una molécula de clorofila, por ejemplo, en el que la modificación enzimática comprende el catabolismo de la molécula de la clorofila. En un aspecto, la actividad de esterasa comprende una actividad de clorofila clorofilido-hidrolasa.

Otro aspecto de la descripción proporciona un polipéptido o péptido aislado, sintético o recombinante que incluye al menos 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95 ó 100 o más bases consecutivas de una secuencia polipeptídica o peptídica de la descripción, secuencias sustancialmente idénticas a ellas, y las secuencias complementarias a ellas. El péptido puede ser, por ejemplo, un fragmento inmunógeno, un motivo (por ejemplo, un sitio de unión), una secuencia señal, una secuencia prepro o dominios catalíticos (CDs) o un sitio activo.

La descripción también proporciona polipéptidos quiméricos (y los ácidos nucleicos que los codifican) que comprenden al menos dos enzimas de la descripción o sus subsecuencias, por ejemplo sitios activos, o dominios catalíticos (CDs). Una proteína quimérica de la descripción (por ejemplo, una proteína de fusión, u otro heterodímero, por ejemplo dos dominios unidos por otros medios, por ejemplo un enlazador, o electrostáticamente) puede comprender un polipéptido (por ejemplo, péptido de sitio activo o de dominio catalítico) de la descripción y otro polipéptido (por ejemplo, péptido de sitio activo o de dominio catalítico) de la descripción u otro polipéptido. Por ejemplo, una proteína quimérica de la descripción puede tener cualquier actividad de un polipéptido implicado en el catabolismo de la clorofila o que tiene una actividad de esterasa (por ejemplo, clorofilasa), por ejemplo, como se describe aquí. En un aspecto, la proteína quimérica de la descripción comprende una fusión de dominios, por ejemplo un dominio único puede mostrar una o cualquier combinación de actividades.

Los polipéptidos de la descripción incluyen enzimas en una forma activa o inactiva. Por ejemplo, los polipéptidos de la descripción incluyen proproteínas antes de la "maduración" o procesamiento de las secuencias prepro, por ejemplo por medio de una enzima de procesamiento de proproteínas, tal como una proproteína convertasa, para generar una proteína mutante "activa". Los polipéptidos de la descripción incluyen enzimas inactivas por otras razones, por ejemplo, antes de la "activación" mediante un evento de procesamiento post-traduccional, por ejemplo una acción de endo- o exo-peptidasa o proteinasa, un suceso de fosforilación, una amidación, una glicosilación o

una sulfuración, un suceso de dimerización, y similares. Los polipéptidos de la descripción incluyen todas las formas activas, incluyendo subsecuencias activas, por ejemplo dominios catalíticos o sitios activos, de las enzimas.

Los métodos para identificar secuencias con dominio "prepro" y secuencias señal son bien conocidos en la técnica; véase, por ejemplo, Van de Ven (1993) Crit. Rev. Oncog. 4 (2):115-136. Por ejemplo, para identificar una secuencia prepro, la proteína se purifica a partir del espacio extracelular y se determina la secuencia N-terminal de las proteínas y se compara con la forma no procesada.

La descripción incluye polipéptidos con o sin una secuencia señal y/o una secuencia prepro. La descripción incluye polipéptidos con secuencias señal y/o secuencias prepro heterólogas. La secuencia prepro (incluyendo una secuencia de la descripción usada como dominio prepro heterólogo) se puede localizar sobre el extremo amino terminal o carboxi terminal de la proteína. La descripción también incluye secuencias señal aisladas o recombinantes, secuencias prepro y dominios catalíticos (por ejemplo, "sitios activos") que comprenden secuencias de la descripción.

El porcentaje de identidad de secuencia se puede encontrar a lo largo de toda la longitud del polipéptido, o la identidad se puede encontrar a lo largo de una región de al menos alrededor de 50, 60, 70, 80, 90, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700 o más restos. Los polipéptidos de la descripción también pueden ser más cortos que la longitud completa de los polipéptidos ejemplares. En aspectos alternativos, se proporcionan aquí polipéptidos (péptidos, fragmentos) que tienen un tamaño que oscila entre alrededor de 5 y la longitud total de un polipéptido, por ejemplo una enzima de la descripción; siendo los tamaños ejemplares de alrededor de 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 100, 125, 150, 175, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, o más restos, por ejemplo restos contiguos de una enzima ejemplar de la descripción.

Los péptidos de la descripción (por ejemplo, una subsecuencia de un polipéptido ejemplar de la descripción) pueden ser útiles, por ejemplo, como sondas de marcaje, antígenos, tolerágenos, motivos, sitios activos de enzimas (por ejemplo, "dominios catalíticos" de enzimas de la descripción), sitios de unión de enzimas de la descripción, secuencias señal y/o dominios prepro.

Los polipéptidos y péptidos de la descripción se pueden aislar a partir de fuentes naturales, pueden ser sintéticos, o pueden ser polipéptidos generados recombinantemente. Los péptidos y proteínas se pueden expresar recombinantemente *in vitro* o *in vivo*. Los péptidos y polipéptidos de la descripción se pueden obtener y aislar usando cualquier método conocido en la técnica. Los polipéptidos y péptidos de la descripción se pueden sintetizar también, completamente o en parte, usando métodos químicos bien conocidos en la técnica. Véanse, por ejemplo, Caruthers (1980) Nucleic Acids Res. Symp. Ser. 215-223; Horn (1980) Nucleic Acids Res. Symp. Ser. 225-232; Banga, A.K., Therapeutic Peptides and Proteins, Formulation, Processing and Delivery Systems (1995) Technomic Publishing Co., Lancaster, PA. Por ejemplo, la síntesis peptídica se puede llevar a cabo usando diversas técnicas en fase sólida (véanse, por ejemplo, Roberge (1995) Science 269:202; Merrifield (1997) Methods Enzymol. 289:3-13) y la síntesis automática se puede lograr, por ejemplo, usando el Sintetizador de Péptidos ABI 431A (Perkin Elmer) de acuerdo con las instrucciones proporcionadas por el fabricante.

Los péptidos y polipéptidos de la descripción también se pueden glucosilar. La glicosilación se puede añadir post-traduccionally ya sea químicamente o mediante mecanismos biosintéticos celulares, en los que estos últimos incorporan el uso de motivos de glicosilación conocidos, que pueden ser nativos para la secuencia o se pueden añadir en forma de un péptido o añadir en la secuencia de ácido nucleico codificante. La glicosilación puede estar enlazada vía O o enlazada vía N.

Los péptidos y polipéptidos de la descripción, como se definen anteriormente, incluyen todas las formas "miméticas" y "peptidomiméticas". Los términos "mimético" y "peptidomimético" se refieren a un compuesto químico sintético que tiene sustancialmente las mismas características estructurales y/o funcionales de los polipéptidos de la descripción. El mimético puede estar compuesto completamente de análogos sintéticos no naturales de aminoácidos, o es una molécula quimérica de aminoácidos peptídicos parcialmente naturales y análogos parcialmente no naturales de aminoácidos. El mimético también puede incorporar cualquier cantidad de sustituciones conservativas de aminoácidos naturales, en tanto que tales sustituciones tampoco alteren sustancialmente la estructura y/o actividad del mimético. En cuanto a los polipéptidos de la descripción que son variantes conservativas, la experimentación habitual determinará si un mimético está dentro del alcance de la descripción, es decir, que su estructura y/o función no esté sustancialmente alterada a partir de un polipéptido ejemplar de la descripción. En un aspecto, una composición mimética se usa en una composición, sistema celular o procedimiento de la descripción (por ejemplo, una célula hospedante que tiene un plásmido que expresa al menos una enzima de la descripción).

Las composiciones miméticas polipeptídicas de la descripción pueden contener cualquier combinación de componentes estructurales no naturales. En aspecto alternativo, las composiciones miméticas de la descripción incluyen uno o todos los siguientes tres grupos estructurales: a) grupos de enlace de restos, distintos de los enlaces de amida naturales ("enlace peptídico"); b) restos no naturales en lugar de restos de aminoácidos de origen natural; o c) restos que inducen un mimetismo estructural secundario, es decir, inducen o estabilizan una estructura secundaria, por ejemplo una vuelta beta, vuelta gamma, lámina beta, conformación de hélice alfa, y similar. Por ejemplo, un polipéptido de la descripción se puede caracterizar como un mimético cuando todos o algunos de sus

restos están unidos por medios químicos distintos de enlaces peptídicos naturales. Los restos peptidomiméticos individuales se pueden unir mediante enlaces peptídicos, otros enlaces químicos o medios de acoplamiento, tales como, por ejemplo, glutaraldehído, ésteres de N-hidroxisuccinimida, maleimidas bifuncionales, N,N'-díciclohexilcarbodiimida (DCC) o N,N'-diisopropilcarbodiimida (DIC). Los grupos enlazantes que pueden ser una alternativa a los enlazamientos con enlaces de amida tradicionales ("enlace peptídico") incluyen, por ejemplo, cetometileno (por ejemplo, $-C(=O)-CH_2-$ para $-C(=O)-NH-$), aminometileno (CH_2-NH), etileno, olefina ($CH=CH$), éter (CH_2-O), tioéter (CH_2-S), tetrazol (CN_4-), tiazol, retroamida, tioamida, o éster (véase, por ejemplo, Spatola (1983) en *Chemistry and Biochemistry of Amino Acids, Peptides and Proteins*, Vol. 7, p. 267-357, "Peptide Backbone Modifications," Marcell Dekker, NY).

Un polipéptido de la descripción también se puede caracterizar como un mimético al contener todos o algunos restos no naturales en lugar de restos de aminoácidos de origen natural. Los restos no naturales están bien descritos en la bibliografía científica y de patentes; más abajo se describen unas pocas composiciones no naturales ejemplares útiles como miméticos de restos de aminoácidos naturales y directrices. Los miméticos de aminoácidos aromáticos se pueden generar sustituyendo, por ejemplo, D- o L-naftilalanina; D- o L-fenilglicina; D- o L-2-tienilalanina; D- o L-1, -2, 3-, o 4-pirenilalanina; D- o L-3-tienilalanina; D- o L-(2-piridinil)-alanina; D- o L-(3-piridinil)-alanina; D- o L-(2-pirazinil)-alanina; D- o L-(4-isopropil)-fenilglicina; D-(trifluorometil)-fenilglicina; D-(trifluorometil)-fenilalanina; D-p-fluoro-fenilalanina; D o L-p-bifenilfenilalanina; D- o L-p-metoxi-bifenilfenilalanina; D- o L-2-indol(alquil)alaninas; y, D- o L-alquilaninas, en los que alquilo puede ser metilo, etilo, propilo, hexilo, butilo, pentilo, isopropilo, isobutilo, sec-isotilo, iso-pentilo, sustituidos o no sustituidos, o aminoácidos no ácidos. Los anillos aromáticos de un aminoácido no natural incluyen, por ejemplo, anillos aromáticos de tiazolilo, tiofenilo, pirazolilo, bencimidazolilo, naftilo, furanilo, pirrolilo, y piridilo.

Los miméticos de aminoácidos ácidos se pueden generar por sustitución mediante, por ejemplo, aminoácidos no carboxilados a la vez que se mantiene una carga negativa; (fosfono)alanina; treonina sulfatada. Los grupos laterales carboxílicos (por ejemplo, aspartilo o glutamilo) también se pueden modificar selectivamente mediante reacción con carbodiimidas ($R^1-N-C-N-R^1$) tales como, por ejemplo, 1-ciclohexil-3(2-morfolinil-(4-etil) carbodiimida o 1-etil-3(4-azonia-4,4-dimetolpentil) carbodiimida. Aspartilo o glutamilo también se puede convertir en restos asparaginilo y glutaminilo mediante reacción con iones amonio. Los miméticos de aminoácidos básicos se pueden generar mediante sustitución con, por ejemplo, (además de lisina y arginina) los aminoácidos ornitina, citrulina, o ácido (guanidino)-acético, o ácido (guanidino)alquil-acético, en los que alquilo es como se define anteriormente. Los derivados de nitrilo (por ejemplo, que contienen el resto $CN-$ en lugar de $COOH$) se pueden sustituir por asparagina o glutamina. Los restos asparaginilo y glutaminilo se pueden desaminar a los restos aspartilo o glutamilo correspondientes. Los miméticos de restos de arginina se pueden generar haciendo reaccionar arginilo con, por ejemplo, uno o más reactivos convencionales, incluyendo, por ejemplo, fenilgloxal, 2,3-butanodiona, 1,2-ciclohexanodiona, o ninhidrina, en un aspecto en condiciones alcalinas. Los miméticos de restos de tirosina se pueden generar haciendo reaccionar tirosilo con, por ejemplo, compuestos de diazonio aromáticos o tetranitrometano. Se puede usar N-acetilimidazol y tetranitrometano para formar especies O-acetilrosilicas y derivados 3-nitro, respectivamente. Los miméticos de restos de cisteína se pueden generar haciendo reaccionar restos de cisteinilo con, por ejemplo, alfa-haloacetatos tales como ácido 2-cloroacético o cloracetamida y aminas correspondientes, para dar derivados carboximetílicos o carboxiamidometílicos. Los miméticos de restos de cisteína también se pueden generar haciendo reaccionar restos cisteinílicos con, por ejemplo, bromo-trifluoroacetona, ácido alfa-bromo-beta-(5-imidozilo)propiónico; fosfato de cloroacetilo, N-alquilmaleimidas, disulfuro de 3-nitro-2-piridilo; disulfuro de metil 2-piridilo; p-cloromercuribenzoato; 2-cloromercuri-4 nitrofenol; o, cloro-7-nitrobenzo-oxa-1,3-diazol. Los miméticos de lisina se pueden generar (y los restos aminotermiales se pueden alterar) haciendo reaccionar lisinilo con, por ejemplo, anhídrido succínico u otros anhídridos de ácidos carboxílicos. Los miméticos de restos de lisina y que contienen otros alfa-amino también se pueden generar mediante reacción con imidoésteres, tales como picolinimato de metilo, fosfato de piridoxal, piridoxal, cloroborohidruro, ácido trinitrobenenosulfónico, O-metilisourea, 2,4-pentanodiona y reacciones catalizadas por transamidasa con glicoxilato. Los miméticos de metionina se pueden generar mediante reacción con, por ejemplo, sulfóxido de metionina. Los miméticos de prolina incluyen, por ejemplo, ácido pipercolico, ácido tiazolidincarboxílico, 3- o 4-hidroxiprolina, deshidroprolina, 3- o 4-metilprolina o 3,3-dimetilprolina. Los miméticos de restos de histidina se pueden generar haciendo reaccionar histidilo con, por ejemplo, procarbonato de dietilo o bromuro de para-bromofenacilo. Otros miméticos incluyen, por ejemplo, los generados mediante hidroxilación de prolina y lisina; la fosforilación de los grupos hidroxilo de restos serilo o treonilo; la metilación de los grupos alfa-amino de lisina, arginina e histidina; la acetilación de la amina N-terminal; la metilación de restos amídicos de cadena principal o la sustitución con aminoácidos N-metílicos, o la amidación de grupos carboxílicos C-terminales.

Un resto, por ejemplo un aminoácido, de un polipéptido de la descripción también se puede sustituir por un aminoácido (o resto peptidomimético) de la quiralidad opuesta. De este modo, cualquier aminoácido de origen natural en la configuración L (que también se puede denominar como R o S, dependiendo de la estructura de la entidad química) se puede sustituir por el aminoácido del mismo tipo estructural químico o un peptidomimético, pero de la quiralidad opuesta, denominado como el D-aminoácido, pero también se puede denominar como la forma R o S.

La descripción también proporciona métodos para modificar los polipéptidos de la descripción mediante procesos naturales, tales como procesamiento post-traducciona (por ejemplo, fosforilación, acilación, etc.), o mediante

técnicas de modificación química, y los polipéptidos modificados resultantes. Las modificaciones se pueden producir en cualquier parte en el polipéptido, incluyendo la cadena principal peptídica, las cadenas laterales de aminoácidos, y los términos amino o carboxilo. Se apreciará que el mismo tipo de modificación puede estar presente en los mismos grados o grados variables en varios sitios en un polipéptido dado. También, un polipéptido dado puede tener muchos tipos de modificaciones. Las modificaciones incluyen acetilación, acilación, ribosilación con ADP, amidación, unión covalente de flavina, unión covalente de resto hemo, unión covalente de un nucleótido o derivado nucleotídico, unión covalente de un lípido o derivado lipídico, unión covalente de un fosfatidilinositol, ciclación por reticulación, formación de enlace de disulfuro, desmetilación, formación de reticulaciones covalentes, formación de cisteína, formación de piroglutamato, formilación, gamma-carboxilación, glicosilación, formación de anclaje de GPI, hidroxilación, yodación, metilación, miristoilación, oxidación, pegilación, procesamiento proteolítico, fosforilación, prenilación, racemización, selenoilación, sulfatación, y adición de aminoácidos mediada por ARN de transferencia a proteína, tal como arginilación. Véanse, por ejemplo, Creighton, T.E., *Proteins - Structure and Molecular Properties* 2ª Ed., W.H. Freeman and Company, Nueva York (1993); *Posttranslational Covalent Modification of Proteins*, B.C. Johnson, Ed., Academic Press, Nueva York, p. 1-12 (1983).

También se pueden usar métodos de síntesis química de péptidos en fase sólida para sintetizar el polipéptido o fragmentos de la descripción. Tal método se conoce en la técnica desde principios de 1960 (Merrifield, R. B., *J. Am. Chem. Soc.*, 85:2149-2154, 1963) (véase también Stewart, J. M. y Young, J. D., *Solid Phase Peptide Synthesis*, 2ª Ed., Pierce Chemical Co., Rockford, Ill., p. 11-12)), y se ha empleado recientemente en kits de síntesis y diseño de péptidos de laboratorio comercialmente disponibles (Cambridge Research Biochemicals). Tales kits de laboratorio comercialmente disponibles han utilizado generalmente las enseñanzas de H. M. Geysen et al, *Proc. Natl. Acad. Sci., USA*, 81:3998 (1984), y proporcionan la síntesis de péptidos sobre las puntas de una multitud de "varillas" o "alfileres", todos los cuales están conectados a una única placa. Cuando se utiliza tal sistema, se invierte y se inserta una placa de varillas o alfileres en una segunda placa de pocillos o reservorios correspondientes, que contiene disoluciones para la unión o anclaje de un aminoácido apropiado a las puntas de los alfileres o varillas. Repitiendo tal etapa del proceso, es decir, invirtiendo e insertando las puntas de las varillas y alfileres en disoluciones apropiadas, se construyen aminoácidos en péptidos deseados. Además, hay disponible un número de sistemas de síntesis de péptidos FMOC disponibles. Por ejemplo, el ensamblaje de un polipéptido o fragmento se puede llevar a cabo sobre un soporte sólido usando un sintetizador de péptidos automatizado de Applied Biosystems, Inc. Modelo 431A™. Tal equipo proporciona un acceso fácil a los péptidos de la descripción, ya sea mediante síntesis directa o mediante síntesis de una serie de fragmentos que se pueden acoplar usando otras técnicas conocidas.

La descripción incluye polipéptidos de la descripción con y sin señal. El polipéptido que comprende una secuencia señal de la descripción puede ser un polipéptido de la descripción u otro polipéptido.

La descripción incluye polipéptidos inmovilizados de la descripción, incluyendo enzimas, anticuerpos y fragmentos de los mismos. La descripción proporciona métodos para inhibir la actividad polipeptídica, por ejemplo usando mutantes negativos dominantes o anticuerpos de la descripción. La descripción incluye heterocomplejos, por ejemplo proteínas de fusión, heterodímeros, etc., que comprenden las enzimas de la descripción.

Los polipéptidos de la descripción pueden tener actividad enzimática en diversas condiciones, por ejemplo extremas en pH y/o temperatura, agentes oxidantes, y similares. La descripción proporciona métodos que conducen a preparaciones de enzima alternativas con diferentes eficacias y estabilidades catalíticas, por ejemplo frente a temperatura, agentes oxidantes y condiciones de lavado variables. En un aspecto, las variantes enzimáticas se pueden producir utilizando técnicas de mutagénesis dirigida al sitio y/o mutagénesis al azar. En un aspecto, se puede usar la evolución dirigida para producir una gran variedad de variantes enzimáticas con especificidades y estabilidad alternativa.

Las proteínas de la descripción también son útiles como reactivos de investigación para identificar moduladores enzimáticos, por ejemplo activadores o inhibidores de la actividad enzimática. En resumen, se añaden muestras de ensayo (compuestos, caldos, extractos, y similares) a ensayos enzimáticos para determinar su capacidad para inhibir la escisión del sustrato. Los inhibidores identificados de esta manera se pueden usar en industria e investigación para reducir o evitar la proteólisis indeseada. Los inhibidores enzimáticos se pueden combinar para incrementar el espectro de actividad.

La descripción también proporciona métodos para descubrir nuevas enzimas que tienen actividad similar a una enzima de la descripción usando los ácidos nucleicos, polipéptidos y anticuerpos de la descripción. En un aspecto, las bibliotecas de fagómidos se criban para el descubrimiento de una nueva enzima a base de expresión. El cribado de las bibliotecas de fagos o fagómidos puede permitir la detección de clones tóxicos, el acceso mejorado al sustrato, una menor necesidad de manipulación mediante ingeniería de un hospedante, evitando el potencial de cualquier desplazamiento resultante de la excisión másica de la biblioteca, y un crecimiento más rápido a densidades bajas de clones. El cribado de las bibliotecas de fagos o fagómidos puede hacerse en fase líquida o en fase sólida. En un aspecto, la descripción proporciona el cribado en fase líquida. Esto da una mayor flexibilidad en las condiciones del ensayo, una flexibilidad adicional de sustratos, una mayor sensibilidad para clones débiles, y una facilidad de automatización con respecto al cribado en fase sólida.

La descripción proporciona métodos de cribado que usan las proteínas y ácidos nucleicos de la descripción y automatización robótica para posibilitar la ejecución de muchos miles de reacciones biocatalíticas y análisis de cribado en un corto período de tiempo, por ejemplo por día, así como asegurar un nivel elevado de exactitud y reproducibilidad (véase la discusión de matrices, más abajo). Como resultado, se puede producir una biblioteca de compuestos derivados en cuestión de semanas. Para enseñanzas posteriores sobre la modificación de moléculas, incluyendo moléculas pequeñas, véase el documento PCT/US94/09174.

Otro aspecto de la descripción es un polipéptido aislado o purificado que comprende la secuencia de uno de la descripción, o fragmentos que comprenden al menos alrededor de 5, 10, 15, 20, 25, 30, 35, 40, 50, 75, 100, o 150 aminoácidos consecutivos del mismo. Como se ha comentado anteriormente, tales polipéptidos se pueden obtener insertando un ácido nucleico que codifica el polipéptido en un vector de manera que la secuencia codificante esté enlazada operablemente a una secuencia capaz de dirigir la expresión del polipéptido codificado en una célula hospedante adecuada. Por ejemplo, el vector de expresión puede comprender un promotor, un sitio de unión al ribosoma para el inicio de la traducción, y un terminador de la transcripción. El vector puede incluir también secuencias apropiadas para amplificar la expresión.

Otro aspecto de la descripción es polipéptidos o fragmentos de los mismos que tienen una identidad (homología) de secuencia al menos alrededor de 50%, al menos alrededor de 55%, al menos alrededor de 60%, al menos alrededor de 65%, al menos alrededor de 70%, al menos alrededor de 75%, al menos alrededor de 80%, al menos alrededor de 85%, al menos alrededor de 90%, al menos alrededor de 95%, o más de alrededor de 95% con uno de los polipéptidos de la descripción, o un fragmento que comprende al menos 5, 10, 15, 20, 25, 30, 35, 40, 50, 75, 100, o 150 aminoácidos consecutivos de los mismos. La identidad (homología) de secuencia se puede determinar usando cualquiera de los programas descritos anteriormente que alinea los polipéptidos o fragmentos que se están comparando y determina el grado de identidad o similitud de aminoácidos entre ellos. Se apreciará que la equivalencia, o identidad, u "homología" de aminoácidos incluye sustituciones de aminoácidos conservativas tales como las descritas anteriormente.

Los polipéptidos o fragmentos que tienen homología con uno de los polipéptidos de la descripción, o un fragmento que comprende al menos alrededor de 5, 10, 15, 20, 25, 30, 35, 40, 50, 75, 100, o 150 aminoácidos consecutivos de los mismos, se pueden obtener aislando los ácidos nucleicos que los codifican usando las técnicas descritas anteriormente.

Alternativamente, los polipéptidos o fragmentos homólogos se pueden obtener a través de procedimientos bioquímicos de enriquecimiento o purificación. La secuencia de polipéptidos o fragmentos potencialmente homólogos se puede determinar mediante ensayos de actividad, electroforesis en gel y/o microsecuenciación. La secuencia del polipéptido o fragmento homólogo probable se puede comparar con uno de los polipéptidos de la descripción, o un fragmento que comprende al menos alrededor de 5, 10, 15, 20, 25, 30, 35, 40, 50, 75, 100, o 150 aminoácidos consecutivos de los mismos, usando cualquiera de los programas descritos anteriormente.

Otro aspecto de la descripción es un ensayo para identificar fragmentos o variantes de la descripción, que conservan la función enzimática de los polipéptidos de la descripción. Por ejemplo, los fragmentos o variantes de los polipéptidos de la descripción se pueden usar para catalizar reacciones bioquímicas, que indican que el fragmento o variante conserva la actividad enzimática de un polipéptido de la descripción.

El ensayo para determinar si los fragmentos de variantes conservan la actividad enzimática de los polipéptidos de la descripción incluye las etapas de: poner en contacto el fragmento polipeptídico o variante con una molécula sustrato en condiciones que permitan al fragmento polipeptídico o variante funcionar y detectar cualquier descenso en el nivel de sustrato o un incremento en el nivel del producto de reacción específico de la reacción entre el polipéptido y sustrato.

Los polipéptidos de la descripción, o fragmentos que comprenden al menos 5, 10, 15, 20, 25, 30, 35, 40, 50, 75, 100, o 150 aminoácidos consecutivos de los mismos, se pueden usar en una variedad de aplicaciones. Por ejemplo, los polipéptidos o fragmentos de los mismos se pueden usar para catalizar reacciones bioquímicas. En consecuencia con un aspecto de la descripción, se proporciona un procedimiento para la utilización de los polipéptidos de la descripción o polinucleótidos que codifican tales polipéptidos para hidrolizar enlaces éster. En tales procedimientos, una sustancia que contiene un enlace éster (por ejemplo, una clorofila) se pone en contacto con uno de los polipéptidos de la descripción, o sus secuencias sustancialmente idénticas en condiciones que faciliten la hidrólisis del enlace de éster.

La presente descripción explota las propiedades catalíticas únicas de las enzimas. Mientras que el uso de biocatalizadores (es decir, enzimas purificadas o brutas, células no vivas o vivas) en transformaciones químicas requiere normalmente la identificación de un biocatalizador particular que reacciona con un compuesto de partida específico, la presente descripción usa biocatalizadores seleccionados y condiciones de reacción que son específicas para grupos funcionales que están presentes en muchos compuestos de partida, tales como pequeñas moléculas. Cada biocatalizador es específico para un grupo funcional, o varios grupos funcionales relacionados, y puede reaccionar con muchos compuestos de partida que contienen este grupo funcional.

Las reacciones biocatalíticas producen una población de derivados a partir de un único compuesto de partida. Estos derivados se pueden someter a otra ronda de reacciones biocatalíticas para producir una segunda población de compuestos derivados. Se pueden producir miles de variaciones del compuesto o pequeña molécula original con cada iteración de la derivatización biocatalítica.

Las enzimas reaccionan en sitios específicos de un compuesto de partida sin afectar al resto de la molécula, un proceso que es muy difícil de lograr usando métodos químicos tradicionales. Este grado elevado de especificidad biocatalítica proporciona el medio para identificar un único compuesto activo en una biblioteca. La biblioteca se caracteriza por la serie de reacciones biocatalíticas usadas para producirla, una denominada “historia biosintética”. El cribado de la biblioteca en busca de actividades biológicas y el trazado de la historia biosintética identifica la secuencia de reacciones específicas que produce el compuesto activo. La secuencia de reacciones se repite, y se determina la estructura del compuesto sintetizado. Este modo de identificación, a diferencia de otros enfoques sintéticos y de cribado, no requiere tecnologías de inmovilización, y los compuestos se pueden sintetizar y ensayar libremente en disolución usando virtualmente cualquier tipo de ensayo de cribado. Es importante señalar que el grado elevado de especificidad de las reacciones enzimáticas sobre grupos funcionales permite el “trazado” de reacciones enzimáticas específicas que constituyen la biblioteca producida biocatalíticamente.

Muchas de las etapas de procedimiento se llevan a cabo usando automatización robótica que permite la ejecución de muchos miles de reacciones biocatalíticas y ensayos de cribado por día, así como aseguran un nivel elevado de exactitud y reproducibilidad. Como resultado, se puede producir una biblioteca de compuestos derivados en cuestión de semanas, que se tardaría años en producir usando los métodos químicos actuales.

En un aspecto particular, la descripción proporciona un método para modificar moléculas pequeñas, que comprende poner en contacto un polipéptido codificado por un polinucleótido descrito aquí, o fragmentos enzimáticamente activos del mismo, con una molécula pequeña para producir una molécula pequeña modificada. Una biblioteca de moléculas pequeñas modificadas se somete a ensayo para determinar si una molécula pequeña modificada está presente en la biblioteca que exhibe una actividad deseada. Una reacción biocatalítica específica que produce la molécula pequeña modificada de actividad deseada se identifica eliminando sistemáticamente cada una de las reacciones biocatalíticas usadas para producir una porción de la biblioteca, y después sometiendo a ensayo las moléculas pequeñas producidas en la porción de la biblioteca en busca de la presencia o ausencia de la molécula pequeña modificada con la actividad deseada. Las reacciones biocatalíticas específicas que producen la molécula pequeña modificada de actividad deseada se repite opcionalmente. Las reacciones biocatalíticas se llevan a cabo con un grupo de biocatalizadores que reaccionan con restos estructurales distintos encontrados en la estructura de una molécula pequeña, cada biocatalizador es específico para un resto estructural o un grupo de restos estructurales relacionados; y cada biocatalizador reacciona con muchas moléculas pequeñas diferentes que contienen el resto estructural distinto.

Secuencias señal, dominios prepro, dominios de unión y dominios catalíticos

La descripción proporciona secuencias señal enzimáticas (por ejemplo, péptidos señal (SPs)), dominios prepro, dominios de unión y dominios catalíticos (CDs) (por ejemplo, sitios activos). Los SPs, dominios prepro y/o CDs de la descripción pueden ser péptidos aislados o recombinantes, o pueden ser parte de una proteína de fusión, por ejemplo en forma de un dominio heterólogo en una proteína quimérica. La descripción proporciona ácidos nucleicos que codifican estos dominios catalíticos (CDs), dominios prepro y secuencias señal (SPs, por ejemplo un péptido que tiene una secuencia que comprende/consiste en restos amino terminales de un polipéptido de la descripción). En un aspecto, la descripción proporciona una secuencia señal que comprende un péptido que comprende/consiste en una secuencia como la mostrada en los restos 1 a 10, 1 a 11, 1 a 12, 1 a 13, 1 a 14, 1 a 15, 1 a 16, 1 a 17, 1 a 18, 1 a 19, 1 a 20, 1 a 21, 1 a 22, 1 a 23, 1 a 24, 1 a 25, 1 a 26, 1 a 27, 1 a 28, 1 a 28, 1 a 30, 1 a 31, 1 a 32, 1 a 33, 1 a 34, 1 a 35, 1 a 36, 1 a 37, 1 a 38, 1 a 39, 1 a 40, 1 a 41, 1 a 42, 1 a 43, 1 a 44, 1 a 45, 1 a 46, 1 a 47, 1 a 48, 1 a 49, 1 a 50, 1 a 51, o 1 a 52 o más, de un polipéptido de la descripción.

En un aspecto, la descripción también proporciona polipéptidos quiméricos (y los ácidos nucleicos que los codifican) que comprenden al menos dos enzimas de la descripción o sus subsecuencias, por ejemplo dominios catalíticos (CDs) o sitios activos. Por ejemplo, una proteína quimérica de la descripción puede tener cualquier combinación de actividades. En un aspecto, la proteína quimérica comprende una fusión de dominios, por ejemplo un dominio individual puede mostrar una o cualquier combinación de actividades (por ejemplo, como una proteína quimérica recombinante).

La descripción también proporciona secuencias señal aisladas, sintéticas o recombinantes que comprenden/qu consisten en una secuencia señal de la descripción, por ejemplo secuencias señal ejemplares como se exponen en la Tabla 1, más abajo, y polipéptidos que comprenden estas secuencias señal. El polipéptido puede ser otra enzima de la descripción, u otro tipo de enzima o polipéptido. Por ejemplo, para ayudar a leer la Tabla 1, la descripción proporciona una secuencia señal aislada, sintética o recombinante como se expone mediante los restos de aminoácidos amino terminales 1 a 21 ("NH₂-MSRVCLPLTLTLALTLASARA") de SEC ID NO:2, codificada, por ejemplo, por SEC ID NO:1, etc.:

Tabla 1

SEC ID NO:	Posición de la secuencia señal (AA= aminoácido)	Secuencia señal
1, 2	AA1-20	MSRVCLPLTLTLALSARA
11, 12		
13, 14		
15, 16		
17, 18	AA1-25	MKKYKTGLVLSGGGTRGFAHLGVIA
19, 20		
3, 4		
5, 6	AA1-25	MRRIVFLYILALLCVSCANRNPSVS
7, 8	AA1-51	MTRKKIGLALSGGAARGFAHLGVLFVFAEHGIPVDFVAGTSAGSFAGAAFA
9, 10	AA1-23	MFNKALPAAAAVAGLFLSTSAMA

Las secuencias señal (SPs) y/o secuencias prepro de la descripción pueden ser péptidos aislados, o secuencias unidas a otra enzima de la descripción, o una proteína heteróloga, por ejemplo como una proteína de fusión (química). En un aspecto, la descripción proporciona polipéptidos que comprenden secuencias señal de la descripción. En un aspecto, los polipéptidos que comprenden secuencias señal SPs y/o prepro de la descripción comprenden secuencias heterólogas a enzimas de la descripción (por ejemplo, una proteína de fusión que comprende una SP y/o prepro de la descripción, y/o secuencias procedentes de otra proteína). En un aspecto, la descripción proporciona una enzima de la descripción con SPs heterólogas y/o secuencias prepro, por ejemplo secuencias con una secuencia señal de levadura. Las enzimas de la descripción pueden comprender una SP heteróloga y/o prepro en un vector, por ejemplo un vector de la serie pPIC (Invitrogen, Carlsbad, CA). En un aspecto, las SPs y/o secuencias prepro de la descripción se identifican tras la identificación de nuevos polipéptidos. Las rutas mediante las cuales las proteínas se clasifican y transportan a su localización celular apropiada se denominan a menudo como rutas de selección de dianas proteicas. Uno de los elementos más importantes en todos estos sistemas de selección de dianas es una secuencia de aminoácidos corta en el término amino de un polipéptido recientemente sintetizado, denominada la secuencia señal. Esta secuencia señal dirige una proteína a su localización apropiada en la célula, y se elimina durante el transporte o cuando la proteína alcanza su destino final. La mayoría de las proteínas lisosómicas, de membrana, o segregadas tienen una secuencia señal amino terminal que las marca para la translocación en la luz del retículo endoplásmico. Se han determinado más de 100 secuencias señal para proteínas en este grupo. Las secuencias señal pueden variar en longitud desde 13 hasta 36 restos de aminoácidos. Los expertos en la técnica conocen diversos métodos de reconocimiento de secuencias señal. Por ejemplo, en un aspecto, nuevos péptidos señal se identifican mediante un método denominado como SignalP. SignalP usa una red neural combinada que reconoce tanto péptidos señal como sus sitios de escisión. (Nielsen, et al., "Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites". Protein Engineering, vol. 10, nº 1, p. 1-6 (1997)).

Se debería entender que, en algunos aspectos, una enzima de la descripción puede no tener SPs y/o secuencias prepro, o uno o más "dominios". En un aspecto, la descripción proporciona una enzima de la descripción que carece de toda o parte de una SP y/o un dominio prepro. En un aspecto, la descripción proporciona una secuencia de ácido nucleico que codifica una secuencia señal (SP) y/o prepro de una enzima de la descripción, enlazada operablemente a una secuencia de ácido nucleico de una enzima de la descripción diferente, u opcionalmente se puede desear una secuencia señal (SPs) y/o dominio prepro de un tipo diferente de proteína.

La descripción también proporciona polipéptidos aislados o recombinantes que comprenden secuencias señal (SPs), dominio prepro y/o dominios catalíticos (CDs) de la descripción y secuencias heterólogas. Las secuencias heterólogas son secuencias no asociadas naturalmente (por ejemplo, a enzimas de la descripción) con una SP, dominio prepro y/o CD. La secuencia a la que la SP, el dominio prepro y/o el CD no están asociados de forma natural puede estar en el extremo amino terminal de las SP, del dominio prepro y/o de los CD, en el extremo carboxi terminal, y/o en ambos extremos tanto de SP y/o CD. En un aspecto, la presente descripción proporciona un polipéptido aislado o recombinante que comprende (o consiste en) un polipéptido que comprende una secuencia señal (SP), dominio prepro y/o un dominio catalítico (CD) de la descripción, con la condición de que no esté asociado con ninguna secuencia a la que está asociado de forma natural. De forma similar, en un aspecto, la

descripción proporciona ácidos nucleicos aislados o recombinantes que codifican estos polipéptidos. De este modo, en un aspecto, el ácido nucleico aislado o recombinante de la descripción comprende una secuencia codificante para una secuencia señal (SP), dominio prepro y/o dominio catalítico (CD) de la descripción y una secuencia heteróloga (es decir, una secuencia no asociada de forma natural con una secuencia señal (SP), dominio prepro y/o dominio catalítico (CD) de la descripción). La secuencia heteróloga puede estar en el extremo 3' terminal, en el extremo 5' terminal, y/o en ambos extremos de la secuencia codificante de SP, del dominio prepro y/o del CD.

Enzimas híbridas (quiméricas) y bibliotecas peptídicas

En un aspecto, la descripción proporciona enzimas híbridas de la descripción y proteínas de fusión, incluyendo bibliotecas peptídicas, que comprenden secuencias de la descripción. Las bibliotecas peptídicas de la invención se pueden usar para aislar moduladores peptídicos (por ejemplo, activadores o inhibidores) de dianas, tales como enzima de la descripción, sus sustratos, etc. Las bibliotecas peptídicas de la descripción se pueden usar para identificar parejas de unión formales de dianas, tales como ligandos, por ejemplo citocinas, hormonas y similares. En un aspecto, la descripción proporciona proteínas quiméricas que comprenden una secuencia señal (SP), dominio prepro y/o dominio catalítico (CD) de la descripción, o una combinación de los mismos, y una secuencia heteróloga (véase anteriormente).

En un aspecto, las proteínas de fusión de la descripción (por ejemplo, el resto peptídico) se estabilizan conformacionalmente (con respecto a péptidos lineales) para permitir una mayor afinidad de unión a las dianas. La descripción proporciona fusiones de enzimas de la descripción y otros péptidos, incluyendo péptidos conocidos y al azar. Se pueden fusionar de tal manera que la estructura de un polipéptido no resulta perturbada significativamente y el péptido es estabilizado conformacionalmente metabólicamente o estructuralmente. Esto permite la creación de una biblioteca de péptidos cuya presencia en las células como su cantidad es monitorizada fácilmente.

Las variantes de la secuencia de aminoácidos de la descripción se pueden caracterizar por medio de una naturaleza predeterminada de la variación, una característica que las aparta de una forma de origen natural, por ejemplo una variación alélica o entre especies de enzimas de la descripción. En un aspecto, las variantes de la descripción exhiben la misma actividad biológica cualitativa que el análogo de origen natural. Alternativamente, las variantes se pueden seleccionar por tener características modificadas. En un aspecto, mientras que el sitio o región para introducir una variación de la secuencia de aminoácidos está predeterminado, la mutación per se no necesita estar predeterminada. Por ejemplo, con el fin de optimizar el comportamiento de una mutación en un sitio dado, se puede llevar a cabo la mutagénesis al azar en el codón o región diana, y las variantes de la enzima expresadas se pueden cribar para determinar la combinación óptima de actividad deseada. Las técnicas para obtener mutaciones de sustitución en sitios predeterminados en el ADN que tiene una secuencia conocida son bien conocidas, como se explica aquí por ejemplo, mutagénesis con cebador M13 y mutagénesis por PCR. El cribado de los mutantes se puede realizar usando, por ejemplo, ensayos de hidrólisis de clorofila, como se describe en el Ejemplo 1, más abajo. En aspectos alternativos, las sustituciones de aminoácidos pueden ser restos individuales; las inserciones pueden ser del orden de alrededor de 1 a 20 aminoácidos, aunque se pueden realizar inserciones considerablemente mayores. Las supresiones pueden oscilar de alrededor de 1 a alrededor de 20, 30, 40, 50, 60, 70 restos o más. Para obtener un derivado final con las propiedades óptimas, se pueden usar sustituciones, supresiones, inserciones o cualquier combinación de las mismas. Generalmente, estos cambios se realizan en unos pocos aminoácidos para minimizar la alteración de la molécula. Sin embargo, en ciertas circunstancias se pueden tolerar cambios más grandes.

La descripción proporciona polipéptidos, por ejemplo enzimas, de la descripción en los que se ha modificado la estructura de la cadena principal del polipéptido, la estructura secundaria o la estructura terciaria, por ejemplo una estructura de hélice alfa o lámina beta. En un aspecto, se ha modificado la carga o la hidrofobia. En un aspecto, se ha modificado el volumen de una cadena lateral. Los cambios sustanciales en la función o la identidad inmunológica se realizan seleccionando sustituciones que son menos conservativas. Por ejemplo, se pueden realizar sustituciones que afectan más significativamente a: la estructura de la cadena principal del polipéptido en el área de la alteración, por ejemplo una estructura de hélice alfa o lámina beta; una carga o sitio hidrófobo de la molécula, que puede estar en un sitio activo; o una cadena lateral. La descripción proporciona sustituciones en el polipéptido de la descripción, en el que (a) un resto hidrófilo, por ejemplo serilo o treonilo, es sustituido por (o mediante) un resto hidrófobo, por ejemplo leucilo, isoleucilo, fenilalanilo, valilo o alanilo; (b) una cisteína o prolina es sustituida por (o mediante) cualquier otro resto; (c) un resto que tiene una cadena lateral electropositiva, por ejemplo lisilo, arginilo, o histidilo, es sustituido por (o mediante) un resto electronegativo, por ejemplo glutamilo o aspartilo; o (d) un resto que tiene una cadena lateral voluminosa, por ejemplo fenilalanina, es sustituido por (o mediante) uno que no tenga una cadena lateral, por ejemplo glicina. Las variantes pueden mostrar la misma actividad biológica cualitativa como enzimas de la descripción, aunque se pueden seleccionar variantes para modificar las características de la enzima según se necesite.

“Aminoácido” o “secuencia de aminoácidos”, como se usa aquí, se refiere a una secuencia oligopeptídica, peptídica, polipeptídica, o proteica, o a un fragmento, porción, o subunidad de cualquiera de estas, y a moléculas de origen natural o sintéticas.

“Aminoácido” o “secuencia de aminoácidos” incluye una secuencia oligopeptídica, peptídica, polipeptídica, o proteica, o a un fragmento, porción, o subunidad de cualquiera de estas, y a moléculas de origen natural o sintéticas. El término “polipéptido”, como se usa aquí, se refiere a aminoácidos unidos entre sí mediante enlaces peptídicos o enlaces peptídicos modificados, es decir, isómeros peptídicos, y pueden contener aminoácidos modificados distintos de los 20 aminoácidos codificados por genes. Los polipéptidos se pueden modificar mediante procesos naturales, tales como procesamiento post-traduccion, o mediante técnicas de modificación química que son bien conocidas en la técnica. Las modificaciones se pueden producir en cualquier parte en el polipéptido, incluyendo la cadena principal peptídica, las cadenas laterales de aminoácidos, y los términos amino o carboxilo. Se apreciará que el mismo tipo de modificación puede estar presente en los mismos grados o grados variables en varios sitios en un polipéptido dado. También, un polipéptido dado puede tener muchos tipos de modificaciones. Las modificaciones incluyen acetilación, acilación, ribosilación con ADP, amidación, unión covalente de flavina, unión covalente de un resto hemo, unión covalente de un nucleótido o derivado nucleotídico, unión covalente de un lípido o derivado lipídico, unión covalente de un fosfatidilinositol, ciclación por reticulación, formación de enlace de disulfuro, desmetilación, formación de reticulaciones covalentes, formación de cisteína, formación de piroglutamato, formilación, gamma-carboxilación, glicosilación, formación de anclaje de GPI, hidroxilación, yodación, metilación, miristoilación, oxidación, pegilación, procesamiento de glucano hidrolasa, fosforilación, prenilación, racemización, selenoilación, sulfatación, y adición de aminoácidos mediada por ARN de transferencia a proteína, tal como arginilación. (Véase Creighton, T.E., *Proteins - Structure and Molecular Properties* 2ª Ed., W.H. Freeman and Company, Nueva York (1993); *Posttranslational Covalent Modification of Proteins*, B.C. Johnson, Ed., Academic Press, Nueva York, p. 1-12 (1983)). Los péptidos y polipéptidos de la descripción también incluyen todas las formas “miméticas” y “peptidomiméticas”, como se describe con más detalle, más abajo.

Adicionalmente, una secuencia de aminoácidos “sustancialmente idéntica” es una secuencia que difiere de una secuencia de referencia en una o más sustituciones, supresiones o inserciones de aminoácidos conservativas o no conservativas, particularmente cuando tal sustitución se produce en un sitio que no es el sitio activo (dominios catalíticos (CDs)) de la molécula, y con la condición de que el polipéptido retenga esencialmente sus propiedades funcionales. Una sustitución de aminoácidos conservativa, por ejemplo, sustituye un aminoácido por otro de la misma clase (por ejemplo, sustitución de un aminoácido hidrófobo, tal como isoleucina, valina, leucina, o metionina, por otro, o sustitución de un aminoácido polar por otro, tal como una sustitución de arginina por lisina, ácido glutámico por ácido aspártico, o glutamina por asparagina). Uno o más aminoácidos se pueden suprimir, por ejemplo, de un polipéptido, dando como resultado la modificación de la estructura del polipéptido, sin alterar significativamente su actividad biológica. Por ejemplo, se pueden eliminar los aminoácidos amino o carboxi terminales que no son necesarios para la actividad biológica de la enzima. Las secuencias polipeptídicas modificadas de la descripción se pueden ensayar en busca de la actividad biológica (por ejemplo, enzimática, o de unión) mediante cualquier número de métodos, incluyendo la puesta en contacto de la secuencia polipeptídica modificada con un sustrato enzimático y determinando si el polipéptido modificado disminuye la cantidad de sustrato específico en el ensayo o incrementa los bioproductos de la reacción enzimática de un polipéptido funcional con el sustrato.

Los “fragmentos”, como se usa aquí, son una porción de una proteína de origen natural que puede existir en al menos dos conformaciones diferentes. Los fragmentos pueden tener la misma o sustancialmente la misma secuencia de aminoácidos que la proteína de origen natural. “Sustancialmente la misma” significa que una secuencia de aminoácidos es en su mayor parte, pero no completamente, la misma, pero retiene al menos una actividad funcional de la secuencia con la que está relacionada. En aspectos alternativos, dos secuencias de aminoácidos son “sustancialmente las mismas” o “sustancialmente homólogas” si tienen al menos alrededor de 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, 60%, 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, o más de identidad de secuencia. También están incluidos los fragmentos que tienen estructuras tridimensionales diferentes a la de la proteína de origen natural; por ejemplo, una molécula “pro-forma”, tal como una proproteína de baja actividad que se puede modificar mediante escisión para producir una enzima madura con una actividad significativamente mayor.

En un aspecto, las enzimas de la descripción comprenden epítomos o etiquetas de purificación, secuencias señal u otras secuencias de fusión, etc. En un aspecto, las enzimas se pueden fusionar a péptido aleatorio para formar un polipéptido de fusión. Por “fusionado” o “operablemente enlazado” se quiere decir aquí que el péptido aleatorio y la enzima se enlazan entre sí, de tal manera para minimizar la alteración de la estabilidad de la estructura de la enzima, por ejemplo retiene la actividad. El polipéptido de fusión (o polinucleótido de fusión que codifica el polipéptido de fusión) puede comprender también componentes adicionales, incluyendo múltiples péptidos en múltiples bucles.

En un aspecto, los péptidos y ácidos nucleicos que los codifican están aleatorizados, o bien completamente aleatorizados o su aleatorización es parcial, por ejemplo en la frecuencia del nucleótido/resto generalmente o por posición. “Aleatorizado” significa que cada ácido nucleico y péptido consiste esencialmente en nucleótidos y aminoácidos al azar, respectivamente. En un aspecto, los ácidos nucleicos que dan lugar a los péptidos se pueden sintetizar químicamente, y de este modo pueden incorporar cualquier nucleótido en cualquier posición. De este modo, cuando los ácidos nucleicos se expresan para formar péptidos, cualquier resto de aminoácido se puede incorporar en cualquier posición. El procedimiento sintético se puede diseñar para generar ácidos nucleicos

aleatorizados, para permitir la formación de todas o de la mayoría de las combinaciones posibles a lo largo de la longitud del ácido nucleico, formando de este modo una biblioteca de ácidos nucleicos aleatorizados. La biblioteca puede proporcionar una población suficientemente diversa estructuralmente de productos de expresión aleatorizados que afecte a un intervalo probabilísticamente suficiente de respuestas celulares para proporcionar una o más células que exhiben una respuesta deseada. De este modo, la descripción proporciona una biblioteca de interacción suficientemente grande de manera que al menos uno de sus miembros tendrá una estructura que le proporcione afinidad por cierta molécula, proteína, u otro factor.

La descripción proporciona un medio para generar polipéptidos quiméricos que pueden codificar polipéptidos híbridos biológicamente activos (por ejemplo enzimas híbridas de la descripción). En un aspecto, los polinucleótidos originales codifican polipéptidos biológicamente activos. El método de la descripción produce nuevos polipéptidos híbridos utilizando procesos celulares que integran la secuencia de los polinucleótidos originales de manera que el polinucleótido híbrido resultante codifica un polipéptido que demuestra actividades derivadas de los polipéptidos biológicamente activos originales. Por ejemplo, los polinucleótidos originales pueden codificar una enzima particular a partir de diferentes microorganismos. Una enzima codificada por un primer polinucleótido de un organismo o variante puede, por ejemplo, funcionar eficazmente en una condición ambiental particular, por ejemplo salinidad elevada. Una enzima codificada por un segundo polinucleótido de un organismo o variante diferente puede funcionar eficazmente en una condición ambiental diferente, tal como temperaturas extremadamente altas. Un polinucleótido híbrido que contiene secuencias del primer y segundo polinucleótidos originales puede codificar una enzima que muestra características de ambas enzimas codificadas por los polinucleótidos originales. De este modo, la enzima codificada por el polinucleótido híbrido puede funcionar eficazmente en condiciones ambientales compartidas por cada una de las enzimas codificadas por el primer y segundo polinucleótidos, por ejemplo salinidad elevada y temperaturas extremas.

Un polipéptido híbrido resultante del método de la descripción puede exhibir actividad enzimática especializada no desplegada por las enzimas originales. Por ejemplo, tras la recombinación y/o el redistribución reductiva de los polinucleótidos que codifican las actividades enzimáticas, el polipéptido híbrido resultante codificado por un polinucleótido híbrido se puede cribar para determinar las actividades enzimáticas especializadas obtenidas de cada una de las enzimas originales, es decir, el tipo de enlace sobre el que actúa la enzima, y la temperatura a la que funciona la enzima. De este modo, por ejemplo, la enzima se puede cribar para averiguar las funcionalidades químicas que distinguen la enzima híbrida de las enzimas originales, tales como especificidad por el sustrato, o temperatura, o pH, o concentración salina a la cual funciona el polipéptido híbrido.

Las fuentes de los polinucleótidos originales se puede aislar de organismos individuales ("aislados"), colecciones de organismos que se han hecho crecer en medios definidos ("cultivos de enriquecimiento"), u organismos no cultivados ("muestras ambientales"). Es muy preferible el uso de un enfoque independiente del cultivo para obtener polinucleótidos que codifican bioactividades novedosas a partir de muestras ambientales, puesto que permite el acceso a recursos no explotados de biodiversidad.

Las "bibliotecas ambientales" se generan a partir de muestras ambientales y representan los genomas colectivos de organismos de origen natural logrados en vectores de clonación que se han propagado en hospedantes procariotas adecuados. Debido a que el ADN clonado se extrae inicialmente directamente de muestras ambientales, las bibliotecas no están limitadas a la pequeña fracción de procariotas que se pueden hacer crecer en cultivo puro. Adicionalmente, una normalización del ADN ambiental presente en estas muestras podría permitir una representación más equitativa del ADN de todas las especies presentes en la muestra original. Esto puede aumentar espectacularmente la eficacia del descubrimiento de genes interesantes a partir de constituyentes mínimos de la muestra que pueden estar infrarrepresentados en varios órdenes de magnitud en comparación con las especies dominantes.

Por ejemplo, las genotecas generada a partir de uno o más microorganismos no cultivados se criban para determinar una actividad de interés. Las rutas potenciales que codifican moléculas bioactivas de interés se capturan en primer lugar en células procariotas en forma de bibliotecas de expresión de genes. Los polinucleótidos que codifican actividades de interés se aíslan a partir de tales bibliotecas y se introducen en una célula hospedante. La célula hospedante se hace crecer en condiciones que promueven la recombinación y/o el redistribución reductiva creando biomoléculas potencialmente activas con actividades novedosas o potenciadas.

Los microorganismos a partir de los cuales se pueden preparar el polinucleótido incluyen microorganismos procariotas, tales como *Eubacteria* y *Archaeobacteria*, y microorganismos eucariotas inferiores tales como hongos, algunas algas y protozoos. Los polinucleótidos se pueden aislar de muestras ambientales, en cuyo caso el ácido nucleico se puede recuperar sin cultivar un microorganismo, o se puede recuperar de uno o más organismos cultivados. En un aspecto, tales microorganismos pueden ser extremófilos, tales como hipertermófilos, psicrófilos, psicrótrofos, halófilos, barófilos y acidófilos. Se pueden usar los polinucleótidos que codifican enzimas aislados de microorganismos extremófilos. Tales enzimas pueden funcionar a temperaturas por encima de 100°C en manantiales calientes terrestres y respiraderos térmicos de las profundidades marinas, a temperaturas por debajo de 0°C en aguas árticas, en el entorno salino saturado de, por ejemplo, el Mar Muerto, a valores de pH alrededor de 0 en depósitos de carbón y manantiales geotérmicos ricos en azufre, o a valores de pH mayores que 11 en lodos de

aguas residuales. Por ejemplo, varias esterasas y lipasas clonadas y expresadas a partir de organismos extremófilos pueden mostrar una actividad elevada en un amplio intervalo de temperaturas y pHs.

Los polinucleótidos seleccionados y aislados como se describen aquí anteriormente se introducen en una célula hospedante adecuada. Una célula hospedante adecuada es cualquier célula que es capaz de promover la recombinación y/o redistribución reductiva. En un aspecto, los polinucleótidos seleccionados ya están en un vector que incluye secuencias de control apropiadas. La célula hospedante puede ser una célula eucariota superior, tal como una célula de mamífero, o una célula eucariota inferior, tal como una célula de levadura, o, en un aspecto, la célula hospedante puede ser una célula procariota, tal como una célula bacteriana. La introducción del constructo en la célula hospedante se puede efectuar mediante transfección con fosfato de calcio, transfección mediada por DEAE-dextrano, o electroporación (Davis et al., 1986).

Como ejemplos representativos de hospedantes apropiados, se pueden mencionar: células bacterianas, tales como *E. coli*, *Streptomyces*, *Salmonella typhimurium*; células fúngicas, tales como levadura; células de insecto tales como *Drosophila* S2 y *Spodoptera* Sf9; células de animales tales como CHO, COS o melanoma de Bowes; adenovirus; y células vegetales. La selección de un hospedante apropiado se considera que está dentro del alcance de los expertos en la técnica a partir de las enseñanzas aquí.

Con referencia particular a diversos sistemas de cultivo de células de mamífero que se pueden emplear para expresar proteína recombinante, los ejemplos de sistemas de expresión de mamíferos incluyen las líneas COS-7 de fibroblastos de riñón de mono, descritas en "SV40-transformed simian cells support the replication of early SV40 mutants" (Gluzman, 1981), y otras estirpes celulares capaces de expresar un vector compatible, por ejemplo, las estirpes celulares C127, 3T3, CHO, HeLa y BHK. Los vectores de expresión de mamíferos comprenderán un origen de replicación, un promotor y potenciador adecuados, y también cualesquiera sitios de unión a ribosoma necesarios, sitio de poliadenilación, sitios dadores y aceptores de ajuste, secuencias de terminación transcripcional, y secuencias no transcritas de flanco en 5'. Se pueden usar secuencias de ADN derivadas del ajuste de SV40 y sitios de poliadenilación para proporcionar los elementos genéticos no transcritos requeridos.

En otro aspecto, se concibe que el método de la presente descripción se puede usar para generar nuevos polinucleótidos que codifican rutas bioquímicas a partir de uno o más operones o racimos génicos o sus porciones. Por ejemplo, las bacterias y muchos eucariotas tienen un mecanismo coordinado para regular genes cuyos productos están implicados en procesos relacionados. Los genes se agrupan, en estructuras denominadas "racimos génicos", en un único cromosoma, o se transcriben juntos bajo el control de una única secuencia reguladora, incluyendo un único promotor que inicia la transcripción de todo el racimo. De este modo, un racimo génico es un grupo de genes adyacentes que son idénticos o están relacionados, habitualmente en cuanto a su función. Un ejemplo de una ruta bioquímica codifica por racimos génicos son los policétidos.

El ADN del racimo génico se puede aislar de diferentes organismos y se puede ligar en vectores, particularmente vectores que contienen secuencias reguladoras de la expresión que pueden controlar y regular la producción de una proteína detectable o una actividad de matriz relacionada con proteínas a partir de los racimos génicos ligados. El uso de vectores que tienen una capacidad excepcionalmente grande para la introducción de ADN exógeno es particularmente apropiado para uso con tales racimos génicos, y se describe a título de ejemplo aquí para incluir el factor f (o factor de fertilidad) de *E. coli*. Este factor f de *E. coli* es un plásmido que afecta a la transferencia de alta frecuencia de sí mismo durante la conjugación, y es ideal para lograr y propagar de forma estable grandes fragmentos de ADN, tales como racimos génicos de muestras microbianas mixtas. Un aspecto es usar vectores de clonación, denominados "fósmidos" o vectores de cromosoma artificial bacteriano (BAC). Estos derivan del factor f de *E. coli* que es capaz de integrar de forma estable grandes segmentos de ADN genómico. Cuando se integran con ADN de una muestra ambiental no cultivada mixta, esto hace posible lograr grandes fragmentos genómicos en forma de una "biblioteca de ADN ambiental" estable. Otro tipo de vector para uso en la presente descripción es un vector cosmídico. Los vectores cosmídicos se diseñaron originalmente para clonar y propagar grandes segmentos de ADN genómico. La clonación en vectores cosmídicos se describe con detalle en Sambrook et al., *Molecular Cloning: A Laboratory Manual*, 2ª Ed., Cold Spring Harbor Laboratory Press (1989). Una vez ligados en un vector apropiado, dos o más vectores que contienen diferentes racimos génicos de policétido sintasa se pueden introducir en una célula hospedante adecuada. Las regiones de homología de secuencia parcial compartida por los racimos génicos promoverán procesos que dan como resultado la reorganización de las secuencias, dando como resultado un racimo génico híbrido. El nuevo racimo génico híbrido se puede cribar entonces en busca de actividades mejoradas no encontradas en los racimos génicos originales.

Por lo tanto, en un aspecto, la descripción se refiere a un método para producir un polipéptido híbrido biológicamente activo y cribar tal polipéptido en busca de una actividad potenciada:

1) introduciendo al menos un primer polinucleótido en enlazamiento operable, y un segundo polinucleótido en enlazamiento operable, compartiendo el al menos primer nucleótido y segundo nucleótido al menos una región de homología de secuencia parcial, en una célula hospedante adecuada;

2) hacer crecer la célula hospedante en condiciones que promueven la reorganización de secuencias, dando como resultado un polinucleótido híbrido en enlazamiento operable;

- 3) expresar un polipéptido híbrido codificado por el polinucleótido híbrido;
- 4) cribar el polipéptido híbrido en condiciones que promueven la identificación de la actividad biológica potenciada; y
- 5) aislar el polinucleótido que codifica el polipéptido híbrido.

5 Los métodos para cribar diversas actividades enzimáticas son conocidos por los expertos en la técnica, y se explican en toda la presente memoria descriptiva. Tales métodos se pueden emplear cuando se aíslan los polipéptidos y polinucleótidos de la descripción.

Metodologías de cribado y dispositivos de monitorización "en línea"

10 Al poner en práctica los métodos de la descripción, se puede usar una variedad de aparatos y metodologías junto con los polipéptidos y ácidos nucleicos de la descripción, por ejemplo para cribar polipéptidos para determinar la actividad enzimática, para cribar compuestos como moduladores potenciales, por ejemplo activadores o inhibidores de la actividad, para determinar anticuerpos que se unen a un polipéptido de la descripción, para determinar ácidos nucleicos que se hibridan a un ácido nucleico de la descripción, para cribar células que expresan un polipéptido de la descripción, y similares. Además de los formatos de matrices descritos con detalle más abajo para cribar muestras,

15 también se pueden usar formatos alternativos para poner en práctica los métodos de la descripción. Tales formatos incluyen, por ejemplo, espectrómetros de masas, cromatógrafos, por ejemplo HPLC de alto rendimiento y otras formas de cromatografía de líquidos, y formatos más pequeños, tales como placas de 1536 pocillos, placas de 384 pocillos y etc. Los aparatos de cribado de alto rendimiento se pueden adaptar y utilizar para poner en práctica los métodos de la descripción, véase, por ejemplo la solicitud de patente U.S. nº 20020001809.

20 Matrices capilares

Los ácidos nucleicos o polipéptidos de la descripción se pueden inmovilizar en o aplicar a una matriz. Las matrices se pueden usar para cribar o monitorizar bibliotecas de composiciones (por ejemplo, pequeñas moléculas, anticuerpos, ácidos nucleicos, etc.) en busca de su capacidad para unirse a o modular la actividad de un ácido nucleico o un polipéptido de la descripción. Las matrices capilares, tales como la GIGAMATRIX™, Diversa Corporation, San Diego, CA, y las matrices descritas en, por ejemplo, la solicitud de patente U.S. nº 20020080350 A1, los documentos WO 0231203 A, WO 0244336 A, proporcionan un aparato alternativo para soportar y cribar muestras. En un aspecto, la matriz capilar incluye una pluralidad de capilares formados en una matriz de capilares adyacentes, en el que cada capilar comprende al menos una pared que define una luz para retener una muestra. La luz puede ser cilíndrica, cuadrada, hexagonal o de cualquier otra forma geométrica con tal que las paredes formen

30 una luz para la retención de un líquido o muestra. Los capilares de la matriz capilar se pueden mantener juntos en íntima proximidad para formar una estructura plana. Los capilares se pueden unir entre sí, fusionándolos (por ejemplo, cuando los capilares están hechos de vidrio), encolándolos, pegándolos, o fijándolos lado con lado. Adicionalmente, la matriz capilar puede incluir material intersticial dispuesto entre capilares adyacentes en la matriz, formando de ese modo un dispositivo plano sólido que contiene una pluralidad de orificios de paso.

35 Una matriz capilar puede estar formada por cualquier número de capilares individuales, por ejemplo un intervalo de 100 a 4.000.000 de capilares. Adicionalmente, una matriz capilar que tenga alrededor de 100.000 capilares individuales o más se puede conformar en el tamaño y forma convencionales de una placa Microliter® para su ajuste en equipos de laboratorio convencionales. Las luces se cargan manual o automáticamente usando cualquier acción o microinyección capilar usando una aguja delgada. Las muestras de interés se pueden retirar posteriormente de los capilares individuales para su análisis o caracterización posterior. Por ejemplo, se coloca una sonda de tipo aguja, delgada, en comunicación fluida con un capilar seleccionado para añadir o retirar material de la luz.

En un ensayo de cribado de un solo recipiente, los componentes del ensayo se mezclan produciendo una disolución de interés, antes de su inserción en la matriz capilar. La luz se llena mediante acción capilar cuando al menos una porción de la matriz se sumerge en una disolución de interés. Las reacciones químicas o biológicas y/o la actividad en cada capilar se monitorizan para determinar los sucesos detectables. Un suceso detectable es referido con frecuencia como un "éxito", que habitualmente se puede distinguir de los capilares que producen "no éxitos" mediante detección óptica. De este modo, las matrices capilares permiten la detección masivamente paralela de "éxitos".

45

En un ensayo de cribado de múltiples recipientes, un polipéptido o ácido nucleico, por ejemplo un ligando, se puede introducir en un primer componente, que es introducido en al menos una porción de un capilar de una matriz capilar. A continuación se puede introducir una burbuja de aire en el capilar detrás del primer componente. A continuación se puede introducir un segundo componente en el capilar, en el que el segundo componente se separa del primer componente mediante la burbuja de aire. El primer y segundo componentes se pueden mezclar a continuación aplicando presión hidrostática a ambos lados de la matriz capilar para destruir la burbuja. Después se monitoriza la

50

55 matriz capilar para determinar un suceso detectable resultante de la reacción o no reacción de los dos componentes.

En un ensayo de cribado de la unión, se puede introducir una muestra de interés en forma de un primer líquido marcado con una partícula detectable en un capilar de una matriz capilar, en el que la luz del capilar se reviste con

un material de unión para unir la partícula detectable a la luz. El primer líquido se puede eliminar a continuación del tubo capilar, en el que la partícula detectable unida se mantiene dentro del capilar, y se puede introducir un segundo líquido en el tubo capilar. El capilar se monitoriza entonces para determinar un suceso detectable resultante de la reacción o no reacción de la partícula con el segundo líquido.

5 Matrices o "biochips"

Los ácidos nucleicos o polipéptidos de la descripción se pueden inmovilizar o aplicar a una matriz. Se pueden usar matrices para cribar o monitorizar bibliotecas de composiciones (por ejemplo, pequeñas moléculas, anticuerpos, ácidos nucleicos, etc.) en busca de su capacidad para unirse a o modular la actividad de un ácido nucleico o un polipéptido de la descripción. Por ejemplo, en un aspecto de la descripción, un parámetro monitorizado es la expresión transcrita de un gen, por ejemplo un gen de la descripción (un ácido nucleico que codifica un polipéptido de la descripción). Uno o más o todos los transcritos de una célula se pueden medir mediante hibridación de una muestra que comprende transcritos de la célula, o ácidos nucleicos representativos de o complementarios a transcritos de una célula, mediante hibridación a ácidos nucleicos inmovilizados en una matriz o "biochip". Mediante el uso de una "matriz" de ácidos nucleicos en un microchip, algunos o todos los transcritos de una célula se pueden cuantificar simultáneamente. Como alternativa, también se pueden usar matrices que comprenden ácido nucleico genómico para determinar el genotipo de una cepa recientemente obtenida mediante ingeniería obtenida mediante los métodos de la descripción. También se pueden usar "matrices polipeptídicas" para cuantificar simultáneamente una pluralidad de proteínas. La presente descripción se puede poner en práctica con cualquier "matriz" conocida, también denominada como "micromatriz" o "matriz de ácido nucleico" o "matriz polipeptídica" o "matriz de anticuerpo" o "biochip", o variación de los mismos. Las matrices son genéricamente una pluralidad de "puntos" o "elementos diana", comprendiendo cada elemento diana una cantidad definida de una o más moléculas biológicas, por ejemplo oligonucleótidos, inmovilizadas sobre un área definida de una superficie de sustrato para la unión específica a una molécula de la muestra, por ejemplo transcritos de ARNm.

Poniendo en práctica los métodos de la descripción, cualquier matriz conocida y/o método para obtener y usar matrices se pueden incorporar en todo o en parte, o sus variaciones, como se describe, por ejemplo, en las patentes U.S. n^{os} 6.277.628; 6.277.489; 6.261.776; 6.258.606; 6.054.270; 6.048.695; 6.045.996; 6.022.963; 6.013.440; 5.965.452; 5.959.098; 5.856.174; 5.830.645; 5.770.456; 5.632.957; 5.556.752; 5.143.854; 5.807.522; 5.800.992; 5.744.305; 5.700.637; 5.556.752; 5.434.049; véanse también, por ejemplo, los documentos WO 99/51773; WO 99/09217; WO 97/46313; WO 96/17958; véanse también, por ejemplo, Johnston (1998) Curr. Biol. 8:R171-R174; Schummer (1997) Biotechniques 23:1087-1092; Kern (1997) Biotechniques 23:120-124; Solinas-Toldo (1997) Genes, Chromosomes & Cancer 20:399-407; Bowtell (1999) Nature Genetics Supp. 21:25-32. Véanse también las solicitudes de patentes U.S. publicadas n^{os} 20010018642; 20010019827; 20010016322; 20010014449; 20010014448; 20010012537; 20010008765.

El término "matriz" o "micromatriz" o "biochip" o "chip", como se usa aquí, es una pluralidad de elementos diana, comprendiendo cada elemento diana una cantidad definida de uno o más polipéptidos (incluyendo anticuerpos) o ácidos nucleicos inmovilizados sobre un área definida de una superficie de sustrato, como se explica con más detalle, más abajo.

Anticuerpos y métodos de cribado a base de anticuerpos

La descripción proporciona anticuerpos aislados o recombinantes que se unen específicamente a un polipéptido de la descripción. Estos anticuerpos se pueden usar para aislar, identificar o cuantificar un polipéptido de la descripción o polipéptidos relacionados. Estos anticuerpos se pueden usar para aislar otros polipéptidos dentro del alcance de la descripción u otros polipéptidos relacionados. Los anticuerpos se pueden diseñar para unirse a un sitio activo de un polipéptido de la descripción. De este modo, la descripción proporciona métodos para inhibir enzimas usando los anticuerpos de la descripción (véase la discusión anterior). La descripción proporciona fragmentos de las enzimas de la descripción, incluyendo fragmentos inmunogénicos de un polipéptido de la descripción. La descripción proporciona composiciones que comprenden un polipéptido o péptido de la descripción y coadyuvantes o portadores y similares.

Los anticuerpos se pueden usar en inmunoprecipitación, tinción, columnas de inmutafinidad, y similares. Si se desea, las secuencias de ácidos nucleicos que codifican antígenos específicos se pueden generar mediante inmunización seguido del aislamiento del polipéptido o ácido nucleico, amplificación o clonación e inmutafización del polipéptido sobre una matriz de la descripción. Como alternativa, los métodos de la descripción se pueden usar para modificar la estructura de un anticuerpo producido por una célula a modificar, por ejemplo se puede incrementar o disminuir una afinidad del anticuerpo. Además, la capacidad para obtener o modificar anticuerpos puede ser un fenotipo manipulado en una célula mediante los métodos de la descripción.

Los métodos de inmunización, producción y aislamiento de anticuerpos (policlonales y monoclonales) son conocidos por los expertos en la técnica, y se describen en la bibliografía científica y de patentes; véanse, por ejemplo, Coligan, CURRENT PROTOCOLS IN IMMUNOLOGY, Wiley/Greene, NY (1991); Stites (eds.) BASIC AND CLINICAL IMMUNOLOGY (7^a ed.) Lange Medical Publications, Los Altos, CA ("Stites"); Goding, MONOCLONAL ANTIBODIES: PRINCIPLES AND PRACTICE (2^a ed.) Academic Press, Nueva York, NY (1986); Kohler (1975) Nature 256:495;

Harlow (1988) ANTIBODIES, A LABORATORY MANUAL, Cold Spring Harbor Publications, Nueva York. Los anticuerpos también se pueden generar *in vitro*, por ejemplo usando bibliotecas de presentación de fagos que expresan sitios de unión a anticuerpos recombinantes, además de los métodos *in vivo* tradicionales que usan animales. Véanse, por ejemplo, Hoogenboom (1997) Trends Biotechnol. 15:62-70; Katz (1997) Annu. Rev. Biophys. Biomol. Struct. 26:27-45.

Los polipéptidos de la descripción o fragmentos que comprenden al menos alrededor de 5, 10, 15, 20, 25, 30, 35, 40, 50, 75, 100, or 150 o más aminoácidos consecutivos del mismo, también se pueden usar para generar anticuerpos que se unen específicamente a los polipéptidos o fragmentos. Los anticuerpos resultantes se pueden usar en procedimientos de cromatografía por inmutafinidad para aislar o purificar el polipéptido, o para determinar si el polipéptido está presente en una muestra biológica. En tales procedimientos, una preparación proteica, tal como un extracto, o una muestra biológica se pone en contacto con un anticuerpo capaz de unirse específicamente a uno de los polipéptidos de la descripción, o fragmentos que comprenden al menos alrededor de 5, 10, 15, 20, 25, 30, 35, 40, 50, 75, 100, or 150 o más aminoácidos consecutivos del mismo.

En procedimientos de inmutafinidad, el anticuerpo se une a un soporte sólido, tal como una perla u otra matriz de columna. La preparación proteica se coloca en contacto con el anticuerpo en condiciones en las que el anticuerpo se une específicamente a uno de los polipéptidos de la descripción, o fragmento del mismo. Después de un lavado para eliminar las proteínas no unidas específicamente, se eluyen los polipéptidos unidos específicamente.

La capacidad de las proteínas en una muestra biológica para unirse al anticuerpo se puede determinar usando cualquiera de una variedad de procedimientos familiares para los expertos en la técnica. Por ejemplo, la unión se puede determinar marcando el anticuerpo con un marcador detectable, tal como un agente fluorescente, un marcador enzimático, o un radioisótopo. Como alternativa, la unión del anticuerpo a la muestra se puede detectar usando un segundo anticuerpo que tiene tal marcador detectable en él. Los ensayos particulares incluyen ensayos de ELISA, ensayos de sándwich, radioinmunoensayos, y transferencias Western.

Los anticuerpos policlonales generados frente a los polipéptidos de la descripción, o fragmentos que comprenden al menos alrededor de 5, 10, 15, 20, 25, 30, 35, 40, 50, 75, 100, or 150 o más aminoácidos consecutivos del mismo, se pueden obtener mediante inyección directa de los polipéptidos en un animal, o administrando los polipéptidos a un animal. Por ejemplo, un animal no humano. El anticuerpo así obtenido se unirá entonces al propio polipéptido. De esta manera, se puede usar incluso una secuencia que codifica sólo un fragmento del polipéptido, para generar anticuerpos que se pueden unir a todo el polipéptido nativo. Tales anticuerpos se pueden usar entonces para aislar el polipéptido de células que expresan ese polipéptido.

Para la preparación de anticuerpos monoclonales, se puede usar cualquier técnica que proporcione anticuerpos producidos mediante cultivos continuos de estirpes celulares. Los ejemplos incluyen la técnica de hibridoma (Kohler and Milstein, Nature, 256:495-497, 1975), la técnica de trioma, la técnica de hibridoma de célula B humana (Kozbor et al., Immunology Today 4:72, 1983), y la técnica de hibridoma de EBV (Cole, et al., 1985, in Monoclonal Antibodies and Cancer Therapy, Alan R. Liss, Inc., p. 77-96).

Las técnicas descritas para la producción de anticuerpos monocatenarios (patente U.S. nº 4.946.778) se pueden adaptar para producir anticuerpos monocatenarios frente a los polipéptidos de la descripción, o fragmentos que comprenden al menos alrededor de 5, 10, 15, 20, 25, 30, 35, 40, 50, 75, 100, or 150 aminoácidos consecutivos del mismo. Como alternativa, se pueden usar ratones transgénicos para expresar anticuerpos humanizados frente a estos polipéptidos o sus fragmentos.

Los anticuerpos generados frente a los polipéptidos de la descripción, o fragmentos que comprenden al menos alrededor de 5, 10, 15, 20, 25, 30, 35, 40, 50, 75, 100, or 150 o más aminoácidos consecutivos del mismo, se pueden usar en el cribado de polipéptidos similares de otros organismos y muestras. En tales técnicas, los polipéptidos procedentes del organismo se ponen en contacto con el anticuerpo, y se detectan aquellos anticuerpos que se unen específicamente al anticuerpo. Para detectar la unión al anticuerpo, se puede usar cualquiera de los procedimientos descritos anteriormente. Uno de tales ensayos de cribado se describe en "Methods for Measuring Cellulase Activities", Methods in Enzymology, Vol 160, p. 87-116.

El término "anticuerpo" incluye un péptido o polipéptido derivado de, modelado tras o codificado sustancialmente por un gen inmunoglobulínico o genes inmunoglobulínicos, o sus fragmentos, capaces de unirse específicamente a un antígeno o epítopo, véase, por ejemplo, Fundamental Immunology, Tercera Edición, W.E. Paul, ed., Raven Press, N.Y. (1993); Wilson (1994) J. Immunol. Methods 175:267-273; Yarmush (1992) J. Biochem. Biophys. Methods 25:85-97. El término anticuerpo incluye porciones que se unen a antígenos, es decir, "sitios de unión a antígenos" (por ejemplo, fragmentos, subsecuencias, regiones que determinan la complementariedad (CDRs)), que retienen la capacidad para unirse a antígeno, incluyendo (i) un fragmento Fab, un fragmento monovalente que consiste en los dominios VL, VH, CL y CH1; (ii) un fragmento F(ab')₂, un fragmento bivalente que comprende dos fragmentos Fab enlazados mediante un puente de disulfuro en la región bisagra; (iii) un fragmento Fb que consiste en los dominios VH y CH1; (iv) un fragmento Fv que consiste en los dominios VL y VH de un solo brazo de un anticuerpo, (v) un fragmento dAb (Ward et al., (1989) Nature 341: 544-546), que consiste en un dominio VH; y (vi) una región que

determina la complementariedad (CDR) aislada. Los anticuerpos monocatenarios también están incluidos mediante la referencia al término “anticuerpo”.

Kits

La descripción proporciona kits que comprenden las composiciones, por ejemplo ácidos nucleicos, casetes de expresión, vectores, células, semillas o plantas o partes de plantas transgénicas polipéptidos (por ejemplo, enzimas implicadas en el catabolismo de la clorofila o que tienen una actividad de esterasa (por ejemplo, clorofilasa)) y/o anticuerpos de la descripción. Los kits también pueden contener material de instrucción que enseña las metodologías y usos industriales de la descripción, como se describe aquí. En un aspecto, los kits se diseñan para adecuarse a los niveles de procesamiento a escala industrial, por ejemplo de alimentos, piensos, aceites, y similares.

Construcción de células completas y medición de parámetros metabólicos

Los métodos de la descripción proporcionan la evolución de células completas, o la construcción de células completas, de una célula para desarrollar una nueva cepa celular que tenga un nuevo fenotipo, por ejemplo una ruta de catabolismo de la clorofila modificada, o una actividad enzimática (por ejemplo, clorofilasa) nueva o modificada, modificando la composición genética de la célula. La composición genética se puede modificar mediante adición a la célula de un ácido nucleico de la descripción, por ejemplo una secuencia codificante para una enzima de la descripción. Véanse, por ejemplo, los documentos WO 0229032; WO 0196551.

Para detectar el nuevo fenotipo, al menos un parámetro metabólico de una célula modificada se monitoriza en la célula en un marco temporal a “tiempo real” o “en línea”. En un aspecto, una pluralidad de células, tal como un cultivo celular, se monitoriza en “tiempo real” o “en línea”. En un aspecto, una pluralidad de parámetros metabólicos se monitoriza en “tiempo real” o “en línea”. Los parámetros metabólicos se pueden monitorizar usando una enzima de la descripción.

El análisis de flujo metabólico (AFM) está basado en un marco bioquímico conocido. Se construye una matriz metabólica linealmente independiente basándose en la ley de conservación de la masa y en la hipótesis del estado pseudoestacionario (HESE) en los metabolitos intracelulares. Al poner en práctica los métodos de la descripción, se establecen redes metabólicas, incluyendo:

- la identidad de todos los sustratos, productos y metabolitos intermedios de la ruta
- la identidad de todas las reacciones químicas que interconvierten los metabolitos de las rutas, la estequiometría de las reacciones de la ruta,
- la identidad de todas las enzimas que catalizan las reacciones, la cinética de la reacción enzimática,
- las interacciones reguladoras entre los componentes de la ruta, por ejemplo interacciones alostéricas, interacciones enzima-enzima, etc.,
- la compartimentalización intracelular de las enzimas o cualquier otra organización supramolecular de las enzimas, y,
- la presencia de cualesquiera gradientes de concentración de metabolitos, enzimas o moléculas efectoras o barreras de difusión para su movimiento.

Una vez que se construye la red metabólica para una cepa dada, se puede introducir la presentación matemática mediante la noción de matriz para estimar los flujos metabólicos intracelulares si están disponibles los datos del metaboloma en línea. El fenotipo metabólico se basa en los cambios de la red metabólica completa en una célula. El fenotipo metabólico se basa en el cambio de utilización de la ruta con respecto a las condiciones ambientales, la regulación genética, el estado de desarrollo y el genotipo, etc. En un aspecto de los métodos de la descripción, después del cálculo del AFM en línea, el comportamiento dinámico de las células, su fenotipo y otras propiedades se analizan investigando la utilización de la ruta. Por ejemplo, si aumenta el suministro de glucosa y disminuye el oxígeno durante la fermentación de la levadura, la utilización de rutas respiratorias se reducirá y/o detendrá, y dominará la utilización de las rutas fermentativas. El control del estado fisiológico de los cultivos celulares se hará posible después del análisis de la ruta. Los métodos descritos aquí pueden ayudar a determinar cómo manipular la fermentación determinando cómo cambiar el suministro de sustrato, la temperatura, el uso de inductores, etc., para controlar el estado fisiológico de las células para moverse a lo largo de la dirección deseable. Al poner en práctica los métodos de la descripción, los resultados del AFM también se pueden comparar con datos del transcriptoma y proteoma para diseñar experimentos y protocolos de ingeniería metabólica o barajado de genes, etc.

Al poner en práctica los métodos de la descripción, se puede conferir y detectar cualquier fenotipo modificado o nuevo, incluyendo características nuevas o mejoradas en la célula. Se puede monitorizar cualquier aspecto del metabolismo o del crecimiento.

Monitorización de la expresión de un transcrito de ARNm

En un aspecto de la descripción, el fenotipo construido comprende incrementar o disminuir la expresión de un transcrito de ARNm (por ejemplo, un mensaje para un polipéptido implicado en el catabolismo de clorofila o que tiene actividad de esterasa (por ejemplo, clorofilasa)) o generar nuevos transcritos en una célula. Esta expresión aumentada o reducida se puede rastrear comprobando la presencia de una enzima de la descripción, o mediante ensayos de la actividad enzimática. Los transcritos de ARNm, o mensajes, también se pueden detectar y cuantificar por medio de cualquier método conocido en la técnica, incluyendo, por ejemplo, transferencias Northern, reacciones de amplificación cuantitativas, hibridación a matrices, y similares. Las reacciones de amplificación cuantitativas incluyen, por ejemplo, PCR cuantitativa, incluyendo, por ejemplo, reacción en cadena de la polimerasa mediante transcripción inversa cuantitativa, o RT-PCR; RT-PCR en tiempo real cuantitativa, o "RT-PCR cinética en tiempo real" (véanse, por ejemplo, Kreuzer (2001) Br. J. Haematol. 114:313-318; Xia (2001) Transplantation 72:907-914).

En un aspecto de la descripción, el fenotipo creado se genera desactivando la expresión de un gene homólogo. La secuencia codificante del gen o uno o más elementos de control de la transcripción se pueden desactivar, por ejemplo, promotores o potenciadores. De este modo, la expresión de un transcrito puede estar anulada completamente o solo disminuida.

En un aspecto de la descripción, el fenotipo construido comprende incrementar la expresión de un gen homólogo. Esto se puede efectuar desactivando un elemento de control negativo, incluyendo un elemento regulador de la transcripción que actúa en cis o trans, o mutagenizando un elemento de control positivo. Uno o más, o todos los transcritos de una célula se pueden medir mediante hibridación de una muestra que comprende transcritos de la célula, o ácidos nucleicos representativos de o complementarios a transcritos de una célula, mediante hibridación a ácidos nucleicos inmovilizados sobre una matriz.

Monitorización de la expresión de polipéptidos, péptidos y aminoácidos

En un aspecto de la descripción, el fenotipo creado comprende incrementar o disminuir la expresión de un polipéptido (por ejemplo, un polipéptido implicado en el catabolismo de clorofila o que tiene actividad de esterasa (por ejemplo, clorofilasa)), o generar nuevos polipéptidos en una célula. Esta expresión aumentada o disminuida se puede rastrear determinando la cantidad de enzima presente o mediante ensayos de la actividad. Los polipéptidos, péptidos y aminoácidos también se pueden detectar y cuantificar por medio de cualquier método conocido en la técnica, incluyendo, por ejemplo, resonancia magnética nuclear (RMN), espectrofotometría, radiografía (radiomarcado de proteínas), electroforesis, electroforesis capilar, cromatografía de líquidos de alta resolución (HPLC), cromatografía de capa fina (TLC), cromatografía de hiperdifusión, diversos métodos inmunológicos, por ejemplo inmunoprecipitación, inmunodifusión, inmunoelectroforesis, radioinmunoensayos (RIA), ensayo inmunoabsorbente ligado a enzima (ELISA), ensayos inmunofluorescentes, electroforesis en gel (por ejemplo, SDS-PAGE), tinción con anticuerpos, clasificador de células activadas por fluorescencia (FACS), pirólisis-espectrometría de masas, Espectrometría de Infrarrojos por Transformada de Fourier, espectrometría Raman, GC-MS, y LC-Electropulverización y espectrometrías de masas por electropulverización en tándem con cap-LC, y similares. También se pueden cribar bioactividades novedosas usando métodos, o variaciones de los mismos, descritos en la patente U.S. nº 6.057.103. Además, como se explica con detalle más abajo, uno o más, o todos los polipéptidos de una célula se pueden medir usando una matriz de proteínas.

Enzimas

La descripción proporciona nuevas composiciones y métodos para tratar enzimáticamente, por ejemplo decolorar o "blanquear", preparaciones de algas, de animales (por ejemplo, peces) y/o vegetales, piensos, alimentos o aceites que comprenden clorofila (la clorofila puede estar en las preparaciones, piensos, alimentos o aceites de forma natural, como un contaminante, como una composición indeseada en un producto procesado, etc.). En un aspecto, las composiciones que contienen clorofila o contaminadas con clorofila, por ejemplo preparaciones de algas, de animales o vegetales, piensos, alimentos o aceites, se tratan enzimáticamente usando una clorofilasa o enzima equivalente. En una composición o método de la descripción se puede usar cualquier polipéptido que tenga una actividad que pueda modificar una clorofila o metabolito de la clorofila.

Clorofilasas

Los polipéptidos y/o péptidos de la descripción pueden tener actividad de esterasa, por ejemplo una clorofilasa o una actividad similar. Los polipéptidos y/o péptidos de la descripción pueden incluir anticuerpos catalíticos, enzimas, sitios activos, y similares. Estos polipéptidos y/o péptidos de la descripción que tienen actividad de esterasa (por ejemplo, clorofilasa) se pueden usar en composiciones o métodos de la descripción. Por ejemplo, en un aspecto, las composiciones y métodos de la descripción tratan enzimáticamente composiciones que contienen clorofila o contaminadas con clorofila al hidrolizar la clorofila (Figura 1A) hasta fitol (Figura 1B) y clorofilida (Figura 1C).

Se puede usar en una composición o método de la descripción cualquier clorofilasa, clase o clorofila clorofilido-hidrolasa o polipéptido que tenga una actividad similar (por ejemplo, clorofila-clorofilido hidrolasa 1 o clase 1, o clorofila-clorofilido hidrolasa 2 o clase 2, véanse, por ejemplo, NCBI P59677_1 y P59678, respectivamente). En una composición o método de la descripción se puede usar cualquier polipéptido (por ejemplo, enzima o anticuerpo

catalítico) que catalice la hidrólisis de un enlace de éster de clorofila para producir clorofilida y fitol. Cualquier polipéptido aislado, recombinante o sintético o quimérico (una combinación de sintético y recombinante) (por ejemplo, enzima o anticuerpo catalítico) se puede usar, por ejemplo se puede usar una clorofilasa, clase o clorofila clorofilido-hidrolasa o polipéptido que tenga una actividad similar en una composición o método de la descripción; véase, por ejemplo, Marchler-Bauer (2003) *Nucleic Acids Res.* 31: 383-387.

En un aspecto, las composiciones y métodos de la descripción se pueden poner en práctica con enzimas como se describe en el documento WO 0229022. Por ejemplo, en un aspecto, las composiciones y métodos de la descripción pueden comprender la expresión recombinante de enzimas, por ejemplo clorofilasas, tales como polinucleótidos que codifican clorofilasas. En un aspecto, el ácido nucleico recombinante se expresa en células completas, extractos celulares o *in vitro*. En un aspecto, el polinucleótido codificante de la enzima se modifica para dar como resultado la producción de niveles alterados de enzima (por ejemplo, clorofilasa) en una célula hospedante transformada.

En un aspecto, las composiciones y métodos de la invención se pueden poner en práctica con enzimas conocidas, tales como clorofilasas (incluyendo clases y clorofila clorofilido-hidrolasas) y polipéptidos relacionados bien conocidos en la técnica. Por ejemplo, se puede usar la clorofilasa de *Arabidopsis thaliana*, como se describe, por ejemplo, en la entrada NCBI NM_123753 (en la que la enzima que tiene una secuencia como se expone en SEC ID NO:22 es codificada, por ejemplo, por SEC ID NO:21):

SEC ID NO:21

```

AAAAAAGTAAAGAAAAGAAAACTAATAAAGAACAAAAAAATGTCTCTTCTTCATCAAGAAACGCC
TTTGAAGATGGCAAATACAAATCAAATCTCTTAACCTGGACTCATCATCTCGTTGCTGCAAAATAACA
CCGTCTTCTAGAGCTTCACCGTCTCCGCCAAAGCAGCTGTTGGTGGCTACGCCGGTGGAGGAAGGAGAT
TATCCGGTGGTGATGCTCCTCCATGGTTACCTTCTCTACAACCTCCTTCTATTCTCAGCTTATGTTGCAT
GTCTCTTCTCATGGCTTCATCCTCATCGCTCCTCAGTTATATAGTATCGCCGGACAGACACAATGGAT
GAGATTAAATCAACGGCGGAGATTATGGATTGGTTATCAGTAGGACTTAATCACTTTCTTCCAGCGCAA
GTAACACCAAACCTATCCAAATTTGCCCTCTCCGGCCATAGCCGCGGTGGCAAAACCGCGTTTGCGGTC
GCCTTAAAGAAATTTGGGTACTCCTCGAATCTAAAGATCTCGACATTGATCGGTATAGATCCAGTCGAT
GGAACAGGGAAAGGGAAACAAACCCCTCCTCCGGTGTGGCTTACCTTCCAAACTCATTTGACCTAGAC
AAAACGCCCTATACTTGTGATCGGTTCCGGGCTTGGTGAAACCGCTCGGAACCCATTATTCCACCGTGT
GCACCTCCCGGAGTGAATCACCGAGAGTTCTTTCCGGAATGTCAAGGTCCAGCATGGCATTTCGTTGCG
AAGGATTATGGGCATTTGGACATGCTTGATGATGATACAAAAGGGATTAGAGGGAAGAGTTCTTATTGT
TTGTGTAAGAATGGTGAAGAGAGGAGACCAATGAGGAGATTCGTTGGTGGACTTGTGTATCATTTTTG
AAGGCTTATTTGGAAGGAGATGATCGTGAATTAGTTAAGATCAAAGATGGGTGTCACGAGGATGTTCCC
GTTGAAATTCAAGAGTTTGAGGTTATCATGTAAACATAAGTTTTCTTTAGGGGCTGGTTTTTCTATTG
TCAATATCATCAGCTTTTGTGCTTATGGTTTTACAACTTATATTGTACAACCTTTTAAGTCACCTCT
TTGCTTATGATATTAACCCGATC

```

SEC ID NO:22

```

MSSSSSRNAFEDGKYKSNLLTLDSSSRCKITPSSRASPSPPKQLLVATPVEEGDYPVVMLLHGYLLYN
SFYSQMLMLHVSSHGFILIAPLYSIAGPDTMDEIKSTAEIMDWLSVGLNHFLPAQVTPNLSKFALSGHS
RGGKTAFVAVALKKFGYSSNLKISTLIGIDPVDGTGKGKQTPPPVLAYLPNSFDLDKTPILVIGSGLGET
ARNPLFPFCAPPVGNHREFFRECQGPWFHFAKDYGHLDMLDDDTKGIRGKSSYCLCKNGEERRPMRRF
VGGLVVSFLKAYLEGDDRELVKIKDGCHEDVPVEIQEFVIM

```

La clorofilasa de *Ginkgo biloba* se puede usar como se describe, por ejemplo, en la entrada NCBI AY292526:

SEC ID NO:23

TTGAAAAACAAAAACGAAGAAGATGAACTCAGTACTTGCACACAGCCATCGGCCATGGTTTTAGTGAAG
GATGTGTTTCAGCGAAGGTCTTTTACCTGTTCAAATCCTCGCAATTCCACAAGCCAACTCATCTCCATGC
TCAAAATTAGCAGACAAAAACGGAAGTGAACACAGCCTTCTCCTTGTGCGCCTCTTAAACCCCTGCTG
ATCGCTCTTCTTCCCAACATGGAGATTATCCTCTCATCCTCTTTTTCCACGGCTATGTACTCTCAAT
TCCTTCTATTCTCAACTCTTGCGCCATGTTGCTTCCCATGGATACATCGCCATAGCTCCTCAGATGTAC
AGTGTAATTGGCCCCAAATACGACTCCAGAAATAGCCGATGCAGCGGCCATTACAGACTGGTTACGAGAT
GGACTCTCGGATAATCTTCCGCAAGCTTTAAACAATCATGTGAGGCCCAATTTTGAGAAATTTGTGCTA
GCGGGGCACTCGCGCGGGGTAAAGTGGCATTGCACTTGCCCTAGGTGAGTCTCGCAGCCATCTTTA
AAGTACTCGGCCCTTGTAGGTCTTGATCCAGTCGATGGAATGGGAAAAGATCAACAAACCAGTCATCCT
ATTCTGTCATACAGAGAGCATTCTTTGATTGTTGGGTATGCCAACATTAGTGGTAGGTTGCGGCCTGGGT
CCGTGCAAAAGAAACCTCTCTTCCCTCCCTGTGCTCCCCAAGGTGTTAACCACCATGATTTCTTCTAC
GAATGTGTCGCTCCTGCCTATCATTTTGTGCTCTGATTATGGGCATCTTGATTTCTTAGACGACGAC
ACCAAAGGAATAAGAGGAAAGGCTACTTATTGCCTCTGTAAGAATGGGGAAGCAAGAGAGCCAATGCGG
AAGTTTAGCGGTGGAATTGTGGTTGCATTTCTTCAAGCATTCTTGGTGATAATCGTGGAGCCCTGAAT
GATATTATGGTTTATCCTTCACATGCTCCAGTCAAGATTGAGCCTCCAGAGTCTTTGGTTACAGAAGAT
GTAAAAATCCCCAGAAGTCAATTATTACGCCGGGCAGTTTGCAGATGATGTACCATGGTATTATGCATT
AAAGGAATGTATTTGTTATTAAAAAAATATTAAGAAGTAAAAAAAAAAAAAAAAA

SEC ID NO:24

MVLVKDVFSEGLPLPVQILAI PQANSSPCSKLADKNGTATTPSPCRPPKPLLIALPSQHGDYPLILFFHG
YVLLNSFYSQLLRHVASHGYIAIAPQMYSVIGPNTTPEIADAAAITDWLRDGLSDNLPQALNNHVRPNF
EKFVLAGHSRGGKVAFALALGRVSQPSLKYSALVGLDPVDGMGKDQQTSHPILSYREHSFDLGMPTLVV
GSGLGPKRNPLFPFCAPQGVNHHDFYECVAPAYHFVASYDGHLDLDDDTKGIRKATYCLCKNGEA
REPMRKFSGGIVVAFQLAFLGDNRGALNDIMVYP SHAPVKIEPPESLVTEDEVKSPEVELLRRAVCR

La clorofilasa de *Brassica oleracea* se puede usar como se describe, por ejemplo, en la entrada NCBI AF337546:

5

SEC ID NO:25

ACACAAAAAATATATAACACAAAGAAATAGAAGAAGGAAAAAATGTCCCCCTCCTTTCTTTTCTTTAC
TTTGTTTTTTGATAAGGAAATGTCTCTTCATCATCAGCAAACCTCTTTGAGGACGGCAAATACAAAAC
AGATCTTTTAAAGTAGGCTTATCATCTTGCTGCTGGAAAAAGCCCTCCTCTTCTCCGACTCCGCAGTC
TCCGCCGAAGAGGCTTTTGGTGGCAACGCCGCTGGAGGAAGGAGAATATCCGGTGGTGATGCTCCTCCA
TGGTTACCTTCTCTACAACTCATTTTATCCCAGCTTATGTTGCATGTCTTCCCATGGCTTCATTGT
CATCGCTCCGAGTTATATAGCATTGCCGGACCAGACACCATGGATGAGATAAAATCAACGGCAGAGAT
TATTGATTGGTTATCGGTTCGACTAAACCACTTTCTTCCACCACAAGTAACACCAAACCTATCCAAGTT
CGCACTCTCCGGCCATAGCCGTGGTGGGAAGACCGCATTTGCCTTGGCCTTAAAGAAATTTGGTACTC
GTCCGACCTAAAGATCTCGGCATTGATAGGTATAGATGTTGGAAGTGTGTTTTGGACAAATGGCTATGG
CCAATATTCGGGTGAATTTTTCGAGCAATTTGATTGTCGAAATGACCGGATTGTGGAATCGTAGGATTC
ATTGTTATGAGCACTATGGTATAGTGAATCATATATCAAAAACGAAGTTCGTTTGAATGAGAAATGAA
AGTCTAAAATAGATTATTTGTAAAATATCTATATTAGAATTATGAGGTAAGAAACCTCTGTGTGTTAAA
ATGGAGAAGTTATAACAAAGTTATAAAAAACTTTGTAAACAATTTGGTGTGTTAGC

SEC ID NO:26

MSPSFLFFTLFLIKEMSSSSSANSFEDGKYKTDLLTVGLSSCCWKKPSSSPTPQSPPKRLLVATPVEEG
EYPVVMLLHGYYLLNSFYSQLMLHVSSHGFIVIA PQLYSIAGPDTMDEIKSTAEIIDWLSVGLNHFLLPP
QVTNLSKFALSGHSRGGKTAFAALALKKFGYSSDLKISALIGIDVGTVFWTNGYGYSGEFFFQFDCRN
DRIVES

10

La clorofilasa de *Citrus sinensis* se puede usar como se describe, por ejemplo, en la entrada NCBI Q9MV14:

SEC ID NO:27

MAAMVDAKPAASVQGTPLLATATLPVFTRGIYSTKRITLETSSPSSPPPKPLIIVTPAGKGTFFNVILF
 LHGTSLSNKSYSKIFDHIASHGFIVVAPQLYTSIPPPSATNELNSAAEVAEWLPQGLQQNLPEANV
 SLVAVMGHSRGGQTAFALSLRYGFGAVIGLDPVAGTSKTTGLDPSILSFDSFDFSIPVTVIGTGLGGVA
 RCITACAPEGANHEEFFNRCKNSSRAHFVATDYGHMDILDDNPSDVKSWALSKYFCKNGNESRDPMRRC
 VSGIVVAFLKDFFYGDAEDFRQILKDPSFAPIKLDSEYIDASSMLTTTHVKV

Preparaciones enzimáticas

Las enzimas usadas en los métodos de la invención se pueden formular o modificar, por ejemplo modificar químicamente, por ejemplo, para potenciar la solubilidad del aceite, la estabilidad, la actividad, o para la inmovilización. Por ejemplo, las enzimas usadas en los métodos de la invención se pueden formular para ser anfipáticas o más lipófilas. Por ejemplo, las enzimas usadas en los métodos de la invención se pueden encapsular, por ejemplo en liposomas o geles, por ejemplo hidrogeles de alginato o perlas de alginato o equivalentes. Las enzimas usadas en los métodos de la invención se pueden formular en sistemas micelares, por ejemplo un medio de sistema micelar ternario (TMS) o sistema micelar inverso (RMS). Las enzimas usadas en los métodos de la invención se pueden formular como se describe en Yi (2002) J. of Molecular Catalysis B: Enzymatic, Vol. 19, nº 0, p. 319-325. Por ejemplo, la enzima anfipática, por ejemplo clorofilasa, en forma de un medio de sistema micelar ternario (TMS) o de sistema micelar inverso (RMS) se puede encapsular en hidrogeles de alginato. En un aspecto, una enzima, por ejemplo una clorofilasa, se prepara en tampón acuoso y se retiene en un hidrogel, por ejemplo TMS/alginato y RMS/alginato. Un enfoque para encapsular enzima, por ejemplo clorofilasa, puede ser el emulsionamiento y/o la gelificación interna del sistema enzima-TMS o RMS.

Las reacciones enzimáticas de los métodos de la invención se pueden llevar a cabo *in vitro*, incluyendo, por ejemplo, matrices capilares, como se explica más abajo, o en sistemas de células completas. En un aspecto, las reacciones enzimáticas de los métodos de la invención se llevan a cabo en una vasija de reacción o en múltiples vasijas. En un aspecto, las reacciones enzimáticas de los métodos de la invención se llevan a cabo en un aparato de refinado de aceite vegetal.

Las composiciones y métodos de la invención se pueden poner en práctica con enzimas inmovilizadas, por ejemplo clorofilasa inmovilizada. La enzima se puede inmovilizar sobre cualquier soporte orgánico o inorgánico. Los soportes inorgánicos ejemplares incluyen alúmina, celita, Dowex-1-cloruro, perlas de vidrio y gel de sílice. Los soportes orgánicos ejemplares incluyen hidrogeles de alginato o perlas de alginato o equivalentes.

En diversos aspectos de la descripción, la inmovilización de la clorofilasa se puede optimizar mediante adsorción física sobre diversos soportes inorgánicos, incluyendo alúmina, celita, Dowex-1-cloruro, perlas de vidrio y gel de sílice. Las enzimas usadas para la práctica de la invención se pueden inmovilizar en diferentes medios, incluyendo agua, disolución de tampón Tris-HCl, y un sistema micelar ternario que contiene disolución de tampón Tris-HCl, hexano y tensioactivo. Se obtuvo la mayor eficiencia de inmovilización (84,56%) y la mayor actividad específica (0,34 μmol de clorofila hidrolizada mg proteína⁻¹ por minuto) cuando se suspendió clorofilasa en disolución de tampón Tris-HCl y se adsorbió sobre gel de sílice.

Aplicaciones industriales y médicas

Los polipéptidos, por ejemplo enzimas de la descripción implicadas en el catabolismo de la clorofila o que tienen una actividad de esterasa (por ejemplo, clorofilasa), se pueden usar en una variedad de aplicaciones médicas industriales, como se describe aquí. Las composiciones y métodos de la descripción se pueden usar junto con cualquier uso industrial o aplicación farmacéutica o médica para el tratamiento de materiales que contienen clorofila, por ejemplo preparaciones vegetales, materiales que comprenden aceite. Por ejemplo, las composiciones y métodos de la invención se pueden usar con procedimientos para convertir un fosfolípido no hidratable en una forma hidratable, para el desgomado del aceite, para el procesamiento de aceites procedentes de plantas, pescado, algas y similares, por nombrar sólo unas pocas aplicaciones. Por ejemplo, los métodos de la descripción se pueden usar con el procesamiento de grasas y aceites como se describe, por ejemplo, en la Publicación de Solicitud de Patente JP H6-306386, que describe la conversión de fosfolípidos presentes en los aceites y grasas en sustancias solubles en agua que contienen grupos ácido fosfórico.

Las composiciones y métodos de la invención se pueden usar conjuntamente con métodos para procesar aceites vegetales, tales como los derivados de o aislados de salvado de arroz, soja, cáñola, palma, semilla de algodón, maíz, pepita de palma, coco, cacahuete, sésamo, girasol. Las composiciones y métodos de la invención se pueden usar conjuntamente con métodos para procesar aceites esenciales, por ejemplo aquellos procedentes de aceites de semillas de frutas, por ejemplo uva, albaricoque, borraja, etc. Las composiciones y métodos de la invención se pueden usar conjuntamente con métodos para procesar aceites y fosfolípidos en formas diferentes, incluyendo formas brutas, desgomadas, gomas, agua de lavado, arcilla, sílice, grasa para jabón, y similares. Las composiciones y métodos de la invención se pueden usar conjuntamente con métodos para procesar aceites con alto contenido en fósforo (por ejemplo, un aceite de haba de soja), aceites de pescado, aceites de animales, aceites vegetales, aceites de algas, y similares.

Las composiciones y métodos de la invención se pueden usar conjuntamente con métodos usar para procesar y obtener aceites comestibles, aceites biodiésel, liposomas para fármacos y cosméticos, fosfolípidos estructurados y lípidos estructurados. Las composiciones y métodos de la invención se pueden usar conjuntamente con métodos para la extracción de aceites. Las composiciones y métodos de la invención se pueden usar conjuntamente con métodos para obtener diversos jabones.

Los métodos pueden comprender modificar el pH (por ejemplo, incrementar el pH) para promover la separación acuosa de clorofilida. De este modo, las composiciones y métodos de la invención también pueden comprender procedimientos de neutralización cáustica, por ejemplo, con condiciones de pH cáustico-neutro. En un aspecto, las composiciones y métodos de la descripción comprenden una etapa de neutralización, por ejemplo, en el tratamiento de “aceites químicamente refinados”, por ejemplo usando clorofilasas, y/o en la separación de clorofilida. Las composiciones y métodos de la descripción pueden comprender modificar el pH para promover la separación acuosa de clorofilida.

En un aspecto, las composiciones y métodos de la descripción comprenden el uso de dispositivos y procedimientos de refinado con sílice adsorbente reducida o libres de adsorbente, que se conocen en la técnica, por ejemplo usando TriSyl Silica Refining Processes (Grace Davison, Columbia, MD), o sílices SORBSIL R™ (INEOS Silicas, Joliet, IL).

Tratamiento enzimático, o procedimientos de “blanqueo” o decoloración

La descripción proporciona nuevas composiciones y métodos para tratar enzimáticamente, por ejemplo decolorar o “blanquear”, preparaciones de algas, de animales (por ejemplo, pescado) y/o vegetales, piensos, alimentos o aceites, como se ilustra en las Figuras 8 a 16. En un aspecto, se tratan alimentos o aceites que contienen clorofila o contaminados con clorofila. Por ejemplo, en aspectos alternativos, los aceites vegetales, incluyendo aceites procesados de oleaginosas, tales como aceite de cáñola (colza) o aceite de haba de soja, o frutos oleosos, tales como aceite de palma, son procesados usando las composiciones y/o métodos de la descripción.

Al menos una etapa en este método ejemplar implica el uso de una enzima, por ejemplo una enzima clorofilasa, que puede hidrolizar clorofila a fitol y clorofilida. En aspectos alternativos, una, varias o todas las etapas usan una enzima. La reacción puede ser *in vitro* o *in vivo*.

La Figura 8 ilustra la reacción de una esterasa ejemplar de la descripción en la degradación de la clorofila – la clorofilasa (clasa) cataliza la hidrólisis de un enlace de éster en la clorofila para producir clorofilida y fitol, en el que la clorofilida entra en la fase acuosa debido a un anillo de porfirina hidrófilo, y el fitol se separa en una fase oleosa (hidrófoba). En un aspecto de un procedimiento de la descripción, el anillo de porfirina hidrófilo se separa con la fracción de goma/agua usando uno cualquiera de los muchos métodos bien conocidos.

La Figura 9 ilustra y compara el procedimiento de decoloración (blanqueo) enzimático tradicional frente a uno ejemplar de la descripción, en el que el procedimiento de blanqueo enzimático puede incorporar una esterasa de la descripción. En el método tradicional, el aceite vegetal bruto se desgoma (opcionalmente, se neutraliza de forma cáustica), se blanquea usando, por ejemplo, adsorción en arcilla con eliminación subsiguiente de la arcilla, y desodorización, para producir aceite “refinado, blanqueado y desodorizado” o RBD. En este procedimiento de blanqueo enzimático ejemplar de la descripción, el aceite vegetal bruto se desgoma (opcionalmente, se neutraliza de forma cáustica), se blanquea usando, por ejemplo, un polipéptido de la descripción, tal como una clorofilasa de la descripción, con separación acuosa subsiguiente de la clorofilida, seguido de la desodorización, para producir un aceite “refinado, blanqueado y desodorizado” o RBD. La necesidad del desgomado depende del contenido de fósforo y de otros factores (todos ellos conocidos en la técnica). Típicamente se desgoman soja y cáñola.

La Figura 10 ilustra un procedimiento de decoloración (blanqueo) enzimático ejemplar de la descripción – un procedimiento combinado de desgomado-blanqueo (“decoloración”). En este procedimiento de blanqueo enzimático ejemplar de la descripción, el aceite vegetal bruto se desgoma y se blanquea enzimáticamente usando un polipéptido de la descripción, tal como una esterasa, por ejemplo una clorofilasa, de la descripción en una etapa, o “un recipiente”. El desgomado puede ser un desgomado “tradicional” o un desgomado enzimático, por ejemplo que implica fosfolípido o fosfolípidos y/o hidrólisis. En un aspecto, el procedimiento ejemplar de la descripción comprende una etapa de separación acuosa subsiguiente para eliminar el producto de reacción clorofilida, goma y/o jabón. En un aspecto, esto es seguido de la desodorización para producir un aceite “refinado, blanqueado y desodorizado” o RBD.

La Figura 11 ilustra un procedimiento de decoloración (blanqueo) enzimático ejemplar de la descripción que combina etapas de desgomado, blanqueo enzimático (“decoloración”) y neutralización cáustica. En este procedimiento de blanqueo enzimático ejemplar de la descripción, el aceite vegetal bruto se desgoma, se neutraliza y se blanquea enzimáticamente usando un polipéptido de la descripción, tal como una esterasa, por ejemplo una clorofilasa, de la descripción en una etapa, o “un recipiente”. El desgomado puede ser un desgomado “tradicional” o un desgomado enzimático, por ejemplo que implica fosfolípido o fosfolípidos y/o hidrólisis. En un aspecto, el procedimiento ejemplar de la descripción comprende una etapa de separación acuosa subsiguiente para eliminar el producto de reacción clorofilida, goma y/o jabón.

La Figura 12 ilustra un procedimiento de decoloración (blanqueo) enzimático ejemplar de la descripción que

comprende la aplicación de un polipéptido de la descripción, tal como una esterasa, por ejemplo una clorofilasa, a una preparación de oleaginosas, seguido de una etapa de separación acuosa subsiguiente (para eliminar, por ejemplo, el producto de reacción clorofilada, o gomas y/o jabones), seguido de los procedimientos ilustrados en las Figuras 9, 10 u 11.

La Figura 13 ilustra un esquema de refinado de oleaginosas general que comprende la extracción, refinado y modificación de una oleaginosas, en el que, además de un polipéptido de la descripción, tal como una esterasa, por ejemplo una clorofilasa, a una oleaginosas en una o varias o todas estas etapas, también se añaden otros polipéptidos y/o sustancias químicas, por ejemplo celulasa, hemicelulasa, proteasa, pectinasa, fosfolipasa A, B, C y/o D, esterasa (por ejemplo, una esterasa selectiva), una lipasa (por ejemplo, 1,3 lipasa), una lipasa selectiva, una clorofilasa conocida, u otra enzima implicada en el catabolismo de la clorofila, y similar.

La Figura 14 ilustra un procedimiento industrial ejemplar de la descripción – un procedimiento de biodesgomado, que comprende el uso de una fosfolipasa A y al menos un polipéptido de la descripción que tiene actividad de enzima clorofilasa. El al menos un polipéptido de la descripción que tiene actividad de clorofilasa se puede añadir a una o a varias o a todas las siguientes etapas: se puede añadir al aceite bruto, en el procedimiento de desgomado o en el aceite desgomado, a un tanque de almacenamiento o de retención, con la fosfolipasa A (por ejemplo, en “el tanque de día” de la figura) y/o el tanque cáustico.

La Figura 15 ilustra otro procedimiento industrial ejemplar de la descripción que comprende el uso de al menos un polipéptido de la descripción que tiene actividad de enzima clorofilasa. El al menos un polipéptido de la descripción que tiene actividad de clorofilasa se puede añadir a una o a varias o a todas las siguientes etapas: se puede añadir al aceite bruto, en el procedimiento de desgomado o en el aceite desgomado, a un tanque de almacenamiento o de retención, a un tanque cáustico y/o a una mezcladora de retención.

La Figura 16 ilustra otro procedimiento industrial ejemplar de la descripción que comprende el uso de al menos un polipéptido de la descripción que tiene actividad de enzima clorofilasa. En este procedimiento ejemplar, se añade fosfolipasa C (PLC) al procedimiento de desgomado o en el aceite desgomado, con la enzima clorofilasa de la descripción. El al menos un polipéptido de la descripción que tiene actividad de clorofilasa se puede añadir a una o a varias o a todas las siguientes etapas: se puede añadir al aceite bruto, en el procedimiento de desgomado o en el aceite desgomado (con una PLC), a un tanque de almacenamiento o de retención, a un tanque cáustico y/o a una mezcladora de retención.

Desgomado de aceites y procesamiento del aceite vegetal

Las composiciones y métodos de la descripción se pueden usar en diversas etapas del procesamiento de aceites vegetales, tales como en la extracción de aceites vegetales, particularmente en la eliminación de “gomas de fosfolípidos” en un procedimiento denominado “desgomado de aceites”.

Las composiciones y métodos de la descripción se pueden usar en métodos para procesar aceites vegetales a partir de diversas fuentes, tales como salvado de arroz, habas de soja, colza, cacahuetes y otras nueces, sésamo, girasol, palma y maíz. Los métodos se pueden usar conjuntamente con procedimientos basados en la extracción como con hexano, con el refinado subsiguiente de los extractos brutos a aceites comestibles. La primera etapa en la secuencia de refinado es el proceso denominado “desgomado”, que sirve para separar fosfátidos mediante adición de agua. El material precipitado mediante desgomado se separa y se procesa posteriormente a mezclas de lecitinas. Las lecitinas comerciales, tales como lecitina de haba de soja y lecitina de girasol, son materiales semisólidos o muy viscosos. Consisten en una mezcla de lípidos polares, principalmente fosfolípidos, y aceite, principalmente triglicéridos. Las composiciones y métodos de la descripción se pueden usar antes o después de cualquier etapa en un procedimiento, o antes o después de cualquier combinación de etapas, o antes o después de todas las etapas, en un procedimiento, por ejemplo antes, durante o tras la extracción mecánica o química, el desgomado y/o blanqueo, y similar.

Las composiciones y métodos de la descripción se pueden usar en (es decir, conjuntamente con) cualquier procedimiento de “desgomado”, incluyendo desgomado con agua, desgomado con aceite ALCON (por ejemplo, para habas de soja), desgomado safinco “superdesgomado”, desgomado UF, desgomado TOP, unidesgomado, desgomado en seco y desgomado con ENZYMAX™. Véanse, por ejemplo, las patentes U.S. n^{os} 6.355.693; 6.162.623; 6.103.505; 6.001.640; 5.558.781; 5.264.367. Las composiciones y métodos de la descripción se pueden usar en cualquier método de procedimiento de aceite, por ejemplo desgomado o procesos equivalentes. Por ejemplo, las composiciones y métodos de la descripción se pueden usar en procedimientos como se describen en las patentes U.S. n^{os} 5.558.781; 5.288.619; 5.264.367; 6.001.640; 6.376.689; en el documento WO 0229022; el desgomado de aceites como se describe, por ejemplo, en el documento WO 98/18912; procedimientos como se describen en la solicitud JP n^o H5-132283 (presentada el 25 de abril de 1993); la solicitud EP número: 82870032.8, y similares. Diversos procedimientos de “desgomado”, incorporados por los métodos de la descripción, se describen en Bockisch, M. (1998) en Fats and Oils Handbook, The extraction of Vegetable Oils (Capítulo 5), 345-445, AOCS Press, Champaign, Illinois. Las composiciones y métodos de la descripción se pueden usar en la aplicación industrial de desgomado enzimático de aceites triglicéridos, como se describe, por ejemplo, en el documento EP 513.709.

En un aspecto, las composiciones y métodos de la descripción se usan para tratar aceites vegetales, por ejemplo aceites brutos, tales como salvado de arroz, soja, cáñola, girasol y similares. En un aspecto, esto mejora la eficiencia del procedimiento de desgomado. En un aspecto, los métodos de la descripción dan como resultado la separación mejorada de la clorofila desde la fase oleosa, por ejemplo durante la centrifugación. La separación mejorada de estas fases puede dar como resultado una eliminación más eficiente de las clorofilas desde el aceite, incluyendo aceites tanto hidratables como no hidratables.

Las composiciones y métodos de la descripción se pueden usar en la aplicación industrial del desgomado enzimático como se describe, por ejemplo, en el documento CA 1102795, que describe un método para aislar lípidos polares a partir de lípidos de cereales mediante la adición de al menos 50% en peso de agua. Este método es un desgomado modificado, en el sentido de que utiliza el principio de añadir agua a una mezcla de aceite bruta.

En un aspecto, la descripción proporciona procedimientos enzimáticos que comprenden el uso de composiciones y métodos de la descripción que comprenden la hidrólisis de fosfolípidos hidratados en un aceite a una temperatura de alrededor de 20°C a 40°C, a un pH alcalino, por ejemplo un pH de alrededor de pH 8 a pH 10, usando un tiempo de reacción de alrededor de 3 a 10 minutos.

En diversos procedimientos ejemplares de la descripción, un número de etapas distintas componen el procedimiento de desgomado que antecede a los procedimientos de blanqueo del núcleo y refinado con desodorización. Estas etapas incluyen calentar, mezclar, retener, separar y secar. Tras la etapa de calentamiento, se añade agua y a menudo ácido, y se mezclan para permitir que la "goma" de fosfolípido insoluble se aglomere en partículas que se pueden separar. Mientras que el agua separa muchos de los fosfátidos en el desgomado, las porciones de los fosfolípidos son fosfátidos no hidratables (NHPs) presentes como sales de calcio o magnesio. Los procedimientos de desgomado se dirigen a estos NHPs mediante la adición de ácido. Tras la hidratación de los fosfolípidos, el aceite se mezcla, se retiene y se separa mediante centrifugación. Finalmente, el aceite se seca y se almacena, se transporta o refina. Las gomas resultantes se procesan posteriormente para productos de lecitina, o se añaden nuevamente a la harina. Como se señala anteriormente, las composiciones y métodos de la descripción se pueden usar antes o después de cualquiera de estas etapas, o antes o después de cualquier combinación de etapas, o antes o después de todas las etapas, en cualquier método de procesamiento.

Al terminar un tratamiento enzimático de la descripción, el líquido tratado (por ejemplo, aceite) se separa con un medio apropiado, tal como un separador centrífugo, y se obtiene el aceite procesado. En un aspecto, los compuestos, producidos mediante modificación enzimática de la clorofila, se transfieren parcial o totalmente a la fase acuosa y se eliminan de la fase oleosa. Al terminar el tratamiento enzimático, si es necesario, el aceite procesado se puede lavar adicionalmente con agua o con ácido orgánico o inorgánico tal como, por ejemplo, ácido acético, ácido fosfórico, ácido succínico, y similar, o con disoluciones salinas.

En un procedimiento ejemplar para el desgomado mediante ultrafiltración, una enzima usada en un método de la descripción se une a un filtro, o la enzima se añade a un aceite antes de la filtración. Las enzimas usadas en composiciones o métodos de la descripción se pueden inmovilizar a cualquier sustrato, por ejemplo filtros, fibras, columnas, perlas, coloides, geles, hidrogeles, mallas, y similares.

Las composiciones y métodos de la descripción se pueden usar para mejorar la extracción de aceite, el desgomado del aceite y la neutralización cáustica del aceite (por ejemplo, aceites vegetales). En un aspecto, en un procedimiento de la descripción, se usa una composición o método de la descripción y al menos un agente degradante de la pared de la célula vegetal (por ejemplo, una celulasa, una hemicelulasa o similar, para reblandecer las paredes e incrementar el rendimiento de la extracción). En un método ejemplar, para mejorar la extracción de aceite y el desgomado del aceite, se usa una fosfolipasa, por ejemplo una fosfolipasa C, u otra hidrolasa (por ejemplo, una celulasa, una hemicelulasa, una esterasa, una proteasa y/o una fosfatasa). Por ejemplo, en un aspecto, durante una etapa de trituración asociada con la producción de aceite (incluyendo, pero sin limitarse a, aceite de haba de soja, de cáñola, de girasol, de salvado de arroz), se puede usar una fosfolipasa u otra enzima. Usando enzimas antes de o en lugar de la extracción con disolventes, es posible incrementar el rendimiento del aceite y reducir la cantidad de fosfolípidos hidratables y no hidratables en el aceite bruto. La reducción global de fosfolípidos en el aceite bruto dará como resultado rendimientos mejorados durante el refinado, con el potencial de eliminar el requisito de una etapa de desgomado adicional antes del blanqueo y desodorización.

Las composiciones y métodos de la descripción también se pueden poner en práctica usando procedimientos como se describen en la patente U.S. n° 5.414.100. Por ejemplo, en un aspecto, los métodos o composiciones comprenden además procedimientos cromatográficos para la desacidificación de aceites vegetales a temperatura ambiente. Estos procedimientos se pueden retroajustar en operaciones de desacidificación usando refinado con miscelas o extracción con disolventes, el aceite vegetal bruto se disuelve en un disolvente tal como alcohol isopropílico, y se hace pasar a través de una columna de alúmina activada (óxido de aluminio) a temperatura ambiente. El procedimiento, que elimina el contacto físico entre tanto el aceite y un reactivo alcalino como el aceite y agua, simplifica procedimientos de blanqueo subsiguientes al eliminar también algunos pigmentos de color. La alúmina gastada se puede reactivar lavándola con una disolución diluida de hidróxido de sodio o hidróxido de potasio.

Las composiciones y métodos de la descripción también se pueden poner en práctica usando procedimientos como se describen en el documento JP57156482, 1982 (solicitud nº JP19810040794 19810320), que describe el refinado de grasas o aceites vegetales como subproductos.

Las composiciones y métodos de la descripción también se pueden poner en práctica usando procedimientos como se describen en la patente U.S. nº 5.315.021. Por ejemplo, en un aspecto, los métodos o composiciones de la descripción se pueden poner en práctica con procedimientos para eliminar impurezas de color de la clorofila a partir de aceites vegetales. Los procedimientos pueden comprender dispersar una fuente de ácido fosfórico en aceite vegetal para formar una mezcla que tiene un contenido de humedad menor que 0,1% en peso, mezcla la cual se mantiene a una temperatura en el intervalo de 70°C a 160°C hasta que se forma un precipitado que contiene impurezas de color de la clorofila. Esto puede ser seguido de la separación del material precipitado a partir del aceite, para eliminar las impurezas de color de la clorofila con el material precipitado, por ejemplo durante el procesamiento convencional del aceite hasta e incluyendo la eliminación de la arcilla de blanqueo a partir del aceite.

Procesamiento enzimático de oleaginosas

Las composiciones y métodos de la descripción se pueden usar para el procesamiento enzimático de oleaginosas, incluyendo pasta de haba de soja, de cáñola (colza), de coco, de aguacate y de oliva. En un aspecto, estos procedimientos de la descripción pueden incrementar el rendimiento del aceite y mejorar la calidad nutricional de las harinas obtenidas. En algunos aspectos, el procesamiento enzimático de oleaginosas usando las enzimas y métodos de la descripción proporcionará beneficios económicos y medioambientales, así como tecnologías alternativas para la extracción de aceite y el procesamiento de alimento para el consumo humano y animal. En aspectos alternativos, los procedimientos de la descripción comprenden además el uso de fosfolipasas, proteasas, fosfatasa, fitasas, xilanasas, amilasas (por ejemplo, α -amilasas), glucanasas (por ejemplo, β -glucanasas), poligalacturonasas, galactolipasas, celulasas, hemicelulasas, pectinasas y otras enzimas degradantes de la pared de las células vegetales, así como preparaciones enzimáticas mixtas y lisados celulares. En aspectos alternativos, los procedimientos de la descripción se pueden poner en práctica junto con otros procedimientos, por ejemplo tratamientos enzimáticos, por ejemplo con carbohidrasas, incluyendo celulasa, hemicelulasa y otras actividades degradantes secundarias, o procedimientos químicos, por ejemplo extracción con hexano de aceite de haba de soja. El tratamiento enzimático puede incrementar la capacidad de extracción del aceite en 8-10% cuando el tratamiento enzimático se lleva a cabo antes de la extracción con disolventes.

En aspectos alternativos, los procedimientos de la descripción se pueden poner en práctica con procedimientos de extracción acuosa. Los métodos de extracción acuosa pueden ser tecnologías alternativas medioambientalmente más limpias para la extracción del aceite. Los procedimientos de la descripción también pueden usar enzimas que hidrolizan los polisacáridos estructurales que forman la pared celular de oleaginosas, o que hidrolizan las proteínas que forman las membranas corporales de células y lípidos, por ejemplo utilizando digestiones que comprenden celulasa, hemicelulasa, y/o protopectinasa para la extracción de aceite a partir de células de haba de soja. En un aspecto, los métodos se ponen en práctica con una enzima de la descripción como se describe por Kasai (2003) J. Agric. Food Chem. 51:6217-6222, que da a conocer que la enzima más eficaz para digerir la pared celular fue celulasa.

En un aspecto, las proteasas se usan en combinación con los métodos de la descripción. Se ha evaluado el efecto combinado de variables operacionales y la actividad enzimática de proteasa y celulasa sobre los rendimientos de la extracción de aceite y proteínas combinado con otros parámetros del procedimiento, tales como la concentración de enzima, tiempo de hidrólisis, tamaño de partículas y relación sólido-líquido. En un aspecto, los métodos de la descripción se ponen en práctica con protocolos como se describe por Rosenthal (2001) Enzyme and Microb. Tech. 28:499-509, que da a conocer que el uso de proteasa puede dar como resultado rendimientos significativamente mayores de aceite y proteína con respecto al control cuando se usa harina tratada con calor.

En un aspecto, la extracción completa de proteínas, pectina, y hemicelulosa se usan en combinación con los métodos de la descripción. La célula vegetal consiste en una serie de polisacáridos asociados a menudo con o sustituidos por proteínas o compuestos fenólicos. La mayoría de estos hidratos de carbono son sólo parcialmente digeridos o pobremente utilizados por las enzimas digestivas. La destrucción de estas estructuras mediante enzimas de procesamiento o degradantes puede mejorar su disponibilidad nutricional. En un aspecto, los métodos de la descripción se ponen en práctica con protocolos como se describe por Ouhida (2002) J. Agric. Food Chem. 50:1933-1938, que da a conocer que se ha logrado una degradación significativa de la celulosa de la pared celular de haba de soja (hasta 20%) tras la extracción completa de proteína, pectina y hemicelulosa.

En un aspecto, los métodos de la descripción comprenden la incorporación de diversos tratamientos enzimáticos en el tratamiento de semillas, por ejemplo semillas de cáñola, comprendiendo estos tratamientos el uso de proteasas, celulasas y hemicelulasas (en diversas combinaciones entre sí y con una o más enzimas de la descripción). Por ejemplo, los métodos pueden comprender tratamientos enzimáticos de semillas de cáñola a 20 a 40 de humedad durante la incubación con enzimas antes de un procedimiento convencional, como se describe, por ejemplo, por Sosulski (1990) Proc. Can. Inst. Food Sci. Technol. 3:656. Los métodos de la descripción pueden comprender además la incorporación de proteasas, α -amilasas, poligalacturonasas (en diversas combinaciones entre sí y con una o más enzimas de la descripción) para hidrolizar material celular en harina de coco y liberar el aceite de coco,

que se puede recuperar mediante centrifugación, como se describe, por ejemplo, por McGlone (1986) J. of Food Sci. 51:695-697. Los métodos de la descripción pueden comprender además la incorporación de pectinasas, α -amilasas, proteasas, celulasas en diferentes combinaciones (entre sí y con una o más enzimas de la descripción) para dar como resultado una mejora significativa del rendimiento (~70% en el mejor caso) durante la extracción enzimática de aceite de aguacate, como se describe, por ejemplo, por Buenrostro (1986) Biotech. Letters 8(7):505-506. En procedimientos de la descripción para la extracción de aceite de oliva, la pasta de oliva se trata con celulasa, hemicelulasa, poligalacturonasa, pectina-metiltransferasa, proteasa y sus combinaciones (entre sí y con una o más enzimas de la descripción), como se describe, por ejemplo, por Montedoro (1976) Acta Vitamin. Enzymol. (Milano) 30:13.

En un aspecto, los métodos de la descripción comprenden además la incorporación de diversos tratamientos enzimáticos en el tratamiento de semillas, por ejemplo semillas de cáñola, comprendiendo estos tratamientos el uso de proteasas, celulasas, y hemicelulasas (en diversas combinaciones entre sí y con una o más enzimas de la descripción). Por ejemplo, los métodos pueden comprender tratamientos enzimáticos de semillas de cáñola a 20 a 40 de humedad durante la incubación con enzimas antes de un procedimiento convencional, como se describe, por ejemplo, por Sosulski (1990) Proc. Can. Inst. Food Sci. Technol. 3:656. Los métodos de la descripción pueden comprender además la incorporación de proteasas, α -amilasas, poligalacturonasas (en diversas combinaciones entre sí y con una o más enzimas de la descripción) para hidrolizar material celular en harina de coco y liberar el aceite de coco, que se puede recuperar mediante centrifugación, como se describe, por ejemplo, por McGlone (1986) J. of Food Sci. 51:695-697. Los métodos de la descripción pueden comprender además la incorporación de pectinasas, α -amilasas, proteasas, celulasas en diferentes combinaciones (entre sí y con una o más enzimas de la descripción) para dar como resultado una mejora significativa del rendimiento (~70% en el mejor caso) durante la extracción enzimática de aceite de aguacate, como se describe, por ejemplo, por Buenrostro (1986) Biotech. Letters 8(7):505-506. En procedimientos de la descripción para la extracción de aceite de oliva, la pasta de oliva se trata con celulasa, hemicelulasa, poligalacturonasa, pectina-metiltransferasa, proteasa y sus combinaciones (entre sí y con una o más enzimas de la descripción), como se describe, por ejemplo, por Montedoro (1976) Acta Vitamin. Enzymol. (Milano) 30:13.

En un aspecto, las composiciones y métodos de la descripción se pueden poner en práctica con métodos como se describe en la patente U.S. nº 6.376.689. Por ejemplo, en un aspecto, las composiciones y métodos de la descripción pueden comprender un procedimiento de desgomado/decoloración ácida en una sola etapa, que elimina los compuestos de tipo clorofila de los aceites vegetales procedentes de semillas, especialmente semillas dañadas por heladas que tienen grandes cantidades de compuestos de tipo clorofila. En un aspecto, los métodos de la descripción comprenden además una mezcla de ácidos sulfúrico y fosfórico acuosos que se mezcla con el aceite para eliminar compuestos de tipo clorofílico a partir del aceite. El aceite purificado puede tener menos de alrededor de 5 ppm de compuestos de tipo clorofila, menos de alrededor de 50 ppm de fósforo, o menos de alrededor de 1,0 por ciento en peso de ácidos grasos libres.

Purificación de fitosteroles a partir de aceites vegetales

Las composiciones (por ejemplo, esterasas) y métodos de la descripción también se pueden usar conjuntamente con métodos y procedimientos para purificar fitosteroles y triterpenos, o esteroides vegetales, a partir de aceites vegetales. Los fitosteroles que se pueden purificar usando métodos de la descripción incluyen β -sitosterol, campesterol, estigmasterol, estigmastanol, β -sitostanol, sitostanol, desmosterol, calinasterol, poriferasterol, clionasterol y brasicasterol. Los esteroides vegetales son productos agrícolas importantes para las industrias de la salud y de la nutrición. De este modo, las composiciones (por ejemplo, esterasas) y métodos de la descripción se pueden usar para obtener emulsionantes para fabricantes cosméticos e intermedios esteroideos y precursores para la producción de fármacos hormonales. Las composiciones (por ejemplo, esterasas) y métodos de la descripción se pueden usar para obtener (por ejemplo, purificar) análogos de fitosteroles y sus ésteres para uso como agentes reductores de colesterol con beneficios sanitarios cardiológicos. Las composiciones (por ejemplo, esterasas) y métodos de la descripción se pueden usar para purificar esteroides vegetales para reducir los niveles séricos de colesterol inhibiendo la absorción de colesterol en la luz intestinal. Las composiciones (por ejemplo, esterasas) y métodos de la descripción se pueden usar para purificar esteroides vegetales que tienen propiedades inmunomoduladoras a concentraciones extremadamente bajas, incluyendo respuesta celular potenciada de linfocitos T, y la capacidad citotóxica de células asesinas naturales contra una estirpe celular de cáncer. Las composiciones (por ejemplo, esterasas) y métodos de la descripción se pueden usar para purificar esteroides vegetales para el tratamiento de tuberculosis pulmonar, artritis reumatoide, manejo de pacientes infectados con VIH, e inhibición de estrés inmunitario, por ejemplo en corredores de maratón.

Las composiciones (por ejemplo, esterasas) y métodos de la descripción se pueden usar para purificar componentes de esteroides presentes en las fracciones de esteroides de aceites vegetales de primera necesidad (por ejemplo, aceites de coco, cáñola, manteca de cacao, maíz, semilla de algodón, lino, oliva, palma, cacahuete, salvado de arroz, alazor, sésamo, haba de soja, girasol), tales como sitosterol (40,2-92,3%), campesterol (2,6-38,6%), estigmasterol (0-31%) y 5-avenasterol (1,5-29%).

Aparato para refinar aceite vegetal

La descripción proporciona producto de fabricación que comprende un sistema de desgomado para el tratamiento enzimático de composiciones que contienen clorofila o contaminadas con clorofila, que comprende (a) un aparato de refinado de aceite vegetal; y (b) un polipéptido que tiene una actividad de clorofilasa operablemente integrado en el aparato de refinado del aceite vegetal, en el que la actividad del polipéptido comprende la catálisis de una reacción que modifica la clorofila, y el aparato de refinado del aceite vegetal puede hacer reaccionar una composición que contiene clorofila o contaminada con clorofila con el polipéptido en condiciones en las que el polipéptido puede catalizar una reacción que modifica la clorofila.

Los productos de fabricación pueden comprender cualquier aparato de refinado de aceite vegetal o su combinación, por ejemplo un expulsor que deja aceite (por ejemplo, de Pennwalt Corp.), o un dispositivo de separación de goma gravitacional.

La descripción proporciona producto de fabricación que comprende enzimas inmovilizadas, por ejemplo una clorofilasa inmovilizada, por ejemplo una esterasa de la descripción. En un aspecto del producto de fabricación, la clorofilasa comprende una clorofilasa inmovilizada en sílice. La sílice comprende un gel de sílice o equivalente. La sílice comprende una sílice TriSyl Silica o una sílice SORBSIL RTM.

En un aspecto, los productos de fabricación de la descripción comprenden aparato para ajustar el pH, por ejemplo incrementar el pH ("tratamiento cáustico"), y después, como alternativa, neutralizar el pH.

EJEMPLOS

Ejemplo 1: Ensayo de actividad de esterasa ejemplar

El siguiente ejemplo demuestra un ensayo de esterasa ejemplar (actividad de clorofilasa) para aislar y caracterizar enzimas de la descripción y los ácidos nucleicos que las codifican, y para determinar si un polipéptido está dentro del alcance de la descripción.

Las esterases se cribaron en busca de actividad sobre la clorofila a partir de espinaca para producir clorofilida. En este ensayo de esterasa ejemplar (actividad de clorofilasa), el formato de cribado de esterasa comprende:

- Placas cribadas por duplicado.
 - Controles positivos (CHLasa) y negativos en cada placa.
 - 1 mM de CHL, 20% de lisado celular, 20% de acetona, pH 7,5, 0,01% de HBT.
 - 24 h de tiempo de incubación a 30°C en la oscuridad.
 - 100 ml de volumen de reacción.
 - Análisis mediante LC-VIS; inyección de 1 ml de muestra.
- Este método de cribado de esterasa usó HPLC para analizar los productos de reacción. La Figura 2 y la Figura 3 ilustran datos que muestran los resultados del ensayo de actividad de esterasa (actividad de clorofilasa) usando las enzimas ejemplares indicadas de la descripción.

Para la HPLC:

Columna: Cromolith SpeedROD RP-18e 50-4, 6 mm (Cat# UM1082/086) Caudal: 1,0 ml/min.; Inyección: 1,0 ml.

A: H ₂ O			
B: MeOH + 1mM NH ₄ OAc			
C: MTBE			
T (min)	A	B	C
0	10%	80%	10%
2,3	10%	80%	10%
2,31	0%	50%	50%
4	0%	50%	50%
4,1	10%	80%	10%

7	10%	80%	10%
---	-----	-----	-----

Señal DAD	L (nm)	Bw	Referencia 1	Bw
1	660	20	710 nm	10

Compuesto	Rt
CHLa	4,20
CHLb	4,15
PHPa	4,30
PHPb	4,25
CHPa	0,85
CHPb	0,80
PHBa	1,00
PHBb	0,95

- 5 Los datos ilustrados en la Figura 2 ilustran niveles crecientes de producto de reacción entre los puntos de tiempo de 24 h y 48 h, en el que los niveles de producto de reacción indican actividad de clorofilasa para SEC ID NO:2, SEC ID NO:4, SEC ID NO:6, SEC ID NO:8, SEC ID NO:12, SEC ID NO:14, SEC ID NO:16, SEC ID NO:18 y SEC ID NO:20. Los datos ilustrados en la Figura 3 ilustran niveles crecientes de producto de reacción entre los puntos de tiempo 24 h y 48 h, en el que los niveles de producto de reacción indican actividad de clorofilasa para SEC ID NO:10.

LISTADO DE SECUENCIAS

10 <110> VERENIUM CORPORATION

<120> COMPOSICIONES Y MÉTODOS PARA LA DECOLORACIÓN ENZIMÁTICA DE CLOROFILA

<130> p/52268.ep02

<140> ep05762779.6 - divisional

<141> 14-06-2005

15 <150> US 60/580,447

<151> 16-06-2004

<160> 27

<170> PatentIn version 3.1

<210> 1

20 <211> 915

<212> ADN

<213> Desconocido

<220>

<223> Obtenido de muestra ambiental

<400> 1

atgtcgcgtg tctgtcttcc actcacactc acactcgcac tcacactctc tgccagagca	60
gcggagccga ccacggtgaa gctctggccc ggcaaagcgc ccggcgagac gaaggacatc	120
ggcccgagga agtatctcga aaccaagaag gggcagcttg acatcaagcg gctcaccaac	180
gtgagcgaac cgaccatcac gatctattcg ccgccaaagg agaaggcgaa cgacacagtg	240
gtgattgtcg cggccggcgg cgggtacaac atcctcgcca tcgaacacga agggaccgat	300
gtctgcgagt ggctgaactc gctgggcgtg accgcgggtgc tgetcaagta tcgcgtgccg	360
cgggcggcca tgcatcgcc ggacaacctc gcgatgattc aggacgcgca gcgcgccatc	420
agtcttggtc gcagtatgca gaaagagttg ggcatcacc caacacgtgt gggaatgctc	480
ggcttttcgg ccggggggaa cctcacccgt tgcaccgcgt tggccgaaaa gcggatgtac	540
gagaacatcg acaagggtga cgaggatttc aactgcacgc cgaacttcgc catcctcgtg	600
taccccgct atctcgtcga gaaggatgga acgctgcgcg ccgagttcaa agtgaagatc	660
gattcgccgc cgatgttctt cgtgcattcg tccgacgaca acgtgagcag tgaaaacagc	720
gtggcgcttt acctggcact caagaagaac aaggtgccgg cggaaatgca cctctacgcc	780
agcggcgggc acggctacgg aatgcgcaag gttccgcac cgtgcgcaag ctggcccgac	840
cgcgcccgcg agtggatgaa agcacaccgg ctgcttgaga aggcccaagcc cgagccggat	900
gggaagaagg agtaa	915

<210> 2

<211> 304

5 <212> PRT

<213> Desconocido

<220>

<223> Obtenido de muestra ambiental

<220>

10 <221> SEÑAL

<222> (1)...(20)

<400> 2

ES 2 540 728 T3

Met Ser Arg Val Cys Leu Pro Leu Thr Leu Thr Leu Ala Leu Thr Leu
 1 5 10 15
 Ser Ala Arg Ala Ala Glu Pro Thr Thr Val Lys Leu Trp Pro Gly Lys
 20 25 30
 Ala Pro Gly Glu Thr Lys Asp Ile Gly Pro Glu Lys Tyr Leu Glu Thr
 35 40 45
 Lys Lys Gly Gln Leu Asp Ile Lys Arg Leu Thr Asn Val Ser Glu Pro
 50 55 60
 Thr Ile Thr Ile Tyr Ser Pro Pro Lys Glu Lys Ala Asn Asp Thr Val
 65 70 75 80
 Val Ile Val Ala Pro Gly Gly Gly Tyr Asn Ile Leu Ala Ile Glu His
 85 90 95
 Glu Gly Thr Asp Val Cys Glu Trp Leu Asn Ser Leu Gly Val Thr Ala
 100 105 110
 Val Leu Leu Lys Tyr Arg Val Pro Arg Arg Pro Met Gln Ser Pro Asp
 115 120 125
 Asn Leu Ala Met Ile Gln Asp Ala Gln Arg Ala Ile Ser Leu Val Arg
 130 135 140
 Ser Met Gln Lys Glu Leu Gly Ile His Pro Thr Arg Val Gly Met Leu
 145 150 155 160
 Gly Phe Ser Ala Gly Gly Asn Leu Thr Ala Cys Thr Ala Leu Ala Glu
 165 170 175
 Lys Arg Met Tyr Glu Asn Ile Asp Lys Val Asp Glu Val Phe Asn Cys
 180 185 190
 Thr Pro Asn Phe Ala Ile Leu Val Tyr Pro Ala Tyr Leu Val Glu Lys
 195 200 205
 Asp Gly Thr Leu Arg Ala Glu Phe Lys Val Lys Ile Asp Ser Pro Pro
 210 215 220
 Met Phe Phe Val His Ser Ser Asp Asp Asn Val Ser Ser Glu Asn Ser
 225 230 235 240
 Val Ala Leu Tyr Leu Ala Leu Lys Lys Asn Lys Val Pro Ala Glu Met
 245 250 255
 His Leu Tyr Ala Ser Gly Gly His Gly Tyr Gly Met Arg Lys Val Pro
 260 265 270
 His Pro Cys Ala Ser Trp Pro Asp Arg Ala Ala Glu Trp Met Lys Ala
 275 280 285
 His Arg Leu Leu Glu Lys Ala Lys Pro Glu Pro Asp Gly Lys Lys Glu
 290 295 300

<210> 3

<211> 927

<212> ADN

<213> Desconocido

5 <220>

<223> Obtenido de muestra ambiental

<400> 3

```

atgtcaettg atccacaaac gaaatttgta ttagcccaat tggetgctgc cgatgctcct    60
ccaatggaga ctttgtcacc ggaaatggct agacaggcat tcatattgcc gcaaggtgca    120
gtggaggaag tagggaaggt agaaaatcga acgattcctg gtcctgcaac ggacatccct    180
gtccgtgttt attaccctaa agaactccag cctgaaaatc ccgcgctagt tttctatcat    240
ggcggcggct gggtaattgg caatttggac tctcatgacg atatttgccg tgctttaacc    300
aacctgcgca actgctgac catttccgtc gactaccgcc tggetccaga gaataaattt    360
cctgctgcgg ttgaggatgc ttatgctgct gcacaatatg tgtatgacca tgcagaagat    420
ttcaaagtcg acaagacccg tattgccgtt ggcggtgaca gtgccggtgg aaatcttgcc    480
gctgctgtga cgaatctggc aaaagacaaa aactcacctt ctatttgttt ccaacttctg    540
atttatccaa gcaccaatgc aggtggtgag ccaacagcat caatggttga gaatgcccat    600
ggctatttct tagaaaaagg cacgatggat tggttccgtg actgctacct gaacagtga    660
gaagacaaac agaatccgct agtctcaccg atgctttatg atgatttcca aggactgcct    720
ccagcaattg tgattactgc cgagtacgat ccattgcgag atgaaggcga agcatatgcc    780
aaaaagctgg gcgaagcagg ggttgcgtgc caaaccatcc gctttgacgg tacgattcat    840
ggctttgtca gcatgtcggc tgtcattagc caagggaagg cggcattgga aaaagcagga    900
gaggctttaa ctaaagcctt tcaataa                                     927

```

<210> 4

10 <211> 308

<212> PRT

<213> Desconocido

<220>

<223> Obtenido de muestra ambiental

15 <400> 4

ES 2 540 728 T3

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

Ala Asp Ala Pro Pro Met Glu Thr Leu Ser Pro Glu Met Ala Arg Gln
20 25 30

Ala Phe Ile Leu Pro Gln Gly Ala Val Glu Glu Val Gly Lys Val Glu
35 40 45

Asn Arg Thr Ile Pro Gly Pro Ala Thr Asp Ile Pro Val Arg Val Tyr
50 55 60

Tyr Pro Lys Glu Leu Gln Pro Glu Asn Pro Ala Leu Val Phe Tyr His
65 70 75 80

Gly Gly Gly Trp Val Ile Gly Asn Leu Asp Ser His Asp Asp Ile Cys
85 90 95

Arg Ala Leu Thr Asn Leu Ala Asn Cys Val Thr Ile Ser Val Asp Tyr
100 105 110

Arg Leu Ala Pro Glu Asn Lys Phe Pro Ala Ala Val Glu Asp Ala Tyr
115 120 125

Ala Ala Ala Gln Tyr Val Tyr Asp His Ala Glu Asp Phe Lys Val Asp
130 135 140

Lys Thr Arg Ile Ala Val Gly Gly Asp Ser Ala Gly Gly Asn Leu Ala
145 150 155 160

Ala Val Val Thr Asn Leu Ala Lys Asp Lys Asn Ser Pro Ser Ile Cys
165 170 175

Phe Gln Leu Leu Ile Tyr Pro Ser Thr Asn Ala Gly Gly Glu Pro Thr
180 185 190

Ala Ser Met Val Glu Asn Ala His Gly Tyr Phe Leu Glu Lys Gly Thr
195 200 205

Met Asp Trp Phe Arg Asp Cys Tyr Leu Asn Ser Glu Glu Asp Lys Gln
210 215 220

Asn Pro Leu Val Ser Pro Met Leu Tyr Asp Asp Phe Gln Gly Leu Pro
225 230 235 240

Pro Ala Ile Val Ile Thr Ala Glu Tyr Asp Pro Leu Arg Asp Glu Gly
245 250 255

Glu Ala Tyr Ala Lys Lys Leu Gly Glu Ala Gly Val Ala Val Gln Thr
260 265 270

Ile Arg Phe Asp Gly Thr Ile His Gly Phe Val Ser Met Ser Ala Val
275 280 285

Ile Ser Gln Gly Lys Ala Ala Leu Glu Lys Ala Gly Glu Ala Leu Thr
290 295 300

Lys Ala Phe Gln
305

<210> 5

<211> 672

<212> ADN

5 <213> Desconocido

<220>

<223> Obtenido de muestra ambiental

<400> 5

atgagaagaa ttgttttctt ttatatcttg gcgttgcctt gcgtatcctg tgcgaaccgg	60
aatccttctg tctcggaccc cgctccgtca ccgacctgg gggagcaatc gtcagcagat	120
ttaccgaaga tcgtagcatt cggcgacagc ctaactgctg gctttggtct tccccagaat	180
gaaagttatc cagcgctttt gcaggaacga ctgagacagg acggttatga ttacgaagtg	240
atcaatgccg gagtttcagg cgataccagc gcgggagggtg tgcgtaggat cgactgggtt	300
ctggatgaga gcgtacgaat tctaattctt gaattgggtg ctaatgattt tcttcgtggc	360
catccagtag ctcagactaa gaaaaacctc gccgttatca tcgaacgtgc tcaggaaaaa	420
aacgtcaggg ttctgctagc agggatgttt gcaccgacga atactggatg ggagtaccag	480
gggcagatcc agcaaagtgt taacgatctt tcgcgagaaa aagccgtacc actaattcca	540
ttctttctcg agggagtagc cggcattcca actctgaatc tagctgatgg cattcatecc	600
aatgcggctg gcacaaagat cgttgcgtgag aatgtgtaca agtatctaaa accgatgctc	660
ccatctgatt aa	672

10 <210> 6

<211> 223

<212> PRT

<213> Desconocido

<220>

15 <223> Obtenido de muestra ambiental

<220>

<221> SEÑAL

<222> (1) ... (25)

<220>

<221> DOMINIO

<222> (44)...(212)

<223> Lipasa/Acilhidrolasa similar a GDSL

5 <400> 6

```

Met Arg Arg Ile Val Phe Leu Tyr Ile Leu Ala Leu Leu Cys Val Ser
1          5          10          15

Cys Ala Asn Arg Asn Pro Ser Val Ser Asp Pro Ala Pro Ser Pro Thr
          20          25          30

Leu Gly Glu Gln Ser Ser Ala Ser Leu Pro Lys Ile Val Ala Phe Gly
          35          40          45

Asp Ser Leu Thr Ala Gly Phe Gly Leu Ser Gln Asn Glu Ser Tyr Pro
          50          55          60

Ala Leu Leu Gln Glu Arg Leu Arg Gln Asp Gly Tyr Asp Tyr Glu Val
65          70          75          80

Ile Asn Ala Gly Val Ser Gly Asp Thr Ser Ala Gly Gly Val Arg Arg
          85          90          95

Ile Asp Trp Val Leu Asp Glu Ser Val Arg Ile Leu Ile Leu Glu Leu
          100          105          110

Gly Ala Asn Asp Phe Leu Arg Gly His Pro Val Ala Gln Thr Lys Lys
          115          120          125

Asn Leu Ala Val Ile Ile Glu Arg Ala Gln Glu Lys Asn Val Arg Val
          130          135          140

Leu Leu Ala Gly Met Phe Ala Pro Thr Asn Thr Gly Trp Glu Tyr Gln
145          150          155          160

Gly Gln Ile Gln Gln Met Phe Asn Asp Leu Ser Arg Glu Lys Ala Val
          165          170          175

Pro Leu Ile Pro Phe Phe Leu Glu Gly Val Ala Gly Ile Pro Thr Leu
          180          185          190

Asn Leu Ala Asp Gly Ile His Pro Asn Ala Ala Gly Thr Lys Ile Val
          195          200          205

Ala Glu Asn Val Tyr Lys Tyr Leu Lys Pro Met Leu Pro Ser Asp
          210          215          220

```

<210> 7

<211> 783

<212> ADN

10 <213> Desconocido

<220>

<223> Obtenido de muestra ambiental

<400> 7

```

atgacacgaa aaaaaatcgg cttggcttta tccggcggcg cggcgcgtgg cttcgcacat    60
ctcggagttt tgaaagtgtt cgccgaacac ggcatgcccg tcgatttcgt cgccggaacg    120
agcgcgggtt cgtttgcggg cgcggctttc gcttcgggtt taagcgtggc ggaaatcatc    180
gaaatgtcaa ggaaaatcag ttggtttcgg atgaccggat ttctgtactc gccgaaaggc    240
ttgtctgtcg acgcgccgat gggagcgttt atcaatcagc attttccgcg caaaaaattt    300
gaagaactgc cgataccggt cgcggaatt acgtgcgac tcgaaaccgg cgaggaaatc    360
gtgcttaag aaaccggcga cgtggcgact gccgtccgtg cgagctgcgc gctgccggc    420
gtcttcgtgc caatcgaata cggcggacgg cgattgatag acggcggcgt cgtgtcgaa    480
gtgccgacga gagccgtcaa aaagctcggc gcggaaatca ttattgcggt tgacgttctg    540
gcgctcggca caacttactg gggttcgcct tccactttgc tcggcatctt tttccaatcg    600
gcgatgatgc ttctccgcgc cgcttcaaaa tctcatcatt accgcgcgag cgtcgtcatc    660
acgccgcaaa tcgtcatct gcgccggac gaaatcagca aaatggacga atttatcaaa    720
gcgggcgaac aagccgcgct tgaaaaagt gacgaaatca aggctttgct cgccgaacaa    780
taa                                                                    783

```

<210> 8

5 <211> 260

<212> PRT

<213> Desconocido

<220>

<223> Obtenido de muestra ambiental

10 <220>

<221> DOMINIO

<222> (8)...(167)

<223> fosfolipasa semejante a patatina

<400> 8

```

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

Gly Phe Ala His Leu Gly Val Leu Lys Val Phe Ala Glu His Gly Ile
          20          25          30

Pro Val Asp Phe Val Ala Gly Thr Ser Ala Gly Ser Phe Ala Gly Ala
          35          40          45

Ala Phe Ala Ser Gly Leu Ser Val Ala Glu Ile Ile Glu Met Ser Arg
50          55          60

```

15

ES 2 540 728 T3

Lys Ile Ser Trp Phe Arg Met Thr Gly Phe Ser Tyr Ser Pro Lys Gly
 65 70 75 80
 Leu Leu Ser Asn Ala Pro Met Gly Ala Phe Ile Asn Gln His Phe Pro
 85 90 95
 Arg Lys Lys Phe Glu Glu Leu Pro Ile Pro Phe Ala Ala Ile Thr Cys
 100 105 110
 Asp Leu Glu Thr Gly Glu Glu Ile Val Leu Lys Glu Thr Gly Asp Val
 115 120 125
 Ala Thr Ala Val Arg Ala Ser Cys Ala Leu Pro Gly Val Phe Val Pro
 130 135 140
 Ile Glu Tyr Gly Gly Arg Arg Leu Ile Asp Gly Gly Val Val Ser Asn
 145 150 155 160
 Val Pro Thr Arg Ala Val Lys Lys Leu Gly Ala Glu Ile Ile Ile Ala
 165 170 175
 Val Asp Val Leu Ala Cys Gly Thr Thr Tyr Trp Gly Ser Pro Ser Thr
 180 185 190
 Leu Leu Gly Ile Phe Phe Gln Ser Ala Met Met Leu Leu Arg Ala Ala
 195 200 205
 Ser Lys Ser His His Tyr Arg Ala Ser Val Val Ile Thr Pro Gln Ile
 210 215 220
 Ala His Leu Arg Pro Asp Glu Ile Ser Lys Met Asp Glu Phe Ile Lys
 225 230 235 240
 Ala Gly Glu Gln Ala Ala Leu Glu Lys Val Asp Glu Ile Lys Ala Leu
 245 250 255
 Leu Ala Glu Gln
 260

<210> 9

<211> 849

<212> ADN

5 <213> Desconocido

<220>

<223> Obtenido de muestra ambiental

<400> 9

atgttcaaca aagcacttcc tgcagcggcc gcggtcgccg gtctgttcct ttccacctcg 60
 gccatggcgc tgcgtgccga tacgcccga acccccttcc cgtcggtgtc cagcttcgaa 120

agggtccggtc cctacagcac caccagccgc agcgaaggcc ccaactgccg gatctatcga 180
 ccgagcaacc tgggccagaa cggcgctccgc caccgggtca tcctctgggg caacggcacc 240
 ggtgccagcc cgaccaccta ccgcggcctg ctcgagcact gggccagcca cggcttcgtg 300
 gtcgcccgcg ccgagacctc caacgccggc tccgggcggc agatgctcaa ctgcctgagc 360
 tacctgcaga ccgaagccgg gcgcagcagc ggcacctacg tcggccgggt caatctcggc 420
 cgcgtcggca cgtccggcca ctgcagggc ggtggcggct cgatcatggc cgggcgcgac 480
 acccggtatca aaaccacgc gcccatgcag cctacaccc tgggcctggg ccatgtgagt 540
 tcctcgaga gccagcagaa cggcccgatg ttcctgatgt ccggcagcct ggacaccctg 600
 gccggcccca ccctgaacca ggcgcgggtc tatcgccggg ccaacgtgcc ggtgttctgg 660
 ggcacctgc gcggcgccag ccacttcgtg ccggtcggca gtgcggcggg ttaccgtggc 720
 ccgtccaccg cctggttcgg ctaccagctg atggacgata cctcggcggc cagccagttc 780
 gtcggcacca actgcggcct gtgcgcgac ttctcctgga ccgacatcca gcgcaagggc 840
 gcgctgtaa 849

<210> 10

<211> 282

<212> PRT

5 <213> Desconocido

<220>

<223> Obtenido de muestra ambiental

<220>

<221> SEÑAL

10 <222> (1)...(23)

<400> 10

Met Phe Asn Lys Ala Leu Pro Ala Ala Ala Val Ala Gly Leu Phe
 1 5 10 15

Leu Ser Thr Ser Ala Met Ala Leu Leu Pro Asp Thr Pro Gly Thr Pro
 20 25 30

Phe Pro Ser Val Ser Ser Phe Glu Arg Ser Gly Pro Tyr Ser Thr Thr
 35 40 45

Ser Arg Ser Glu Gly Pro Asn Cys Arg Ile Tyr Arg Pro Ser Asn Leu
 50 55 60

Gly Gln Asn Gly Val Arg His Pro Val Ile Leu Trp Gly Asn Gly Thr
 65 70 75 80

Gly Ala Ser Pro Thr Thr Tyr Arg Gly Leu Leu Glu His Trp Ala Ser
 85 90 95

ES 2 540 728 T3

His Gly Phe Val Val Ala Ala Ala Glu Thr Ser Asn Ala Gly Ser Gly
100 105 110

Arg Glu Met Leu Asn Cys Leu Ser Tyr Leu Gln Thr Glu Ala Gly Arg
115 120 125

Ser Ser Gly Thr Tyr Val Gly Arg Leu Asn Leu Gly Arg Val Gly Thr
130 135 140

Ser Gly His Ser Gln Gly Gly Gly Gly Ser Ile Met Ala Gly Arg Asp
145 150 155 160

Thr Arg Ile Lys Thr Thr Ala Pro Met Gln Pro Tyr Thr Leu Gly Leu
165 170 175

Gly His Val Ser Ser Ser Gln Ser Gln Gln Asn Gly Pro Met Phe Leu
180 185 190

Met Ser Gly Ser Leu Asp Thr Leu Ala Gly Pro Thr Leu Asn Gln Ala
195 200 205

Pro Val Tyr Arg Arg Ala Asn Val Pro Val Phe Trp Gly Thr Leu Arg
210 215 220

Gly Ala Ser His Phe Val Pro Val Gly Ser Ala Gly Gly Tyr Arg Gly
225 230 235 240

Pro Ser Thr Ala Trp Phe Arg Tyr Gln Leu Met Asp Asp Thr Ser Ala
245 250 255

Arg Ser Gln Phe Val Gly Thr Asn Cys Gly Leu Cys Arg Asp Phe Ser
260 265 270

Trp Thr Asp Ile Gln Arg Lys Gly Ala Leu
275 280

<210> 11

<211> 672

<212> ADN

5 <213> Desconocido

<220>

<223> Obtenido de muestra ambiental

<400> 11

gtgtccgccc ccgaagccgc gagcgtggag catgcggacc ccggcggggt cgccaccctg	60
gcggaacgcc ccgcggcgcg tccgcgcac gtcttctctgg gcgacagcct gaccgccggg	120
tacggactgg cgcgggagca gtccgtgcc gcgctactcc agcggcggt cgaccgggag	180
gggtatgagt acgaagtcgt caatgcgggc gtctcggggc acacgtccgc cggcgggctg	240

agcgggtcgc actggtcgct cgacggcgag gtcgcectgc tgggtggtgga gtcgggcgcc 300
aacgacgggc tgcggggcct gcccggtgcc gcgatgaage gcaacctgga cgccatcacc 360
acgcgcgccc gcgcgcgggg catcaccgtg gtgctcgccg gcattggaagc gccgccgaac 420
tacggcgccg cctacacgac cgagttccgg caggcggttc acgatctcgc ccgtaccac 480
gacgtgccgt tcgtgccgtt ctttctcgaa ggggtagccg gcctgccgca cctcaacacc 540
gccgacggca tccatcccaa cgccgagggc gcgcgcgtgg tcgaggccaa cgtctggcag 600
gtgctggaac cgttactcga cgagcctggc cgggcggccg cctcgcgccg gcagcccgcg 660
gatggccgct ga 672

<210> 12

<211> 223

<212> PRT

5 <213> Desconocido

<220>

<223> Obtenido de muestra ambiental

<220>

<221> DOMINIO

10 <222> (30)...(198)

<223> Lipasa/Acilhidrolasa similar a GDLS

<400> 12

Met	Ser	Ala	Pro	Glu	Ala	Ala	Ser	Val	Glu	His	Ala	Asp	Pro	Gly	Gly
1				5					10					15	
Val	Ala	Thr	Leu	Ala	Glu	Arg	Pro	Ala	Ala	Arg	Pro	Arg	Ile	Val	Phe
			20					25					30		
Leu	Gly	Asp	Ser	Leu	Thr	Ala	Gly	Tyr	Gly	Leu	Ala	Arg	Glu	Gln	Ser
		35					40					45			
Val	Pro	Ala	Leu	Leu	Gln	Arg	Arg	Leu	Asp	Arg	Glu	Gly	Tyr	Glu	Tyr
	50					55					60				
Glu	Val	Val	Asn	Ala	Gly	Val	Ser	Gly	Asp	Thr	Ser	Ala	Gly	Gly	Leu
65					70					75					80
Ser	Arg	Leu	Asp	Trp	Ser	Leu	Asp	Gly	Glu	Val	Ala	Leu	Leu	Val	Val
			85						90					95	
Glu	Leu	Gly	Ala	Asn	Asp	Gly	Leu	Arg	Gly	Leu	Pro	Val	Ser	Ala	Met
			100					105					110		
Lys	Arg	Asn	Leu	Asp	Ala	Ile	Ile	Thr	Arg	Ala	Arg	Ala	Arg	Gly	Ile
		115					120					125			
Thr	Val	Val	Leu	Ala	Gly	Met	Glu	Ala	Pro	Pro	Asn	Tyr	Gly	Ala	Ala
	130					135						140			

Tyr Thr Thr Glu Phe Arg Gln Ala Phe His Asp Leu Ala Arg Thr His
145 150 155 160

Asp Val Pro Phe Val Pro Phe Phe Leu Glu Gly Val Ala Gly Leu Pro
165 170 175

His Leu Asn Ile Ala Asp Gly Ile His Pro Asn Ala Glu Gly Ala Arg
180 185 190

Val Val Glu Ala Asn Val Trp Gln Val Leu Glu Pro Leu Leu Asp Glu
195 200 205

Pro Gly Arg Ala Ala Ala Ser Arg Ala Gln Pro Ala Asp Gly Arg
210 215 220

<210> 13

<211> 900

<212> ADN

5 <213> Desconocido

<220>

<223> Obtenido de muestra ambiental

<400> 13

atgaaggcaa gctgccgcga catcgaaatc gaatacgaag cattcggcca tccggacgac	60
ccggccatcg tgctgatcat gggactcggg ggccagctca tctgtggcc ggaagcgttc	120
tgccgcatgc tggccgacgc tggctactac gtggtgcgct tcgacaaccg cgacatcggc	180
ctgtcgacgc atctcgatca cttccccgc cccaacctgc cgtcgcgcgc gtcgcgcag	240
gcgctgcgcc tgccagtctg cgccagttac acgctcgacg acatggcgga cgacgtcgcg	300
ggcctgcttg atgcactgaa catcaagcag gcgcacgctc tcggcggtgc gatgggcggc	360
atgatcgccc agctgctggc cgcacggcac gcgaccgcgc tgcgcagcct gacctgtctg	420
atgaccacca gcggcgcgcg caacgttccg ggcccctcac tcggcatgcg catggaaatg	480
atccgtcgac cgcgcgacac ctgcgcgag gggctgatcc gccatggtat gcgtacctgg	540
cggatcatcg gcagcccgcg gtacccgaag ccggaagcgc aactgcgccg catcgtcgcc	600
gagggctttg accgcgcggt tcacccggcc ggtttcatgc gccagctgca cgccgttctc	660
gcggcaccga gccgcgcacc cctgctgccg cgcaccaaac agccggccga cgtcattcac	720
ggcgacgccg acctgctggt accagtggca gcggcacgtg atctggtgcg ccgcctgccg	780
aacgccacgc tcgacatcgt gccgggcatg gggcatgact tcccgaccga gatcatgccg	840
cgtatcgcgc gccgcattgt cgaaaccgcg gcacgcgacc cgcagcggct cgccgcctag	900

10 <210> 14

<211> 299

<212> PRT

<213> Desconocido

<220>

15 <223> Obtenido de muestra ambiental

<220>

<221> DOMINIO

<222> (49)...(288)

<223> plegamiento de hidrolasa alfa/beta

5 <400> 14

Met Lys Ala Ser Cys Arg Asp Ile Glu Ile Glu Tyr Glu Thr Phe Gly
1 5 10 15

His Pro Asp Asp Pro Ala Ile Val Leu Ile Met Gly Leu Gly Gly Gln
20 25 30

Leu Ile Leu Trp Pro Glu Ala Phe Cys Arg Met Leu Ala Asp Ala Gly
35 40 45

His Tyr Val Val Arg Phe Asp Asn Arg Asp Ile Gly Leu Ser Thr His
50 55 60

Leu Asp His Leu Pro Arg Pro Asn Leu Pro Leu Ala Ala Leu Arg Gln
65 70 75 80

Ala Leu Arg Leu Pro Val Arg Ala Ser Tyr Thr Leu Asp Asp Met Ala
85 90 95

Asp Asp Val Ala Gly Leu Leu Asp Ala Leu Asn Ile Lys Gln Ala His
100 105 110

Val Val Gly Val Ser Met Gly Gly Met Ile Ala Gln Leu Leu Ala Ala
115 120 125

Arg His Ala Thr Arg Val Arg Ser Leu Thr Leu Leu Met Thr Thr Ser
130 135 140

Gly Ala Arg Asn Val Pro Gly Pro Ser Leu Gly Met Arg Met Glu Met
145 150 155 160

Ile Arg Arg Pro Arg Asp Thr Ser Arg Glu Gly Leu Ile Arg His Gly
165 170 175

Met Arg Thr Trp Arg Ile Ile Gly Ser Pro Thr Tyr Pro Lys Pro Glu
180 185 190

Ala Glu Leu Arg Arg Ile Val Ala Glu Gly Phe Asp Arg Ala Phe His
195 200 205

Pro Ala Gly Phe Met Arg Gln Leu His Ala Val Leu Ala Ala Pro Ser
210 215 220

ES 2 540 728 T3

Arg Ala Pro Leu Leu Pro Arg Ile Lys Gln Pro Ala Asp Val Ile His
225 230 235 240

Gly Asp Ala Asp Leu Leu Val Pro Val Ala Ala Ala Arg Asp Leu Val
245 250 255

Arg Arg Leu Pro Asn Ala Thr Leu Asp Ile Val Pro Gly Met Gly His
260 265 270

Asp Phe Pro Thr Glu Ile Met Pro Arg Ile Ala Arg Arg Ile Val Glu
275 280 285

Thr Ala Ala Arg Asp Pro Gln Arg Leu Ala Ala
290 295

<210> 15

<211> 759

<212> ADN

5 <213> Desconocido

<220>

<223> Obtenido de muestra ambiental

<400> 15

atgggcaaac aaatgcctta tagagtgatt tttccagtaa actaccaaac ttccaaaaaa	60
cgctatccag tgatttacct tctgcacggc ctttcgggca attacaaaaa ctggaccgag	120
aaaaccaaat tgacaaaata tgcaaccgaa tacaacttct tgattatcac agttgaaggc	180
gagaacgggtt ggtattctga cagtaaaatc aaacctaaca ggctttacga aagctacata	240
atccaggaac tgatacctga agttgataaa aagtttcgca caatagcaga tcgcaatcac	300
agagctatcg caggtctttc gatgggtggt tacggtgcga taaagtttgg cttgaaatat	360
cctgaaaaat tcgctttgat tgggtccttt agcggcgctc ttgcggcaac ttcaatcaag	420
gaaggcacag cgcttgagtg gattacaaaa accatcaatg atgctttcgg tcctgaagga	480
agcgaatcca gaaaagaaaa cgatattttc cagattgcgc gcgacctgaa tgatgaacaa	540
ataaagaaac ttccctttat ttactttgat tgcggaaccg aagattttct gttcaaagac	600
aatcaggatt tcatgaaact acttggtgaa aagagaatca aacacgaata tagacaaaag	660
cccggcacc c atacgtggga atactgggat agccaaataa aagaattctt gggtttagcc	720
aacaaattta tagaaaaaac tgctttatgg cgatactaa	759

10 <210> 16

<211> 252

<212> PRT

<213> Desconocido

<220>

15 <223> Obtenido de muestra ambiental

<220>

<221> DOMINIO

<222> (1)...(240)

<223> Esterasa putativa

<400> 16

```

Met Gly Lys Gln Met Pro Tyr Arg Val Ile Phe Pro Val Asn Tyr Gln
1          5          10          15

Thr Ser Lys Lys Arg Tyr Pro Val Ile Tyr Leu Leu His Gly Leu Ser
20          25          30

Gly Asn Tyr Lys Asn Trp Thr Glu Lys Thr Lys Leu Thr Lys Tyr Ala
35          40          45

Thr Glu Tyr Asn Phe Leu Ile Ile Thr Val Glu Gly Glu Asn Gly Trp
50          55          60

Tyr Ser Asp Ser Lys Ile Lys Pro Asn Arg Leu Tyr Glu Ser Tyr Ile
65          70          75          80

Ile Gln Glu Leu Ile Pro Glu Val Asp Lys Lys Phe Arg Thr Ile Ala
85          90          95

Asp Arg Asn His Arg Ala Ile Ala Gly Leu Ser Met Gly Gly Tyr Gly
100         105         110

Ala Ile Lys Phe Gly Leu Lys Tyr Pro Glu Lys Phe Ala Leu Ile Gly
115         120         125

Ser Phe Ser Gly Ala Leu Ala Ala Thr Ser Ile Lys Glu Gly Thr Ala
130         135         140

Leu Glu Trp Ile Thr Lys Thr Ile Asn Asp Ala Phe Gly Pro Glu Gly
145         150         155         160

Ser Glu Ser Arg Lys Glu Asn Asp Ile Phe Gln Ile Ala Arg Asp Leu
165         170         175

Asn Asp Glu Gln Ile Lys Lys Leu Pro Phe Ile Tyr Phe Asp Cys Gly
180         185         190

Thr Glu Asp Phe Leu Phe Lys Asp Asn Gln Asp Phe Met Lys Leu Leu
195         200         205

Val Glu Lys Arg Ile Lys His Glu Tyr Arg Gln Lys Pro Gly Thr His
210         215         220

Thr Trp Glu Tyr Trp Asp Ser Gln Ile Lys Glu Phe Leu Gly Leu Ala
225         230         235         240

Asn Lys Phe Ile Glu Lys Thr Ala Leu Trp Arg Tyr
245         250

```

5

<210> 17

<211> 753

<212> ADN

<213> Desconocido

<220>

<223> Obtenido de muestra ambiental

<400> 17

```

atgaaaaagt ataaaacagg cctgggttta agtgggtggcg gaacacgtgg ttttgcgcat    60
ctgggagtg tagcagctct ttacgataag ggaattaaac ccgacataat ttcgggtaca    120
agtgctggtg ccattgttgg cgcttttatt gctgccggga aaaatccaca cgacgttgtg    180
gagattttta aaaagggatc gtttttcaat tacacaaaac tacagatacc ccgcgacggt    240
ttgatgaaac tggatggact gaaggaattg tttcaaaagg aaattcatgt aaaaaacctc    300
gaagagcttg aaatccccct ttttattgcc atttcgaatt taaataaagg aaccgtggaa    360
tacagaaata gcggtctttt ggttgaaact gtgcttgcc catcttcgat cccatactt    420
tttgctccgg ttttaatcgg cgacgatttg tatgttgacg ggggattaat ggataacatt    480
ccggttgaa ccatcaaaact cgatttgtaa caaatcattg tttcaaatat cagtccgatt    540
aatcccggtg aaaaaattaa aaatctgatt cacattgcta ctcgactttt ttatatgagt    600
gtaaacgcca acatgaaaca ggttaaaaaa tattccaccc atttatattga accggacgga    660
atagatactt acgaaatttt aagccgcacc cagccgatg agttgtatga acttggatat    720
5  aattcgacaa ttaaaatcct gaattcgac taa                                753

```

<210> 18

<211> 250

<212> PRT

<213> Desconocido

10 <220>

<223> Obtenido de muestra ambiental

<220>

<221> DOMINIO

<222> (8)...(166)

15 <223> Fosfolipasa semejante a patatina

<400> 18

```

Met Lys Lys Tyr Lys Thr Gly Leu Val Leu Ser Gly Gly Gly Thr Arg
1           5           10           15

Gly Phe Ala His Leu Gly Val Ile Ala Ala Leu Tyr Asp Lys Gly Ile
          20           25           30

Lys Pro Asp Ile Ile Ser Gly Thr Ser Ala Gly Ala Ile Val Gly Ala

```

ES 2 540 728 T3

35	40	45
Phe Ile Ala Ala Gly Lys Asn Pro His Asp Val Val Glu Ile Phe Lys 50 55 60		
Lys Gly Ser Phe Phe Asn Tyr Thr Lys Leu Gln Ile Pro Arg Asp Gly 65 70 75 80		
Leu Met Lys Leu Asp Gly Leu Lys Glu Leu Phe Gln Lys Glu Ile His 85 90 95		
Val Lys Asn Leu Glu Glu Leu Glu Ile Pro Leu Phe Ile Ala Ile Ser 100 105 110		
Asn Leu Asn Lys Gly Thr Val Glu Tyr Arg Asn Ser Gly Leu Leu Gly 115 120 125		
Glu Thr Val Leu Ala Ser Ser Ser Ile Pro Ile Leu Phe Ala Pro Val 130 135 140		
Leu Ile Gly Asp Asp Leu Tyr Val Asp Gly Gly Leu Met Asp Asn Ile 145 150 155 160		
Pro Val Glu Pro Ile Lys Leu Asp Cys Glu Gln Ile Ile Val Ser Asn 165 170 175		
Ile Ser Pro Ile Asn Pro Val Glu Lys Ile Lys Asn Leu Ile His Ile 180 185 190		
Ala Thr Arg Thr Phe Tyr Met Ser Val Asn Ala Asn Met Lys Gln Val 195 200 205		
Lys Lys Tyr Ser Thr His Tyr Ile Glu Pro Asp Gly Ile Asp Thr Tyr 210 215 220		
Glu Ile Leu Ser Arg Thr His Ala Asp Glu Leu Tyr Glu Leu Gly Tyr 225 230 235 240		
Asn Ser Thr Ile Lys Ile Leu Asn Ser His 245 250		

<210> 19

<211> 1071

<212> ADN

5 <213> Desconocido

<220>

<223> Obtenido de muestra ambiental

<400> 19

atgggtgcga ataatccggt cacaaaattt gtaatggata aggggcaggg cactgcagca 60

cgtttgctgg attgcctgcc gagctttgca caagaacgtc tggtaaaagc attggattat 120
ccttatgact atcctgattt ggatccttat ataaaatgct tgatggcgat tcagattaaa 180
cagggagagc acagttttat tagtgcagat gcggtgcagt cacgtcaact gtttgatgaa 240
cgcatgaaag ctattcaggc gaaaccgacg cgggtcaagg cagtcgagga tctgcgtttg 300
ccattgcaaa atggcactat ctttgcccgca cattatcatc cggcaccaca gaaaaaattg 360
ccgatgggtca ttttctatca tgggtggtgca tttatagtgg gtggtctgga tacgcatgat 420
gagttctgcc gtttattggc ggtgcatgcc aagggtgcagg tactcagcgt ggcttatccg 480
ctaacgcccc aatacagtc tttgcagatg gtacaggtct gtgaagatgc tctggcttgg 540
gtacatcaaa acatcaagca gttgaaaatc tataaaaacc agattgtagt ggcaggggat 600
agtgcaggtg gaaatctggc ggcagtggtt gcgcagcggg gtgccgataa aatctatgca 660
cctcgcgcac agttattgct ttaccggct gtcgatttta aaagccgcca tccttctttt 720
tatgcataca atcagggcct ggtactttca gctcaggata ttgatctggt gaccaagcta 780
tatgctgaaa cacatcaggt cgaactggat gatccgctga tttcaccgac ctatggtgaa 840
ctgaagaatc tgccgcctgc ctatgtgatt accgcccgtc atgatgtgtt gcatgatgag 900
ggttcaatct atgcgctgaa gttacgtgag aatggggtgc gagtgtatta tcaggagtat 960
acggatcagg cccacggttt tatcaatctg accccaattc ataaacgttc gaaaaagcag 1020
gtcattgaac tcagcaagaa tttccgtaaa ttcttgata aaaagatctg a 1071

<210> 20

<211> 356

<212> PRT

5 <213> Desconocido

<220>

<223> Obtenido de muestra ambiental

<400> 20

Met Gly Ala Asn Asn Pro Val Thr Lys Phe Val Met Asp Lys Gly Gln
1 5 10 15

Gly Thr Ala Ala Arg Leu Leu Asp Cys Leu Pro Ser Phe Ala Gln Glu
20 25 30

Arg Leu Val Lys Ala Leu Asp Tyr Pro Tyr Asp Tyr Pro Asp Leu Asp
35 40 45

Pro Tyr Ile Lys Cys Leu Met Ala Ile Gln Ile Lys Gln Gly Glu His
50 55 60

Ser Phe Ile Ser Ala Asp Ala Val Gln Ser Arg Gln Leu Phe Asp Glu
65 70 75 80

ES 2 540 728 T3

Arg Met Lys Ala Ile Gln Ala Lys Pro Thr Pro Val Lys Ala Val Glu
 85 90 95
 Asp Leu Arg Leu Pro Leu Gln Asn Gly Thr Ile Phe Ala Arg His Tyr
 100 105 110
 His Pro Ala Pro Gln Lys Lys Leu Pro Met Val Ile Phe Tyr His Gly
 115 120 125
 Gly Ala Phe Ile Val Gly Gly Leu Asp Thr His Asp Glu Phe Cys Arg
 130 135 140
 Leu Leu Ala Val His Ala Lys Val Gln Val Leu Ser Val Ala Tyr Pro
 145 150 155 160
 Leu Thr Pro Glu Tyr Ser Pro Leu Gln Met Val Gln Val Cys Glu Asp
 165 170 175
 Ala Leu Ala Trp Val His Gln Asn Ile Lys Gln Leu Lys Ile Tyr Lys
 180 185 190
 Asn Gln Ile Val Val Ala Gly Asp Ser Ala Gly Gly Asn Leu Ala Ala
 195 200 205
 Val Val Ala Gln Arg Ser Ala Asp Lys Ile Tyr Ala Pro Arg Ala Gln
 210 215 220
 Leu Leu Leu Tyr Pro Ala Val Asp Phe Lys Ser Arg His Pro Ser Phe
 225 230 235 240
 Tyr Ala Tyr Asn Gln Gly Leu Val Leu Ser Ala Gln Asp Ile Asp Leu
 245 250 255
 Val Thr Lys Leu Tyr Ala Glu Thr His Gln Val Glu Leu Asp Asp Pro
 260 265 270
 Leu Ile Ser Pro Thr Tyr Gly Glu Leu Lys Asn Leu Pro Pro Ala Tyr
 275 280 285
 Val Ile Thr Ala Arg His Asp Val Leu His Asp Glu Gly Ser Ile Tyr
 290 295 300
 Ala Leu Lys Leu Arg Glu Asn Gly Val Arg Val Tyr Tyr Gln Glu Tyr
 305 310 315 320
 Thr Asp Gln Ala His Gly Phe Ile Asn Leu Thr Pro Ile His Lys Arg
 325 330 335
 Ser Lys Lys Gln Val Ile Glu Leu Ser Lys Asn Phe Arg Lys Phe Leu
 340 345 350
 Asp Lys Lys Ile
 355

<210> 21

<211> 1127

5 <212> ADN

<213> *Arabidopsis thaliana*

<400> 21

```

aaaaaaagta aagaaaagaa aaactaataa agaacaaaaa aaatgtcctc ttcttcatca      60
agaaacgcct ttgaagatgg caaatacaaa tcaaatctct taaccttga ctcacatct      120
cgttgctgca aaataacacc gtcttctaga gcttcaccgt ctccgccaaa gcagctgttg      180
gtggctacgc cgggtggagga aggagattat cgggtggtga tgctcctcca tggttacctt      240
ctctacaact ccttctattc tcagcttatg ttgcatgtct cttctcatgg cttcatcctc      300
atcgctcctc agttatatag tatcgccgga ccagacacaa tggatgagat taaatcaacg      360
gcggagatta tggattggtt atcagtagga cttaatcact ttcttcacgc gcaagtaaca      420
ccaaacctat ccaaatttgc cctctccggc catagccgcg gtggcaaac cgcgtttgcg      480
gtcgcttaa agaaatttgg gtactcctcg aatctaaaga tctcgacatt gatcgggata      540
gatccagtcg atggaacagg gaaagggaaa caaacccctc ctccggtgtt ggcttacctt      600
ccaaactcat ttgacctaga caaacgcct atacttgtga tcggttcggg gcttggtgaa      660
accgctcgga acccattatt cccaccgtgt gcacctccc gagtgaatca ccgagagttc      720
tttcgggaat gtcaaggtec agcatggcat ttcgttgca aggattatgg gcatttggac      780
atgcttgatg atgatacaaa agggattaga gggaaagatt cttattgttt gtgtaagaat      840
ggtgaagaga ggagaccaat gaggagattc gttggtggac ttgttgtatc atttttgaag      900
gcttatttgg aaggagatga tcgtgaatta gttaagatca aagatgggtg tcacgaggat      960
gttcccgttg aaattcaaga gtttgagggt atcatgtaaa cataagtttt tctttagggg    1020
ctgggttttc tattgtcaat atcatcagct tttgttgctt atgggttttac aaacttatat    1080
tgtacaactc ttaagtcac ctctttgctt atgatattaa cccgac      1127

```

<210> 22

5 <211> 318

<212> PRT

<213> *Arabidopsis thaliana*

<400> 22

```

Met Ser Ser Ser Ser Ser Arg Asn Ala Phe Glu Asp Gly Lys Tyr Lys
1           5           10           15

Ser Asn Leu Leu Thr Leu Asp Ser Ser Ser Arg Cys Cys Lys Ile Thr
          20           25           30

```

ES 2 540 728 T3

Pro Ser Ser Arg Ala Ser Pro Ser Pro Pro Lys Gln Leu Leu Val Ala
 35 40 45
 Thr Pro Val Glu Glu Gly Asp Tyr Pro Val Val Met Leu Leu His Gly
 50 55 60
 Tyr Leu Leu Tyr Asn Ser Phe Tyr Ser Gln Leu Met Leu His Val Ser
 65 70 75 80
 Ser His Gly Phe Ile Leu Ile Ala Pro Gln Leu Tyr Ser Ile Ala Gly
 85 90 95
 Pro Asp Thr Met Asp Glu Ile Lys Ser Thr Ala Glu Ile Met Asp Trp
 100 105 110
 Leu Ser Val Gly Leu Asn His Phe Leu Pro Ala Gln Val Thr Pro Asn
 115 120 125
 Leu Ser Lys Phe Ala Leu Ser Gly His Ser Arg Gly Gly Lys Thr Ala
 130 135 140
 Phe Ala Val Ala Leu Lys Lys Phe Gly Tyr Ser Ser Asn Leu Lys Ile
 145 150 155 160
 Ser Thr Leu Ile Gly Ile Asp Pro Val Asp Gly Thr Gly Lys Gly Lys
 165 170 175
 Gln Thr Pro Pro Pro Val Leu Ala Tyr Leu Pro Asn Ser Phe Asp Leu
 180 185 190
 Asp Lys Thr Pro Ile Leu Val Ile Gly Ser Gly Leu Gly Glu Thr Ala
 195 200 205
 Arg Asn Pro Leu Phe Pro Pro Cys Ala Pro Pro Gly Val Asn His Arg
 210 215 220
 Glu Phe Phe Arg Glu Cys Gln Gly Pro Ala Trp His Phe Val Ala Lys
 225 230 235 240
 Asp Tyr Gly His Leu Asp Met Leu Asp Asp Asp Thr Lys Gly Ile Arg
 245 250 255
 Gly Lys Ser Ser Tyr Cys Leu Cys Lys Asn Gly Glu Glu Arg Arg Pro
 260 265 270
 Met Arg Arg Phe Val Gly Gly Leu Val Val Ser Phe Leu Lys Ala Tyr
 275 280 285
 Leu Glu Gly Asp Asp Arg Glu Leu Val Lys Ile Lys Asp Gly Cys His
 290 295 300
 Glu Asp Val Pro Val Glu Ile Gln Glu Phe Glu Val Ile Met
 305 310 315

<210> 23

<211> 1157

5 <212> ADN

<213> *Ginkgo biloba*

<400> 23

ttgaaaaaca aaaacgaaga agatgaactc agtacttgca cacagccatc ggccatgggt	60
ttagtgaagg atgtgttcag cgaaggctct ttacctgttc aaatcctcgc aattccacaa	120
gccaactcat ctccatgctc aaaattagca gacaaaaacg gaactgcaac cacgccttct	180
ccttgctggc ctccataaac cctgctgac gctcttcctt cccaacatgg agattatcct	240
ctcctcctct ttttccacgg ctatgtactc ctcaattcct tctattctca actcttgccg	300
catgttgctt cccatggata catcgccata gctcctcaga tgtacagtgt aattggccca	360
aatacgactc cagaaatagc cgatgcagcg gccattacag actgggtacg agatggactc	420
tccgataatc ttccgcaagc tttaaacaat catgtgagcg ccaattttga gaaatttggt	480
ctagcggggc actcgcgcgg gggtaaagtg gcatttgac ttgccctagg tcgagtctcg	540
cagccatctt taaagtactc ggcccttgta ggtcttgatc cagtcgatgg aatgggaaaa	600
gatcaacaaa ccagtcctcc tattctgtca tacagagagc attcctttga tttgggtatg	660
ccaacattag tggtaggttc gggcctgggt ccgtgcaaaa gaaaccctct cttccctccc	720
tgtgctcccc aagggtgtaa ccaccatgat ttcttctacg aatgtgtcgc tcctgcctat	780
cattttgttg cctctgatta tgggcatctt gatttcttag acgacgacac caaaggaata	840
agaggaaagg ctacttattg cctctgtaag aatggggaag caagagagcc aatgcggaag	900
tttagcgggt gaattgtggt tgcatttctt caagcatttc ttggtgataa tcgtggagcc	960
ctgaatgata ttatgggtta tccttcacat gctccagtca agattgagcc tccagagtct	1020
ttggttacag aagatgtaaa atccccagaa gtcgaattat tacgccgggc agtttgcaga	1080
tgatgtacca tggattatg cattaaagga atgtatttgt tattaataaa atattaagaa	1140
gtaaaaaaaa aaaaaaa	1157

<210> 24

5 <211> 342

<212> PRT

<213> *Ginkgo biloba*

<400> 24

Met	Val	Leu	Val	Lys	Asp	Val	Phe	Ser	Glu	Gly	Pro	Leu	Pro	Val	Gln
1				5					10					15	

ES 2 540 728 T3

Ile Leu Ala Ile Pro Gln Ala Asn Ser Ser Pro Cys Ser Lys Leu Ala
20 25 30

Asp Lys Asn Gly Thr Ala Thr Thr Pro Ser Pro Cys Arg Pro Pro Lys
35 40 45

Pro Leu Leu Ile Ala Leu Pro Ser Gln His Gly Asp Tyr Pro Leu Ile
50 55 60

Leu Phe Phe His Gly Tyr Val Leu Leu Asn Ser Phe Tyr Ser Gln Leu
65 70 75 80

Leu Arg His Val Ala Ser His Gly Tyr Ile Ala Ile Ala Pro Gln Met
85 90 95

Tyr Ser Val Ile Gly Pro Asn Thr Thr Pro Glu Ile Ala Asp Ala Ala
100 105 110

Ala Ile Thr Asp Trp Leu Arg Asp Gly Leu Ser Asp Asn Leu Pro Gln
115 120 125

Ala Leu Asn Asn His Val Arg Pro Asn Phe Glu Lys Phe Val Leu Ala
130 135 140

Gly His Ser Arg Gly Gly Lys Val Ala Phe Ala Leu Ala Leu Gly Arg
145 150 155 160

Val Ser Gln Pro Ser Leu Lys Tyr Ser Ala Leu Val Gly Leu Asp Pro
165 170 175

Val Asp Gly Met Gly Lys Asp Gln Gln Thr Ser His Pro Ile Leu Ser
180 185 190

Tyr Arg Glu His Ser Phe Asp Leu Gly Met Pro Thr Leu Val Val Gly
195 200 205

Ser Gly Leu Gly Pro Cys Lys Arg Asn Pro Leu Phe Pro Pro Cys Ala
210 215 220

Pro Gln Gly Val Asn His His Asp Phe Phe Tyr Glu Cys Val Ala Pro
225 230 235 240

Ala Tyr His Phe Val Ala Ser Asp Tyr Gly His Leu Asp Phe Leu Asp
245 250 255

Asp Asp Thr Lys Gly Ile Arg Gly Lys Ala Thr Tyr Cys Leu Cys Lys
260 265 270

Asn Gly Glu Ala Arg Glu Pro Met Arg Lys Phe Ser Gly Gly Ile Val
275 280 285

Val Ala Phe Leu Gln Ala Phe Leu Gly Asp Asn Arg Gly Ala Leu Asn
290 295 300

Asp Ile Met Val Tyr Pro Ser His Ala Pro Val Lys Ile Glu Pro Pro
305 310 315 320

Glu Ser Leu Val Thr Glu Asp Val Lys Ser Pro Glu Val Glu Leu Leu
325 330 335

Arg Arg Ala Val Cys Arg
340

<210> 25

<211> 884

<212> ADN

5 <213> *Brassica oleracea*

<400> 25

acacaaaaaa	atatataaca	caaagaaata	gaagaaggaa	aaaatgtccc	cctcctttct	60
tttctttact	ttgtttttga	taaaggaaat	gtcctcttca	tcatcagcaa	actcctttga	120
ggacggcaaa	tacaaaacag	atcttttaac	agtaggctta	tcatcttgct	gctggaaaaa	180
gccctcctct	tctccgactc	cgcagtctcc	gccgaagagg	cttttggtgg	caacgccggt	240
ggaggaagga	gaatatccgg	tggtgatgct	cctccatggg	taccttctct	acaactcatt	300
ttattcccag	cttatgttgc	atgtctcttc	ccatggcttc	attgtcatcg	ctccgcagtt	360
atatagcatt	gccggaccag	acaccatgga	tgagataaaa	tcaacggcag	agattattga	420
ttggttatcg	gtcggactaa	accactttct	tccaccacaa	gtaacaccaa	acctatccaa	480
gttcgcactc	tccggccata	gccgtggtgg	gaagaccgca	tttgccttgg	ccttaaagaa	540
atttgataac	tcgtccgacc	taaagatctc	ggcattgata	ggtatagatg	ttggaactgt	600
tttttggaac	aatggctatg	gccaatattc	cgggtgaattt	ttcgagcaat	ttgattgtcg	660
aatgaccggg	attgtggaat	cgtaggattc	attgttatga	gcactatggg	atagtgtaat	720
catatatcaa	aaacgaagtt	cgtttgaatg	agaaatgaaa	gtctaaaata	gattatttgt	780
aaaatatcta	tattagaatt	atgaggtaag	aaacctcttg	tgtttaaaat	ggagaagtta	840
taacaaagtt	ataaaaaact	ttgtaaacaa	tttggtgtgt	tagc		884

<210> 26

<211> 213

10 <212> PRT

<213> *Brassica oleracea*

<400> 26

Met Ser Pro Ser Phe Leu Phe Phe Thr Leu Phe Leu Ile Lys Glu Met

ES 2 540 728 T3

[illegible]

<210> 27

<211> 329

<212> PRT

5 <213> *Citrus sinensis*

<400> 27

Met Ala Ala Met Val Asp Ala Lys Pro Ala Ala Ser Val Gln Gly Thr
1 5 10 15

Pro Leu Leu Ala Thr Ala Thr Leu Pro Val Phe Thr Arg Gly Ile Tyr
 20 25 30
 Ser Thr Lys Arg Ile Thr Leu Glu Thr Ser Ser Pro Ser Ser Pro Pro
 35 40 45
 Pro Pro Lys Pro Leu Ile Ile Val Thr Pro Ala Gly Lys Gly Thr Phe
 50 55 60
 Asn Val Ile Leu Phe Leu His Gly Thr Ser Leu Ser Asn Lys Ser Tyr
 65 70 75 80
 Ser Lys Ile Phe Asp His Ile Ala Ser His Gly Phe Ile Val Val Ala
 85 90 95
 Pro Gln Leu Tyr Thr Ser Ile Pro Pro Pro Ser Ala Thr Asn Glu Leu
 100 105 110
 Asn Ser Ala Ala Glu Val Ala Glu Trp Leu Pro Gln Gly Leu Gln Gln
 115 120 125
 Asn Leu Pro Glu Asn Thr Glu Ala Asn Val Ser Leu Val Ala Val Met
 130 135 140
 Gly His Ser Arg Gly Gly Gln Thr Ala Phe Ala Leu Ser Leu Arg Tyr
 145 150 155 160
 Gly Phe Gly Ala Val Ile Gly Leu Asp Pro Val Ala Gly Thr Ser Lys
 165 170 175
 Thr Thr Gly Leu Asp Pro Ser Ile Leu Ser Phe Asp Ser Phe Asp Phe
 180 185 190
 Ser Ile Pro Val Thr Val Ile Gly Thr Gly Leu Gly Gly Val Ala Arg
 195 200 205
 Cys Ile Thr Ala Cys Ala Pro Glu Gly Ala Asn His Glu Glu Phe Phe
 210 215 220
 Asn Arg Cys Lys Asn Ser Ser Arg Ala His Phe Val Ala Thr Asp Tyr
 225 230 235 240
 Gly His Met Asp Ile Leu Asp Asp Asn Pro Ser Asp Val Lys Ser Trp
 245 250 255
 Ala Leu Ser Lys Tyr Phe Cys Lys Asn Gly Asn Glu Ser Arg Asp Pro
 260 265 270

ES 2 540 728 T3

Met Arg Arg Cys Val Ser Gly Ile Val Val Ala Phe Leu Lys Asp Phe
 275 280 285

Phe Tyr Gly Asp Ala Glu Asp Phe Arg Gln Ile Leu Lys Asp Pro Ser
 290 295 300

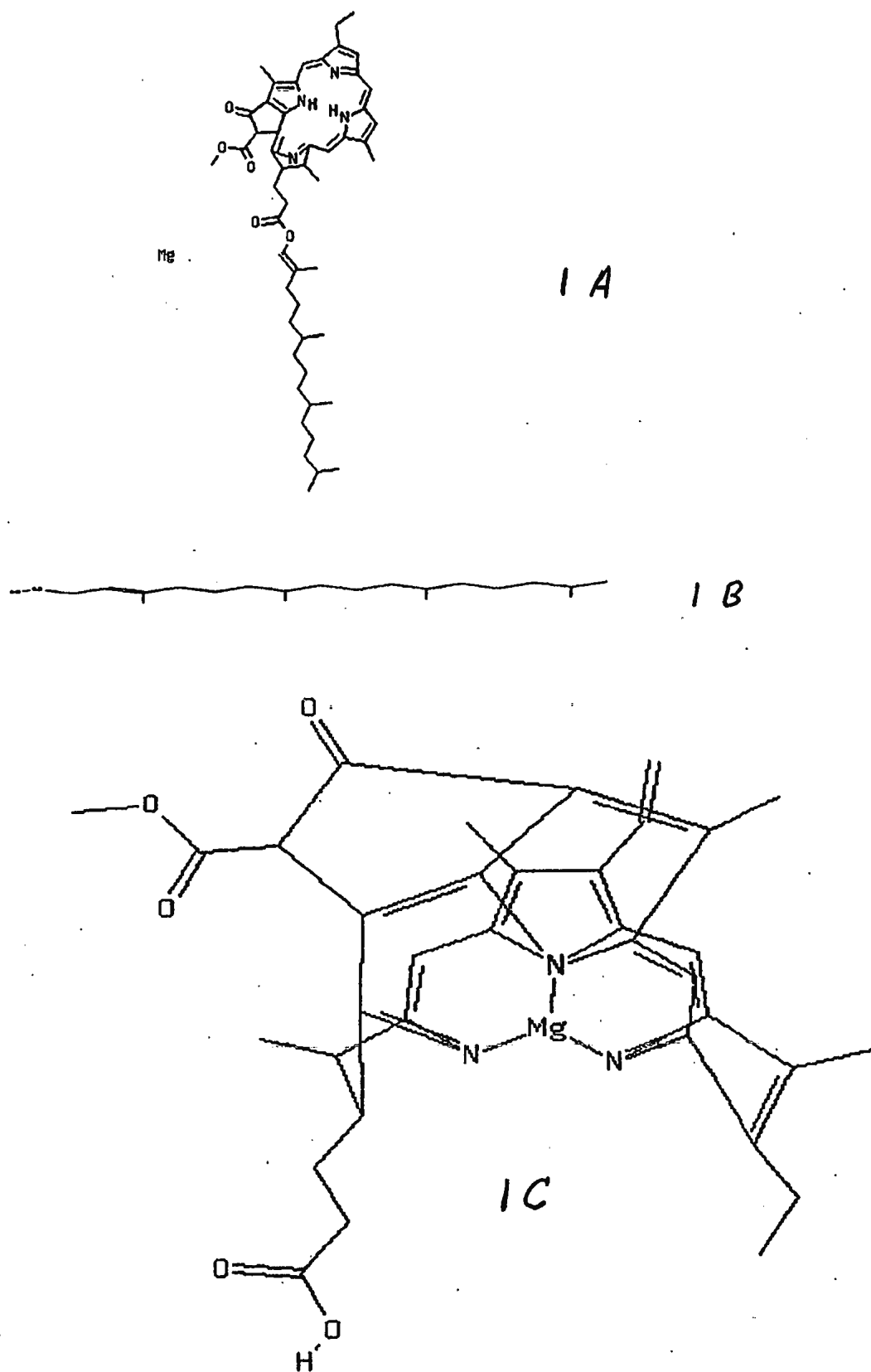
Phe Ala Pro Ile Lys Leu Asp Ser Val Glu Tyr Ile Asp Ala Ser Ser
 305 310 315 320

Met Leu Thr Thr Thr His Val Lys Val
 325

REIVINDICACIONES

1. Uso de al menos un polipéptido según un polipéptido de SEC ID NO:22, o un polipéptido codificado por la secuencia de ácido nucleico de SEC ID NO:21, que tiene una actividad de clorofilasa, en un tratamiento enzimático que es una decoloración de una composición que contiene feofitina en condiciones en las que el polipéptido puede catalizar la hidrólisis de la feofitina para generar una feoforbida y un fitol; y la feoforbida y el fitol se separan.
2. El uso de la reivindicación 1, en el que el polipéptido se inmoviliza en: una matriz, una célula, un metal, una resina, un polímero, un material cerámico, un vidrio, un microelectrodo, una partícula grafitica, una perla, un gel, una placa, un tubo capilar, un soporte inorgánico o soporte orgánico, y opcionalmente el soporte comprende: una alúmina, celite, Dowex-1-cloruro, perlas de vidrio, sílice, gel de sílice, hidrogel de alginato, o perla de alginato.
3. El uso de cualquiera de las reivindicaciones 1 ó 2, en el que la composición que contiene feofitina:
 - (a) deriva de una planta, un animal o un alga, o una mezcla de los mismos,
 - (b) comprende un material vegetal, aceite vegetal o extracto vegetal,
 - (c) comprende un aceite vegetal o un aceite de semillas,
 - (d) comprende un aceite de palma o un aceite de cánola,
 - (e) comprende un aceite bruto o un aceite refinado,
 - (f) comprende una preparación de aceite bruto sin diluir, o
 - (g) comprende una preparación de algas.
4. El uso de cualquiera de las reivindicaciones 1 a 3, en el que al menos una etapa del tratamiento se lleva a cabo en una vasija de reacción, y opcionalmente la vasija de reacción comprende un dispositivo de separación de goma gravitacional o un tanque de retención.
5. El uso de cualquiera de las reivindicaciones 1 a 3, en el que al menos una etapa del tratamiento se lleva a cabo en un extracto celular o en una célula completa.
6. El uso de cualquiera de las reivindicaciones 1 a 5, en el que el tratamiento comprende además modificar el pH para promover la separación acuosa de una feoforbida.
7. El uso de cualquiera de las reivindicaciones 1 a 6, en el que el tratamiento se combina con una etapa de neutralización cáustica.
8. El uso de cualquiera de las reivindicaciones 1 a 7, en el que se usa polipéptido con una fosfolipasa, y opcionalmente la fosfolipasa es una fosfolipasa C.
9. El uso de cualquiera de las reivindicaciones 1 a 8, en el que el tratamiento está incluido en un procedimiento de desgomado.
10. El uso de cualquiera de las reivindicaciones 1 a 9, en el que el tratamiento comprende además la eliminación de clorofila o feofitina residual; plaguicidas; o hidrocarburos aromáticos policíclicos.
11. El uso de cualquiera de las reivindicaciones 1 a 10, en el que el tratamiento enzimático de una composición que contiene feofitina se combina con un procedimiento de desgomado y una etapa de neutralización cáustica.

Figura 1



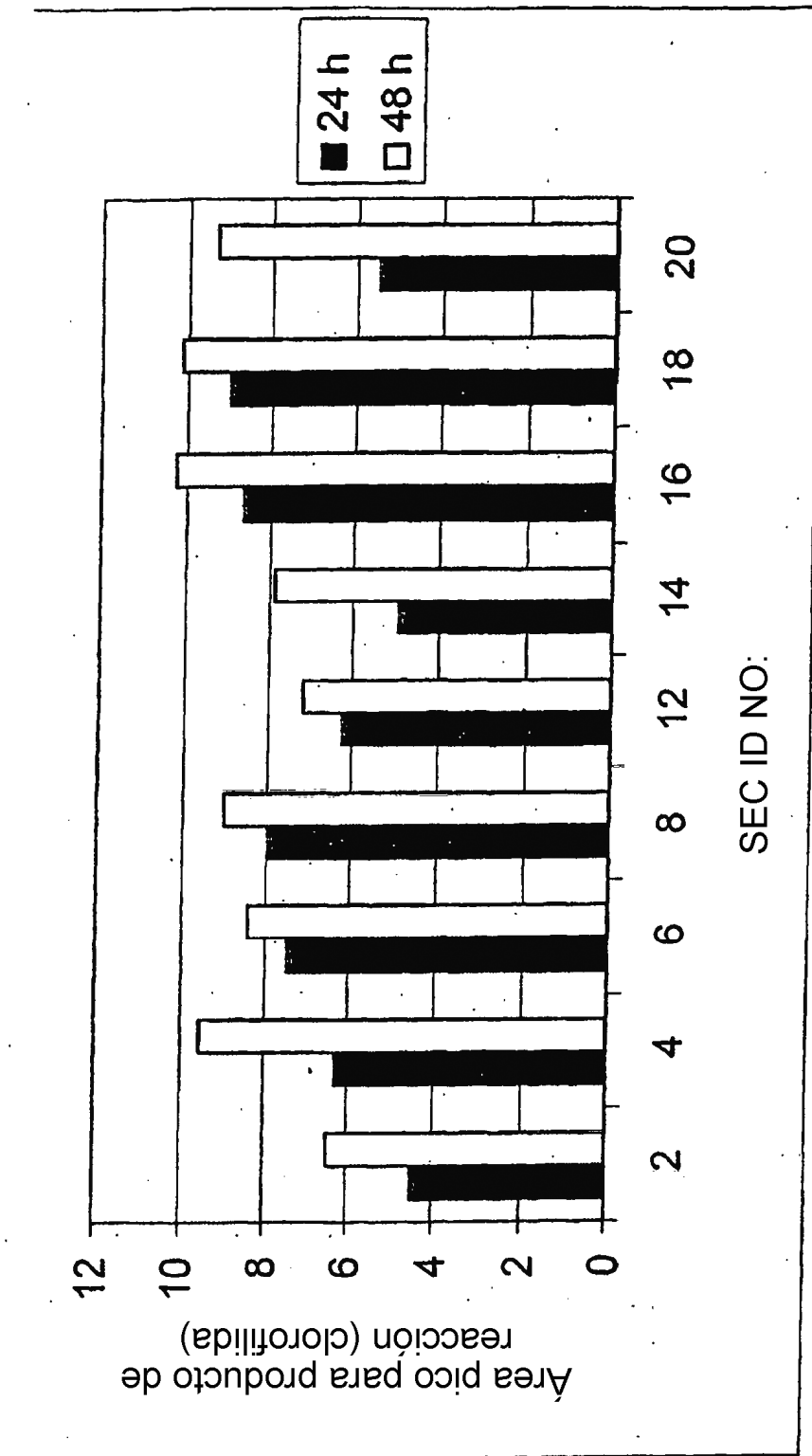


Figura 2

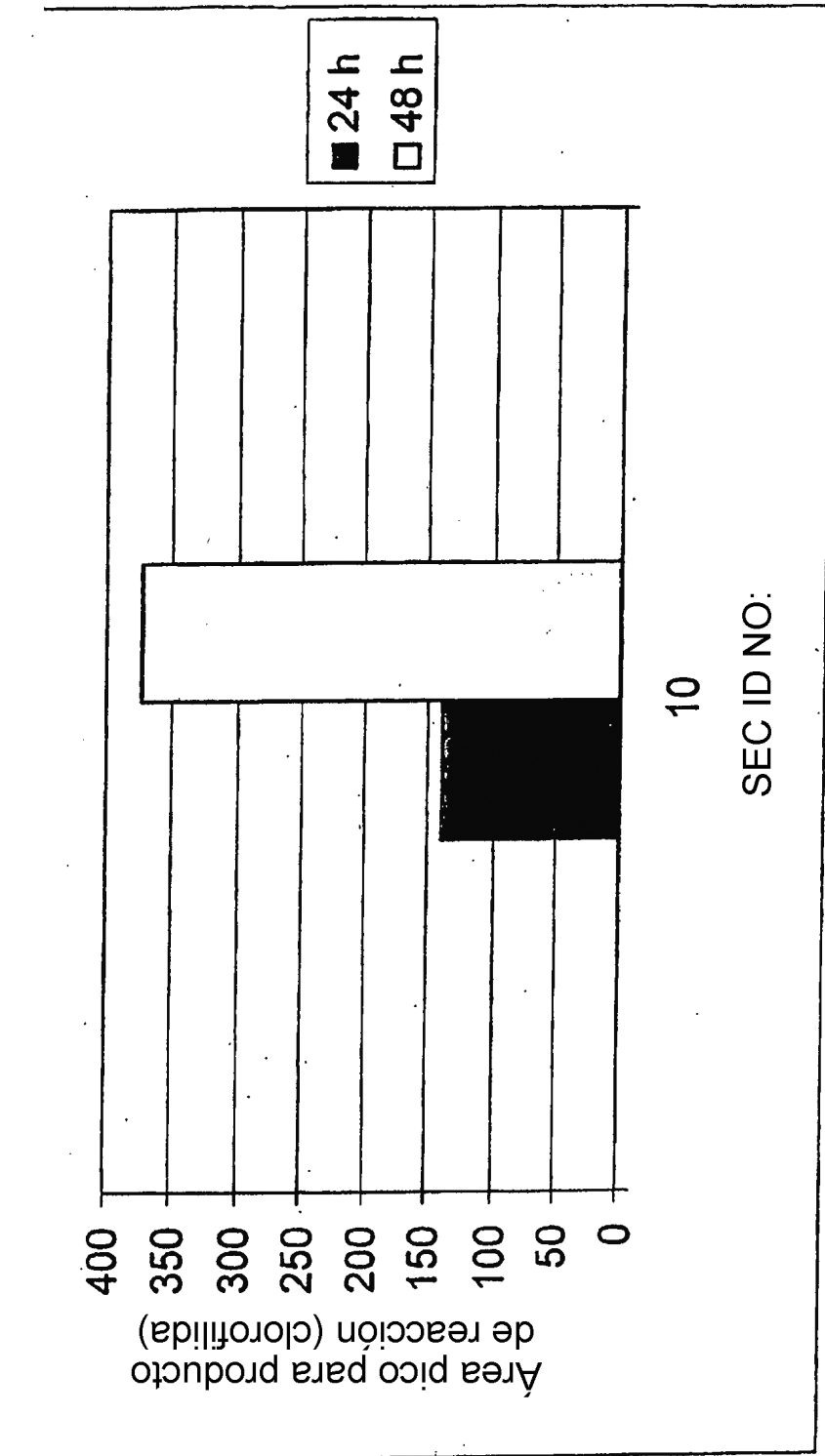


Figura 3

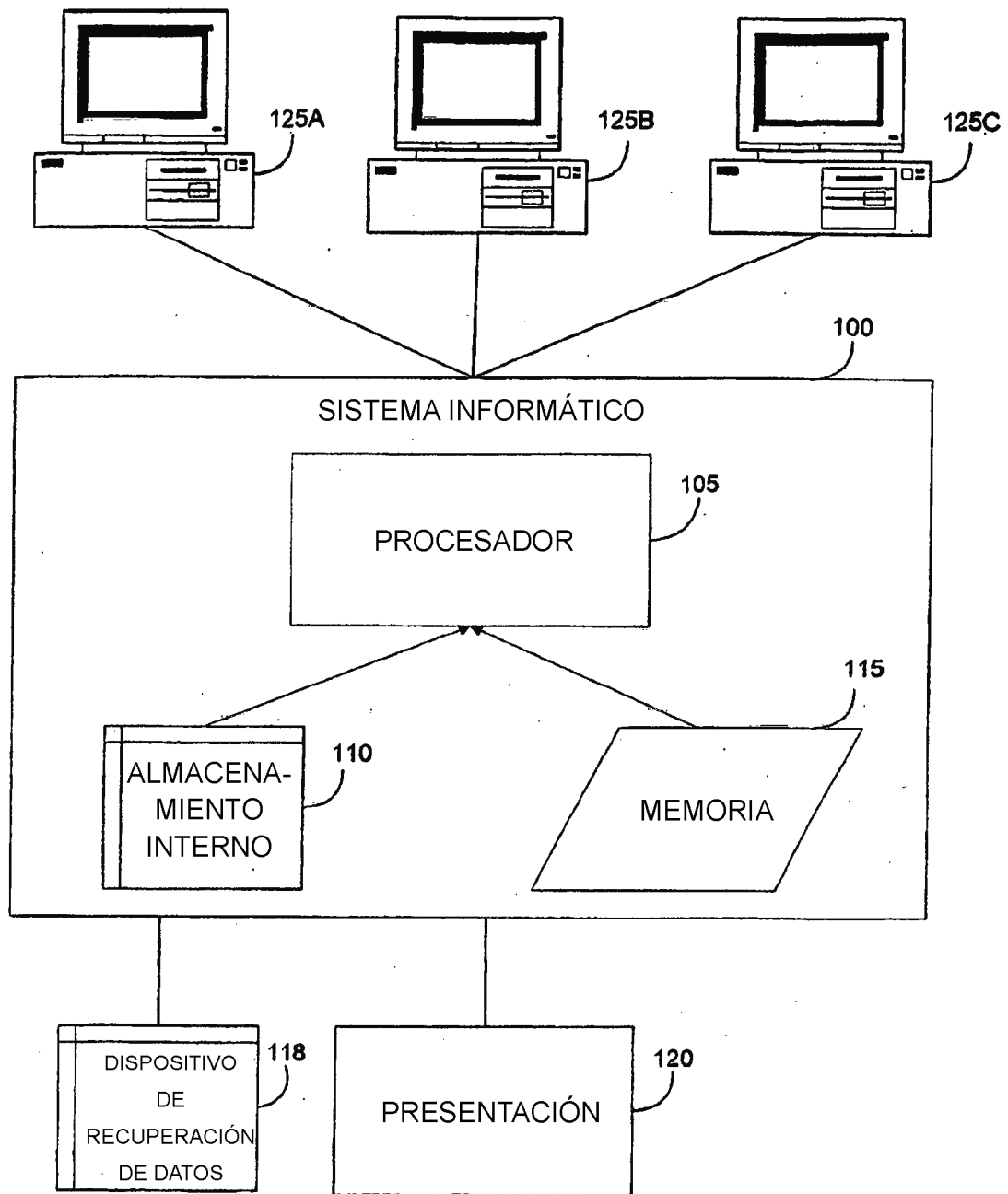


Figura 4

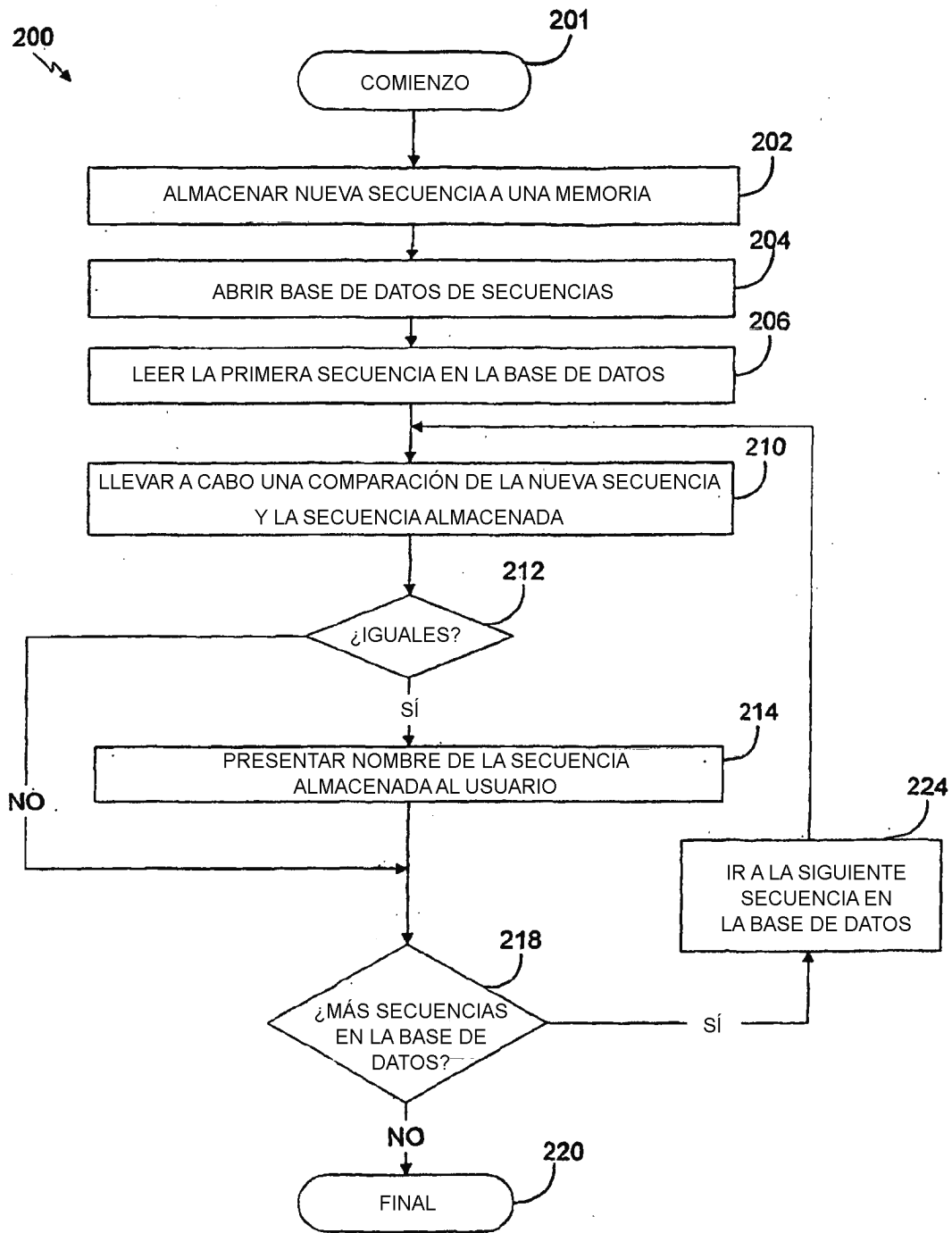


Figura 5

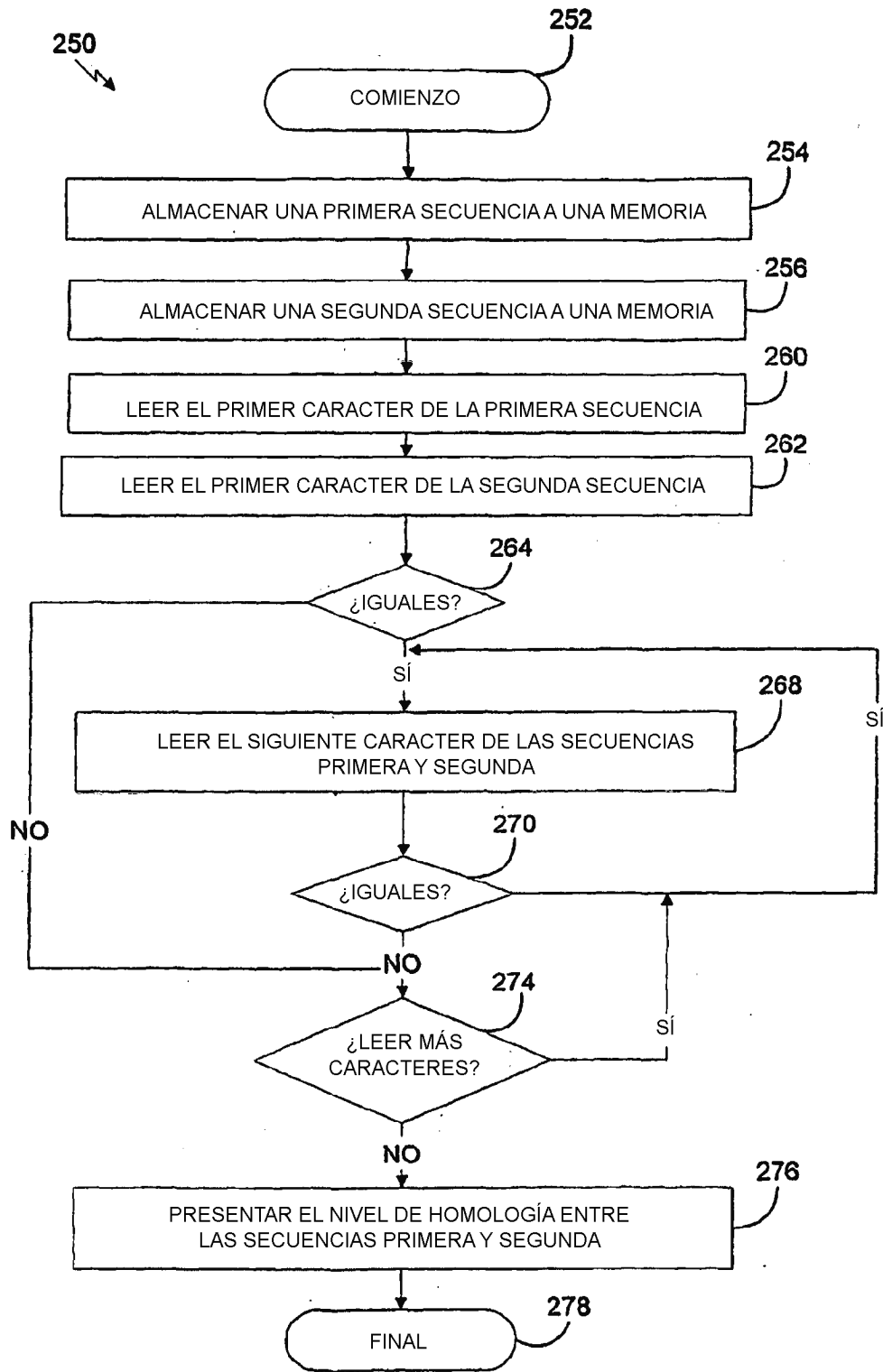


FIGURA 6

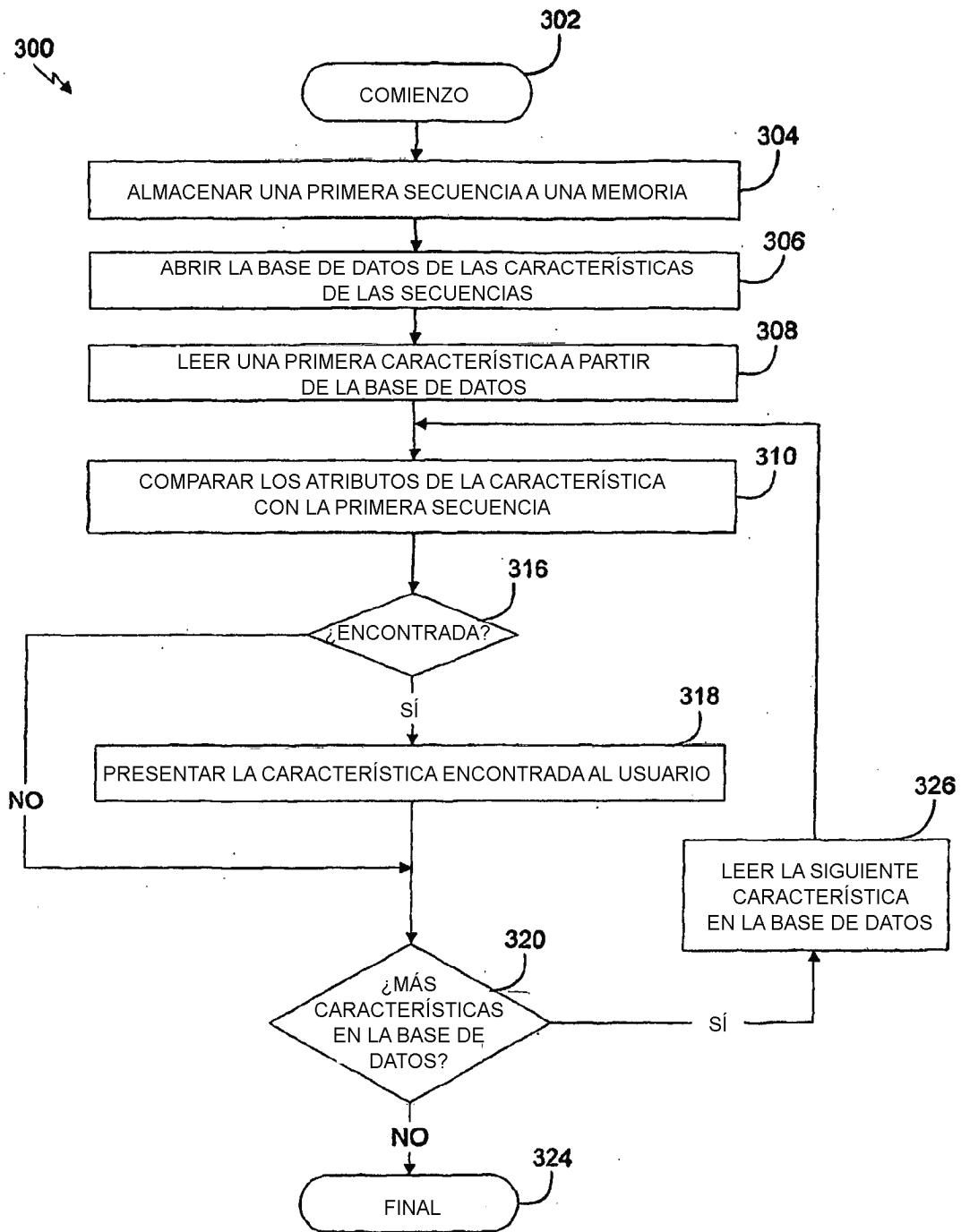


Figura 7

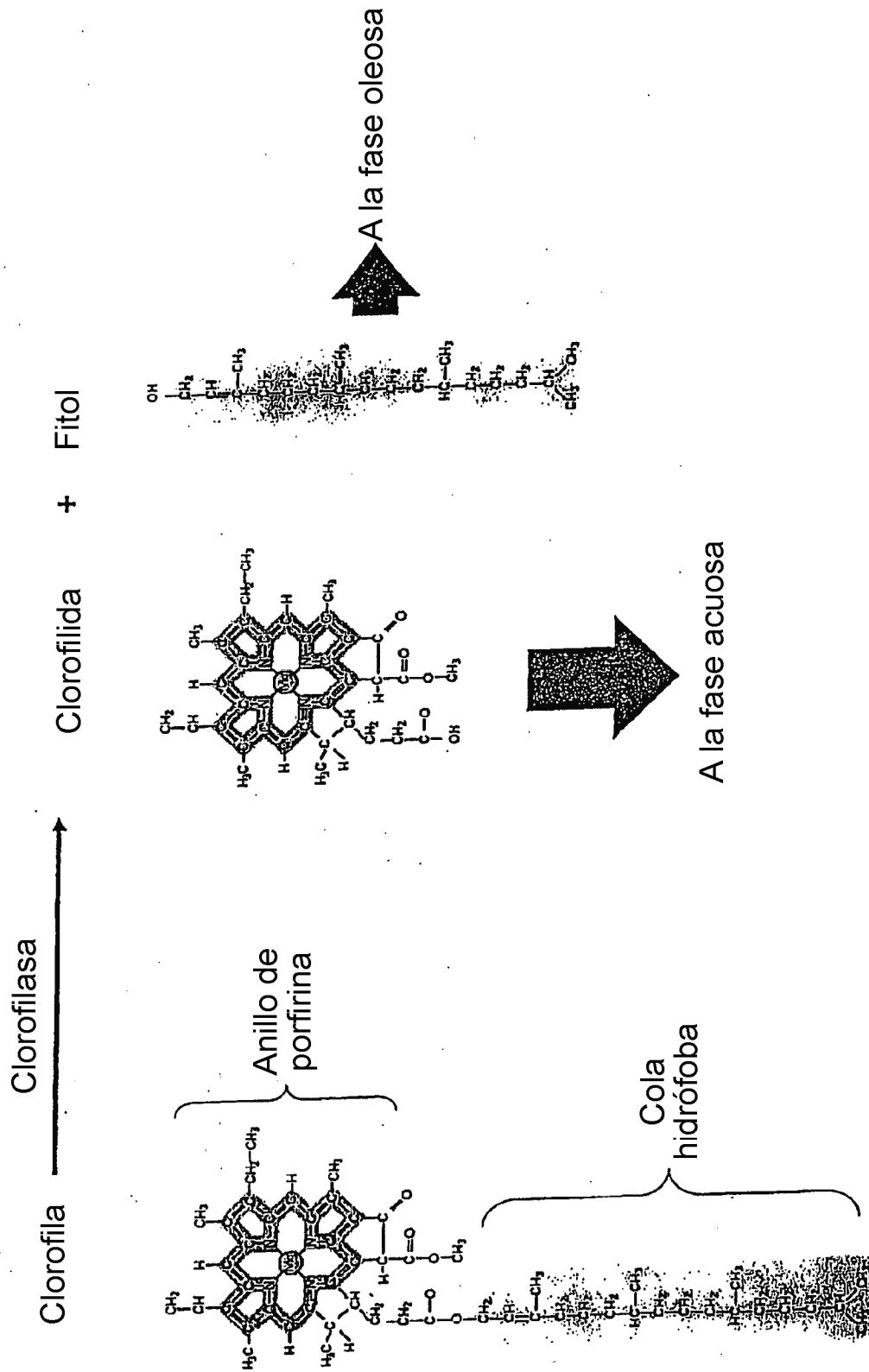


FIGURA 8

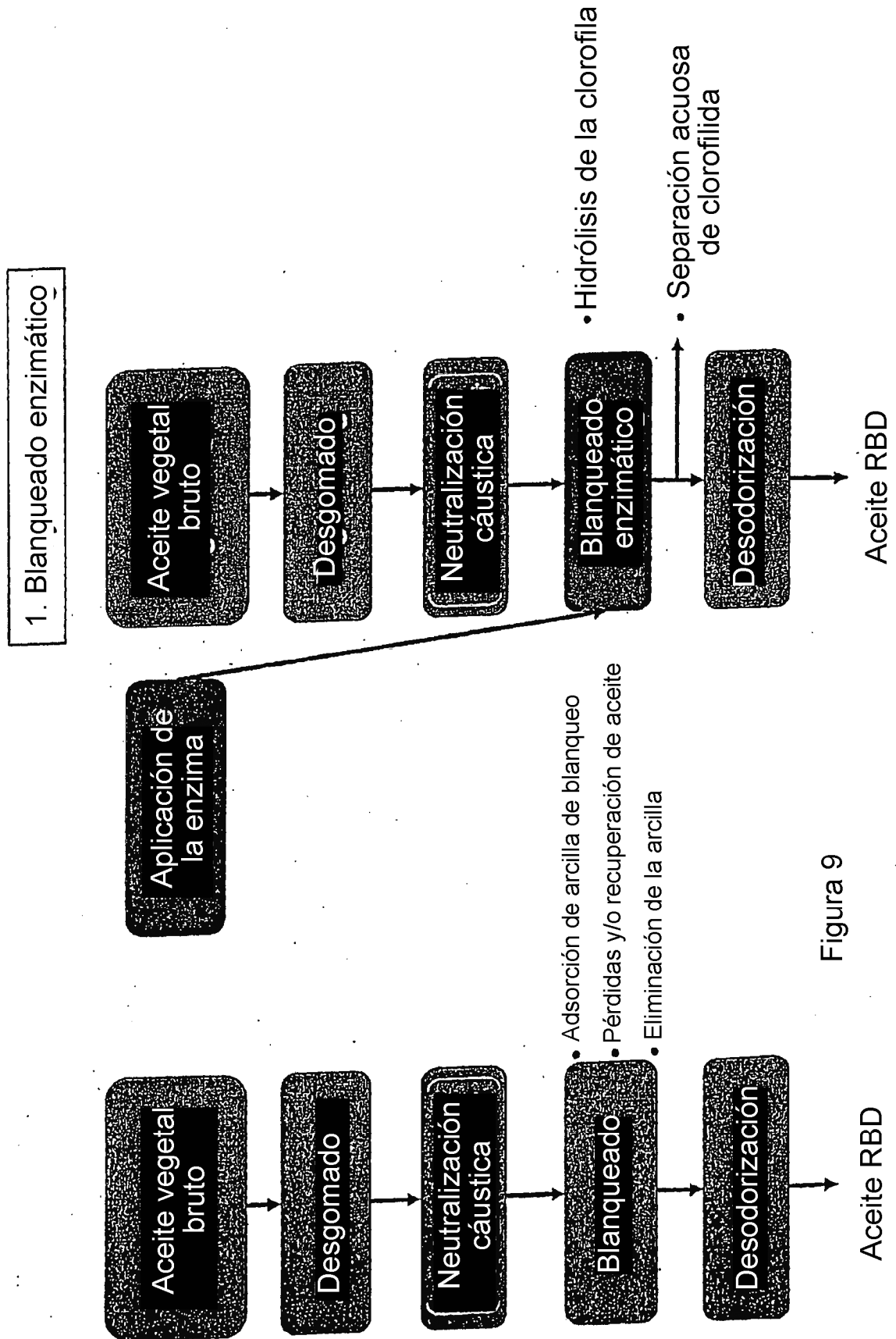


Figura 9

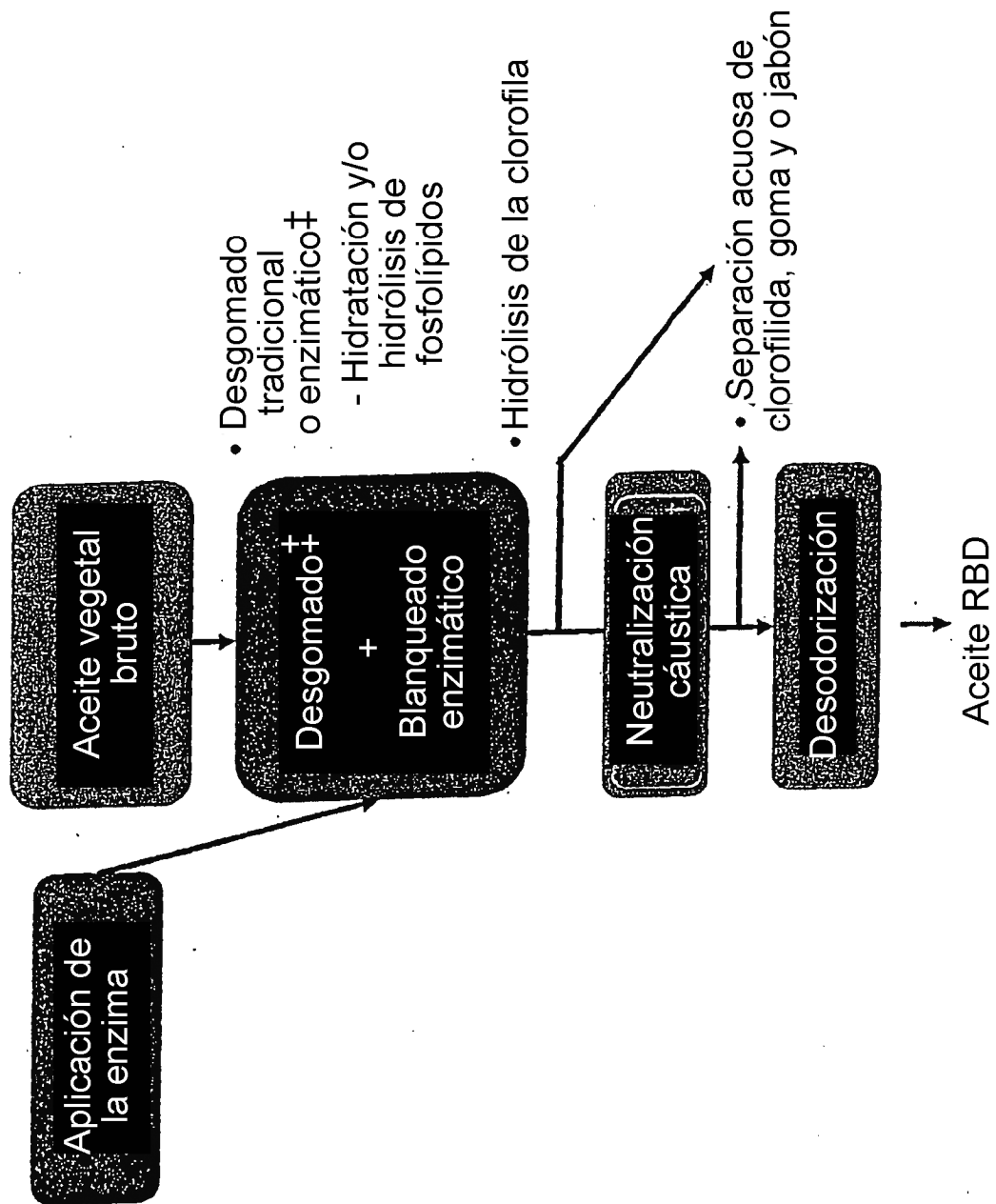


Figura 10

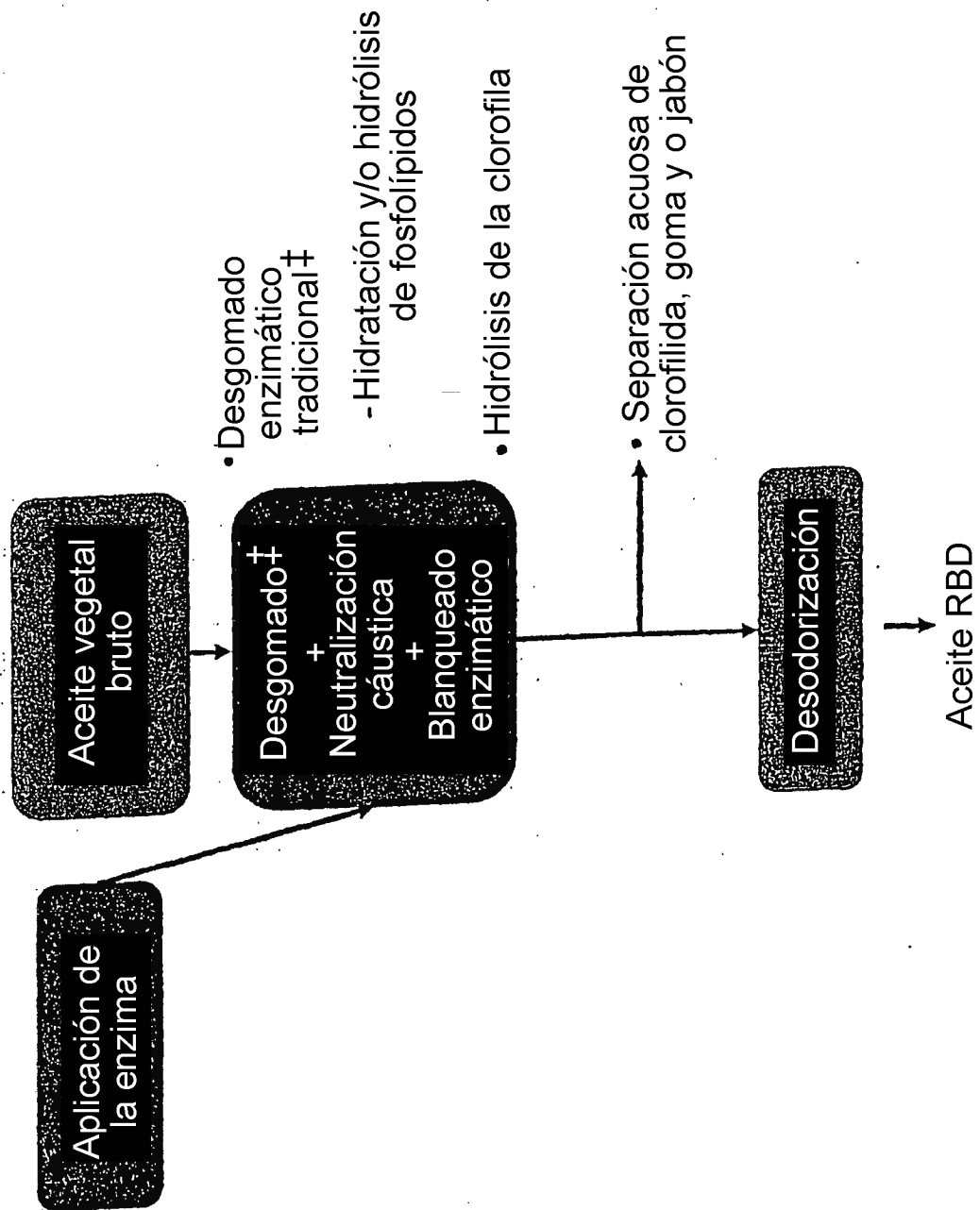


Figura 11

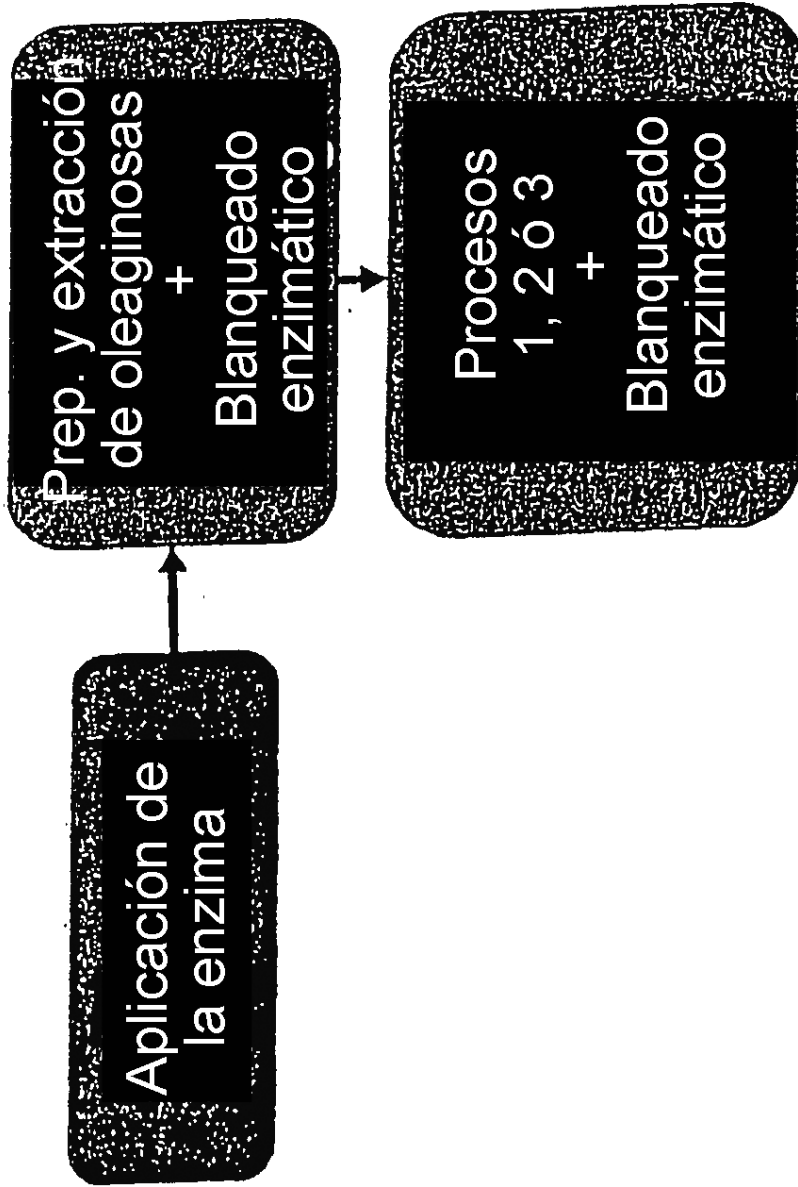


Figura 12

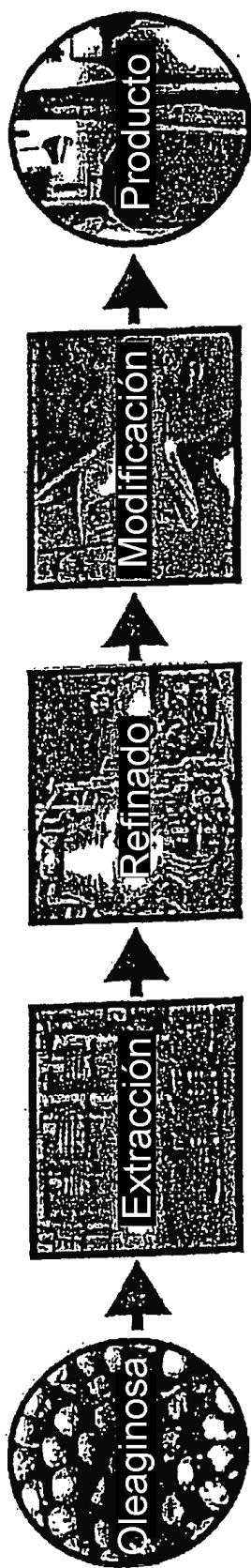


Figura 13

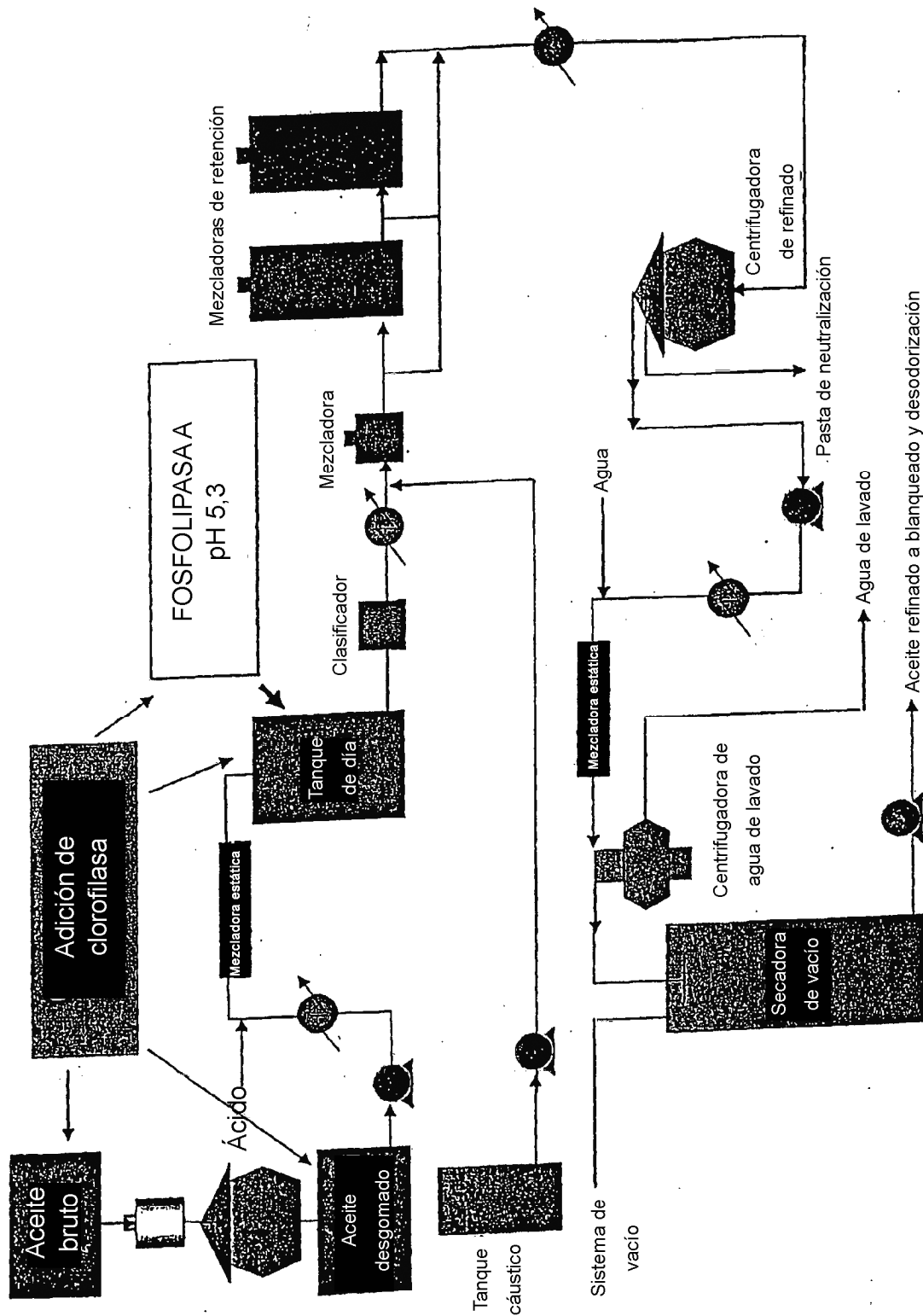


Figura 14

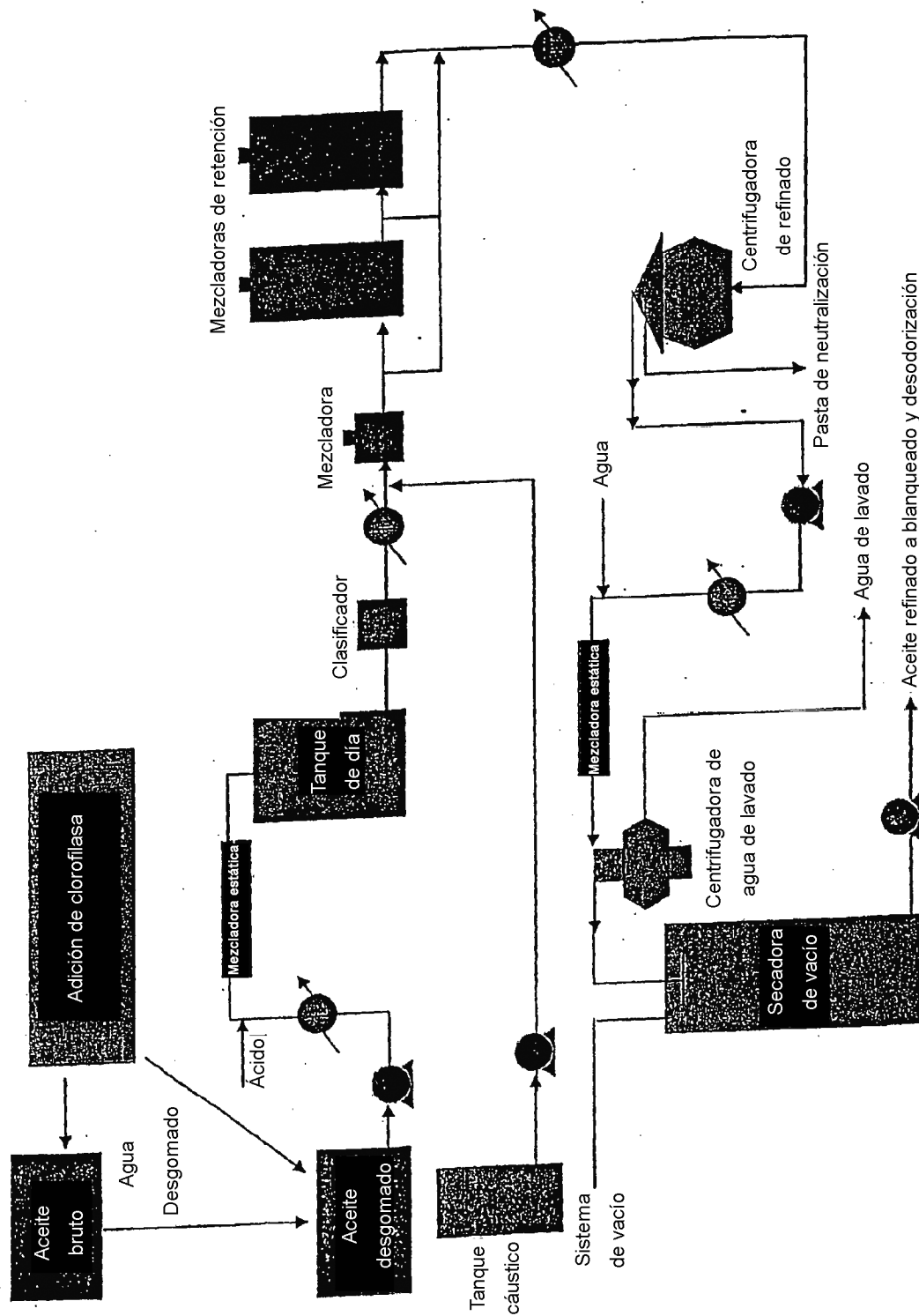


Figura 15

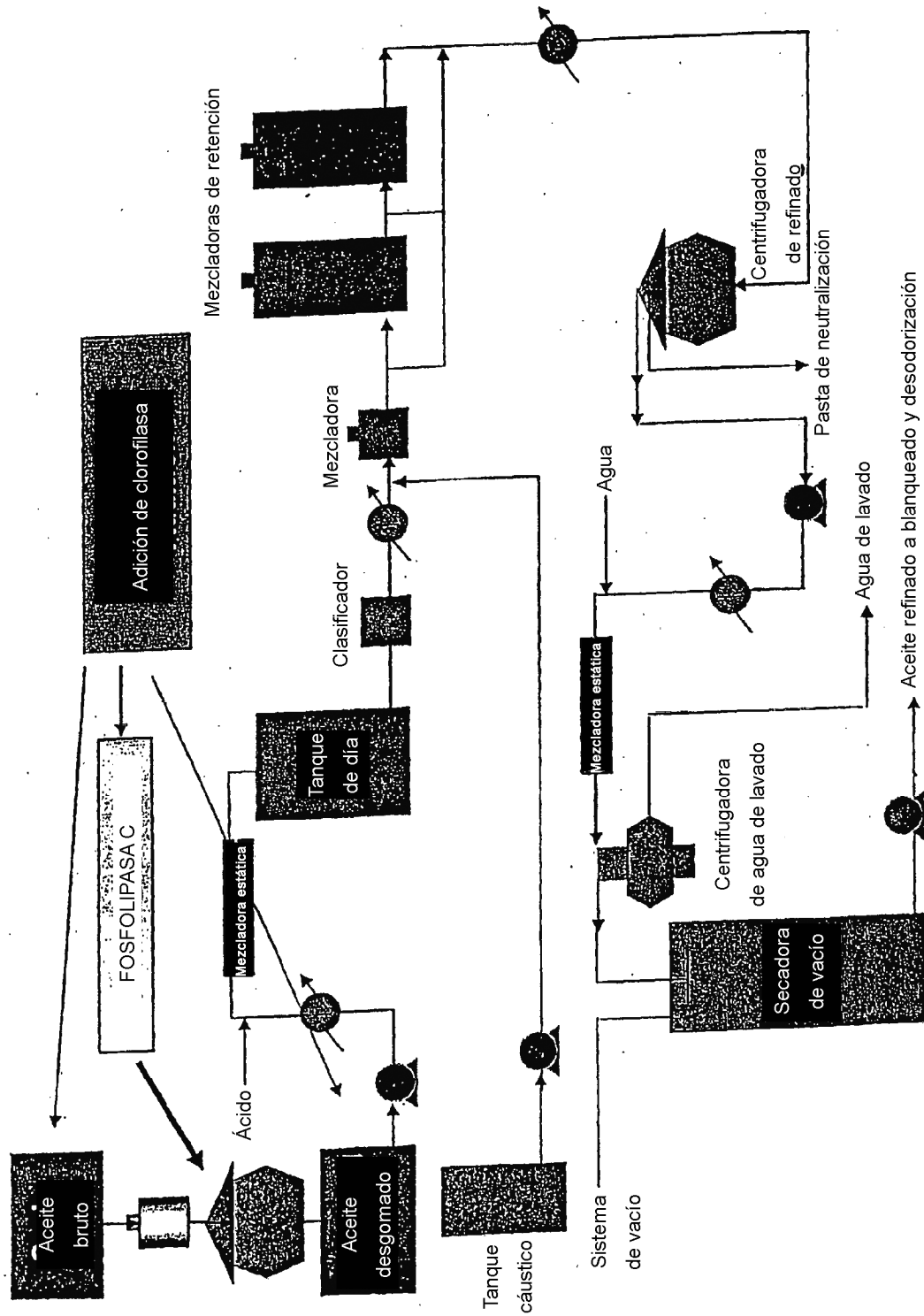


Figura 16