

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 461 857**

51 Int. Cl.:

C12N 9/18 (2006.01)

A23L 1/27 (2006.01)

C11B 3/00 (2006.01)

A23K 1/165 (2006.01)

A23L 1/227 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **14.06.2005 E 05762779 (6)**

97 Fecha y número de publicación de la concesión europea: **26.02.2014 EP 1791853**

54 Título: **Composiciones y métodos para la decoloración enzimática de clorofila**

30 Prioridad:

16.06.2004 US 580447 P

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

21.05.2014

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

ES 2 461 857 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

Composiciones y métodos para la decoloración enzimática de clorofila

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 clorofila (por ejemplo, una clorofilasa) para la modificación enzimática de una clorofila, 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 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.

La degradación de la clorofila en aceite de cáñola con clorofilasa derivada del alga *Phaeodactylum tricornutum* es conocida desde Khamessan y Kermasha, J. Food Biochemistry 1996 (20), p. 311-328. La actividad de clorofilasa descrita aquí parece estar influida por la presencia de un tensioactivo, un disolvente orgánico y la concentración de aceite de cáñola.

35 El documento WO 2005/032496 describe polinucleótidos que codifican polipéptidos que tienen actividad de hidrolasa. El documento WO 2005/032496 no describe el uso de hidrolasas en la degradación de clorofila.

SUMARIO

40 La invención proporciona composiciones y métodos para el tratamiento enzimático (“blanqueo” o “decoloración”) de composiciones que contienen clorofila o que están contaminadas con clorofila, 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. El “blanqueo enzimático” de las composiciones y métodos de la invención comprende el uso de una enzima modificadora de la clorofila, por ejemplo un polipéptido que tiene actividad de clorofilasa, incluyendo clusas y clorofila clorofilido-hidrolasas, y péptidos relacionados, o cualquier
45 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. 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.

50 En los métodos y procedimientos de la invención, se añade en cualquier momento y en cualquier lugar en el método o procedimiento, por ejemplo como se discute aquí, una clorofilasa, que puede ser una nueva clorofilasa de la invención, o una enzima conocida, incluyendo clusas y clorofila clorofilido-hidrolasas y péptidos relacionados, o una combinación de los mismos, o cualquier enzima catabólica de la clorofila. Por ejemplo, la clorofilasa (que puede ser una nueva clorofilasa de la invención, 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
55 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, se describe un método o procedimiento en el que la clorofilasa (de la invención, o conocida) y/o cualquier enzima catabólica de la clorofila se pueden añadir en cualquier combinación de estas etapas, o en todas estas etapas.

5 La invención 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.

10 La invención 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 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.

15 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 los componentes de la composición se eliminan con cantidades significativamente menores de arcilla de blanqueo y/o al menos algún otro adsorbente (por ejemplo, una sílice). 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 como se describen aquí pueden reducir la pérdida de rendimiento por arrastre y división de grasas atribuida a la catálisis por condiciones de arcilla/blanqueo.

20 En un método (reacción) ilustrado ejemplar de la invención, una clorofilasa cataliza la hidrólisis de clorofila para generar clorofilida, que 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. Se describen composiciones y métodos en los que un procedimiento de separación acuoso puede eliminar parcial o completamente la necesidad de agentes adsorbentes. Sin embargo, 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. La sílice puede comprender una sílice TriSylo una sílice SORBSIL R™.

25 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 tiene una actividad de clorofilasa o feofitinas (que puede ser una nueva clorofilasa de la invención, 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.

30 En un aspecto, las composiciones y métodos de la invención se ponen en práctica como o con procedimientos industriales, por ejemplo blanqueo o neutralización cáustica o desgomado del aceite. El uso de las composiciones y métodos como se describen aquí facilita la reducción de la cantidad de o la 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 y que alteran el aspecto y/o estabilidad del aceite acabado. De este modo, en la práctica de la invención, al eliminar parcial o completamente la necesidad de adsorbentes, se pueden 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, y 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 vertido de adsorbentes de blanqueo.

35 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 invención, o una enzima conocida, o una combinación de las mismas) se pueden emplear en cualquier punto en un procedimiento de desgomado (por ejemplo desgomado enzimático). Por ejemplo, los péptidos 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.

5 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 invención, al menos una etapa se lleva a cabo en un extracto celular. Como alternativa a 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.

10 Se describen aquí 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 invención, 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.

15 Se describen aquí 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 invención, 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.

20 La invención proporciona procedimientos de desgomado que comprenden una etapa para el blanqueo 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 invención, 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.

25 Existe un segundo éster en las clorofilas y feofitinas – un éster metílico. Los métodos 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.

30 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.

35 Se describen aquí además métodos y procedimientos que 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 invención, o una enzima conocida, o una combinación de las mismas). En aspectos alternativos, los métodos de la invención 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.

40 En los métodos y procedimientos, el polipéptido es una esterasa (por ejemplo, una enzima de la invención), 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 pueden 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 los métodos, el polipéptido comprende además un liposoma, un hidrogel o un gel.

45 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.

50 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. Los métodos, las composiciones que contienen clorofila o contaminadas con clorofila comprenden un tejido o tela. 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 Los métodos descritos aquí 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 invención), 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. El producto de fabricación, el aparato de refinado de aceite vegetal comprenden 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 Se describen aquí tratamientos enzimáticos que comprenden detergentes 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 invención), 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 Se describen aquí 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 invención), 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 invención proporciona el uso de un polipéptido codificado por un ácido nucleico aislado, sintético o recombinante que comprende una secuencia de ácido nucleico que tiene al menos alrededor de 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 invención, por ejemplo SEC ID NO:9.

35 Las identidades de secuencia se determinan mediante análisis con un algoritmo de comparación de secuencias o mediante inspección visual. 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 descritos aquí también incluyen ácidos nucleicos aislados, sintéticos o recombinantes que codifican un polipéptido de la invención, 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 El polipéptido usado en la presente invención tiene 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.

50 El ácido nucleico aislado, sintético o recombinante de la invención 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.

55 Un ácido nucleico aislado, sintético o recombinante de la invención 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. 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.

5

Se describen aquí 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 invención, 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 invención codifica un polipéptido que tiene una actividad de enzima esterasa, incluyendo una actividad de clorofilasa (una clasa), 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. La actividad de esterasa descrita aquí 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, o toda la longitud del gen o transcrito. 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.

10

15

Se describe aquí 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 clasa)), 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 invención, o sus fragmentos o subsecuencias, en el que la sonda identifica el ácido nucleico mediante unión o hibridación.

20

Se describe aquí 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 clasa)), 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 invención, en el que las identidades de secuencia se determinan mediante análisis con un algoritmo de comparación de secuencias o mediante inspección visual.

25

30

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 invención, o una subsecuencia de la misma.

35

Se describen aquí ácidos nucleicos que codifican proteínas (por ejemplo, enzimas), incluyendo los polipéptidos de la invención, generadas mediante amplificación, por ejemplo reacción en cadena de la polimerasa (PCR), usando un par de cebadores de amplificación de la invención. La invención 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 invención. La invención 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 invención. El par de cebadores de amplificación amplifica un ácido nucleico de una librería, por ejemplo una librería génica, tal como una librería medioambiental.

40

45

Se describen aquí casetes de expresión que comprenden un ácido nucleico de la invención o una subsecuencia del mismo. 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.

50

La descripción proporciona vehículos de clonación que comprenden un casete de expresión (por ejemplo, un vector) de la invención o un ácido nucleico de la invención.

La descripción proporciona célula transformada que comprende un ácido nucleico de la invención o un casete de expresión (por ejemplo, un vector), o un vehículo de clonación. 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. Se describe aquí 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 invención. Se describen aquí métodos para inhibir la traducción de un mensaje enzimático (de una enzima de la invención) en una célula, que comprenden administrar a la célula o expresar a la célula un oligonucleótido antisentido que comprende una secuencia de ácido nucleico complementaria a o capaz de hibridarse en condiciones restrictivas a un ácido

55

nucleico de la invención. 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.

Los métodos para inhibir la traducción de un mensaje enzimático en una célula 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 invención. La invención proporciona moléculas de ARN inhibidor (ARNi) bicatenario que comprenden una subsecuencia de una secuencia de la invención. El ARNi tiene una longitud de alrededor de 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25 o más nucleótidos dúplex. La invención proporciona métodos para inhibir la expresión de un polipéptido (por ejemplo, una enzima de la invención) 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 invención.

Se describe aquí 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 identidad, o una identidad de secuencia completa (100%) con un polipéptido o péptido ejemplar de la invención, tal como 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% idéntica a SEC ID NO: 10 que se usa en la invención, 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. 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 descritas aquí 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, 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 invención incluyen secuencia codificada por un ácido nucleico de la invención. Las secuencias polipeptídicas o peptídicas ejemplares de la invención incluyen polipéptidos o péptidos especialmente unidos mediante un anticuerpo de la invención. 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.

Un polipéptido usado en la invención 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. La actividad de esterasa comprende una actividad de clorofila clorofilido-hidrolasa.

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

La actividad enzimática de un polipéptido descrito aquí, por ejemplo el polipéptido usado en la presente invención, puede ser termoestable. El polipéptido descrito aquí, por ejemplo el polipéptido usado en la invención, 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. La actividad enzimática de un polipéptido usado en la invención puede ser 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. 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.

El polipéptido aislado, sintético o recombinante puede comprender el polipéptido descrito aquí que carece de una secuencia señal. El polipéptido aislado, sintético o recombinante puede comprender el polipéptido descrito aquí que comprende una secuencia señal heteróloga.

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

5 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. Como alternativa, un polipéptido descrito aquí
 10 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. Un polipéptido de la presente 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. Como alternativa, un polipéptido descrito aquí 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 invención proporciona polipéptidos inmovilizados de la invención, un polipéptido codificado por un ácido nucleico de la invención, o un polipéptido que comprende un polipéptido de la invención 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.

15 La presente 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 invención; (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.

20 La presente descripción proporciona métodos para producir un polipéptido recombinante, que comprenden las etapas de: (a) proporcionar un ácido nucleico de la invención 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. 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
 25 polipéptido recombinante en una célula transformada.

La presente 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 descrito aquí; o un polipéptido codificado por un ácido nucleico descrito aquí; (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).

35 La presente 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 descrito aquí; o un polipéptido codificado por un ácido nucleico descrito aquí; (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
 40 incremento en la cantidad de un producto de reacción identifica un sustrato de un polipéptido implicado en el catabolismo de la clorofila o que tiene una actividad de esterasa (por ejemplo, clorofilasa).

La presente 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 invención (por ejemplo, un polipéptido codificado por un ácido nucleico descrito aquí). 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. Como alternativa, el algoritmo de comparación de secuencias comprende un programa de ordenador que indica polimorfismos. El sistema de ordenador puede comprender además un identificador que
 45 identifica uno o más características en dicha secuencia. La presente descripción proporciona medios legibles por ordenador que tienen almacenados en ellos una secuencia polipeptídica o una secuencia de ácido nucleico descrita aquí. La presente 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 invención; 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 invención; 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. El método puede
 50
 55
 60

comprender además un identificador que identifica una o más características en una secuencia. El método puede comprender leer la primera secuencia usando un programa de ordenador, e identificar una o más características en la secuencia.

5 La presente 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 descrito aquí; (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. 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 descrita aquí. El par de cebadores de amplificación es un par de amplificación descrito aquí.

20 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 descrito aquí, 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. 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.

35 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 invención; 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. 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 de ligación sintética (SLR) o una combinación de los mismos. Como alternativa, 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.

50 El método se puede repetir iterativamente hasta que se produce un polipéptido impliado 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. El polipéptido variante es termotolerante, y retiene cierta actividad después de exponerlo a una temperatura elevada. Como alternativa, 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. 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. 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 aumentar la termotolerancia o termoestabilidad de un polipéptido descrito aquí, o un polipéptido codificado por un ácido nucleico descrito aquí, comprendiendo el método la glucosilación de

un polipéptido que comprende al menos treinta aminoácidos contiguos de un polipéptido de la presente descripción; o un polipéptido codificado por una secuencia de ácido nucleico de la invención, incrementando de ese modo la termotolerancia o termoestabilidad del polipéptido. 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.

5 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 descrito aquí, o una secuencia de ácido nucleico descrita aquí, 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.

10 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 descrito aquí; (b) hacer crecer la planta en condiciones en las que la secuencia de ácido nucleico heteróloga se exprese en la célula vegetal. La invención 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 invención; (b) hacer crecer la planta en condiciones en las que la secuencia de ácido nucleico heteróloga se exprese en la célula vegetal. Como alternativa, una enzima usada en la invención se puede preparar mediante la expresión de un polinucleótido descrito aquí en un organismo seleccionado tal como una bacteria, una levadura, una planta, un insecto, un hongo y un animal. Los organismos ejemplares para expresar polipéptidos de la invención pueden ser *S. pombe*, *S. cerevisiae*, *Pichia* sp., por ejemplo *Pichia pastoris*, *E. coli*, *Streptomyces* sp., *Bacillus* sp. y *Lactobacillus* sp.

15 También se describe aquí un método para obtener un polipéptido descrito aquí. 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. Como alternativa, se describe un método para obtener un polipéptido o péptido. 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.

20 También se describe un método para generar una variante, que incluye obtener un ácido nucleico que tiene una secuencia descrita aquí, secuencias esencialmente idénticas a ella, secuencias complementarias a una secuencia de la invención, 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.

25 La descripción proporciona sistemas biosintéticos para el catabolismo de la clorofila, que comprenden al menos una enzima descrita aquí. 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 descrita aquí. El sistema comprende una pluralidad de ácidos nucleicos que codifican enzimas, en el que las enzimas están implicadas en el catabolismo de la clorofila. La pluralidad de ácidos nucleicos que codifican enzimas comprende todas las enzimas en una ruta de catabolismo de la clorofila. La pluralidad de ácidos nucleicos que codifican enzimas está contenida en al menos un plásmido, casete de expresión o vector de expresión.

30 El sistema biosintético de la invención 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*.

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

50 DESCRIPCIÓN DE LOS DIBUJOS

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

55 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 invención, 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 invención, como se describe con detalle más abajo.

- La Figura 5 es un diagrama de flujo que ilustra un procedimiento de la invención 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.
- 5 La Figura 6 es un diagrama de flujo que ilustra 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 procedimiento identificador 300 para detectar la presencia de una característica en una secuencia, como se describe con detalle más abajo.
- 10 La Figura 8 ilustra la reacción de una esterasa ejemplar de la invención 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 invención, como se describe con detalle más abajo.
- La Figura 10 ilustra una reacción de decoloración (blanqueo) enzimática ejemplar de la invención, como se describe con detalle más abajo.
- 15 La Figura 11 ilustra un procedimiento de decoloración (blanqueo) enzimática ejemplar de la invención, 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 invención, como se describe con detalle más abajo.
- 20 La Figura 13 ilustra un esquema de refinado de oleaginosas ejemplar, que comprende la extracción, el refinado y la modificación de una oleaginosas usando una esterasa de la invención, como se describe con detalle más abajo.
- La Figura 14 ilustra un procedimiento industrial ejemplar de la invención – un procedimiento de biodesgomado -, que comprende usar al menos un polipéptido de la invención, como se describe con detalle más abajo.
- 25 La Figura 15 ilustra otro procedimiento industrial ejemplar de la invención, que comprende usar al menos un polipéptido de la invención, como se describe con detalle más abajo.
- La Figura 16 ilustra otro procedimiento industrial ejemplar de la invención, que comprende usar al menos un polipéptido de la invención que tiene actividad de enzima clorofilasa.
- 30 Símbolos de referencia similares en los diversos dibujos indican elementos similares.

DESCRIPCIÓN DETALLADA

- La presente invención 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
- 35 "blanqueo enzimático" o "decoloración") de clorofila usado en las composiciones y métodos de la invención 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.
- 40 Por ejemplo, en un aspecto, la invención 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 invención proporciona métodos de blanqueo enzimático que usan una enzima clorofilasa para la hidrólisis enzimática de una clorofila o cualquier anillo de
- 45 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
- 50 La descripción proporciona ácidos nucleicos aislados, recombinantes y sintéticos (por ejemplo, un ácido nucleico ejemplar de la invención, incluyendo SEC ID NO:1, SEC ID NO:3, SEC ID NO:5, SEC ID NO:7, SEC ID NO:9, SEC

5 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 invención, por ejemplo las secuencias de aminoácidos ejemplares como se exponen 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), tales como ácidos nucleicos que codifican un polipéptido de SEC ID NO:10 usado en la presente invención. También se describen casetes de expresión tales como vectores de expresión, que comprenden ácidos nucleicos de la invención, que incluyen polinucleótidos que codifican los polipéptidos de la invención. La descripción también muestra métodos para descubrir nuevas secuencias polipeptídicas usando los ácidos nucleicos de la invención. La descripción también incluye métodos para inhibir la expresión de genes, transcritos y polipéptidos usando los ácidos nucleicos de la invención. También se proporcionan métodos para modificar los ácidos nucleicos de la invención, por ejemplo, mediante reensamblaje de ligación sintética, sistema de evolución dirigida optimizada y/o mutagénesis de saturación.

15 Los ácidos nucleicos usados en la invención se pueden obtener, aislar y/o manipular, por ejemplo, mediante clonación y expresión de librerías de ADNc, amplificación de ADN mensajero o genómico mediante PCR, y similares.

Al poner en práctica los métodos de la invención, los genes homólogos se pueden modificar manipulando un ácido nucleico molde, como se describe aquí. La invención 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.

20 La descripción también proporciona un ácido nucleico aislado que comprende una de las secuencias de la invención, 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 invención. 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 ácidos nucleicos aislados pueden comprender ARN.

25 Los ácidos nucleicos aislados descritos aquí se pueden usar para preparar uno de los polipéptidos para uso según la invención, 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 invención.

30 El ácido nucleico aislado que codifica uno de los polipéptidos de la invención puede incluir, pero no está limitado a: sólo la secuencia codificante de una de un ácido nucleico de la invención 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 "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.

35 Alternativamente, las secuencias de ácido nucleico par uso según la invención 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 invención. Como se usa en la presente memoria, 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.

40 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 invención 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.

55 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 invención 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. 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 invención. 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., págs. 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 invención, ya sea mediante síntesis directa o mediante síntesis de una serie de fragmentos que se pueden acoplar usando otras técnicas conocidas.

El término "variante" hace referencia a polinucleótidos o polipéptidos de la invención 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 invención. 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 sitio 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.

Técnicas generales

La presente invención 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 invención (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 invención, 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 invención) 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. nº 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 Laborator, (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. Theor and Nucleic Acid Preparation, Tijssen, ed. Elsevier, N.Y. (1993).

Otro medio útil para obtener y manipular ácidos nucleicos usados para poner en práctica los métodos de la invención 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 invención incluyen bibliotecas genómicas o de ADNc contenidas, por ejemplo, en cromosomas artificiales de mamífero (MACs), véanse, por ejemplo, las patentes U.S. n^{os} 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 invención 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 invención (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.

Un ácido nucleico que codifica un polipéptido de la invención para uso según la invención se ensambla en una fase apropiada con una secuencia líder capaz de dirigir secreción del polipéptido traducido o un fragmento del mismo.

Secuencias de control transcripcionales y traduccionales

La descripción proporciona secuencias de ácido nucleico (por ejemplo, ADN) de la invención 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.

Vectores de expresión y vehículos de clonación

Se proporcionan vectores de expresión y vehículos de clonación que comprenden ácidos nucleicos para uso según la invención, por ejemplo secuencias que codifican las enzimas de la invención. Los vectores de expresión y los vehículos de clonación de la invención 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 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. 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.

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 ácidos nucleicos para uso según la invención 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 ilustrativo 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. 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.

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, 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*.

Además, los 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*.

El ácido nucleico que codifica uno de los polipéptidos usados en la invención, 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 invención, 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 por 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.

Una variedad de vectores de clonación y de expresión para uso con hospedantes procariontes y eucariotas son descritos por Sambrook, et al., Molecular Cloning: A Laboratory y Manual, 2a Ed., Cold Spring Harbor, N.Y., (1989).

Células hospedantes y célula transformadas

La descripción también proporciona una célula transformada que comprende una secuencia de ácido nucleico de la invención, por ejemplo una secuencia que codifica una enzima de la invención, o un vector de la invención. La célula hospedante puede ser cualquiera de las células hospedantes familiares para los expertos en la técnica, incluyendo células procariontes, células eucariotas, tales como células bacterianas, célula fúngicas, células de levadura, células de mamífero, células de insecto, o células vegetales. Las células bacterianas ilustrativas 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 ilustrativas incluyen S2 de *Drosophila* y Sf9 de *Spodoptera*. Las células de levadura ilustrativas incluyen *Pichia pastoris*, *Saccharomyces cerevisiae* o *Schizosaccharomyces pombe*. Las células animales ilustrativas 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 concretos 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)).

5 Los ácidos nucleicos o vectores como se describen aquí 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 ilustrativos 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 retrovímica) 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.

15 Cuando sea apropiado, las células hospedantes manipuladas mediante ingeniería se pueden cultivar en medios nutrientes convencionales según sea apropiado para activar promotores, seleccionar transformantes o amplificar los genes de la invención. 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.

20 Las células hospedantes que contienen los polinucleótidos de interés, por ejemplo ácidos nucleicos de la invención, 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 identifica 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.

25 La descripción proporciona un método para sobreexpresar una enzima recombinante en una célula, que comprende expresar un vector que comprende un ácido nucleico para uso según la invención, 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%, 97%, 98%, 99%, o más de identidad de secuencia con una secuencia ejemplar de la invención, tal como SEC ID NO: 9, 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 ácido nucleico de la invención. 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. La sobreexpresión 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. 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 para uso según la invención.

45 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*, célula fúngicas, tales como levadura, células de insecto tales como S2 de *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.

50 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 concretos 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)).

55 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 invención también pueden incluir o no un resto de aminoácido de metionina inicial.

Amplificación de Ácidos Nucleicos

Al poner en practica la invención, los ácidos nucleicos para uso según la invención y los ácidos nucleicos que codifican las enzimas para uso según la invención, o los ácidos nucleicos modificados para uso según la invención, se pueden reproducir mediante amplificación. La amplificación también se puede usar para clonar o modificar los ácidos nucleicos de la invención. De este modo, la invención proporciona pares de secuencias de cebadores de amplificación para amplificar ácidos nucleicos de la invención. 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.

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, Kwoh (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, 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 invención (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. Según la invención, se usa un ácido nucleico comprende una secuencia que tiene al menos 90%, 91 %, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% de identidad o más, o identidad de secuencia completa (100%) con un ácido nucleico de SEC ID NO: 9. 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 descrito aquí. En la invención se usa un polipéptido que tiene al menos 90%, 91 %, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% de identidad de secuencia con SEC ID NO: 10. 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 en uso según la invención 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 invención, y secuencias esencialmente idénticas a éstas. Las secuencias homólogas y los fragmentos de las secuencias de ácido nucleico de la invención 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 invención se pueden representar en el formato de un solo carácter tradicional (Véase la contraportada de Stryer, Lubert. Biochemistry, 3a 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 esta parte 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

5 "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.

10 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

15 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

20 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

25 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"), BLIMSP (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 Multiseuencia 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

40 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.

55

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 (HSP) 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 HSP más largos que las

60

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, por ejemplo menor que alrededor de 0,01, y por ejemplo menor que alrededor de 0,001.

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:

- (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;
- (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.

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 se obtiene preferiblemente a partir de una base de datos de secuencias de proteínas o de ácidos nucleicos. 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. 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). 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.

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

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. Como alternativa, 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. Las secuencias son sustancialmente idénticas a lo largo de toda la longitud de las regiones codificantes.

Sistemas informáticos y productos de programas informáticos

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

5 En consecuencia, la presente 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 invención. 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 pueda 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 invención.

10 Los polipéptidos de la descripción incluyen las secuencias polipeptídicas descritas aquí, por ejemplo las secuencias ejemplares descritas, 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 descrita aquí.

15 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 invención. Se apreciará que los códigos del polipéptido como se muestra en las secuencias de aminoácidos de la invención 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.

20 Una secuencia de ácido nucleico o polipeptídica como se describe aquí 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 invención, una o más de las secuencias polipeptídicas de la invención. Se describe aquí un medio legible por ordenador que tiene registradas en él al menos 2, 5, 10, 15, o 20 o más secuencias de ácido nucleico.

25 Se describe aquí un medio legible por ordenador que tiene registradas en él una o más de las secuencias de ácidos nucleicos de la invención. Se describe además un medio legible por ordenador que tiene registradas en él una o más de las secuencias polipeptídicas de la invención. También se describe 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.

30 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.

35 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.

40 Se describen además 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 invención, o una secuencia polipeptídica de la invención. 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.

El sistema informático 100 incluye un procesador 105 conectado a un bus que está conectado a una memoria principal 115 (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. 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. 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 registros de 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 invención, o una secuencia polipeptídica de la invención, (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.

El sistema informático 100 puede comprender adicionalmente un algoritmo de comparación de secuencias para comparar una secuencia de ácido nucleico de la invención, o una secuencia polipeptídica de la invención, 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 invención, o una secuencia polipeptídica de la invención, 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, se describe aquí un sistema informático que comprende un procesador, un dispositivo de almacenamiento de datos que tiene almacenado en él una secuencia de ácido nucleico descrita aquí, o una secuencia polipeptídica descrita aquí, un dispositivo de almacenamiento de datos recuperables que tiene almacenado en él secuencias nucleotídicas o secuencias polipeptídicas descritas aquí, 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 aquí, o una secuencia polipeptídica descrita aquí, o puede identificar motivos estructurales en secuencias que se comparan con estos códigos de ácido nucleico y códigos polipeptídicos. 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 invención, o las secuencias polipeptídicas de la invención.

Se describe además un método para determinar el nivel de homología entre una secuencia de ácido nucleico de la invención, o una secuencia polipeptídica de la invención, 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 invención, o las secuencias polipeptídicas descritas aquí, a través del uso del programa informático, y determinando la homología entre el código de ácido nucleico o el código polipeptídico y las secuencias nucleotídicas o las secuencias polipeptídicas de referencia.

La Figura 6 es un diagrama de flujo que ilustra 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 puede ser 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 invención, con una o más secuencias nucleotídicas de referencia con el fin de determinar si el código de ácido nucleico de la invención 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 invención. El programa informático puede ser un programa que determina si

una secuencia de ácido nucleico de la invención contiene un polimorfismo de un solo nucleótido (SNP) con respecto a una secuencia nucleotídica de referencia.

En consecuencia, se describe un método para determinar si una secuencia de ácido nucleico descrita aquí 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. 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 invención 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.

El sistema a base de ordenador puede comprender además un identificador para identificar características dentro de una secuencia de ácido nucleico descrita aquí o una secuencia polipeptídica descrita aquí.

Un "identificador" se refiere a uno o más programas que identifica ciertas características dentro de una secuencia de ácido nucleico de la invención, o de una secuencia polipeptídica de la invención. El identificador puede comprender un programa que identifica un marco de lectura abierto en una secuencia de ácido nucleico descrita aquí.

La Figura 7 es un diagrama de flujo que ilustra 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 enzimáticos, motivos de hélice-vuelta-hélice, u otros motivos conocidos por los expertos en la técnica.

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.

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.

En consecuencia, se describe además un método para identificar una característica dentro de una secuencia de ácido nucleico de la invención, o una secuencia polipeptídica de la invención, que comprende leer el código o códigos del ácido o ácidos nucleicos o el código o códigos del polipéptido o polipéptidos mediante el uso de un programa informático que identifica características en ellos, e identificar características en el código o códigos del ácido o ácidos nucleicos con el programa informático. 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 invención, o las secuencias polipeptídicas de la invención, 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.

Una secuencia de ácido nucleico descrita aquí, o una secuencia polipeptídica descrita aquí, 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 invención, o una secuencia polipeptídica de la invención, 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 invención, o una

secuencia polipeptídica de la invención. Se pretende que la siguiente lista no limite la invención sino que proporcione apoyo a los programas y bases de datos que son útiles con las secuencias de ácido nucleico descritas aquí, o las secuencias polipeptídicas descritas aquí.

5 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),
10 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.),
15 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 Genseqn. Muchos otros programas y bases de
datos serán manifiestos para un experto en la técnica dada la presente descripción.

20 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

25 La presente descripción proporciona ácidos nucleicos aislados, sintéticos o recombinantes que se hibridan en
condiciones restrictivas a una secuencia ejemplar descrita aquí (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
30 condiciones restrictivas bajas, incluyendo las condiciones de elevada restricción y de restricción reducida descritas
aquí. 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 invención, como se explica más abajo.

35 Los ácidos nucleicos de la invención 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 invención, 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.

40 Los ácidos nucleicos 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. Como alternativa, 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.

45 Como alternativa, los ácidos nucleicos 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). Los ácidos nucleicos de la presente 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.

50 . Véase Sambrook, Tijssen y Ausubel para una descripción de un tampón SSC y condiciones equivalentes.

55 Estos métodos se pueden usar para aislar ácidos nucleicos de la invención. 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 invención 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 invención.

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 invención, 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. La sonda comprende al menos 10 bases consecutivas de un ácido nucleico de la invención. 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 invención. Las sondas identifican un ácido nucleico mediante unión y/o hibridación.

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* 2a Ed., Cold Spring Harbor Laboratory Press (1989).

Inhibición de la expresión de enzimas

La descripción proporciona ácidos nucleicos complementarios a (por ejemplo, secuencias antisentido a) los ácidos nucleicos descritos aquí, 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.

Oligonucleótidos antisentido

Se describen aquí 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 invención. 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.

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 invención 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 invención (véase, por ejemplo, Gold (1995) *J. of Biol. Chem.* 270:13581-13584).

Modificación de ácidos nucleicos

La presente descripción proporciona métodos para generar variantes de los ácidos nucleicos para uso según la invención, 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. 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 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. nº 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. nº 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. nºs 6.291.242; 6.287.862; 6.287.861; 5.955.358; 5.830.721; 5.824.514; 5.811.238; 5.605.793.

Las siguientes publicaciones describen una variedad de procedimientos y/o métodos de recombinación recursivos que se pueden incorporar en los métodos de la invención: 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; Crameri (1998) "DNA shuffling of a family of genes from diverse species accelerates directed evolution" *Nature* 391:288-291; Crameri (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; Crameri 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; Crameri 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 from 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 y Smith (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 y Smith (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 et al. (1985) "The use of phosphorothioate-modified DNA in restriction enzyme reactions to prepare nicked DNA" *Nucl. Acids Res.* 13: 8749-

8764; Taylor et al. (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 et al. (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 con saltos (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 et al. (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 et al. (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 invención 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. nº 5.605.793 de Stemmer (25 de febrero de 1997), "Methods for In Vitro Recombination"; Patente U.S. nº 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. nº 5.830.721 de Stemmer et al. (3 de noviembre de 1998), "DNA Mutagenesis by Random Fragmentation and Reassembly"; Patente U.S. nº 5.834.252 de Stemmer, et al. (10 de noviembre de 1998) "End-Complementary Polymerase Reaction"; Patente U.S. nº 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 Cramerí, "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 Cramerí "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 Cramerí, "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 Zarling, "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. nº (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 Crameri 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. No. 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. No. 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 de ligación sintética (SLR), o una combinación de los mismos, que se usan para modificar los ácidos nucleicos de la invención 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^{os} 6.361.974; 6.280.926; 5.939.250.

Mutagénesis de saturación de sitio génico (GSSM)

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 invención, 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. 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.

También se describe 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 (GSSMTM)). 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.

Reensamblaje de ligación sintética (SLR)

La descripción proporciona un sistema de modificación génica no estocástico denominado "reensamblaje de ligación sintética", o simplemente "SLR", un "proceso de evolución dirigida", para generar polipéptidos, por ejemplo enzimas de la invención, 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 de construcción 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 n^o (USSN) 09/332.835 titulada "Synthetic Ligation Reassembly in Directed Evolution" y presentada el 14 de junio de 1999 ("USSN 09/332.835"). 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.

Sistema de evolución dirigida optimizada

5 La invención proporciona un sistema de modificación génica no estocástico denominado “sistema de evolución dirigida optimizada” para generar polipéptidos, por ejemplo enzimas o anticuerpos de la invención, con propiedades nuevas o alteradas. La evolución dirigida optimizada se refiere al uso de ciclos repetidos de reclasificación 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.

Determinación de los sucesos de cruzamiento

15 Se describe aquí 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. El procesamiento descrito aquí se lleva a cabo preferiblemente en MATLAB® (The Mathworks, Natick, Massachusetts), un lenguaje de programación y entorno de desarrollo para computación técnica.

Procesos iterativos

20 En la práctica de la invención, 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 invención), 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.

25 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.

Barajado *in vivo*

35 El barajado *in vivo* de las moléculas se usa en métodos de la invención que proporcionan variantes de polipéptidos para uso según la invención, 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 homología, 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.

Producción de variantes de secuencias

45 La descripción también proporciona métodos adicionales para obtener variantes de secuencias de las secuencias de ácido nucleico en uso según la invención. La descripción también proporciona métodos adicionales para aislar polipéptidos. La invención proporciona variantes de secuencias génicas (por ejemplo, un gen, ADNc o mensaje), 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.

50 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, MgCl₂, MnCl₂, 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 MgCl₂, 0,5 mM de MnCl₂, 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 MgCl₂, 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. Por ejemplo, se pueden incluir oligonucleótidos en las reacciones de PCR. Como alternativa, 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*. Por ejemplo, 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.

5 Por ejemplo, se crean variantes usando mutagénesis de conjunto exponencial. La mutagénesis de conjunto exponencial es un proceso para generar librerías 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 Res.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.

10 Por ejemplo, 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 químicas 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".

15 Las variantes de los polipéptidos en uso según la invención pueden ser variantes en las que uno o más de los restos de aminoácidos de los polipéptidos de las secuencias de la invención 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.

20 La descripción proporciona realizaciones alternativas de los polipéptidos de la invención (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.

25 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.

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

35 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.

40 Los fragmentos, derivados y análogos retienen la misma función o actividad biológica que los polipéptidos que se usan en la invención. 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

45 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. 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. 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 invención incluyen bacterias, levaduras, hongos, células vegetales, células de insecto y células de mamífero. De este modo, la invención 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 gasserii*, *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*, *Saccharomyces 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 invención 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 invención 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 invención), 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 invención. La descripción también proporciona métodos para obtener y usar estos animales no humanos transgénicos.

. 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; véase Baguisi (1999) Nat. Biotechnol. 17:456-461, que demuestra la producción de cabras transgénicas. "Animales a los que se les ha suprimido un gen" también se puede usar para la práctica de los métodos de la invención.

Plantas y semillas transgénicas

La invención 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 descrita aquí. La descripción también proporciona productos vegetales, por ejemplo aceites, semillas, hojas, extractos y similares, que comprenden un ácido nucleico y/o un polipéptido descrito aquí. 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 descrito aquí se puede construir según cualquier método conocido en la técnica. Véase, por ejemplo, la patente U.S. nº 6.309.872.

La descripción también proporciona "plantas a las que se les ha suprimido un gen", 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 "a las que se les ha suprimido un gen" 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 invención 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 invención 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 invención, 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 invención 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.

La obtención de plantas o semillas transgénicas comprende incorporar secuencias de la invención 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.

La descripción proporciona la transformación de plantas monocotiledóneas usando los ácidos nucleicos de la invención, 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.

La obtención de plantas o semillas transgénicas comprende incorporar secuencias de la invención 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.

La descripción proporciona la transformación de plantas monocotiledóneas usando los ácidos nucleicos de la invención, 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.

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., *Protoplasts 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 medioambientales 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 invención conduce a cambios fenotípicos, las plantas que comprenden los ácidos nucleicos recombinantes de la invención se pueden cruzar sexualmente con una segunda planta para obtener un producto final. De este modo, la semilla descrita aquí se puede derivar de un cruce entre dos plantas transgénicas de la invención, o un cruce entre una planta de la invención y otra planta. Los efectos deseados (por ejemplo, expresión de los polipéptidos de la invención para producir una planta en la que está alterado el comportamiento de la floración) se pueden potenciar cuando ambas plantas parentales expresan los polipéptidos de la invención. Los efectos deseados se pueden pasar a las futuras generaciones de plantas por medios de propagación estándar.

Los ácidos nucleicos y polipéptidos descritos aquí se expresan o insertan en cualquier planta o semilla. Las plantas transgénicas de la invención pueden ser dicotiledóneas o monocotiledóneas. Los ejemplos de plantas transgénicas monocotiledóneas de la invención son hierbas tales como poa común (hierva 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 invención son tabaco, legumbres, tales como lupinos, palata, 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 invención 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*.

Como alternativa, los ácidos nucleicos se expresan en plantas que contienen células de fibras, incluyendo, por ejemplo, algodón, árbol de algodón de seda (*Kapok*, *Ceiba pentandra*), *Chilopsis linearis*, árbol de creosota, *Krascheninnikovia lanata*, balsa, *Boehmeria nivea*, *Hibiscus cannabinus*, *Cannabis sativa*, *Hibiscus sabdariffa*, yute, sisal, abacá y lino. En realizaciones alternativas, las plantas transgénicas de la invención 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 invención. 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 invención 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

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 invención, 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.

Un polipéptido que se usa en la invención 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. Por ejemplo, la actividad de esterasa comprende una actividad de clorofila clorofilido-hidrolasa.

Como alternativa 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 descrita aquí, 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 descritas aquí o sus subsecuencias, por ejemplo sitios activos, o dominios catalíticos (CDs). Una proteína quimérica descrita aquí (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) descrito aquí y otro polipéptido (por ejemplo, péptido de sitio activo o de dominio catalítico) descrito aquí u otro polipéptido. Por ejemplo, una proteína quimérica descrita aquí 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. La proteína quimérica descrita aquí comprende una fusión de dominios, por ejemplo un dominio único puede mostrar una o cualquier combinación de actividades.

Los polipéptidos descritos aquí incluyen enzimas en una forma activa o inactiva. Por ejemplo, los polipéptidos descritos aquí 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 descritos aquí incluyen enzimas inactivas por otras razones, por ejemplo, antes de la "activación" mediante un evento de procesamiento post-traduccion, por ejemplo una acción de endo- o exo-peptidasa o proteinasa, un suceso de fosforilación, una amidación, una glucosilación o una sulfuración, un suceso de dimerización, y similares. Los polipéptidos descritos aquí 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, p. ej., 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 invención 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 descritas aquí.

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 invención también pueden ser más cortos que la longitud completa de los péptidos 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 invención; 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 ilustrativa descrita aquí.

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

Los polipéptidos y péptidos en uso según la invención 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 en uso según la invención se pueden obtener y aislar usando cualquier método conocido en la técnica. Los polipéptidos y péptidos que se usan aquí según la invención 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 en uso según la invención también se pueden glucosilar. La glucosilación se puede añadir post-traduccionalmente ya sea químicamente o mediante mecanismos biosintéticos celulares, en los que estos últimos incorporan el uso de motivos de glucosilación 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 glucosilación puede estar enlazada vía O o enlazada vía N.

Los péptidos y polipéptidos 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. El mimético puede estar compuesto completamente de análogos sintéticos no naturales de aminoácidos, o es una molécula química 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 invención que son variantes conservativas, la experimentación habitual determinará si un mimético está dentro del alcance de la invención, es decir, que su estructura y/o función no esté sustancialmente alterada a partir de un polipéptido ejemplar de la invención. Por ejemplo, una composición mimética se usa en una composición, sistema celular o procedimiento de la invención (por ejemplo, una célula hospedante que tiene un plásmido que expresa al menos una enzima de la invención).

La descripción también proporciona métodos para modificar los polipéptidos de la invención mediante procesos naturales, tales como procesamiento post-traduccional (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. 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 invención. 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 invención, 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 con y sin señal. El polipéptido que comprende una secuencia señal puede ser un polipéptido descrito aquí u otro polipéptido.

La descripción incluye polipéptidos inmovilizados en un uso según la invención, 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 invención. La descripción incluye heterocomplejos, por ejemplo proteínas de fusión, heterodímeros, etc., que comprenden las enzimas de la invención.

Los polipéptidos en uso según la invención pueden tener actividad enzimática en diversas condiciones, por ejemplo extremas en pH y/o temperatura, agentes oxidantes, y similares. Se describen 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. Las variantes enzimáticas se pueden producir utilizando técnicas de mutagénesis dirigida al sitio y/o mutagénesis al azar. Se puede usar la evolución dirigida para producir una gran variedad de variantes enzimáticas con especificidades y estabilidad alternativa.

Los polipéptidos o fragmentos que tienen homología con uno de los polipéptidos de la invención, 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 invención, 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.

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

El ensayo para determinar si los fragmentos de variantes conservan la actividad enzimática de los polipéptidos de la invención 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 descritos aquí, 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, se proporciona un procedimiento para la utilización de los polipéptidos de la invención o polinucleótidos que codifican tales polipéptidos para hidrolizar enlaces glucosídicos. En tales procedimientos, una sustancia que contiene un enlace glucosídico (por ejemplo, una clorofila) se pone en contacto con uno de los polipéptidos de la invención en condiciones que faciliten la hidrólisis del enlace de éster.

Secuencias señal, dominios prepro 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 invención 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 invención). La presente 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 invención.

- 5 La descripción también proporciona polipéptidos quiméricos (y los ácidos nucleicos que los codifican) que comprenden al menos dos enzimas de la invención o sus subsecuencias, por ejemplo dominios catalíticos (CDs) o sitios activos. Por ejemplo, una proteína quimérica de la invención puede tener cualquier combinación de actividades. 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).
- 10 La descripción también proporciona secuencias señal aisladas, sintéticas o recombinantes que comprenden/que consisten en una secuencia señal descrita aquí, 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 descrita aquí, u otro tipo de enzima o polipéptido. Por ejemplo, para ayudar a leer la Tabla 1, la invención proporciona una secuencia señal aislada, sintética o recombinante como se expone mediante los restos de aminoácidos amino terminales 1 a 21 ("NH₂-MSRVCLPLTLTLALSARA") 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, 11, 12, 13, 14, 15, 16	AA1-20	MSRVCLPLTLTLALSARA
17, 18, 19, 20, 3, 4	AA1-25	MKKYKTGLVLSGGGTRGFAHLGVIA
5, 6	AA1-25	MRRIVFLYILALLCVSCANRNPSVS
7, 8	AA1-51	MTRKKIGLALSOGGAARGFAHLGVLFVFAEHGIPVDFVAGTSAGSFAGAAFA
9, 10	AA1-23	MFNKALPAAAAVAGLFLSTSAMA

- 20 Las secuencias señal (SPs) y/o secuencias prepro descritas aquí pueden ser péptidos aislados, o secuencias unidas a otra enzima de la invención, o una heteróloga, por ejemplo como una proteína de fusión (quimérica). Se proporcionan además polipéptidos que comprenden secuencias señal descritas aquí. Los polipéptidos que comprenden secuencias señal SPs y/o prepro descritas aquí comprenden secuencias heterólogas a una enzima descrita aquí (por ejemplo, una proteína de fusión que comprende una SP y/o prepro de la invención, y/o secuencias procedentes de otra proteína). La descripción proporciona una enzima descrita aquí con SPs heterólogas y/o
- 25 secuencias prepro, por ejemplo secuencias con una secuencia señal de levadura. Las enzimas de la invención pueden comprender una SP heteróloga y/o prepro en un vector, por ejemplo un vector de la serie pPIC (Invitrogen, Carlsbad, CA). Las SPs y/o secuencias prepro de la invención 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
- 30 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
- 35 100 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, 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)).
- 40

Se debería entender que una enzima puede no tener SPs y/o secuencias prepro, o uno o más "dominios". La descripción proporciona una enzima que carece de toda o parte de una SP y/o un dominio prepro. La descripción proporciona una secuencia de ácido nucleico que codifica una secuencia señal (SP) y/o prepro de una enzima de la

invención, enlazada operablemente a una secuencia de ácido nucleico de una enzima de la invención 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 invención y secuencias heterólogas. Las secuencias heterólogas son secuencias no asociadas naturalmente (por ejemplo, a enzimas de la invención) 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. 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 invención, con la condición de que no esté asociado con ninguna secuencia a la que está asociado de forma natural. De forma similar, la descripción proporciona ácidos nucleicos aislados o recombinantes que codifican estos polipéptidos. De este modo, el ácido nucleico aislado o recombinante descrito aquí comprende una secuencia codificante para una secuencia señal (SP), dominio prepro y/o dominio catalítico (CD) de la invención 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 invención). 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 librerías peptídicas

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

“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” incluyen 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-traduccional, o mediante técnicas de modificación química que son bien conocidas en la técnica (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)).

Las enzimas descritas aquí comprenden epítopos o etiquetas de purificación, secuencias señal u otras secuencias de fusión, etc. Las xilanasas se pueden fusionar a péptido aleatorio para formar un polipéptido de fusión.

Los microorganismos a partir de los cuales se pueden preparar el polinucleótido incluyen microorganismos procariontes, tales como *Eubacteria* y *Archaeobacteria*, y microorganismos eucariotes inferiores tales como hongos, algunas algas y protozoos. Los polinucleótidos se pueden aislar de muestras medioambientales, en cuyo caso el ácido nucleico se puede recuperar sin cultivar un microorganismo, o se puede recuperar de uno o más organismos cultivados. 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 reclasificación reductiva. 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 la célula hospedante puede ser una célula procarionte, 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 flanqueo 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.

Por lo tanto, la descripción se refiere a un método para producir un polipéptido híbrido biológicamente activo y cribar el 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.

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 invención.

Metodologías de cribado y dispositivos de monitorización "en línea"

Al poner en práctica los métodos de la invención, se puede usar una variedad de aparatos y metodologías junto con los polipéptidos y ácidos nucleicos que se usan en la invención, 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 descrito aquí, para determinar ácidos nucleicos que se hibridan a un ácido nucleico descrito aquí, para cribar células que expresan un polipéptido de la invención, y similares. Además de los formatos de matrices descritos con detalle más abajo para cribar muestras, también se pueden usar formatos alternativos para poner en práctica los métodos de la invención. 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 invención, véase, por ejemplo la solicitud de patente U.S. nº 20020001809.

Matrices capilares

Los ácidos nucleicos o polipéptidos que se usan en la invención se pueden inmovilizar en o aplicar a una matriz. Las matrices se pueden usar para cribar o monitorizar librerías 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 invención. 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. 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 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.

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 Microtiter® 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".

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 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.

Matrices o "biochips"

Los ácidos nucleicos o polipéptidos que se usan en la invención se pueden inmovilizar o aplicar a una matriz. Se pueden usar matrices para cribar o monitorizar librerías 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 invención. Por ejemplo, un parámetro monitorizado es la expresión transcrita de un gen, por ejemplo un gen de la invención (un ácido nucleico que codifica un polipéptido de la invención). 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 descritos aquí. También se pueden usar "matrices polipeptídicas" para cuantificar simultáneamente una pluralidad de proteínas. La presente invención 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.

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 invención. Estos anticuerpos se pueden usar para aislar, identificar o cuantificar un polipéptido que se usa en la invención o polipéptidos relacionados. Estos anticuerpos se pueden usar para aislar otros polipéptidos dentro del alcance de la invención u otros polipéptidos relacionados. Los anticuerpos se pueden diseñar para unirse a un sitio activo de un polipéptido de la invención.

Los anticuerpos se pueden usar en inmunoprecipitación, tinción, columnas de inmovilización, y similares. Si se desea, las secuencias de ácidos nucleicos que codifican antígenos específicos se pueden generar mediante inmunización seguida del aislamiento del polipéptido o ácido nucleico, amplificación o clonación e inmovilización del polipéptido sobre una matriz de la invención. Como alternativa, los métodos de la invención 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 invención.

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 (7ª ed.) Lange Medical Publications, Los Altos, CA ("Stites"); Goding, MONOCLONAL ANTIBODIES: PRINCIPLES AND PRACTICE (2ª 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 librerías 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.

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 invención, como se describe aquí. 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 descritos aquí 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 invención, por ejemplo una secuencia codificante para una enzima descrita aquí. 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". Una pluralidad de células, tal como un cultivo celular, se monitoriza en "tiempo real" o "en línea". Una pluralidad de parámetros metabólicos se monitoriza en "tiempo real" o "en línea". Los parámetros metabólicos se pueden monitorizar usando una enzima descrita aquí.

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 invención, 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. Los métodos descritos aquí, 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 descritos aquí, 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 descritos aquí, 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.

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 tal composición o método 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 pueden tener actividad de esterasa, por ejemplo una clorofilasa o una actividad similar. Los polipéptidos y/o péptidos pueden incluir anticuerpos catalíticos, enzimas, sitios activos, y similares. Estos polipéptidos y/o péptidos que tienen actividad de esterasa (por ejemplo, clorofilasa) se pueden usar en composiciones o métodos descritos aquí. Por ejemplo, en un aspecto, las composiciones y métodos 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 descrito aquí 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 descrito aquí 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 descrito aquí; véase, por ejemplo, Marchler-Bauer (2003) *Nucleic Acids Res.* 31: 383-387.

Las composiciones y métodos descritos aquí se pueden poner en práctica con enzimas como se describe en el documento WO 0229022. Por ejemplo, las composiciones y métodos pueden comprender la expresión recombinante de enzimas, por ejemplo clorofilasas, tales como polinucleótidos que codifican clorofilasas. El ácido nucleico recombinante se expresa en células completas, extractos celulares o *in vitro*. 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.

Las composiciones y métodos descritos aquí 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

AAAAAAGTAAAGAAAAGAAAACTAATAAAGAACAAAAAATGTCCTCTTCTTC
ATCAAGAAACGCCTTTGAAGATGGCAAATACAAATCAAATCTCTTAACCTTGGAC
TCATCATCTCGTTGCTGCAAAATAACACCGTCTTCTAGAGCTTCACCGTCTCCGC
CAAAGCAGCTGTTGGTGGCTACGCCGGTGGAGGAAGGAGATTATCCGGTGGTG
ATGCTCCTCCATGGTTACCTTCTCTACAACCTCCTTCTATTCTCAGCTTATGTTGC
ATGTCTCTTCTCATGGCTTCATCCTCATCGCTCCTCAGTTATATAGTATCGCCGG
ACCAGACACAATGGATGAGATTAATCAACGGCGGAGATTATGGATTGGTTATC
AGTAGGACTTAATCACTTTCTCCAGCGCAAGTAACACCAAACCTATCCAAATTT
GCCCTCTCCGGCCATAGCCGCGGTGGCAAACCGCGTTTGGCGTCCGCTTAAA
GAAATTTGGGTA CTCTCGAATCTAAAGATCTCGACATTGATCGGTATAGATCCA
GTCGATGGAACAGGGAAAGGGAAACAAACCCCTCCTCCGGTGTGGCTTACCT
TCCAAACTCATTTGACCTAGACAAAACGCCTATACTTGTGATCGGTTCCGGGGCT
TGGTGAACCGCTCGGAACCCATTATCCCACCGTGTGCACCTCCCGGAGTGA
ATCACCGAGAGTTCTTTCCGGGAATGTCAAGGTCCAGCATGGCATTTCGTTGCGA
AGGATTATGGGCATTTGGACATGCTTGTGATGATACAAAAGGGATTAGAGGGGA
AGAGTTCTTATTGTTTGTGTAAGAATGGTGAAGAGAGGAGACCAATGAGGAGAT
TCGTTGGTGGACTTGTGATCATTGTTTGAAGGCTTATTTGGAAGGAGATGATCG
TGAATTAGTTAAGATCAAAGATGGGTGTACGAGGATGTTCCCGTTGAAATTCAA
GAGTTTGAGGTTATCATGTAAACATAAGTTTTTTCTTTAGGGGCTGGTTTTTTCTATT
GTCAATATCATCAGCTTTTGTGCTTATGGTTTTACAACTTATATTGTACAACCTC
TTTAAGTCACCTCTTTGCTTATGATATTAACCCGATC

SEC ID NO:22

MSSSSSRNAFEDGKYKSNLLTDSSSRCKITPSSRASPPKQLLVATPVEEGDY
PVVMLLHGYYLNSFYSQLMLHVSSHGFILIPQLYSIAGPDTMDEIKSTAEIMDWLS
VGLNHFLPAQVTPNLSKFA LSGHSRGGKTAFVALKKFGYSSNLKISTLIGIDPVDGT
GKGKQTPPPVLAYLPNSFDLTKPILVIGSGLGETARNPLFPPCAPPGVNHREFFRE
CQGPWFHFAKDYGHLDMLDDDTKGIRGKSSYCLCKNGEERRPMRRFVGGGLVVS
FLKAYLEGDDRELVKIKDGGHEDVPVEIQEFVIM

5

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

SEC ID NO:23

TTGAAAAACAAAAACGAAGAAGATGAACTCAGTACTTGACACACAGCCATCGGCC
ATGGTTTTAGTGAAGGATGTGTTACGCGAAGGTCCTTTACCTGTTCAAATCCTCG
CAATTCCACAAGCCAACTCATCTCCATGCTCAAAATTAGCAGACAAAAACGGAA
CTGCAACCACGCCTTCTCCTTGTGCGGCCTCCTAAACCCCTGCTGATCGCTCTTC
CTTCCCAACATGGAGATTATCCTCTCATCCTCTTTTTCCACGGCTATGTACTCCT
CAATTCCTTCTATTCTCAACTCTTGCGCCATGTTGCTTCCCATGGATACATCGCC
ATAGCTCCTCAGATGTACAGTGAATTGGCCCAAATACGACTCCAGAAATAGCC
GATGCAGCGGCCATTACAGACTGGTTACGAGATGGACTCTCGGATAATCTTCCG
CAAGCTTTAAACAATCATGTGAGGCCCAATTTTGAGAAATTTGTGCTAGCGGGG
CACTCGCGCGGGGGTAAAGTGGCATTTCACACTTGCCTAGGTGAGTCTCGCA
GCCATCTTTAAAGTACTCGGCCCTGTAGGTCTTGATCCAGTCGATGGAATGGG
AAAAGATCAACAAACCAGTCATCCTATTCTGTCATACAGAGAGCATTCTTTGAT
TTGGGTATGCCAACATTAGTGGTAGGTTGGGGCCTGGGTCCGTGCAAAAGAAA
CCCTCTCTCCCTCCCTGTGCTCCCAAGGTGTTAACCACCATGATTTCTTCTAC
GAATGTGTCGCTCCTGCCTATCATTTTGTTGCCTCTGATTATGGGCATCTTGATT
TCTTAGACGACGACACCAAAGGAATAAGAGGAAAGGCTACTTATTGCCTCTGTA
AGAATGGGGAAGCAAGAGAGCCAATGCGGAAGTTTAGCGGTGGAATTGTGGTT
GCATTTCTTCAAGCATTCTTGGTGATAATCGTGGAGCCCTGAATGATATTATGG
TTTATCCTTACATGCTCCAGTCAAGATTGAGCCTCCAGAGTCTTTGGTTACAGA
AGATGTAATAATCCCAGAAGTCGAATTATTACGCCGGGCAGTTTGCAGATGATG
TACCATGGTATTATGCATTAAGGAATGTATTTGTTATTAATAAAAAATATTAAGAAG
TAAAAAAAAAAAAAAAAA

SEC ID NO:24

MVLVKDVFSEGPLPVQILAIPQANSSPCSKLADKNGTATTPSPCRPPKPLLIALPSQH
GDYPLILFFHGYVLLNSFYSQLLRHVASHGYIAIAPQMYSVIGPNTTPEIADAAITDW
LRDGLSDNLPQALNNHVRPNFEKFLVLAGHSRGGKVAFALALGRVSQPSLKYSALVG
LDPVDGMGKDQQTSHPILSYREHSFDLGMPTLVVGSGLGPCKRNPLFPPCAPQGV
NHHDFFYECVAPAYHFVASDYGHLDLDDTKGIRGKATYCLCKNGEAREPMRKF
SGGIVVAFQLQAFLGDNRGALNDIMVYPHAPVKIEPPESLVTEDEVKSPEVELLRRAV
CR

5

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

SEC ID NO:25

ACACAAAAAATATATAACACAAAGAAATAGAAGAAGGAAAAAATGTCCCCCTCC
 TTTCTTTTCTTTACTTTGTTTTGATAAAGGAAATGTCTCTTCATCATCAGCAA
 CTCCTTTGAGGACGGCAAATACAAAACAGATCTTTAACAGTAGGCTTATCATCT
 TGCTGCTGGAAAAAGCCCTCCTCTTCTCCGACTCCGCAGTCTCCGCCGAAGAG
 GCTTTTGGTGGCAACGCCGGTGGAGGAAGGAGAATATCCGGTGGTGTGCTCC
 TCCATGGTTACCTTCTCTACAACCTCATTTTATTCCCAGCTTATGTTGCATGTCTCT
 TCCCATGGCTTCATTGTCATCGCTCCGCAGTTATATAGCATTGCCGGACCAGAC
 ACCATGGATGAGATAAAATCAACGGCAGAGATTATTGATTGGTTATCGGTCGGA
 CTAACCACCTTTCTTCCACCACAAGTAACACCAAACCTATCCAAGTTCGCACTCT
 CCGGCCATAGCCGTGGTGGGAAGACCGCATTGCTTGGCCTTAAAGAAATTT
 GGATACTCGTCCGACCTAAAGATCTCGGCATTGATAGGTATAGATGTTGGAAGT
 GTTTTTGGACAAATGGCTATGGCCAATATTCCGGTGAATTTTTCGAGCAATTTG
 ATTGTGCAAAATGACCGGATTGTGGAATCGTAGGATTGATTGTTATGAGCACTATG
 GTATAGTGTAATCATATATCAAAAACGAAGTTCGTTTGAATGAGAAATGAAAGTC
 TAAAATAGATTATTTGAAAATATCTATATTAGAATTATGAGGTAAGAAACCTCTT
 GTGTTTAAATGGAGAAGTTATAACAAAGTTATAAAAAACTTTGTAACAATTTGG
 TGTGTTAGC

SEC ID NO:26

MSPSFLFFTLFLIKEMSSSSANSFEDGKYKTDLLTVGLSSCCWKKPSSSPTPQSP
 KRLLVATPVEEGEYPVVMLLHGYYLNSFYSQMLHVSSHGFIVAPQLYSIAGPDT
 MDEIKSTAEIIDWLSVGLNHFLPPQVTPNLSKFALSGHSRGGKTAFALALKKFGYSS
 DLKISALIGIDVGTVFWTNGYGQYSGEFFEQFDCRNDRIVES

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

SEC ID NO:27

MAAMVDAKPAASVQGTPLLATATLPVFTRGIYSTKRITLETSSPSSPPPKPLIIVTPA
 GKGTFNVLFLHGTSLSNKSYSKIFDHIASHGFIVVAPQLYTSIPPPSATNELNSAAEV
 AEWL PQGLQQNLPENTEANVSLVAVMGHSRGGQTAFALSLRYGFGAVIGLDPVAG
 TSKTTGLDPSILSFDSDFSIPVTVIGTGLGGVARCITACAPEGANHEEFFNRCKNSS
 RAHFVATDYGHMDILDDNPSDVKSWALSKYFCKNGNESRDPMRRCVSGIVVAFLK
 DFFYGDAEDFRQILKDPSPFAPIKLDSEYIDASSMLTTTHVKV

Preparaciones enzimáticas

- 10 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. 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. Las reacciones enzimáticas de los métodos de la invención se llevan a cabo en una vasija de reacción o en múltiples vasijas. Por ejemplo, las reacciones enzimáticas de los métodos de la invención se llevan a cabo en un aparato de refinado de aceite vegetal.

5 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.

10 En diversos aspectos de la invención, 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.

15 Aplicaciones industriales y médicas

Los polipéptidos, por ejemplo enzimas de la invención 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 invención 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 invención 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.

30 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 salcado 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.

40 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.

45 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. Las composiciones y métodos de la invención 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 invención pueden comprender modificar el pH para promover la separación acuosa de clorofilida.

50 Las composiciones y métodos de la invención 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

55 La invención 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, 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 invención.

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. Como alternativa, una, varias o todas las etapas usan una enzima. La reacción puede ser *in vitro* o *in vivo*.

5 La Figura 8 ilustra la reacción de una esterasa ejemplar de la invención 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 procedimiento de la invención, el anillo de porfirina hidrófilo se separa con la fracción de goma/agua usando uno cualquiera de los muchos métodos bien conocidos.

10 La Figura 9 ilustra y compara el procedimiento de decoloración (blanqueo) enzimático tradicional frente a uno ejemplar de la invención, en el que el procedimiento de blanqueo enzimático puede incorporar una esterasa de la invención. 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 invención, el aceite vegetal bruto se desgoma (opcionalmente, se neutraliza de forma cáustica), se blanquea usando, por ejemplo, un polipéptido de la invención, tal como una clorofilasa de la invención, 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.

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

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

30 La Figura 12 ilustra un procedimiento de decoloración (blanqueo) enzimático ejemplar de la invención que comprende la aplicación de un polipéptido de la invención, 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 clorofilida, o gomas y/o jabones), seguido de los procedimientos ilustrados en las Figuras 9, 10 u 11.

35 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 invención, 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.

40 La Figura 14 ilustra un procedimiento industrial ejemplar de la invención – un procedimiento de biodesgomado, que comprende el uso de una fosfolipasa A y al menos un polipéptido de la invención que tiene actividad de enzima clorofilasa. El al menos un polipéptido de la invención 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.

45 La Figura 15 ilustra otro procedimiento industrial ejemplar de la invención que comprende el uso de al menos un polipéptido de la invención que tiene actividad de enzima clorofilasa. El al menos un polipéptido de la invención 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 de retención, a un tanque cáustico y/o a una mezcladora de retención.

La Figura 16 ilustra otro procedimiento industrial ejemplar de la invención que comprende el uso de al menos un polipéptido de la invención 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 invención. El al menos un polipéptido de la invención 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 invención 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 invención se pueden usar en métodos para procesar aceites vegetales a partir de diversas fuentes, tales como salvado de arroz, habas de soja, colza, cacahuets 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 invención 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 invención 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 invención se pueden usar en cualquier método de procedimiento de aceite, por ejemplo desgomado o procesos equivalentes. Por ejemplo, las composiciones y métodos de la invención 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 invención, 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 invención se pueden usar en la aplicación industrial de desgomado enzimático de aceites triglicéridos, como se describe, por ejemplo, en el documento EP 513709.

Las composiciones y métodos de la invención se usan para tratar aceites vegetales, por ejemplo aceites brutos, tales como salvado de arroz, soja, cáñola, girasol y similares. Esto mejora la eficiencia del procedimiento de desgomado. Los métodos de la invención 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 invención 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.

La descripción proporciona procedimientos enzimáticos que comprenden el uso de composiciones y métodos de la invención 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 invención, 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 gommas 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 invención 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.

5 Al terminar un tratamiento enzimático de la invención, el líquido tratado (por ejemplo, aceite) se separa con un medio apropiado, tal como un separador centrífugo, y se obtiene el aceite procesado. 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.

10 En un procedimiento ejemplar para el desgomado mediante ultrafiltración, una enzima usada en un método de la invención 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 invención se pueden inmovilizar a cualquier sustrato, por ejemplo filtros, fibras, columnas, perlas, coloides, geles, hidrogeles, mallas, y similares.

15 Las composiciones y métodos de la invención 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 procedimiento de la invención, se usa una composición o método de la invención y al menos un agente degradante de la pared de la célula vegetal (por ejemplo, una celulosa, una hemicelulosa 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, 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.

20

25

Las composiciones y métodos de la invención también se pueden poner en práctica usando procedimientos como se describen en la patente U.S. nº 5.414.100. Por ejemplo, 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.

30

35 Las composiciones y métodos de la invención 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 invención también se pueden poner en práctica usando procedimientos como se describen en la patente U.S. nº 5.315.021. Por ejemplo, los métodos o composiciones de la invención 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.

40

45

Procesamiento enzimático de oleaginosas

Las composiciones y métodos de la invención 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. Estos procedimientos de la invención pueden incrementar el rendimiento del aceite y mejorar la calidad nutricional de las harinas obtenidas. El procesamiento enzimático de oleaginosas usando las enzimas y métodos de la invención 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. Como alternativa, los procedimientos de la invención comprenden además el uso de fosfolipasas, proteasas, fosfatasa, fitasas, xilanasas, amilasas (por ejemplo, □-amilasas), glucanasas (por ejemplo, □-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. Como alternativa, los procedimientos de la invención 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

50

55

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.

Los procedimientos de la invención 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 invención 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. Los métodos se ponen en práctica con una enzima de la invención 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.

Las proteasas se usan en combinación con los métodos de la invención. 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. Los métodos descritos aquí 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.

La extracción completa de proteínas, pectina, y hemicelulosa se usan en combinación con los métodos de la invención. 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. Los métodos descritos aquí 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.

Los métodos descritos aquí 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 invención). 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 invención pueden comprender además la incorporación de proteasas, α -amilasas, poligalacturonasas (en diversas combinaciones entre sí y con una o más enzimas de la invención) 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 invención pueden comprender además la incorporación de pectinasas, α -amilasas, proteasas, celulasas en diferentes combinaciones (entre sí y con una o más enzimas de la invención) 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 invención 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 invención), como se describe, por ejemplo, por Montedoro (1976) *Acta Vitamin. Enzymol.* (Milano) 30:13.

Los métodos de la invención 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 invención). 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 invención pueden comprender además la incorporación de proteasas, α -amilasas, poligalacturonasas (en diversas combinaciones entre sí y con una o más enzimas de la invención) 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 invención pueden comprender además la incorporación de pectinasas, α -amilasas, proteasas, celulasas en diferentes combinaciones (entre sí y con una o más enzimas de la invención) 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 invención 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 invención), como se describe, por ejemplo, por Montedoro (1976) *Acta Vitamin. Enzymol.* (Milano) 30:13.

Las composiciones y métodos de la invención se pueden poner en práctica con métodos como se describe en la patente U.S. nº 6.376.689. Por ejemplo, las composiciones y métodos de la invención 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. Los métodos descritos aquí comprenden además una mezcla de ácidos sulfúrico y fosfórico acuosos que se mezcla con el aceite para eliminar compuestos de tipo clorofilico 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.

5 Purificación de fitosteroles a partir de aceites vegetales

Las composiciones (por ejemplo, esterasas) y métodos de la invención 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 invención 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 invención 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 invención 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 invención 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 invención 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 invención 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 invención 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 usada en la invención. El 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.

Los productos de fabricación comprenden aparato para ajustar el pH, por ejemplo incrementar el pH ("tratamiento cáustico"), y después, como alternativa, neutralizar el pH.

La invención se describirá adicionalmente con referencia a los siguientes ejemplos; sin embargo, se ha de entender que la invención no está limitada a tales ejemplos.

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 invención y los ácidos nucleicos que las codifican, y para determinar si un polipéptido está dentro del alcance de la invención.

Las esterasas 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:

ES 2 461 857 T3

- 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.
- 5
- 100 ml de volumen de reacción.
 - Análisis mediante LC-VIS; inyección de 1 ml de muestra.

El 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 descritas aquí.

10 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: H2O				
B: MEOH + 1mM				
NH4OAc				
C: MTBE				
T (min)		A	B	C
0		10%	80%	10%
2,3		10%	80%	10%
2,3				
1		0%	50%	50%
4		0%	50%	50%
4,1		10%	80%	10%
7		10%	80%	10%
Señal DAD	L (nm)	Bw	Referencia I	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

Compuesto	Rt
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

- <110> DIVERSA CORPORATION LAM, David WEINER, David HITCHMAN, Timothy BARTON, Nelson Robert BURK, Mark J.
- 10 <120> COMPOSICIONES Y MÉTODOS PARA LA DECOLORACIÓN ENZIMÁTICA DE CLOROFILA
- <130> 564462013240
- <140> No asignado todavía
- <141> Concurrentemente con la presente
- <150> US 60/580,447
- 15 <151> 2004-06-16
- <160> 27
- <170> PatentIn version 3.1
- <210> 1
- <211> 915
- 20 <212> ADN
- <213> Desconocido
- <220>
- <223> Obtenido de muestra medioambiental
- <400> 1

ES 2 461 857 T3

atgtcgcgtg tctgtcttcc actcacactc acactcgcac tcacactctc tgccagagca 60
 gcggagccga ccacgggtgaa gctctggccc ggcaaagcgc ccggcgagac gaaggacatc 120
 ggcccggaga agtatctcga aaccaagaag gggcagcttg acatcaagcg gctcaccaac 180
 gtgagcgaac cgaccatcac gatctattcg ccgccaaggg agaaggcgaa cgacacagtg 240
 gtgatgtcgc cgcccggcgg cgggtacaac atcctcgcca tcgaacacga agggaccgat 300
 gtctgcgagt ggtgaactc gctgggcgtg accgcggtgc tgctcaagta tcgctgccc 360
 cggcgcccca tgcagtcgcc ggacaacctc gcgatgattc aggacgcgca gcgcgccatc 420
 agtcttgctc gcagtatgca gaaagagttg ggcattcacc caacacgtgt gggaatgctc 480
 ggcttttcgg ccggggggaa cctcaccgct tgcaccgctg tggccgaaa gcggatgtac 540
 gagaacatcg acaaggtgga cgaggattc aactgcacgc cgaacttcgc catcctcgtg 600
 taccocgctg atctcgtcga gaaggatgga acgctgcgcg ccgagttcaa agtgaagatc 660
 gattcgcgcg cgatgttctt cgtgcattcg tccgacgaca acgtgagcag tgaaaacagc 720
 gtggcgcttt acctggcact caagaagaac aaggtgccgg cggaaatgca cctctacgcc 780
 agcggcgggc acggctacgg aatgcgcaag gtteccgcatc cgtgcgcaag ctggcccggc 840
 cgcgcgcggg agtggatgaa agcacaccgg ctgcttgaga agccaagcc cgagccggat 900
 gggaagaagg agtaa 915

<210> 2

<211> 304

<212> PRT

5 <213> Desconocido

<220>

<223> Obtenido de muestra medioambiental

<220>

<221> SEÑAL

10 <222> (1)...(20)

<400> 2

ES 2 461 857 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

ES 2 461 857 T3

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

5 <213> Desconocido

<220>

<223> Obtenido de muestra medioambiental

<400> 3

atgtcacttg atccacaaac gaaatttcta ttagcccaat tggctgctgc cgatgctcct 60
 ccaatggaga ctttgtcacc ggaaatggct agacagcat tcatattgcc gcaaggtgca 120
 gtggaggaag tagggaaggt agaaaatcga acgattcctg gtcttgcaac ggacatccct 180
 gtccgtgttt attaccctaa agaactccag cctgaaaatc cgcgctagt tttctatcat 240
 ggcggcggct gggtaattgg caatttggac tctcatgacg atatttgccg tgctttaacc 300
 aacctcgcca actgcgtgac catttccgtc gactaccgcc tggctccaga gaataaattt 360
 cctgctgcgg ttgaggatgc ttatgctgct gcacaatatg tgtatgacca tgcagaagat 420
 ttcaaagtcg acaagaccgg tattgcccgtt ggcgggtgaca gtgccggtgg aaatcttgcc 480
 gctgtcgtga cgaatctggc aaaagacaaa aactcacctt ctatttgttt ccaacttctg 540
 atttatccaa gcaccaatgc aggtggtgag ccaacagcat caatggttga gaatgccat 600
 ggctatttct tagaaaaagg cagcatggat tggttccgtg actgctacct gaacagtga 660
 gaagacaaac agaatccgct agtctcaccg atgctttatg atgatttcca aggactgcct 720
 ccagcaattg tgattactgc cgagtacgat ccattgagag atgaaggcga agcatatgcc 780
 aaaaagctgg gcaagcagc ggttgctgtc caaacatcc gctttgacgg tacgattcat 840
 ggctttgtca gcatgtcggc tgtcattagc caaggaagc cggcattgga aaaagcagga 900
 gaggctttaa ctaaagcctt tcaataa 927

10

<210> 4

<211> 308

ES 2 461 857 T3

<212> PRT

<213> Desconocido

<220>

<223> Obtenido de muestra medioambiental

5

<400> 4

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

ES 2 461 857 T3

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 medioambiental

<400> 5

atgagaagaa ttgttttctt ttatattctg gcggttctct gcgtatcctg tgcgaaccgg 60
aatccttctg tctcggaccc cgctccgtca ccgaccttgg gggagcaatc gtcagcgagt 120
ttaccgaaga tcgtagcatt cggcgacagc ctaactgctg gctttggtct ttcccagaat 180
gaaagtatc cagcgctttt gcaggaacga ctgagacagg acggttatga ttacgaagt 240
atcaatgccg gagtttcagg cgataccagc gcggggcggg tgcgtaggat cgactggggt 300
ctggatgaga gcgtacgaat tctaattctt gaattgggtg ctaatgattt tcttctgtggc 360
catccagtag ctcagactaa gaaaaacctc gccgttatca tcgaacgtgc tcaggaaaaa 420
aacgtcaggg ttctgctagc agggatgttt gcaccgacga atactggatg ggagtaccag 480
gggcagatcc agcaaagtgt taacgatctt tcgcgagaaa aagccgtacc actaattcca 540
ttctttctcg agggagttagc cggcattcca actctgaatc tagctgatgg cattcatccc 600
aatgcggctg gcacaaagat cgttgctgag aatgtgtaca agtatctaaa accgatgctc 660
ccatctgatt aa 672

10 <210> 6

<211>223

<212> PRT

<213> Desconocido

<220>

<223> Obtenido de muestra medioambiental

<220>

<221> SEÑAL

5 <222> (1)...(25)

<220>

<221> DOMINIO

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

<223> Lipasa/Acilhidrolasa semejante a GDLSL

10 <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

ES 2 461 857 T3

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

5 <213> Desconocido

<220>

<223> Obtenido de muestra medioambiental

<400> 7

```

atgacacgaa aaaaaatcgg cttggcttta tccggcggcg cggcgcgtgg cttcgccat      60
ctcggagttt tgaaagtttt cgccgaacac ggcatcccg tcgatttcgt cgccggaacg      120
agcgcggtt cgtttgccgg cgccgcttcc gcttccggtt taagcgtggc ggaaatcatc      180
gaaatgtcaa ggaaaatcag ttggtttcgg atgaccgat tttcgtactc gccgaaaggc      240
ttgctgtcga acgcgccgat gggagcgttt atcaatcagc attttccgcg caaaaaattt      300
gaagaactgc cgataccggt cgccgcaatt acgtgcatc tcgaaaccgg cgaggaaatc      360
gtgcttaaag aaaccggcga cgtggcgact gccgtccgtg cgagctgcgc gctgcccggc      420
gtcttcgtgc caatcgaata cggcggacgg cgattgatag acggcggcgt cgtgtcgaac      480
gtgccgacga gagccgtcaa aaagctcggc gcggaatca ttattgcggt tgacgttctg      540
gcgtgcggca caacttactg gggttcgcct tccacttgc tcggcatctt tttccaatcg      600
gcgatgatgc ttctccgccc cgcttcaaaa tctcatcatt accgcgcgag cgtcgtcatc      660
acgccgcaaa tcgctcatct gcgcccgac gaaatcagca aatggacga atttatcaaa      720
gcgggcgaac aagccgcgct tgaaaaagtt gacgaaatca aggctttgct cgccgaacaa      780
taa                                                                                   783
    
```

10 <210> 8

<211> 260

<212> PRT

<213> Desconocido

<220>

15 <223> Obtenido de muestra medioambiental

<220>

<221> DOMINIO

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

<223> Fosfolipasa semejante a patatina

20 <400> 8

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

ES 2 461 857 T3

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

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

5 <212> ADN

<213> Desconocido

<220>

<223> Obtenido de muestra medioambiental

<400> 9

```

atgttcaaca aagcacttcc tgcagcggcc gcggtcgcgc gtctgttctt ttccacctcg      60
gccatggcgc tgcctgccga tacgcccgga acccccttcc cgtcgggtgc cagcttcgaa      120
aggtccggtc cctacagcac caccagccgc agcgaaggcc ccaactgccg gatctatcga      180
ccgagcaacc tgggccagaa cggcgtccgc caccgggtca tcctctgggg caacggcacc      240
ggtgccagcc cgaaccaccta ccgcggcctg ctcgagcaact gggccagcca cggcttcgtg      300
gtcgcggcgc ccgagacctc caacgccggc tccggggcgc agatgctcaa ctgcctgagc      360
tacctgcaga ccgaagccgg gcgcagcagc ggcacctacg tcggccgggt caatctcggc      420
cgcgtcggca cgtccggcca ctgcagggc ggtggcggct cgatcatggc cgggcgcgac      480
accggatca aaaccaccgc gcccatgcag ccctacacc tgggcctggg ccatgtgagt      540
tcctcgcaga gccagcagaa cggcccgatg ttctgatgt cggcagcct ggacaccctg      600
gccggcccga ccctgaacca ggcgccggtc tatcgccggg ccaacgtgcc ggtgttctgg      660
ggcaccctgc gcgggccag ccacttcgtg ccggtcggca gtgccggcgg ttaccgtggc      720
ccgtccaacc cctggttccg ctaccagctg atggacgata cctcggcgcg cagccagttc      780
gtcggcacca actgoggcct gtgccgcgac ttctcctgga ccgacatcca ggcgaagggc      840
gcgctgtaa                                     849
    
```

<210> 10

5 <211> 282

<212> PRT

<213> Desconocido

<220>

<223> Obtenido de muestra medioambiental

10 <220>

<221> SEÑAL

<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
    
```

15

ES 2 461 857 T3

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

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 medioambiental

<400> 11

ES 2 461 857 T3

gtgtccgccc ccgaagccgc gagcgtggag catgcggacc cggcggggt cgccaccctg 60
gcggaacgcc ccgcggcgcg tccgcgcata gtcttctctg gcgacagcct gaccgccggg 120
tacggactgg cgcgggagca gtccgtgccc gcgtactcc agcggcggct cgaccgggag 180
gggtatgagt acgaagtcgt caatgcgggc gtctcgggag acacgtccgc cggcgggctg 240
agccggctcg actggtcgtt cgacggcgag gtcgccctgc tgggtggtgga gctcggcgcc 300
aacgacgggc tgcgcggcct gcccggtgcc gcgatgaagc gcaacctgga cgccatcatc 360
acgcgcgccc gcgcgcgggg catcaccgtg gtgctcgccg gcatggaagc gccgcccgaac 420
tacggcgcgg cctacacgac cgagttccgg caggcgtttc acgatctcgc ccgtaccac 480
gacgtgccgt tcgtgccgtt ctttctcgaa ggggtagccg gcctgcccga cctcaacatc 540
gccgacggca tccatcccaa cgccgagggc gcgcgcgtgg tcgaggccaa cgtctggcag 600
gtgctggaac cgttactcga cgagcctggc cggcgggcgc cctcgcgcgc gcagcccgcg 660
gatggccgct ga 672

<210> 12

<211> 223

<212> PRT

5 <213> Desconocido

<220>

<223> Obtenido de muestra medioambiental

<220>

<221> DOMINIO

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

<223> Lipasa/Acilhidrolasa semejante a GDSL

<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

ES 2 461 857 T3

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 medioambiental

<400> 13

atgaaggcaa gctgccgcga catcgaatc gaatacgaaa cattcggcca tccggacgac 60

ccggccatcg tgctgatcat gggactcggg gccagctca tctgtggcc ggaagcgttc 120

tgccgatgc tggcgcgcg tggtcactac gtggtgcgct tegacaaccg cgacatcggc 180

ctgtgcagc atctgatca ccttccccgc cccaacctgc cgctcgccgc gctccgccag 240

gcgctgcgcc tgccagttcg cgccagttac atgctcgacg acatggcggg cgacgtcgcg 300

ggcctgcttg atgcactgaa catcaagcag gcgcacgtcg tcggcgtgtc gatgggcggc 360

atgatgcccc agctgtgtgc cgcacggcac gcgaccgcg tgcgcagcct gaccttgctg 420

atgaccacca gcggcgcgcg caacgttccg ggcccctcac tcggcatgcg catggaaatg 480

atccgtcgac cgcgcgcac ctcgcgcgag gggctgatcc gccatggtat gcgtacctgg 540

ES 2 461 857 T3

cggatcatcg gcagcccgac gtaccggaag cgggaagccg aactgcgccg catcgtcgcc 600
 gagggtttg accgcgcggt tcaccgggcc ggtttcatgc gccagctgca cgccgttctc 660
 gcggcaccga gccgcgcacc cctgctgccg cgcatcaaac agccggccga cgtcattcac 720
 ggcgacgccg acctgctggt accagtggca gcggcagctg atctggtgcg ccgcctgccg 780
 aacgccacgc tcgacatcgt gccgggcatg gggcatgact tcccgaccga gatcatgccg 840
 cgtatcgcgc gccgcattgt cgaaaccgcg gcacgcgacc cgcagcggct cgccgcctag 900

<210> 14

<211> 299

<212> PRT

5 <213> Desconocido

<220>

<223> Obtenido de muestra medioambiental

<220>

<221> DOMINIO

10 <222> (49)...(288)

<223> plegamiento de hidrolasa alfa/beta

<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

ES 2 461 857 T3

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

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 medioambiental

<400> 15

atgggcaaac aatgcctta tagagtgatt tttccagtaa actaccaaac ttcaaaaaa 60
cgctatccag tgatttacct tctgcacggc ctttcgggca attacaaaaa ctggaccgag 120
aaaaccaa at tgacaaaata tgcaaccgaa tacaacttct tgattatcac agttgaaggc 180
gagaacgggtt ggtattctga cagtaaaatc aaacctaaca ggctttacga aagctacata 240
atccaggaac tgatacctga agttgataaa aagtttcgca caatagcaga tcgcaatcac 300
agagctatcg caggtctttc gatgggtggt tacggtgcga taaagtttg cttgaaatat 360
cctgaaaaat tcgctttgat tgggtccttt agcggcgctc ttgcggcaac ttcaatcaag 420
gaaggcacag cgcttgagtg gattacaaaa accatcaatg atgctttcgg tcctgaagga 480

ES 2 461 857 T3

agcgaatcca gaaaagaaaa cgatattttc cagattgcmc gcgacctgaa tgatgaacaa 540
 ataaagaaac ttccctttat ttactttgat tgcggaaccg aagattttct gttcaaagac 600
 aatcaggatt tcatgaaact acttggtgaa aagagaatca aacacgaata tagacaaaag 660
 cccggcacc c atactggga atactgggat agccaaataa aagaattctt gggtttagcc 720
 aacaaattta tagaaaaaac tgctttatgg cgatactaa 759

<210> 16

<211> 252

<212> PRT

5 <213> Desconocido

<220>

<223> Obtenido de muestra medioambiental

<220>

<221> DOMINIO

10 <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
						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

ES 2 461 857 T3

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

<210> 17

<211> 753

<212> ADN

5 <213> Desconocido

<220>

<223> Obtenido de muestra medioambiental

<400> 17

```

atgaaaaagt ataaaacag cctggtttta agtgggtggcg gaacacgtgg ttttgcgcat    60
ctgggagtga tagcagctct ttacgataag ggaattaaac cgcacataat ttcgggtaca    120
agtgctgggtg ccattggttg cgcttttatt gctgccggga aaaatccaca cgacgttggtg    180
gagattttta aaaagggatc gtttttcaat tacacaaaac tacagatacc cgcgcacggt    240
ttgatgaaac tggatggact gaaggaattg tttcaaaagg aaattcatgt aaaaaacctc    300
gaagagcttg aaatccccct ttttattgcc atttogaatt taaataaagg aaccgtggaa    360
tacagaaata gcggtctttt gggtgaaact gtgcttgccct catcttcgat ccccatactt    420
tttgtccggg ttttaatcgg cgacgatttg tatgttgacg ggggattaat ggataacatt    480
ccggttgaac ccatcaaact cgattgtgaa caaatcattg tttcaaatat cagtccgatt    540
aatcccgttg aaaaaattaa aaatctgatt cacattgcta ctcgtacttt ttatatgagt    600
gtaaacgcca acatgaaaca ggtaaaaaaa tattccacc attatattga accggacgga    660
atagatactt acgaaatfff aagccgcacc cacgccgatg agttgtatga acttggatat    720
aattcgacaa ttaaaatcct gaattcgcac taa                                753
    
```

10 <210> 18

<211> 250

<212> PRT

<213> Desconocido

<220>

<223> Obtenido de muestra medioambiental

<220>

<221> DOMINIO

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

5 <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
 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

ES 2 461 857 T3

Lys Lys Tyr Ser Thr His Tyr Phe 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 medioambiental

<400> 19

```

atgggtgcga ataatccggt cacaaaattht gtaatggata aggggcaggg cactgcagca      60
cgtttgctgg attgcctgcc gagctttgca caagaacgtc tggtaaaagc attggattat      120
ccttatgact atcctgattt ggatccttat ataaaatgct tgatggcgat tcagattaaa      180
caggagagac acagttttat tagtgacgat gcggtgcagt cacgtcaact gtttgatgaa      240
cgcatgaaag ctattcagac gaaaccgacg ccggtcaagg cagtcgagga tctgcgtttg      300
ccattgcaaa atggcactat ctttgcccga cattatcatc cggcaccaca gaaaaaattg      360
ccgatggtca ttttctatca tgggtgtgca tttatagtgg gtggtctgga tacgcatgat      420
gagttctgcc gtttattggc ggtgcatgcc aaggtgcagg tactcagcgt ggcttatccg      480
ctaaccgccc aatacagtc tttgcagatg gtacaggtct gtgaagatgc tctggcttgg      540
gtacatcaaa acatcaagca gttgaaaatc tataaaaacc agattgtagt ggcaggggat      600
agtgcaggtg gaaatctggc ggcagtggtt gcgcagcggg gtgccgataa aatctatgca      660
cctcgcgcac agttattgct ttaccggtct gtcgatttta aaagccgcca tccttctttt      720
tatgcataca atcagggcct ggtactttca gctcaggata ttgatctggt gaccaagcta      780
tatgtgaaa cacatcaggt cgaactggat gatccgctga tttcaccgac ctatggtgaa      840
ctgaagaatc tgccgcctgc ctatgtgatt accgcccgtc atgatgtggt gcatgatgag      900
ggttcaatct atgcgctgaa gttacgtgag aatggggtgc gagtgtatta tcaggagtat      960
acggatcagg cccacggttt tatcaatctg accccaattc ataaacgttc gaaaaagcag     1020
gtcattgaac tcagcaagaa tttccgtaaa ttctggata aaaagatctg a                1071
    
```

10 <210> 20

<211> 356

<212> PRT

<213> Desconocido

<220>

15 <223> Obtenido de muestra medioambiental

ES 2 461 857 T3

<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

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

ES 2 461 857 T3

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 tcaaactctc taaccttgga ctcatcatct 120
 cgttgctgca aaataacacc gtcttctaga gcttcaccgt ctccgcaaaa gcagctgttg 180
 gtggctacgc cgggtggagga aggagattat ccggtggtga tgctcctcca tggttacctt 240
 ctctacaact ccttctattc tcagcttatg ttgcatgtct cttctcatgg cttcatcctc 300
 atcgtcctc agttatatag tatcgccgga ccagacacaa tggatgagat taaatcaacg 360
 gcggagatta tggattgggt atcagtagga cttaatcact ttcttcagc gcaagtaaca 420
 ccaaacttat ccaaatttgc cctctccggc catagccgcg gtggcaaac cgcgtttgcg 480
 gtcgccttaa agaaatttgg gtactcctcg aatctaaaga tctcgacatt gatcggata 540
 gatccagtcg atggaacagg gaaagggaaa caaacccctc ctccggtggt ggcttacctt 600
 ccaaactcat ttgacctaga caaacgcct atacttgtga tcggttcggg gcttggtgaa 660
 accgctcggg acccattatt cccaccgtgt gcacctccc gagtgaatca ccgagagttc 720
 tttcgggaat gtcaagggtcc agcatggcat ttcggttggg aggattatgg gcatttgac 780
 atgcttgatg atgatacaaa agggattaga gggaagagtt cttattggtt gtgtaagaat 840

ES 2 461 857 T3

gggtgaatgaga ggaagaccat gaggaatatt gttggtggac ttgttgatc atttttgaag 900
gcttatttgg aaggagatga tcgtgaatta gttaagatca aagatgggtg tcacgaggat 960
gttcccgttg aaattcaaga gtttgaggtt atcatgtaaa cataagtttt tctttagggg 1020
ctggtttttc tattgtcaat atcatcagct tttgttgctt atggttttac aaacttatat 1080
tgtacaactc ttaagtcac ctctttgctt atgatattaa cccgatc 1127

<210> 22

<211> 318

<212> PRT

5 <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
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

ES 2 461 857 T3

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

<212> ADN

5 <213> *Ginkgo biloba*

<400> 23

ttgaaaaaca aaaacgaaga agatgaactc agtacttgca cacagccatc ggccatggtt	60
ttagtgaagg atgtgttcag cgaaggctct ttacctgttc aaatcctcgc aattccacaa	120
gccaactcat ctccatgctc aaaattagca gacaaaaacg gaactgcaac cagccttct	180
ccttgtcggc ctccctaaacc cctgctgac gctcttctct cccaacatgg agattatcct	240
ctcatcctct ttttccacgg ctatgtactc ctcaattcct tctattctca actcttgcgc	300
catgttgctt cccatggata catcgccata gctcctcaga tgtacagtgt aattggccca	360
aatacgactc cagaaatagc cgatgcagcg gccattacag actggttacg agatggactc	420
tcggataatc ttccgcaagc tttaaacaat catgtgaggc ccaattttga gaaatttgtg	480
ctagcggggc actcgcgcgg gggtaaagtg gcatttgcac ttgccctagg tcgagtctcg	540
cagccatctt taaagtactc ggcccttgta ggtcttgatc cagtcgatgg aatgggaaaa	600
gatcaacaaa ccagtcaccc tattctgtca tacagagagc attcctttga tttgggtatg	660
ccaacattag tggtaggttc gggcctgggt ccgtgcaaaa gaaaccctct cttcctccc	720

ES 2 461 857 T3

tgggctccc' aggtgttaa ccaaccatgat ttcttctacg aatgtgtcgc tcctgcctat 780
 cattttgttg cctctgatta tgggcatctt gatttcttag acgacgacac caaaggaata 840
 agaggaaagg ctacttattg cctctgtaag aatggggaag caagagagcc aatgcggaag 900
 tttagcggtg gaattgtggt tgcatttctt caagcatttc ttggtgataa tcgtggagcc 960
 ctgaatgata ttatggttta tccttcacat gctccagtca agattgagcc tccagagtct 1020
 ttggttacag aagatgtaaa atccccagaa gtcgaattat tacgccgggc agtttgcaga 1080
 tgatgtacca tgggtattatg cattaagga atgtatttgt tattaataaaa atattaagaa 1140
 gtaaaaaaaaa aaaaaaa 1157

<210> 24

<211> 342

<212> PRT

5 <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
 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

ES 2 461 857 T3

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 gtctcttca tcatcagcaa actcctttga 120
 ggacggcaaa taaaaacag atcttttaac agtaggetta tcatcttgct gctggaaaaa 180
 gccctcctct tctccgactc cgcagtctcc gccgaagagg cttttggtgg caacgccggt 240
 ggaggaagga gaatatccgg tggatgatgct cctccatggt taccttctct acaactcatt 300

ES 2 461 857 T3

ttatccag cttatgttc atgtcttc ccatggcttc attgtcatcg ctccgcagtt 360
 atatagcatt gccggaccag acaccatgga tgagataaaa tcaacggcag agattattga 420
 ttggtatcg gtcggactaa accactttct tccaccacaa gtaacaccaa acctatccaa 480
 gttcgcactc tccggccata gccgtggtgg gaagaccgca tttgccttgg ccttaaagaa 540
 atttggatac tcgtccgacc taaagatctc ggcattgata ggtatagatg ttggaactgt 600
 tttttggaca aatggctatg gccaatattc cggatgaattt ttcgagcaat ttgattgtcg 660
 aatgaccgg attgtggaat cgtaggattc attgttatga gcactatggt atagtgtaat 720
 catatatcaa aaacgaagt cgtttgaatg agaaatgaaa gtctaaaata gattatttgt 780
 aaaatatcta tattagaatt atgaggtgaa aaacctcttg tgtttaaaat ggagaagtta 840
 taacaaagtt ataaaaaact ttgtaaacaa tttggtgtgt tagc 884

<210> 26

<211> 213

<212> PRT

5 <213> *Brassica oleracea*

<400> 26

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

ES 2 461 857 T3

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

Ala Leu Ala Leu Lys Lys Phe Gly Tyr Ser Ser Asp Leu Lys Ile Ser
 165 170 175

Ala Leu Ile Gly Ile Asp Val Gly Thr Val Phe Trp Thr Asn Gly Tyr
 180 185 190

Gly Gln Tyr Ser Gly Glu Phe Phe Glu Gln Phe Asp Cys Arg Asn Asp
 195 200 205

Arg Ile Val Glu Ser
 210

<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

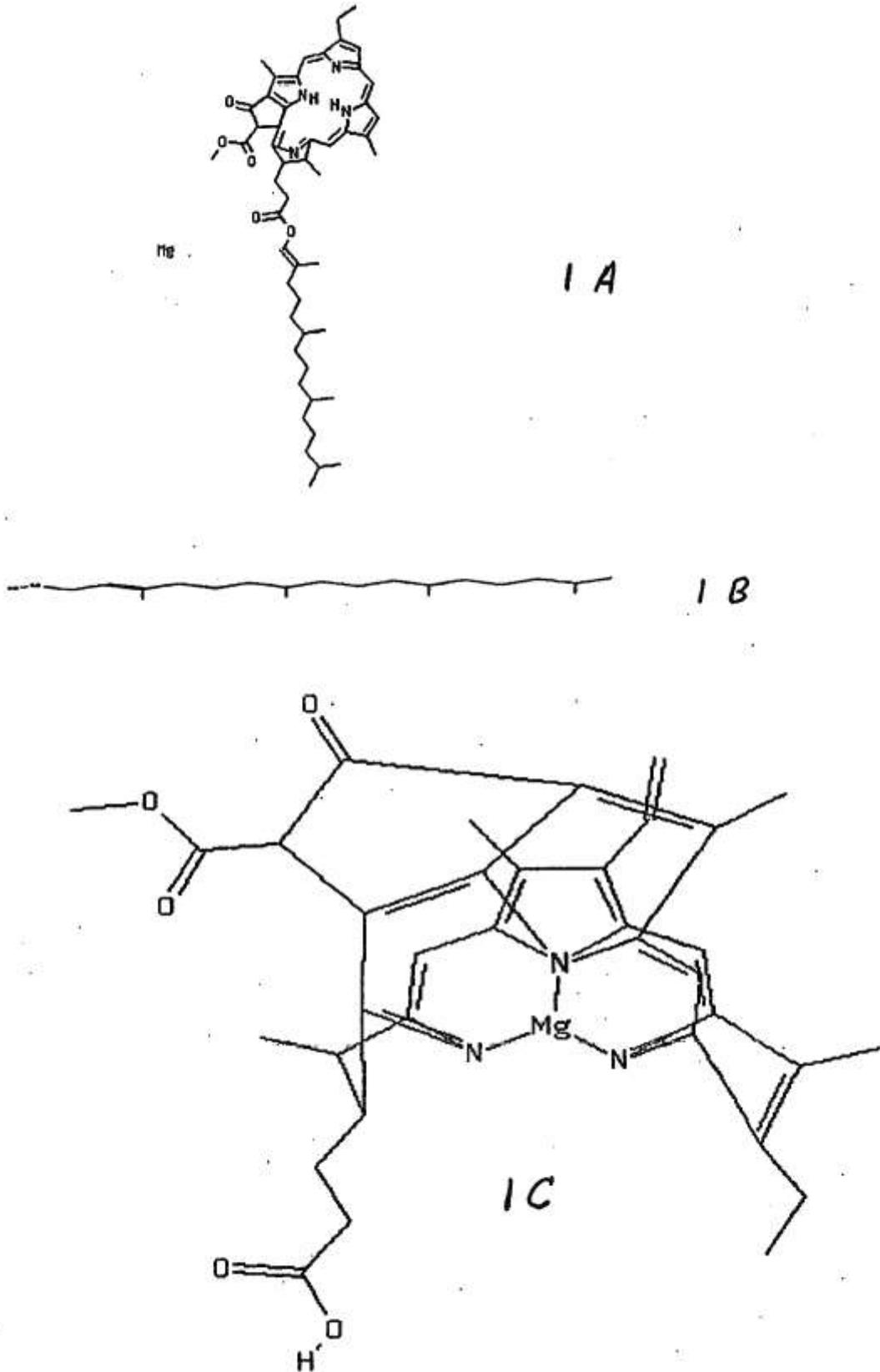
Asn Leu Pro Glu Asn Thr Glu Ala Asn Val Ser Leu Val Ala Val Met
 130 135 140

Gly His Ser Arg Gly Gly Gln Thr Ala Phe Ala Leu Ser Leu Arg Tyr

REIVINDICACIONES

- 5 1. Uso de al menos un polipéptido como se expone en SEC ID NO:10, o de al menos un polipéptido que es una variante de SEC ID NO:10, y la variante es al menos 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, o 99% idéntica a SEC ID NO:10, que tiene una actividad de clorofilasa, en un tratamiento enzimático que es una decoloración de una composición que contiene clorofila en condiciones en las que el polipéptido puede catalizar la hidrólisis de la clorofila para generar una clorofilida y un fitol; y la clorofilida y el fitol se separan.
- 10 2. Uso de al menos un polipéptido como se expone en SEC ID NO:10, o de al menos un polipéptido que es una variante de SEC ID NO:10, y la variante es al menos 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, o 99% idéntica a SEC ID NO:10, 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.
- 15 3. El uso de cualquiera de las reivindicaciones 1 ó 2, 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 gráfica, 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.
- 20 4. El uso de cualquiera de las reivindicaciones 1 a 3, en el que la composición que contiene clorofila o 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 cáñola,
 - (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.
- 25 5. El uso de cualquiera de las reivindicaciones 1 a 4, 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.
- 30 6. El uso de cualquiera de las reivindicaciones 1 a 4, en el que al menos una etapa del tratamiento se lleva a cabo en un extracto celular o en una célula completa.
7. El uso de cualquiera de las reivindicaciones 1 a 6, en el que el polipéptido se usa con una lipoxigenasa.
8. El uso de una cualquiera de las reivindicaciones 1 a 7, en el que el tratamiento comprende además modificar el pH para promover la separación acuosa de una clorofilida o feoforbida.
- 35 9. El uso de cualquiera de las reivindicaciones 1 a 8, en el que el tratamiento se combina con una etapa de neutralización cáustica.
10. El uso de cualquiera de las reivindicaciones 1 a 9, en el que se usa polipéptido con una fosfolipasa, y opcionalmente la fosfolipasa es una fosfolipasa C.
- 40 11. El uso de cualquiera de las reivindicaciones 1 a 10, en el que el tratamiento está incluido en un procedimiento de desgomado.
12. El uso de cualquiera de las reivindicaciones 1 a 11, en el que el tratamiento comprende además la eliminación de clorofila o feofitina residual; plaguicidas; o hidrocarburos aromáticos policíclicos.
13. El uso de cualquiera de las reivindicaciones 1 a 12, en el que el polipéptido es codificado por un ácido nucleico como se expone en SEC ID NO:9.
- 45 14. El uso de cualquiera de las reivindicaciones 1 a 13, en el que el tratamiento enzimático de una composición que contiene clorofila o feofitina se combina con un procedimiento de desgomado y una etapa de neutralización cáustica.
15. El uso del polipéptido de cualquiera de las reivindicaciones 1 a 12, en el que el polipéptido es codificado por una variante del ácido nucleico como se expone en SEC ID NO:9, y la variante es al menos 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, o 99% idéntica a SEC ID NO:9.

Figura 1



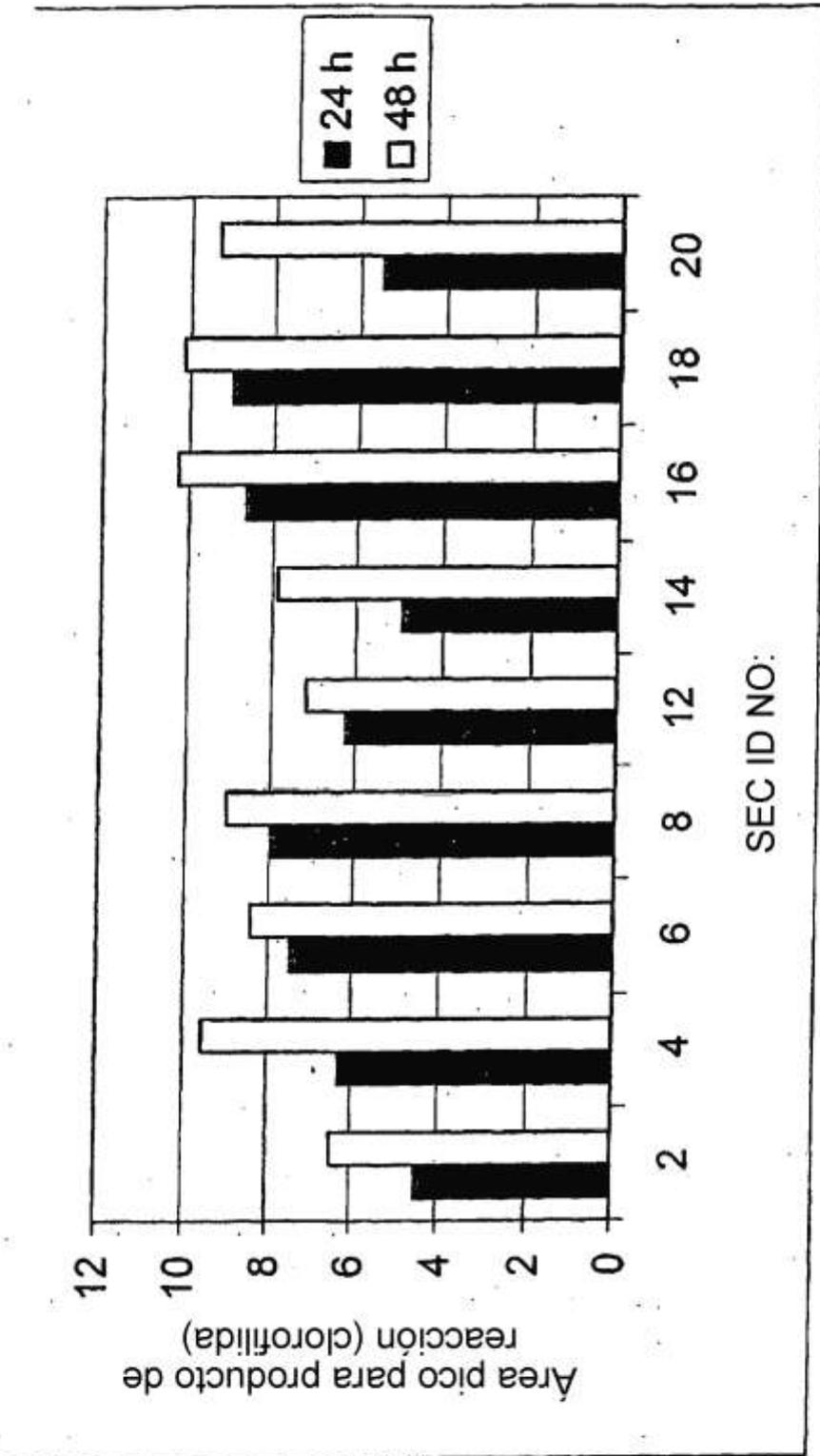


Figura 2

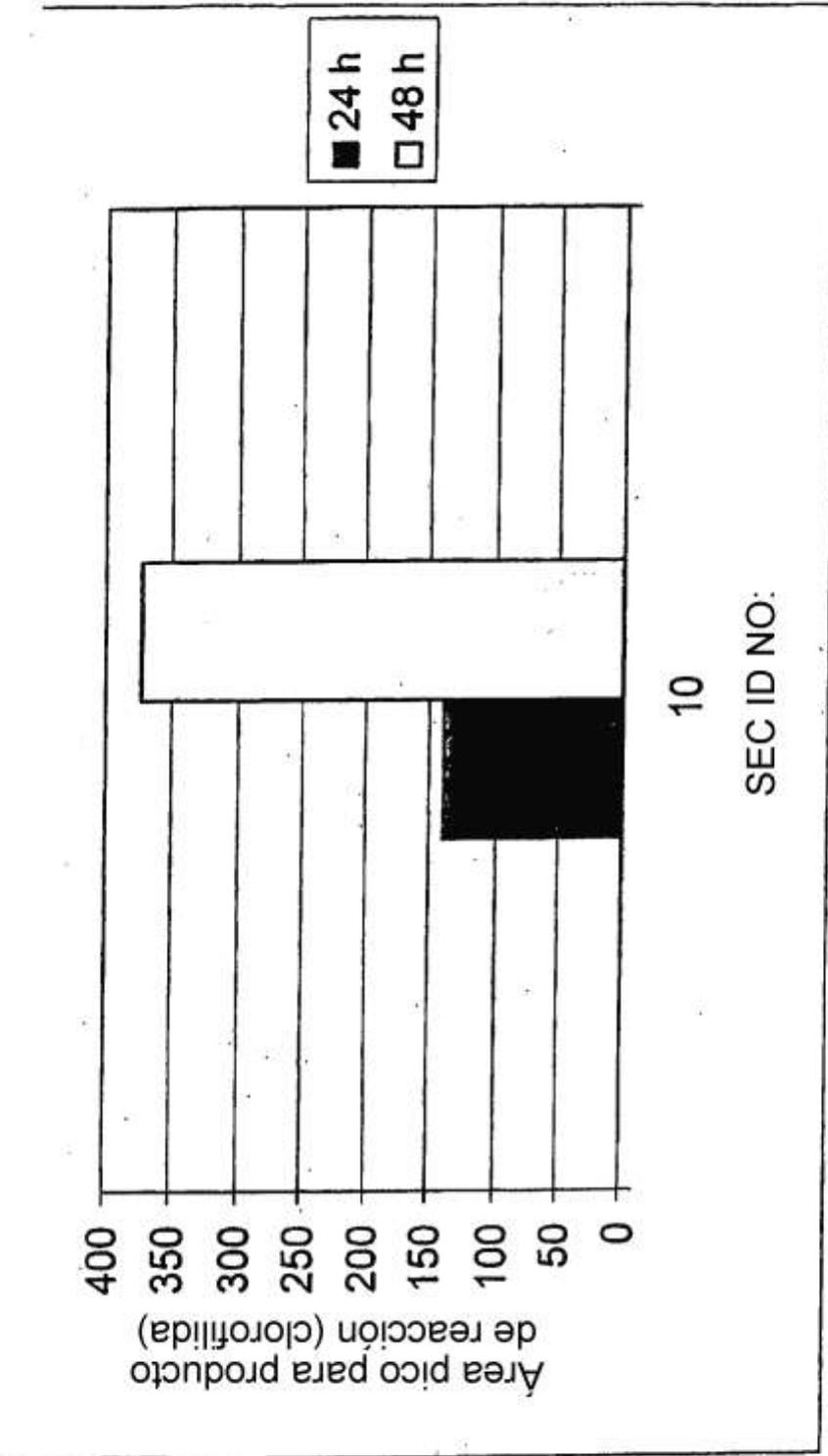


Figura 3

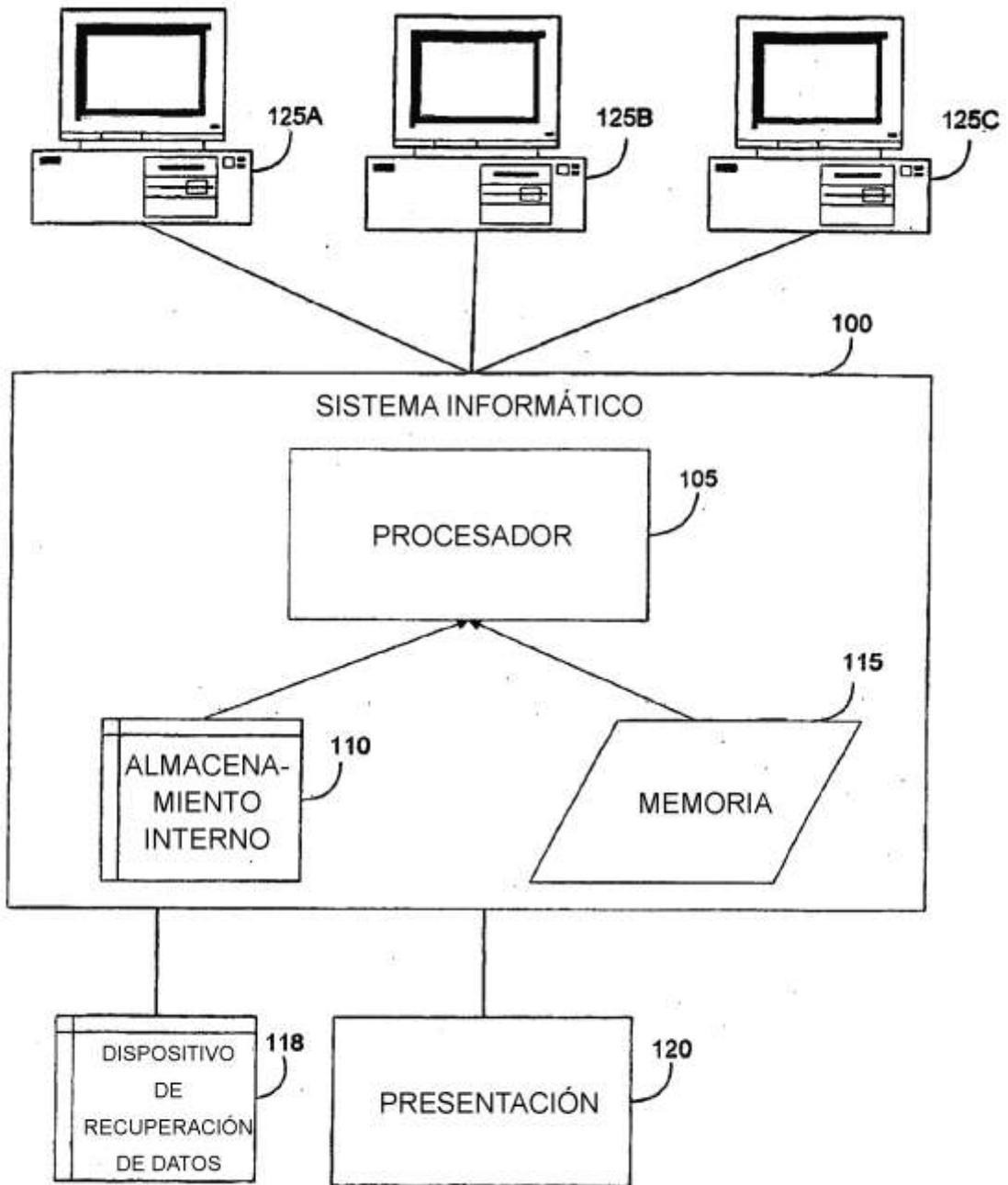


Figura 4

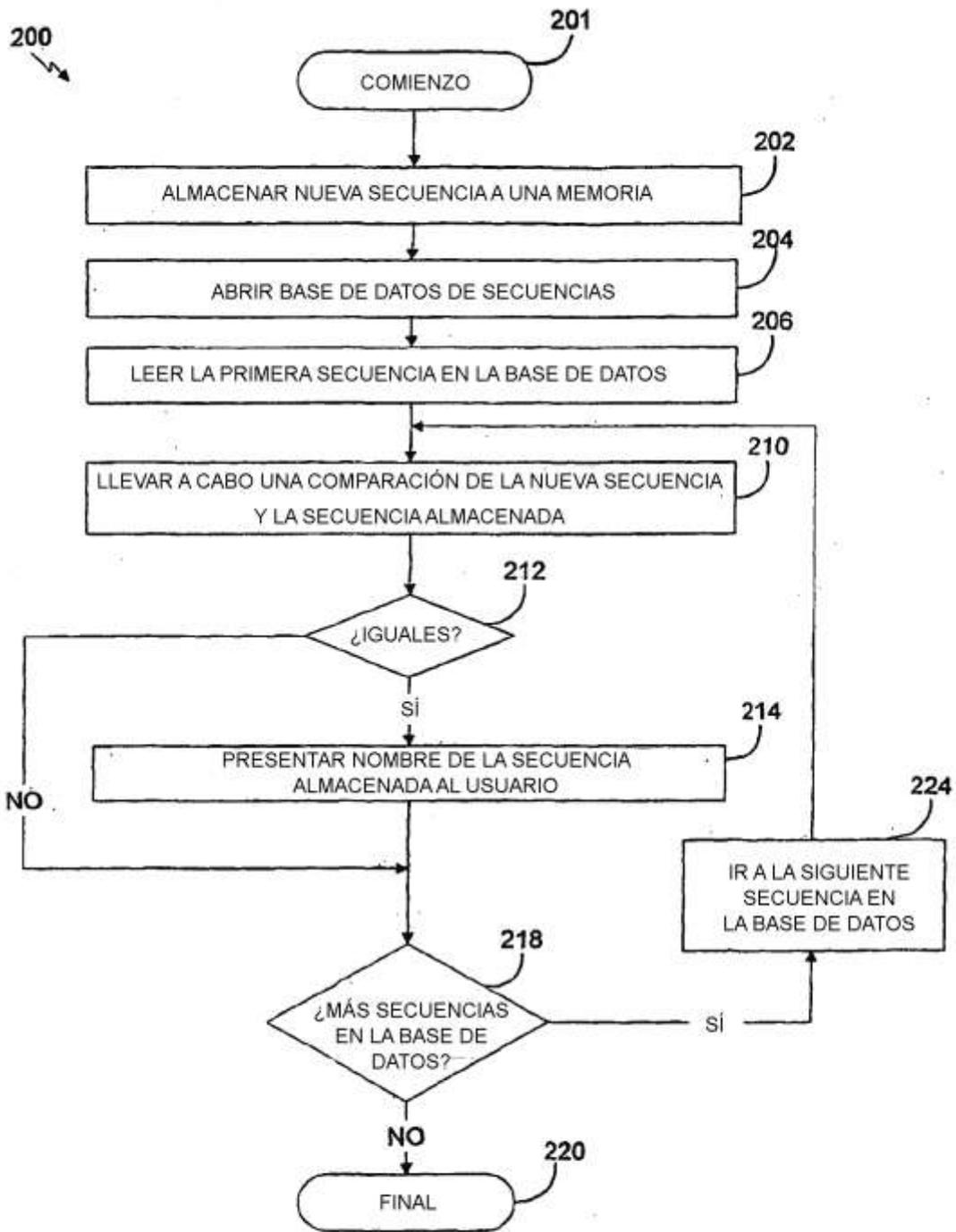


Figura 5

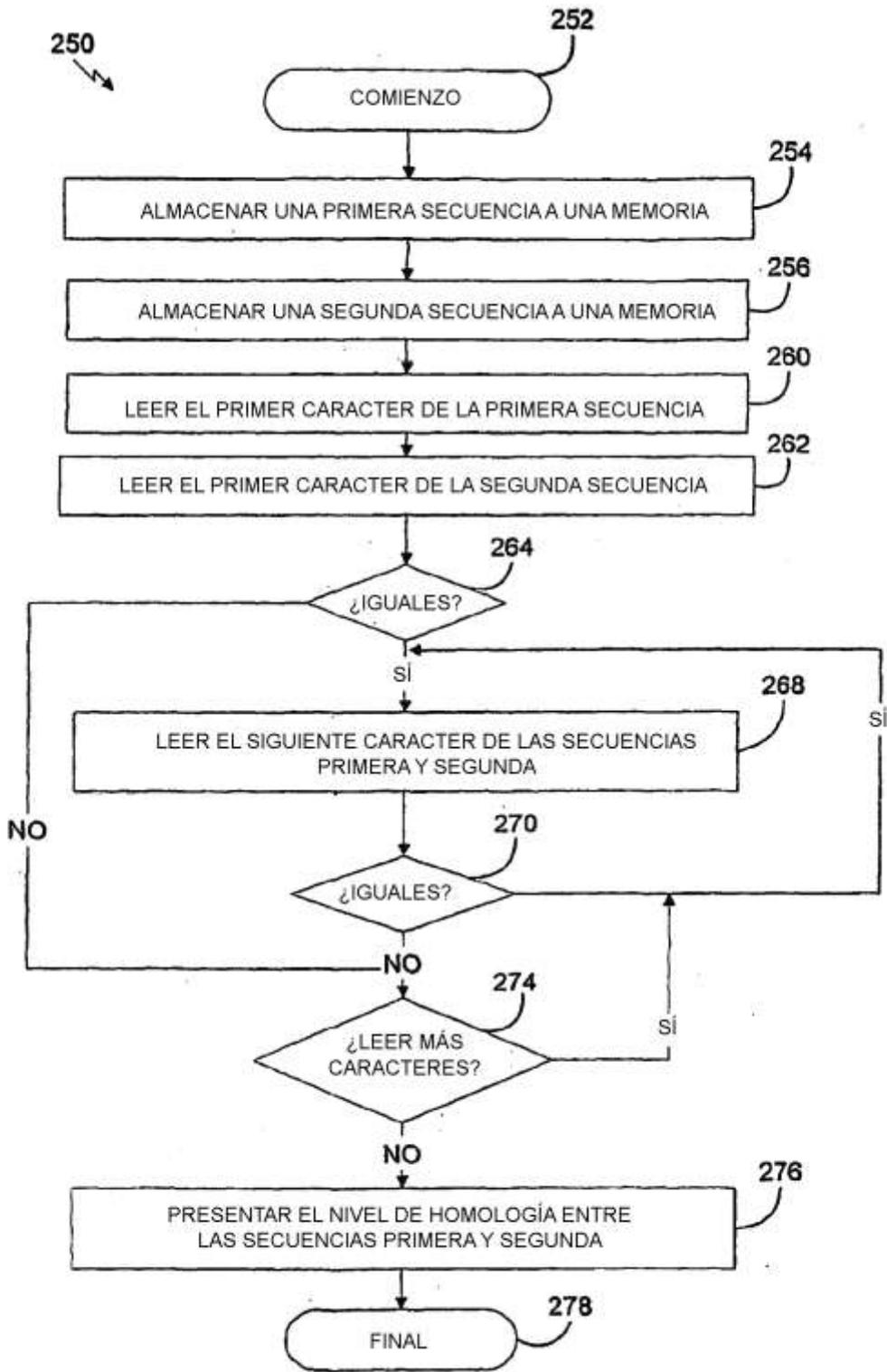


FIGURA 6

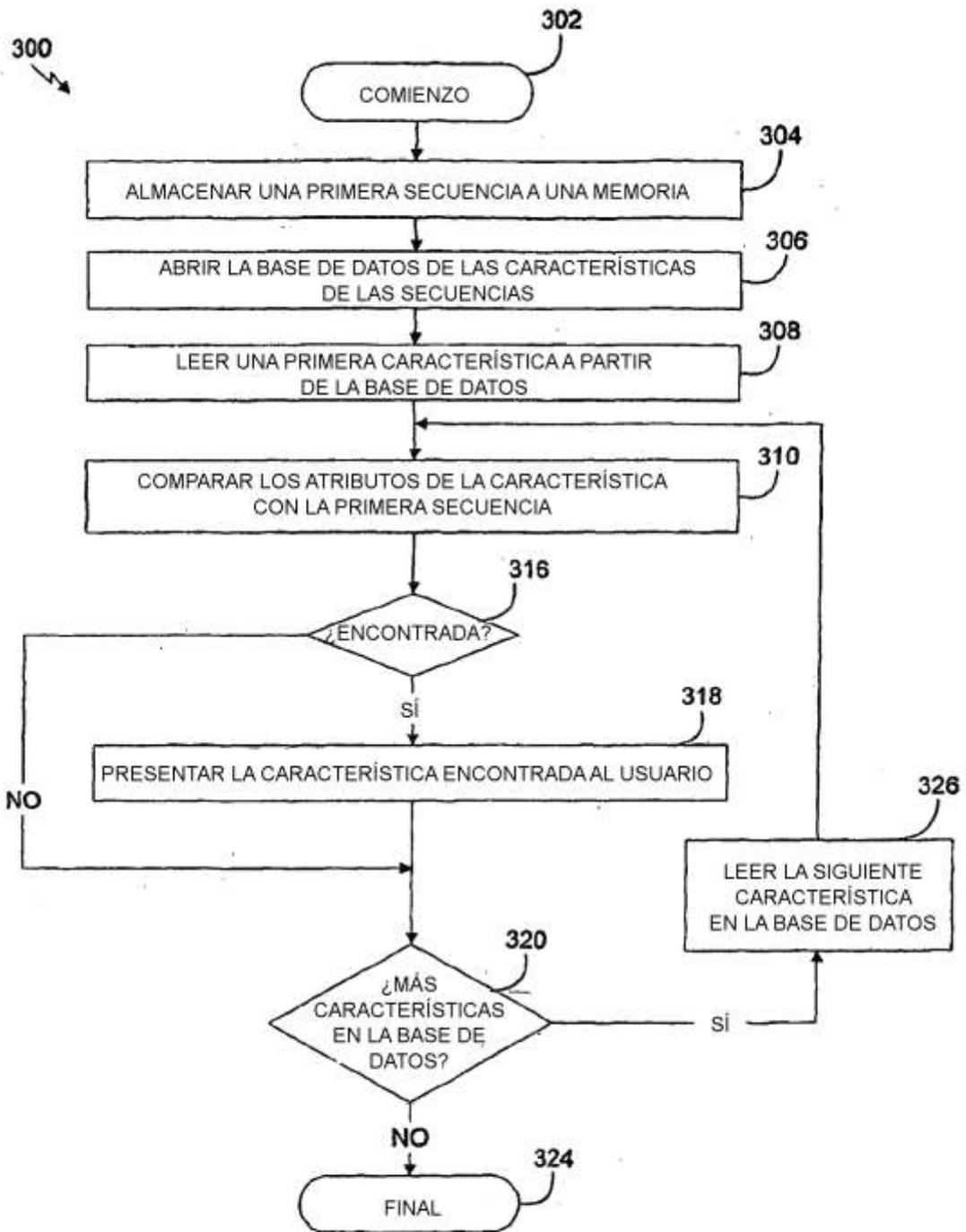


Figura 7

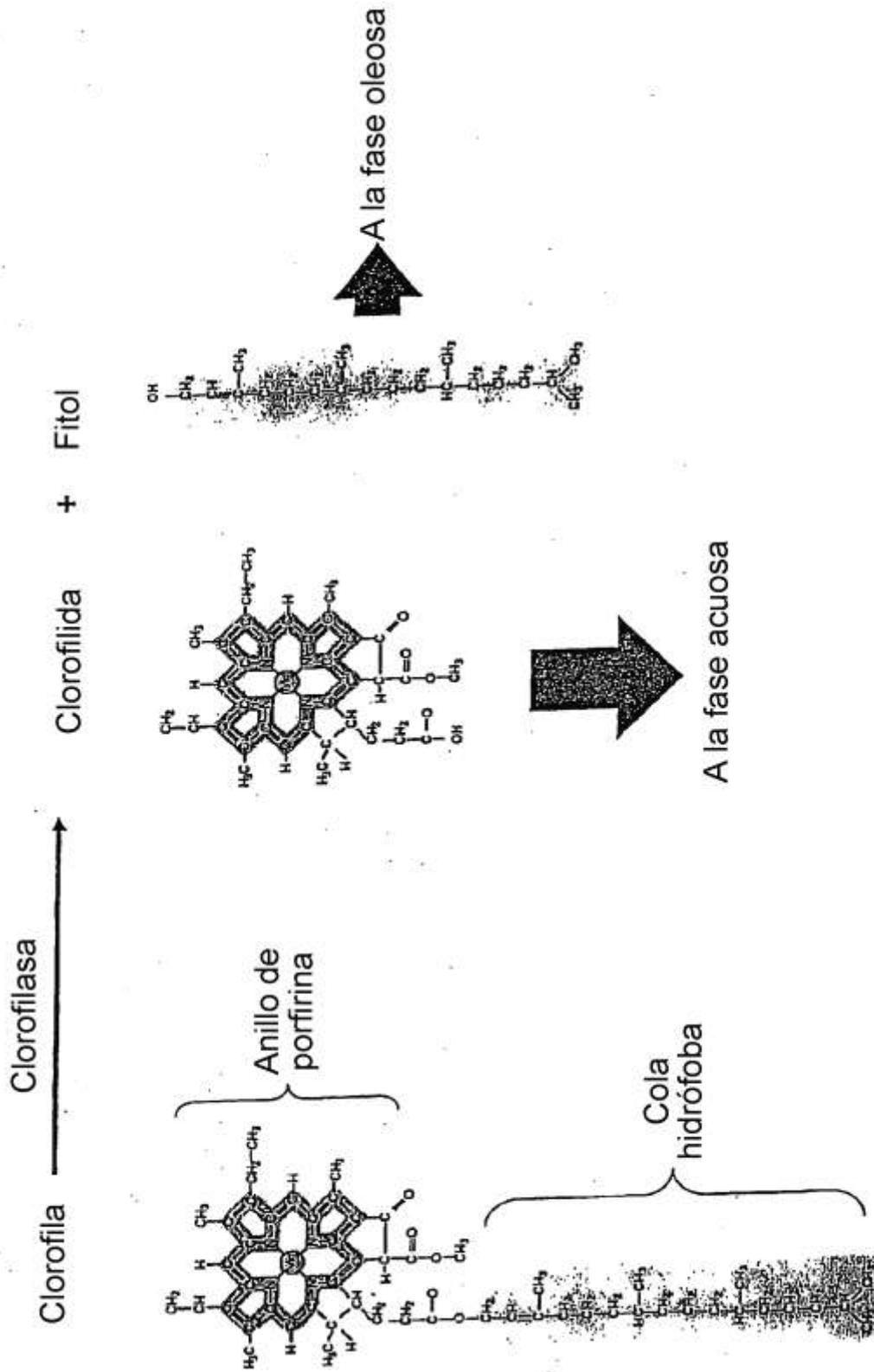


FIGURA 8

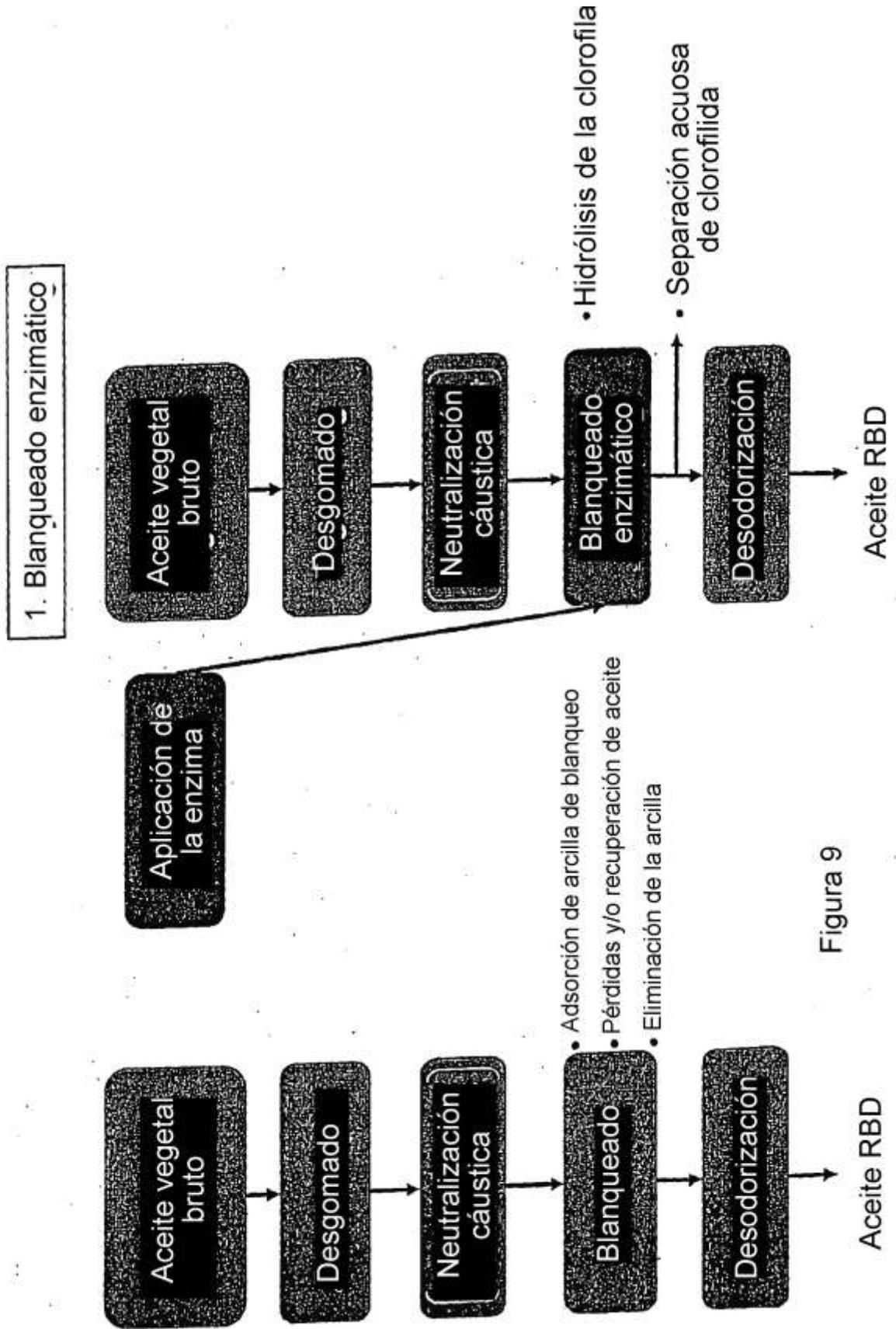


Figura 9

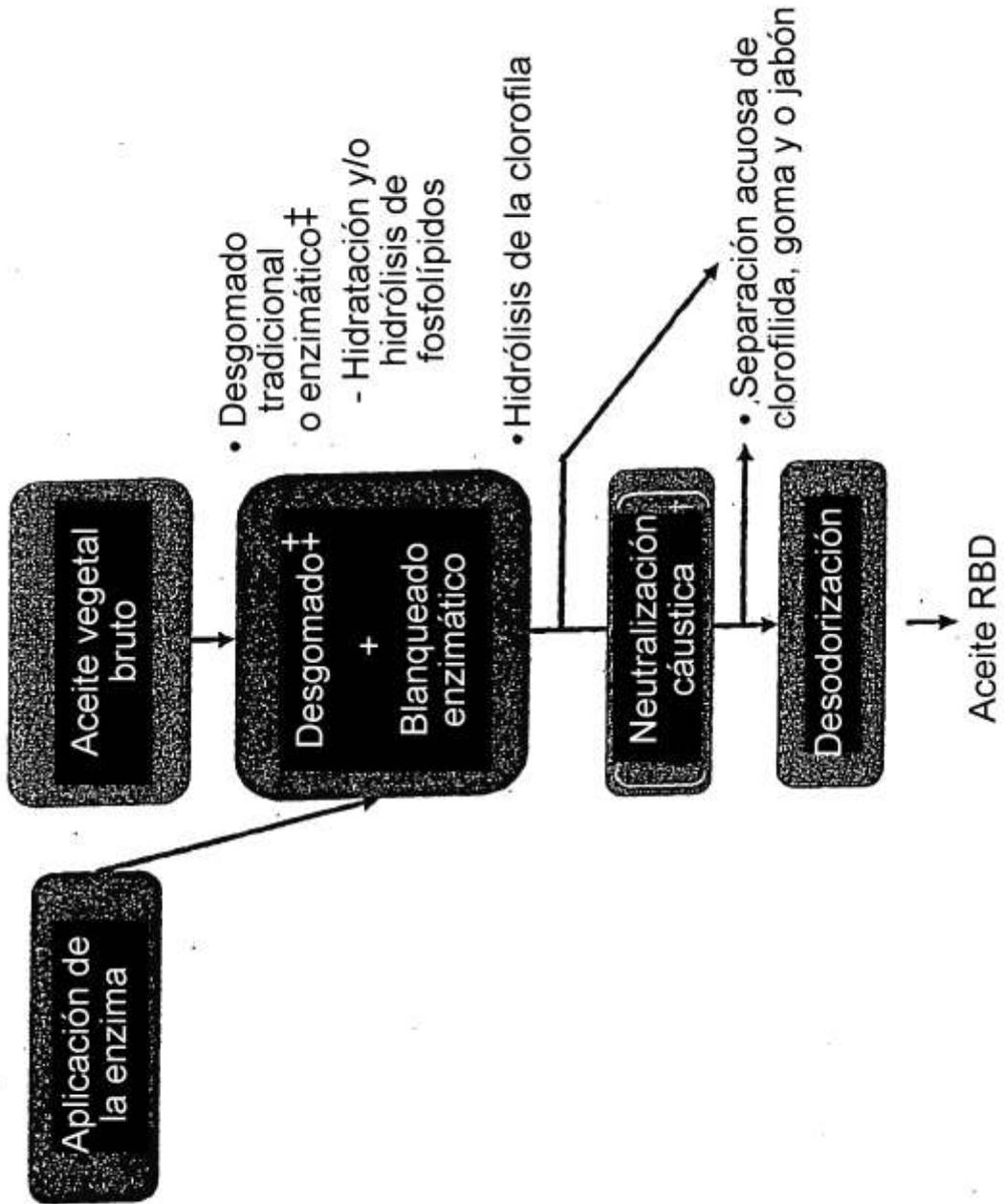


Figura 10

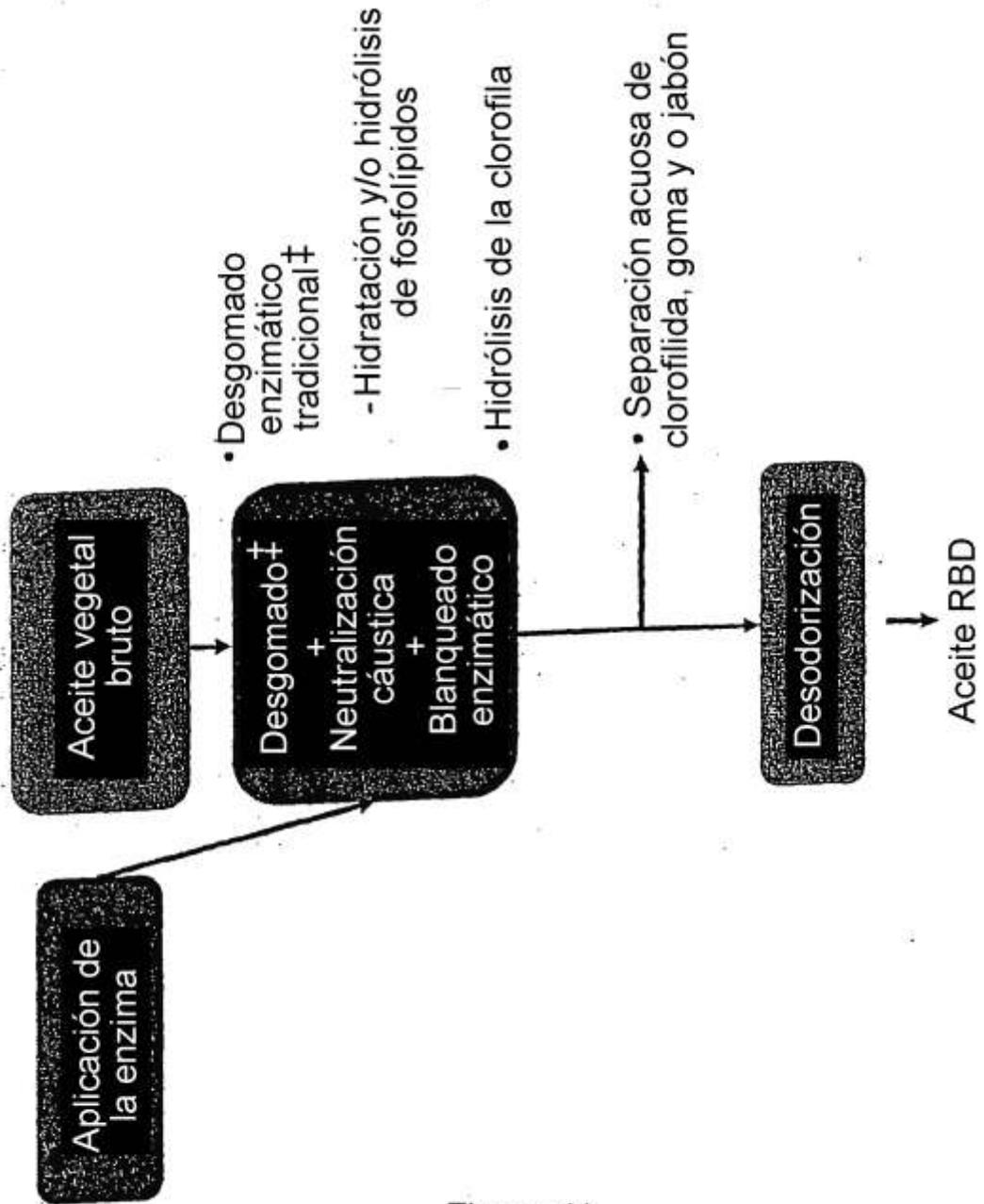


Figura 11

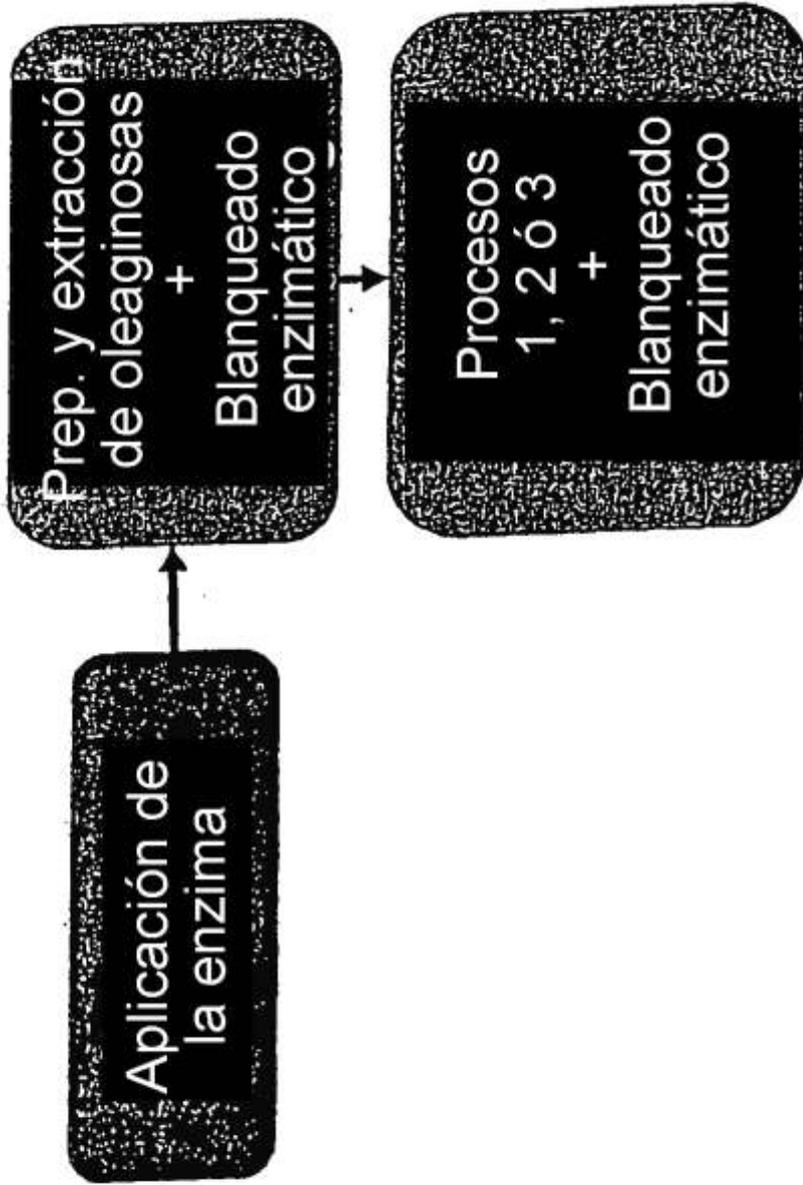


Figura 12

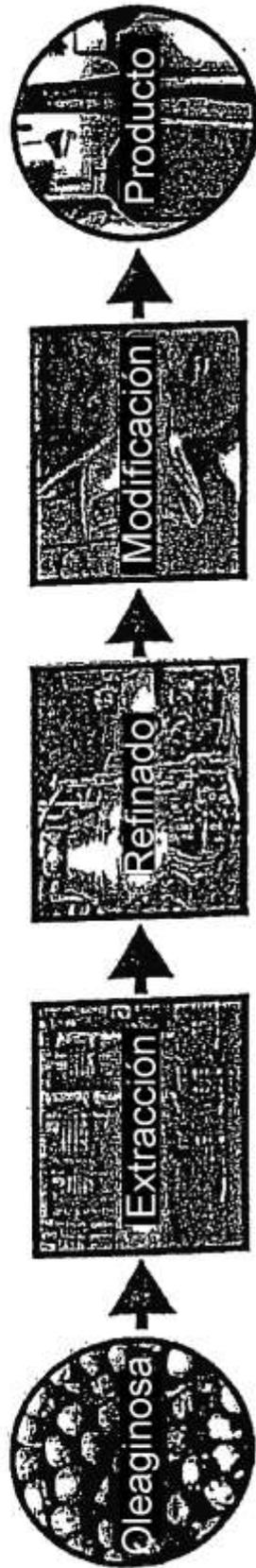


Figura 13

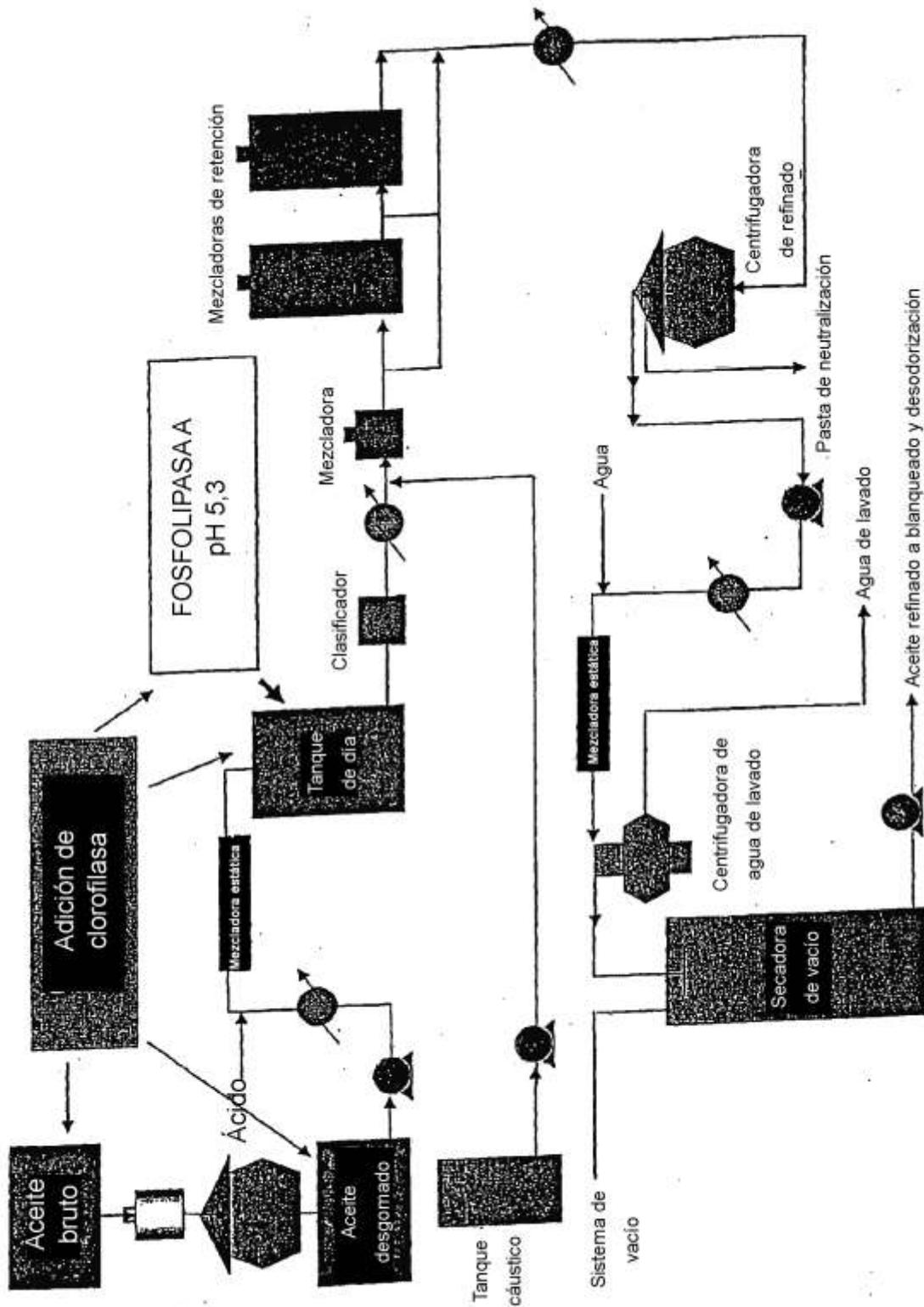


Figura 14

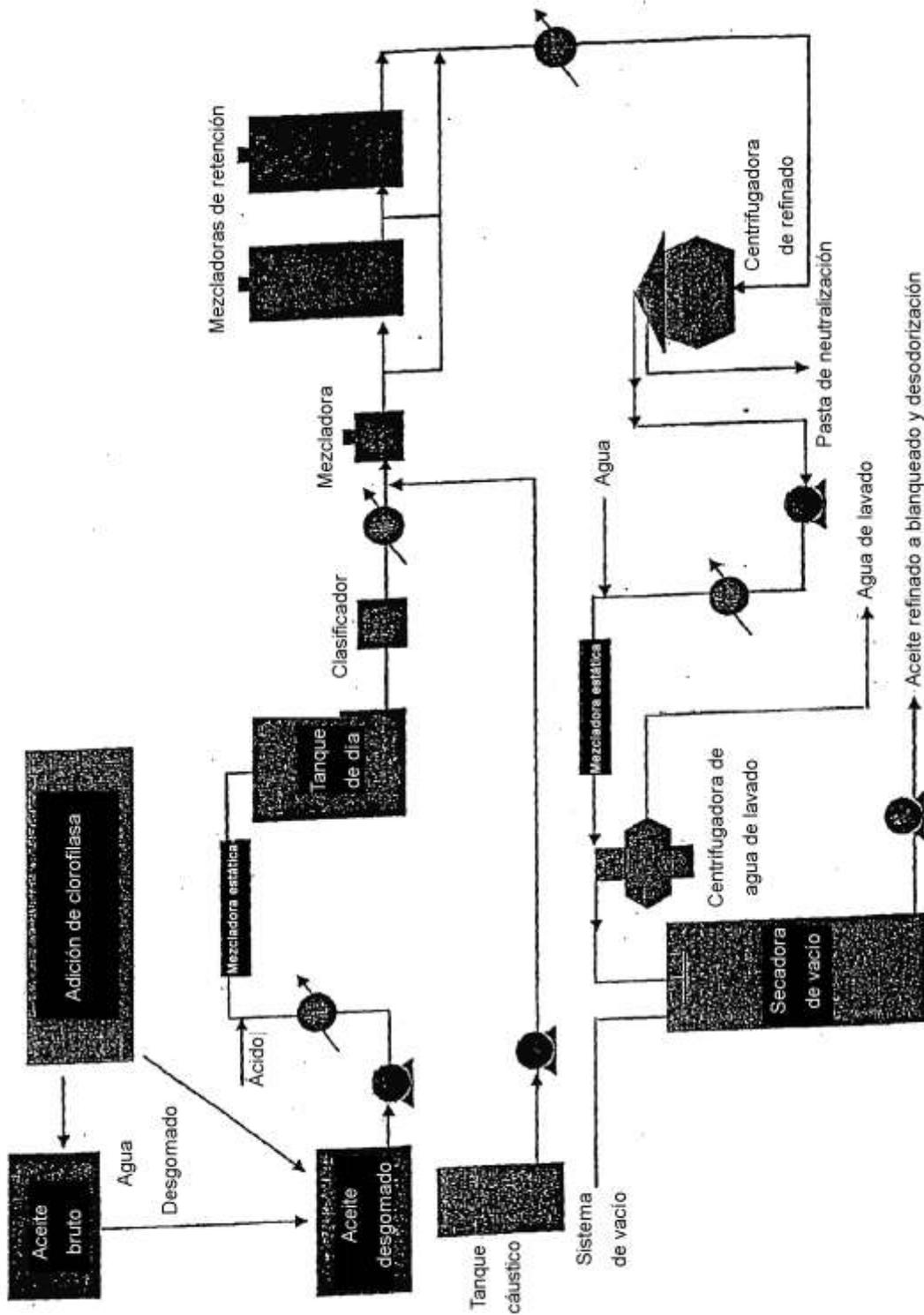


Figura 15

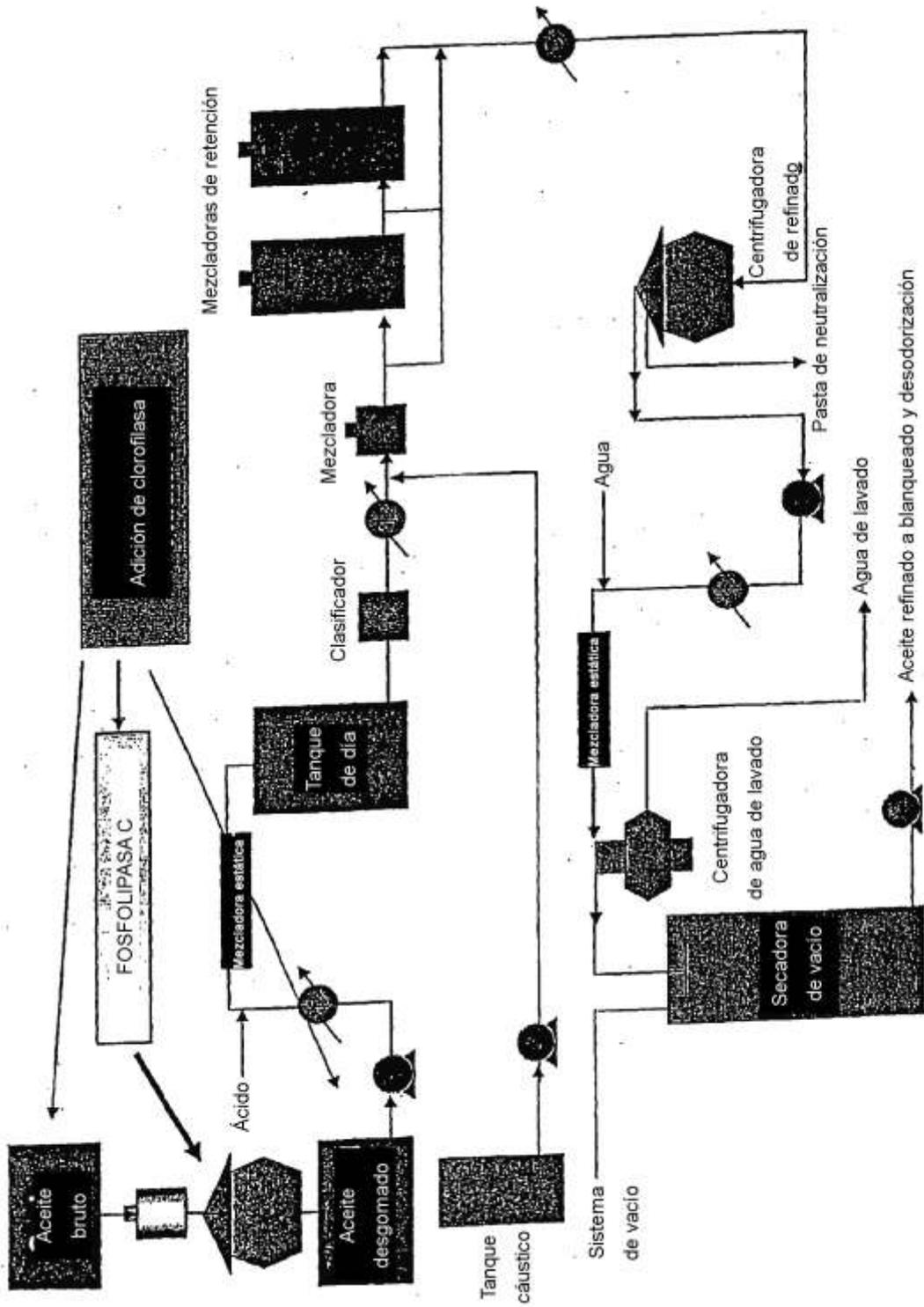


Figura 16