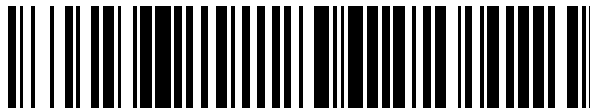


19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 401 795**

51 Int. Cl.:

C12N 9/24 (2006.01)

A23K 1/00 (2006.01)

A23K 1/165 (2006.01)

A23K 1/18 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **02.07.2004 E 04777548 (1)**

97 Fecha y número de publicación de la concesión europea: **20.02.2013 EP 1641910**

54 Título: **Glucanasas, ácidos nucleicos que las codifican y métodos para eleborarlas y utilizarlas**

30 Prioridad:

02.07.2003 US 484725 P

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

24.04.2013

73 Titular/es:

**SYNGENTA PARTICIPATIONS AG (33.3%)
Schwarzwaldallee 215
4058 Basel, CH;
VERENIUM CORPORATION (33.3%) y
BP CORPORATION NORTH AMERICA INC.
(33.3%)**

72 Inventor/es:

**STEER, BRIAN;
CALLEN, WALTER;
HEALEY, SHAUN y
PULLIAM, DERRICK,SYNGENTA
BIOTECHNOLOGY, INC.**

74 Agente/Representante:

UNGRÍA LÓPEZ, Javier

ES 2 401 795 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

Glucanasas, Ácidos Nucleicos que las codifican y métodos para elaborarlas y utilizarlas

5 Referencia Cruzada a la Lista de Secuencias Presentada en un Disco Compacto

Esta solicitud incluye un disco compacto (presentado por cuadruplicado) que contiene una lista de secuencias. El contenido completo de la lista de secuencias se incorpora a la presente en su totalidad como referencia. Esta lista de secuencias se identifica en el disco compacto como sigue.

10

Nombre del Archivo	Fecha de Creación	Tamaño (bytes)
Listing.txt	2 de Julio de 2004	1.562.624

Campo de la Invención

15

Esta invención se refiere generalmente a enzimas, polinucleótidos que codifican las enzimas, al uso de tales polinucleótidos y polipéptidos y más específicamente a polipéptidos (p. ej., enzimas, anticuerpos) que tienen una actividad glucanasa, p. ej., endoglucanasa, p. ej. que cataliza la hidrólisis de enlaces endo- β -1,4- y/o β -1,3-glucanasa internos. En un aspecto la actividad endoglucanasa (p. ej., actividad endo-1,4-beta-D-glucan 4-glucano hidrolasa) que comprende la hidrólisis de los enlaces beta-D-glicosídicos 1,4- y/o β -1,3- en la celulosa, los derivados de celulosa (p. ej., carboximetilcelulosa e hidroxietilcelulosa), liquenina, enlaces beta-1,4- en glucanos beta-1,3-

20

mixtos, tales como beta-D-glucanos y/o xiloglucanos de cereales y otro material vegetal u orgánico que contenga porciones celulósicas.

Antecedentes

25

Las endoglucanasas (p. ej., endo-beta-1,4-glucanasas, EC 3.2.1.4; endo-beta-1,3(1)-glucanasas, EC 3.2.1.6; endo-beta-1,3-glucanasas, EC 3.2.1.39) hidrolizan enlaces β -1,4- y/o β -1,3-glucosídicos en celulosa y glucano para producir glucosa y oligómeros de glucosa de peso molecular más bajo. Los glucanos son polisacáridos formados a partir de D-glucopiranososa unida por 1,4- β - y/o 1,3-glicósido. Las endoglucanasas tienen un valor comercial considerable, utilizándose en la industria alimentaria, para panificación y tratamiento de frutas y vegetales, descomposición de residuos agrícolas, en la fabricación de piensos para animales (p. ej., piensos para pollos), en la producción de pasta de papel y papel, en manufactura textil y productos de limpieza domésticos e industriales. Las endoglucanasas son producidas por hongos y bacterias.

30

35

Los beta-glucanos son los polisacáridos principales de los cereales aparte del almidón. El contenido en glucanos puede variar significativamente dependiendo de la variedad y de las condiciones de crecimiento. Las propiedades fisicoquímicas de este polisacárido son tales que éste da lugar a soluciones viscosas o incluso geles en condiciones oxidativas. Además, los glucanos tienen una alta capacidad de unirse a agua. Todas estas características presentan problemas para varias industrias incluyendo la elaboración de cerveza, panificación, nutrición animal. En las aplicaciones para la elaboración de cerveza, la presencia de glucano da como resultado problemas en la filtrabilidad de la malta y la formación de turbidez. En las aplicaciones para panificación (especialmente para galletas y galletas saladas), los glucanos pueden crear masas pegajosas que son difíciles de trabajar y reducen el tamaño de la galleta. Además, este carbohidrato está implicado en la rehidratación rápida del producto horneado dando como resultado la pérdida del carácter crujiente y la reducción de su vida útil. Para aplicaciones de alimentación de animales monogástricos con dietas de cereales, el beta-glucano es el factor que más contribuye a la viscosidad del contenido de los intestinos y por lo tanto afecta adversamente a la digestibilidad del alimento y al ritmo de crecimiento del animal. Para animales rumiantes, estos beta-glucanos representan componentes sustanciales de la ingesta de fibra y la digestión más completa de los glucanos facilitaría eficiencias de conversión de alimentos superiores. Para las endo-glucanasas para alimentación animal es deseable que sean activas en el estómago del animal.

50

Las endoglucanasas también son importantes para la digestión de la celulosa, un glucano unido por beta-1,4- encontrado en todo el material vegetal. La celulosa es el polisacárido más abundante en la naturaleza. Las enzimas comerciales que digieren la celulosa tienen utilidad en la industria de la pasta de papel y el papel, en la manufactura textil y en productos de limpieza domésticos e industriales.

55

Las publicaciones comentadas en la presente memoria se proporcionan únicamente para su descripción antes de la fecha de presentación de la presente solicitud. Nada en la presente memoria debe ser considerado como una admisión de que la invención no está titulada antes de tal descripción en virtud de la invención anterior.

60

Compendio de la Invención

Las características de la invención se presentan en las reivindicaciones adjuntas.

5 La invención proporciona ácidos nucleicos aislados, sintéticos o recombinantes que comprenden una secuencia de ácido nucleico que tiene una identidad de secuencia de al menos aproximadamente 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, o más, o una identidad de secuencia completa (100%) con el SEQ ID NO: 37, donde el ácido nucleico codifica al menos un polipéptido que tiene (i) una actividad glucanasa, p.ej., endoglucanasa, (ii) una actividad inmunogénica y puede generar un anticuerpo que se une específicamente a un polipéptido que tiene una secuencia expuesta en el SEQ ID NO: 38. Las identidades de secuencia se determinan por medio del análisis con un algoritmo de comparación de secuencias o por medio de inspección visual.

Los ácidos nucleicos ilustrativos de la invención también incluyen ácidos nucleicos aislados, sintéticos o recombinantes que codifican un polipéptido de la invención que tiene una secuencia expuesta en el SEQ ID NO: 38.

15 La invención proporciona ácidos nucleicos aislados, sintéticos o recombinantes que comprenden una secuencia de ácido nucleico que hibrida en condiciones restrictivas con el complemento de un ácido nucleico que comprende el SEQ ID NO: 37, donde el ácido nucleico codifica un polipéptido que tiene una actividad glucanasa, y las condiciones restrictivas incluyen una etapa de lavado que comprende un lavado en 0,2X SSC a una temperatura de 65°C durante 15 minutos.

20 Asimismo se proporcionan ácidos nucleicos aislados, sintéticos o recombinantes que comprenden un ácido nucleico complementario a cualquiera de los ácidos nucleicos anteriormente mencionados de la invención.

25 En un aspecto, el polipéptido tiene una actividad glucanasa, p. ej., endoglucanasa, que cataliza la hidrólisis de enlaces endo- β -1,4- y/o 1,3-glucanasa internos.

30 En un aspecto, el algoritmo de comparación de secuencias es un algoritmo BLAST versión 2.2.2 donde una configuración de filtrado se ajusta a blastall -p blastp -d "nr pataa" -F F, y todas las demás opciones se ajustan por defecto.

Otro aspecto de la invención es una sonda para identificar un ácido nucleico que codifica un polipéptido con una actividad glucanasa, donde la sonda puede comprender al menos 10 bases consecutivas, o al menos de 10 a 50, de 20 a 60, de 30 a 70, de 40 a 80, de 60 a 100, o de 50 a 150 bases consecutivas, de una secuencia que comprende (a) el SEQ ID NO: 37, o (b) una secuencia de la invención, donde la sonda identifica el ácido nucleico mediante unión o hibridación. La sonda puede comprender un oligonucleótido que comprende al menos aproximadamente de 10 a 50, aproximadamente 20 a 60, aproximadamente 30 a 70, aproximadamente 40 a 80, aproximadamente 60 a 100, bases consecutivas de una secuencia que comprende una secuencia de la invención, o fragmentos o subsecuencias de la misma.

40 En un aspecto, la actividad glucanasa de la invención comprende una actividad endoglucanasa, p. ej., una actividad endo-1,4-beta-D-glucan 4-glucano hidrolasa). En un aspecto, la actividad endoglucanasa comprende catalizar la hidrólisis de los enlaces 1,4-beta-D-glicosídicos. En un aspecto, la actividad glucanasa, p. ej., endoglucanasa, comprende una actividad endo-1,4- y/o 1,3-beta-endoglucanasa o una actividad endo- β -1,4-glucanasa. En un aspecto, la actividad glucanasa (p. ej., actividad endo-1,4-beta-D-glucan 4-glucano hidrolasa) comprende la hidrólisis de los enlaces 1,4-beta-D-glicosídicos en la celulosa, los derivados de celulosa (p. ej., carboximetilcelulosa e hidroxietilcelulosa), liquenina, enlaces beta-1,4- en glucanos beta-1,3-mixtos, tales como beta-D-glucanos de cereales y otro material vegetal que contenga porciones celulósicas.

50 En un aspecto, la actividad glucanasa comprende hidrolizar un glucano u otro polisacárido para producir un polisacárido u oligómero de peso molecular más bajo. En un aspecto, el glucano comprende un beta-glucano, tal como un beta-glucano soluble en agua. El beta-glucano soluble en agua puede comprender una masa o un producto de panificación.

55 En un aspecto, la actividad glucanasa, p. ej., endoglucanasa, comprende catalizar la hidrólisis de glucanos en una bebida o un pienso (p. ej., un pienso para animales, tal como un pienso para pollos) o un producto alimenticio. La bebida, pienso o producto alimenticio pueden comprender un pienso para animales a base de cereales, una malta o una cerveza, una fruta o una hortaliza. En un aspecto, la invención proporciona un alimento, pienso (p. ej., un pienso para animales, tal como un pienso para pollos), un líquido, p. ej., una bebida (tal como un zumo de fruta o una cerveza) o un precursor de bebida (p. ej., una malta), que comprende un polipéptido de la invención. El alimento puede ser una masa o un producto de panificación. La bebida o el precursor de bebida pueden ser un zumo de fruta, una cerveza o una malta. En un aspecto, la invención proporciona un método para la clarificación de un líquido, p. ej., un zumo, tal como un zumo de fruta, o una cerveza, mediante el tratamiento del líquido con una enzima de la invención.

Los métodos para el acondicionamiento de la masa comprenden poner en contacto una masa o un producto de panificación con al menos un polipéptido de la invención en condiciones suficientes para acondicionar la masa. Los métodos para la producción de bebidas comprenden la administración de al menos un polipéptido de la invención a una bebida o un precursor de bebida en condiciones suficientes para disminuir la viscosidad de la bebida.

5 En un aspecto, la actividad glucanasa, p. ej., endoglucanasa, comprende catalizar la hidrólisis de glucanos en una célula, p. ej., una célula vegetal o una célula microbiana.

10 El ácido nucleico aislado, sintético o recombinante puede codificar un polipéptido que tiene una actividad glucanasa, p. ej., endoglucanasa que es termoestable. El polipéptido puede conservar una actividad glucanasa u otra actividad en condiciones que comprenden un intervalo de temperatura de entre aproximadamente 37°C y aproximadamente 95°C; entre aproximadamente 55°C y aproximadamente 85°C, entre aproximadamente 70°C y aproximadamente 95°C, o, entre aproximadamente 90°C y aproximadamente 95°C.

15 El ácido nucleico aislado, sintético o recombinante puede codificar un polipéptido que tiene una actividad glucanasa, p. ej., endoglucanasa que es termotolerante. El polipéptido puede conservar una actividad glucanasa u otra después de su exposición a una temperatura en el intervalo de más de 37°C a aproximadamente 95°C o cualquiera en el intervalo de más de 55°C a aproximadamente 85°C. El polipéptido puede conservar una actividad glucanasa u otra después de su exposición a una temperatura en el intervalo entre aproximadamente 1°C y aproximadamente 5°C, entre aproximadamente 5°C y aproximadamente 15°C, entre aproximadamente 15°C y aproximadamente 25°C, entre aproximadamente 25°C y aproximadamente 37°C, entre aproximadamente 37°C y aproximadamente 95°C, entre aproximadamente 55°C y aproximadamente 85°C, entre aproximadamente 70°C y aproximadamente 75°C, o entre aproximadamente 90°C y aproximadamente 95°C, o más. En un aspecto, el polipéptido conserva una actividad glucanasa u otra después de su exposición a una temperatura en el intervalo entre aproximadamente 90°C y aproximadamente 95°C a pH 4,5.

20 Un par de cebadores de amplificación, para amplificar un ácido nucleico que codifica un polipéptido que tiene una actividad glucanasa, puede ser capaz de amplificar un ácido nucleico que comprende una secuencia de la invención, o fragmentos o subsecuencias del mismo. Uno o cada miembro del par de secuencias de cebadores de amplificación puede comprender un oligonucleótido que comprende al menos aproximadamente 10 a 50 bases consecutivas de la secuencia, o aproximadamente 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30 o más bases consecutivas de la secuencia.

35 El par de cebadores puede comprender un primer miembro que tiene una secuencia como se expone mediante aproximadamente los primeros (los 5') 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30 o más residuos de un ácido nucleico de la invención, y un segundo miembro que tiene una secuencia como se expone mediante aproximadamente los primeros (los 5') 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30 o más residuos de la hebra complementaria del primer miembro.

40 La invención proporciona ácidos nucleicos que codifican glucanasas, p. ej., endoglucanasas generados mediante amplificación, p. ej., reacción en cadena de la polimerasa (PCR), utilizando un par de cebadores de amplificación de la invención. La invención proporciona glucanasas generadas mediante amplificación, p. ej. reacción en cadena de la polimerasa (PCR), utilizando un par de cebadores de amplificación. La invención proporciona métodos para elaborar glucanasas mediante amplificación, p. ej., reacción en cadena de la polimerasa (PCR), utilizando un par de cebadores de amplificación. En un aspecto, el par de cebadores de amplificación amplifica un ácido nucleico de una biblioteca, p. ej., una biblioteca génica, tal como una biblioteca ambiental.

45 La invención proporciona métodos para amplificar un ácido nucleico que codifica un polipéptido que tiene una actividad glucanasa, p. ej. endoglucanasa que comprende la amplificación de un ácido nucleico molde con un par de secuencias cebadoras de amplificación capaces de amplificar una secuencia de ácido nucleico de la invención, o fragmentos o subsecuencias de la misma.

50 La invención proporciona casetes de expresión que comprenden un ácido nucleico de la invención o una subsecuencia del mismo. En un aspecto, el casete de expresión puede comprender el ácido nucleico que está unido operablemente a un promotor. El promotor puede ser un promotor viral, bacteriano, de mamífero o vegetal. En un aspecto, el promotor vegetal puede ser un promotor de patata, arroz, maíz, trigo, tabaco o cebada. El promotor puede ser un promotor constitutivo. El promotor constitutivo puede comprender CaMV35S. En otro aspecto, el promotor puede ser un promotor inducible. En un aspecto, el promotor puede ser un promotor específico de tejido o un promotor regulado ambientalmente o regulado evolutivamente. De este modo, el promotor puede ser, p. ej., un promotor específico de semilla, específico de hoja, específico de raíz, específico de tallo o inducido por abscisión. En un aspecto, el casete de expresión puede comprender adicionalmente un vector de expresión vegetal o de virus vegetal.

- 5 La invención proporciona vehículos de clonación que comprenden un casete de expresión (p. ej., un vector) de la invención o un ácido nucleico de la invención. El vehículo de clonación puede ser un vector viral, un plásmido, un fago, un fagémido, un cósmido, un fósido, un bacteriófago o un cromosoma artificial. El vector viral puede comprender un vector de adenovirus, un vector retroviral o un vector viral adenoasociado. El vehículo de clonación puede comprender un cromosoma artificial bacteriano (BAC), un plásmido, un vector derivado del bacteriófago P1-(PAC), un cromosoma artificial de levadura (YAC), o un cromosoma artificial de mamífero (MAC).
- 10 La invención proporciona una célula transformada que comprende un ácido nucleico de la invención o un casete de expresión (p. ej., un vector) de la invención, o un vehículo de clonación de la invención. En un aspecto, la célula transformada puede ser una célula bacteriana, una célula de mamífero, una célula fúngica, una célula de levadura, una célula de insecto o una célula vegetal. En un aspecto, la célula vegetal puede ser una célula de cereal, patata, trigo, arroz, maíz, tabaco o cebada.
- 15 La invención proporciona animales no humanos transgénicos que comprenden un ácido nucleico de la invención o un casete de expresión (p. ej., un vector) de la invención. En un aspecto, el animal es un ratón.
- 20 La invención proporciona plantas transgénicas que comprenden un ácido nucleico de la invención o un casete de expresión (p. ej., un vector) de la invención. La planta transgénica puede ser una planta de cereal, una planta de maíz, una planta de patata, una planta de tomate, una planta de trigo, una planta de semillas oleaginosas, una planta de colza, una planta de soja, una planta de arroz, una planta de cebada o una planta de tabaco.
- 25 La invención proporciona semillas transgénicas que comprenden un ácido nucleico de la invención o un casete de expresión (p. ej., un vector) de la invención. La semilla transgénica puede ser una planta de cereal, una semilla de maíz, un grano de trigo, una semilla oleaginosa, una semilla de colza, una semilla de soja, una semilla carnosa de palma, una semilla de girasol, una semilla de sésamo, un cacahuete o una semilla de planta de tabaco.
- 30 Un oligonucleótido antisentido puede comprender una secuencia de ácido nucleico complementaria a o susceptible de hibridación en condiciones restrictivas con un ácido nucleico de la invención. Los métodos para inhibir la traducción de un mensaje de glucanasa, p. ej. endoglucanasa 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 susceptible de hibridación en condiciones restrictivas con un ácido nucleico de la invención. En un aspecto, el oligonucleótido antisentido está entre 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 de longitud.
- 35 Los métodos para inhibir la traducción de un mensaje de glucanasa, p. ej. endoglucanasa 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 hibridar en condiciones restrictivas con un ácido nucleico de la invención. Las moléculas de ARN inhibidor (ARNi) de doble hebra pueden comprender una subsecuencia de una secuencia de la invención. En un aspecto, el ARNi tiene alrededor de 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25 o más nucleótidos dúplice de longitud. Los métodos para inhibir la expresión de una glucanasa, p. ej. endoglucanasa en una célula comprenden administrar a la célula o expresar en la célula un ARN inhibidor (ARNi) de doble hebra, donde el ARN comprende una subsecuencia de una secuencia de la invención.
- 40 Los métodos para inhibir la expresión de una glucanasa, p. ej. endoglucanasa en una célula comprenden administrar a la célula o expresar en la célula un ARN inhibidor (ARNi) de doble hebra, donde el ARN comprende una subsecuencia de una secuencia de la invención.
- 45 La invención proporciona un polipéptido aislado, sintético o recombinante que comprende una secuencia de aminoácidos que tiene al menos 90%, 91 %, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, o más, o completa (100%) de identidad de secuencia con el SEQ ID NO: 38 y las identidades de secuencia se determinan mediante análisis con un algoritmo de comparación de secuencias o mediante una inspección visual.
- 50 Las secuencias polipeptídicas o peptídicas ilustrativas de la invención incluyen una secuencia codificada por un ácido nucleico de la invención. Las secuencias polipeptídicas o peptídicas ilustrativas de la invención incluyen polipéptidos o péptidos unidos específicamente por un anticuerpo de la descripción.
- 55 En un aspecto, un polipéptido de la invención tiene al menos un actividad glucanasa, p. ej., endoglucanasa.
- 60 En un aspecto, la actividad endoglucanasa comprende actividad endo-1,4-beta-D-glucano 4-glucano hidrolasa. En un aspecto, la actividad endoglucanasa comprende hidrólisis catalizadora de conexiones 1,4-beta-D-glicosídicas o conexiones 1,3-beta-D-glicosídicas. En un aspecto, la actividad endoglucanasa comprende una actividad endo-1,4-beta-endoglucanasa o una actividad endo-β-1,4-glucanasa, actividad endo-1,3-beta-endoglucanasa o actividad endo-β-1,3-glucanasa. En un aspecto, la actividad glucanasa (p. ej., actividad endo-1,4 y/o 1,3-beta-D-glucano 4-glucano hidrolasa) comprende hidrólisis de conexiones 1,4-beta-D-glicosídicas en celulosa, derivados de celulosa (p. ej., carboximetilcelulosa e hidroxietilcelulosa) liquenina, enlaces beta-1,4- y/o 1,3- en beta-1,3 glucanos mixtos, tales como beta-D-glucanos o xiloglucanos de cereales y otros materiales vegetales que contienen partes celulósicas.

- 5 Un polipéptido o péptido aislado, sintético o recombinante puede incluir al menos 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95 o 100 o más bases consecutivas de una secuencia polipeptídica o peptídica de la invención, secuencias sustancialmente idénticas a la misma, y secuencias complementarias a la misma. El péptido puede ser, p. ej., un fragmento inmunogénico, un motivo (p. ej., un sitio de unión), una secuencia señal, una secuencia prepro o un dominio catalítico (CD) o sitio activo.
- 10 La invención proporciona ácidos nucleicos aislados, sintéticos o recombinantes que comprenden una secuencia que codifica un polipéptido que tiene una actividad glucanasa, p. ej., endoglucanasa y una secuencia señal, donde el ácido nucleico comprende una secuencia de la invención. La secuencia señal puede estar derivada de otra glucanasa, mananasa, o xilanasa o una no glucanasa, etc., es decir una enzima heteróloga. La invención proporciona ácidos nucleicos aislados, sintéticos o recombinantes que comprenden una secuencia que codifica un polipéptido que tiene una actividad glucanasa, p. ej., endoglucanasa, donde la secuencia no contiene una secuencia señal y el ácido nucleico comprende una secuencia de la invención.
- 15 En un aspecto, la actividad glucanasa, p. ej., endoglucanasa, comprende catalizar la hidrólisis de conexiones 1,4-beta-D-glicosídicas o conexiones 1,3-beta-D-glicosídicas. En un aspecto, la actividad endoglucanasa comprende una actividad endo-1,4-beta-endoglucanasa.
- 20 En un aspecto, la actividad endoglucanasa comprende hidrolizar un glucano para producir un polisacárido u oligómero de peso molecular más bajo. En un aspecto, el glucano comprende un beta-glucano, tal como un beta-glucano soluble en agua. El beta-glucano soluble en agua puede comprender una masa o un producto de panificación.
- 25 En un aspecto, la actividad glucanasa comprende hidrolizar polisacáridos que comprenden D-glucopiranosas conectadas por 1,4-β-glicósidos. En un aspecto, la actividad glucanasa comprende hidrolizar celulosa. En un aspecto, la actividad glucanasa comprende hidrolizar celulosa en una pasta de madera o papel o un producto de papel.
- 30 En un aspecto, la actividad glucanasa comprende catalizar la hidrólisis de un glucano u otro carbohidrato en un pienso (p. ej., un pienso para animales, tal como pienso para pollos) o un producto alimenticio. El pienso o producto alimenticio pueden comprender un pienso para animales a base de cereales, una malta o una cerveza, una fruta o una hortaliza.
- 35 En un aspecto, la actividad glucanasa comprende catalizar la hidrólisis de un glucano u otro carbohidrato en una célula, p. ej., una célula vegetal, una célula fúngica, o una célula microbiana (p. ej., bacteriana).
- 40 En un aspecto, la actividad glucanasa, p. ej., endoglucanasa es termoestable. El polipéptido puede conservar una actividad glucanasa en condiciones que comprenden un intervalo de temperatura de entre aproximadamente 1°C y aproximadamente 5°C, entre aproximadamente 5°C y aproximadamente 15°C, entre aproximadamente 15°C y aproximadamente 25°C, entre aproximadamente 25°C y aproximadamente 37°C, entre aproximadamente 37°C y aproximadamente 95°C, entre aproximadamente 55°C y aproximadamente 85°C, entre aproximadamente 70°C y aproximadamente 75°C, o entre aproximadamente 90°C y aproximadamente 95°C, o más. En otro aspecto, la actividad glucanasa, p. ej., endoglucanasa puede ser termotolerante. El polipéptido puede conservar una actividad glucanasa después de su exposición a una temperatura en el intervalo de más de 37°C a aproximadamente 95°C, o en el intervalo de más de 55°C a aproximadamente 85°C. En otro aspecto, el polipéptido puede conservar una actividad glucanasa después de su exposición a una temperatura en el intervalo de más de 90°C a aproximadamente 95°C a pH 4,5.
- 50 En un aspecto, el polipéptido aislado, sintético o recombinante puede comprender el polipéptido de la invención que carece de una secuencia señal. En un aspecto, el polipéptido aislado, sintético o recombinante puede comprender el polipéptido de la invención que comprende una secuencia señal heteróloga, tal como una secuencia señal de glucanasa, o mananasa, xilanasa heteróloga o una secuencia señal no de glucanasa, mananasa o xilanasa.
- 55 En un aspecto, la invención proporciona proteínas quiméricas que comprenden un primer dominio que comprende una secuencia señal de la invención y al menos un segundo dominio. La proteína puede ser una proteína de fusión. El segundo dominio puede comprender una enzima. La enzima puede ser una glucanasa, p. ej., una glucanasa, una mananasa, o una xilanasa.
- 60 La invención proporciona polipéptidos quiméricos que comprenden al menos un primer dominio que comprende un péptido señal (SP), una secuencia prepro y/o un dominio catalítico (CD) de la invención y al menos un segundo dominio que comprende un polipéptido o péptido heterólogo, donde el polipéptido o péptido heterólogo no está asociado naturalmente con el péptido señal (SP), la secuencia prepro y/o el dominio catalítico (CD). En un aspecto, el polipéptido o péptido heterólogo no es una glucanasa, una mananasa, o una xilanasa. El polipéptido o péptido

heterólogo puede ser amino terminal con respecto al extremo carboxi o con respecto a ambos extremos del péptido señal (SP), la secuencia prepro y/o el dominio catalítico (CD).

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

10 Las secuencias señal aisladas, sintéticas o recombinantes (p. ej., péptidos señal) pueden consistir en o comprender una secuencia mostrada en los residuos 1 a 14, 1 a 15, 1 a 16, 1 a 17, 1 a 18, 1 a 19, 1 a 20, 1 a 21, 1 a 22, 1 a 23, 1 a 24, 1 a 25, 1 a 26, 1 a 27, 1 a 28, 1 a 28, 1 a 30, 1 a 31, 1 a 32, 1 a 33, 1 a 34, 1 a 35, 1 a 36, 1 a 37, 1 a 38, 1 a 40, 1 a 41, 1 a 42, 1 a 43 o 1 a 44, de un polipéptido de la invención.

15 Las secuencias señal aisladas, sintéticas o recombinantes (p. ej., péptidos señal) pueden consistir en o comprender una secuencia mostrada en la Tabla 3, de más abajo.

20 La actividad glucanasa, p. ej. endoglucanasa puede comprender una actividad específica a alrededor de 37°C en el intervalo de alrededor de 1 a alrededor de 1200 unidades por miligramo de proteína, o, alrededor de 100 a alrededor de 1000 unidades por miligramo de proteína. La actividad glucanasa, p. ej. endoglucanasa puede comprender una actividad específica de alrededor de 100 a alrededor de 1000 unidades por miligramo de proteína, o, de alrededor de 500 a alrededor de 750 unidades por miligramo de proteína. Alternativamente, la actividad glucanasa puede comprender una actividad específica a 37°C en el intervalo de alrededor de 1 a alrededor de 750 unidades por miligramo de proteína, o, de alrededor de 500 a alrededor de 1200 unidades por miligramo de proteína. La actividad glucanasa puede comprender una actividad específica a 37°C en el intervalo de alrededor de 1 a alrededor de 500 unidades por miligramo de proteína, o, de alrededor de 750 a alrededor de 1000 unidades por miligramo de proteína. La actividad glucanasa puede comprender una actividad específica a 37°C en el intervalo de alrededor de 1 a alrededor de 250 unidades por miligramo de proteína. Alternativamente, la actividad glucanasa, p. ej., endoglucanasa puede comprender una actividad específica a 37°C en el intervalo de alrededor de 1 a alrededor de 100 unidades por miligramo de proteína. La termotolerancia puede comprender la retención de al menos la mitad de la actividad específica de la glucanasa a 37°C después de ser calentada a la temperatura elevada. Alternativamente, la termotolerancia puede comprender retención de actividad específica a 37°C en el intervalo de alrededor de 1 a alrededor de 1200 unidades por miligramo de proteína, o, de alrededor de 500 a alrededor de 1000 unidades por miligramo de proteína, después de ser calentada a la temperatura elevada. O, la termotolerancia puede comprender retención de actividad específica a 37°C en el intervalo de alrededor de 1 a alrededor de 500 unidades por miligramo de proteína después de ser calentada a la temperatura elevada.

35 La invención proporciona el polipéptido aislado o recombinante de la invención, donde el polipéptido comprende al menos un sitio de glicosilación. En un aspecto, la glicosilación puede ser una glicosilación ligada a N. En un aspecto, el polipéptido puede ser glicosilado después de ser expresado en *P. pastoris* o *S. pombe*.

40 El polipéptido puede conservar una actividad glucanasa, p. ej., endoglucanasa en condiciones que comprenden un pH de alrededor de 6,5, pH 6, pH 5,5, pH 5, pH 4,5 o pH 4. O, el polipéptido puede conservar una actividad glucanasa en condiciones que comprenden un pH de alrededor de 7, pH 7,5 pH 8,0, pH 8,5, pH 9, pH 9,5, pH 10, pH 10,5 o pH 11. El polipéptido puede conservar una actividad glucanasa después de exposición a condiciones que comprenden un pH de alrededor de 6,5, pH 6, pH 5,5, pH 5, pH 4,5 o pH 4. O, el polipéptido puede conservar una actividad glucanasa después de exposición a condiciones que comprenden un pH de alrededor de 7, pH 7,5 pH 8,0, pH 8,5, pH 9, pH 9,5, pH 10, pH 10,5 o pH 11.

45 La invención proporciona preparaciones de proteína que comprenden un polipéptido de la invención, donde la preparación de proteína comprende un líquido, un sólido o un gel.

50 Los heterodímeros pueden comprender un polipéptido de la invención y una segunda proteína o dominio. El segundo miembro del heterodímero puede ser una glicanasa diferente, una enzima diferente u otra proteína. El segundo dominio puede ser un polipéptido y el heterodímero puede ser una proteína de fusión. El segundo dominio puede ser un epítipo o una etiqueta. Los homodímeros pueden comprender un polipéptido de la invención.

55 Los polipéptidos inmovilizados que tienen una actividad glucanasa, p. ej., endoglucanasa, pueden comprender un polipéptido 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. El polipéptido puede ser inmovilizado sobre una célula, un metal, una resina, un polímero, una cerámica, un vidrio, un microelectrodo, una partícula graffica, una cuenta, un gel, una placa, una matriz o un tubo capilar.

Las matrices pueden comprender un ácido nucleico inmovilizado de la invención. Las matrices pueden comprender un anticuerpo de la descripción.

5 Los anticuerpos aislados, sintéticos o recombinantes se pueden unir específicamente a un polipéptido de la invención o a un polipéptido codificado por un ácido nucleico de la invención. El anticuerpo puede ser un anticuerpo monoclonal o policlonal. Los hibridomas pueden comprender tal anticuerpo, p. ej., un anticuerpo que se une específicamente a un polipéptido de la invención o a un polipéptido codificado por un ácido nucleico de la invención.

10 Un método para aislar o identificar un polipéptido que tiene actividad glucanasa, p. ej., endoglucanasa comprende las etapas de: (a) proporcionar un anticuerpo de la descripción; (b) proporcionar una muestra que comprende polipéptidos; y (c) poner en contacto la muestra de la etapa (b) con el anticuerpo de la etapa (a) en condiciones en las que el anticuerpo se puede unir específicamente al polipéptido, aislando o identificando de ese modo un polipéptido que tiene una actividad glucanasa.

15 Los métodos para elaborar un anticuerpo anti-glucanasa comprenden administrar a un animal no humano un ácido nucleico de la invención o un polipéptido de la invención o subsecuencias de los mismos en una cantidad suficiente para generar una respuesta inmunitaria humoral, elaborando de ese modo un anticuerpo anti-glucanasa. Los métodos para elaborar una respuesta inmunitaria humoral o celular anti-glucanasa comprenden administrar a un animal no humano un ácido nucleico de la invención o un polipéptido de la invención o subsecuencias de los mismos en una cantidad suficiente para generar una respuesta inmunitaria.

20 La invención proporciona métodos para producir un polipéptido recombinante que comprenden las etapas de: (a) proporcionar un ácido nucleico de la invención conectado operablemente a un promotor; y (b) expresar el ácido nucleico de la etapa (a) en condiciones que permitan la expresión del polipéptido, produciendo de ese modo un polipéptido recombinante. En un aspecto, el método puede comprender adicionalmente transformar una célula anfitriona con el ácido nucleico de la etapa (a) seguido de la expresión del ácido nucleico de la etapa (a), produciendo de ese modo un polipéptido recombinante en una célula transformada.

25 Los métodos para identificar un polipéptido que tiene una actividad glucanasa, p. ej., endoglucanasa comprenden las siguientes etapas: (a) proporcionar un polipéptido de la invención; o un polipéptido codificado por un ácido nucleico de la invención; (b) proporcionar un sustrato de glucanasa, p. ej., endoglucanasa; 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 un descenso en la cantidad de sustrato o un incremento en la cantidad de un producto de reacción, donde un descenso en la cantidad del sustrato o un incremento en la cantidad del producto de reacción detecta un polipéptido que tiene actividad glucanasa.

30 Los métodos para identificar un sustrato de glucanasa, p. ej., endoglucanasa comprenden las siguientes etapas: (a) proporcionar un polipéptido de la invención; o un polipéptido codificado por un ácido nucleico de la invención; (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 un descenso en la cantidad de sustrato o un incremento en la cantidad de producto de reacción, donde un descenso en la cantidad de sustrato o un incremento en la cantidad de un producto de reacción identifica el sustrato de ensayo como un sustrato de glucanasa.

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

40 Los métodos para identificar un modulador de una actividad glucanasa, p. ej., endoglucanasa comprenden las siguientes etapas: (a) proporcionar un polipéptido de la invención o un polipéptido codificado por un ácido nucleico de la invención; (b) proporcionar un compuesto de ensayo; (c) poner en contacto el polipéptido de la etapa (a) con el compuesto de ensayo de la etapa (b) y medir una actividad de la glucanasa, donde un cambio en la actividad glucanasa medida en presencia del compuesto de ensayo en comparación con la actividad en ausencia del compuesto de ensayo proporciona una determinación de que el compuesto de ensayo modula la actividad glucanasa. En un aspecto, se puede medir la actividad glucanasa proporcionando un sustrato de glucanasa y detectando un descenso en la cantidad del sustrato o un incremento en la cantidad de un producto de reacción, o, un incremento en la cantidad del sustrato o un descenso en la cantidad de un producto de reacción. Un descenso en la cantidad del sustrato o un incremento en la cantidad del producto de reacción con el compuesto de ensayo en comparación con la cantidad de sustrato o producto de reacción sin el compuesto de ensayo identifican el compuesto de ensayo como un activador de la actividad glucanasa. Un incremento en la cantidad del sustrato o un descenso en la cantidad del producto de reacción con el compuesto de ensayo en comparación con la cantidad de

sustrato o producto de reacción sin el compuesto de ensayo identifican el compuesto de ensayo como un inhibidor de la actividad glucanasa.

Los sistemas informáticos pueden comprender un procesador y un dispositivo para el almacenamiento de datos en donde dicho dispositivo de almacenamiento de datos tiene almacenadas una secuencia de polipéptido o una secuencia de ácido nucleico de la invención (p. ej., un polipéptido codificado por un ácido nucleico de la invención). El sistema informático puede comprender adicionalmente un algoritmo de comparación de secuencias y un dispositivo de almacenamiento de datos que tiene al menos una secuencia de referencia almacenada. El algoritmo de comparación de secuencias puede comprender un programa informático que indica polimorfismos. El sistema informático puede comprender adicionalmente un identificador que identifica una o más características de dicha secuencia. Los medios legibles por ordenador tienen almacenadas una secuencia de polipéptido o una secuencia de ácido nucleico de la invención. Los métodos para identificar una característica en una secuencia comprenden las etapas de (a) leer la secuencia utilizando un programa informático que identifica una o más características en una secuencia, en donde la secuencia comprende una secuencia de polipéptido o una secuencia de ácido nucleico de la invención; y (b) identificar una o más características en la secuencia con el programa informático. Los métodos para comparar una primera secuencia con una segunda secuencia comprenden las etapas de: (a) leer la primera secuencia y la segunda secuencia a través del uso de un programa informático que compara secuencias, en donde la primera secuencia comprende una secuencia de polipéptido 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 informático. La etapa para determinar las diferencias entre la primera secuencia y la segunda secuencia puede comprender adicionalmente la etapa de identificación de polimorfismos. El método puede comprender adicionalmente un identificador que identifica una o más características en una secuencia. El método puede comprender leer la primera secuencia utilizando un programa informático e identificar una o más características en la secuencia.

Los métodos para aislar o recuperar un ácido nucleico que codifica un polipéptido que tiene una actividad glucanasa a partir de una muestra ambiental comprende las etapas de: (a) proporcionar una sonda polinucleotídica que comprende un ácido nucleico de la invención; (b) aislar un ácido nucleico a partir de la muestra ambiental o tratar la muestra ambiental de manera que el ácido nucleico en la muestra sea accesible para la hibridación a la sonda polinucleotídica de la etapa (a); (c) combinar el ácido nucleico aislado, sintético de la muestra ambiental tratada de la etapa (b) con la sonda polinucleotídica de la etapa (a); y (d) aislar el ácido nucleico que hibrida específicamente con la sonda polinucleotídica de la etapa (a), aislando o recuperando de ese modo un ácido nucleico que codifica un polipéptido que tiene una actividad glucanasa a partir de una muestra ambiental. La muestra ambiental puede comprender una muestra de agua, una muestra de líquido, una muestra de suelo, una muestra de aire o una muestra biológica. En un aspecto, la muestra biológica se puede obtener de una célula bacteriana, una célula de protozoo, una célula de insecto, una célula de levadura, una célula vegetal, una célula fúngica o una célula de mamífero.

La invención proporciona métodos para generar una variante de un ácido nucleico que codifica un polipéptido que tiene una actividad glucanasa que comprenden 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 de sus combinaciones, para generar una variante del ácido nucleico molde. En un aspecto, el método puede comprender adicionalmente expresar el ácido nucleico variante para generar un polipéptido de xilanas variante. Las modificaciones, adiciones o deleciones se pueden introducir mediante un método que comprende PCR propensa a error, barajado, mutagénesis dirigida a oligonucleótidos, PCR de ensamblaje, mutagénesis por PCR sexual, mutagénesis *in vivo*, mutagénesis por inserción de casetes, mutagénesis de ensamblaje recursivo, mutagénesis de conjunto exponencial, mutagénesis de sitio específico, reensamblaje de genes, Mutagénesis por Saturación de Sitio de Genes (GSSM™), reensamblaje de ligación sintética (SLR) o una de sus combinaciones. Las modificaciones, adiciones o supresiones se introducen mediante un procedimiento que comprende recombinación, recombinación de secuencia recursiva, mutagénesis de ADN modificada por fosfotioato, mutagénesis de molde que contiene uracilo, mutagénesis de dúplex discontinuo, mutagénesis por reparación de apareamiento erróneo, mutagénesis de cepa anfitriona de reparación defectuosa, mutagénesis química, mutagénesis radiogénica, mutagénesis por supresión, mutagénesis por restricción-selección, mutagénesis por restricción-purificación, síntesis artificial de genes, mutagénesis de ensamblaje, creación de multímeros de ácidos nucleicos quiméricos y una de sus combinaciones.

En un aspecto, el método puede ser repetido iterativamente hasta que se produzca una glucanasa que tenga una actividad alterada o diferente o una estabilidad alterada o diferente de la de un polipéptido codificado por el ácido nucleico molde. En un aspecto, el polipéptido de xilanas variante es termotolerante, y conserva cierta actividad después de ser expuesto a una temperatura elevada. En otro aspecto, el polipéptido de glucanasa variante tiene un aumento de glicosilación en comparación con la glucanasa codificada por un ácido nucleico molde. Alternativamente, el polipéptido de glucanasa variante tiene actividad glucanasa a alta temperatura, donde la glucanasa codificada por el ácido nucleico molde no es activa a la alta temperatura. En un aspecto, el método puede ser repetido iterativamente hasta que se produzca una glucanasa que codifique una secuencia que tenga un uso de codón alterado a partir de la del ácido nucleico molde. En otro aspecto, el método puede ser repetido iterativamente

hasta que se produzca un gen de glucanasa que tenga un nivel mayor o menor de expresión del mensaje o estabilidad que el del ácido nucleico molde.

5 La invención proporciona métodos para modificar codones en un ácido nucleico que codifica un polipéptido que tiene una actividad glucanasa para incrementar su expresión en una célula anfitriona, comprendiendo el método las siguientes etapas: (a) proporcionar un ácido nucleico de la invención que codifica un polipéptido que tiene una actividad glucanasa; y, (b) identificar un codón no preferido o menos preferido en el ácido nucleico de la etapa (a) y reemplazarlo por un codón preferido o utilizado neutramente que codifica el mismo aminoácido que el codón reemplazado, donde un codón preferido es un codón representado en exceso en secuencias codificantes en genes en la célula anfitriona y un codón no preferido o menos preferido es un codón infrarrepresentado en secuencias codificantes en genes en la célula anfitriona, modificando de ese modo el ácido nucleico para incrementar su expresión en una célula anfitriona.

15 La invención proporciona métodos para modificar codones en un ácido nucleico que codifica un polipéptido que tiene una actividad glucanasa; comprendiendo el método las siguientes etapas: (a) proporcionar un ácido nucleico de la invención; y, (b) identificar un codón en el ácido nucleico de la etapa (a) y reemplazarlo por un codón diferente que codifica el mismo aminoácido que el codón reemplazado, modificando de ese modo codones en un ácido nucleico que codifica una glucanasa.

20 La invención proporciona métodos para modificar codones en un ácido nucleico que codifica un polipéptido que tiene una actividad glucanasa para incrementar su expresión en una célula anfitriona, comprendiendo el método las siguientes etapas: (a) proporcionar un ácido nucleico de la invención que codifica un polipéptido de glucanasa; y, (b) identificar un codón no preferido o menos preferido en el ácido nucleico de la etapa (a) y reemplazarlo por un codón preferido o utilizado neutramente que codifica el mismo aminoácido que el codón reemplazado, donde un codón preferido es un codón representado en exceso en secuencias codificantes en genes en la célula anfitriona y un codón no preferido o menos preferido es un codón infrarrepresentado en secuencias codificantes en genes en la célula anfitriona, modificando de ese modo el ácido nucleico para incrementar su expresión en una célula anfitriona.

30 La invención proporciona métodos para modificar un codón en un ácido nucleico que codifica un polipéptido que tiene una actividad glucanasa para disminuir su expresión en una célula anfitriona, comprendiendo el método las siguientes etapas: (a) proporcionar un ácido nucleico de la invención; y (b) identificar al menos un codón preferido en el ácido nucleico de la etapa (a) y reemplazarlo por un codón no preferido o menos preferido que codifica el mismo aminoácido que el codón reemplazado, donde un codón preferido es un codón representado en exceso en secuencias codificantes en genes en una célula anfitriona y un codón no preferido o menos preferido es un codón infrarrepresentado en secuencias codificantes en genes en la célula anfitriona, modificando de ese modo el ácido nucleico para disminuir su expresión en una célula anfitriona. En un aspecto, la célula anfitriona puede ser una célula bacteriana, una célula fúngica, una célula de insecto, una célula de levadura, una célula vegetal o una célula de mamífero.

40 Los métodos pueden producir una biblioteca de ácidos nucleicos que codifican una pluralidad de sitios activos de glucanasa modificados (dominios catalíticos (CD)) o sitios de unión al sustrato, donde los sitios activos modificados o los sitios de unión al sustrato se obtienen de un primer ácido nucleico que comprende una secuencia que codifica un primer sitio activo o un primer sitio de unión al sustrato, comprendiendo el método las siguientes etapas: (a) proporcionar un primer ácido nucleico que codifica un primer sitio activo o un primer sitio de unión al sustrato, donde la primera secuencia de ácido nucleico comprende una secuencia que hibrida en condiciones restrictivas con un ácido nucleico de la invención, y el ácido nucleico codifica un sitio activo de glucanasa o un sitio de unión al sustrato de glucanasa; (b) proporcionar un conjunto de oligonucleótidos mutagénicos que codifican variantes de aminoácidos de origen natural en una pluralidad de codones elegidos como diana en el primer ácido nucleico; y, (c) utilizar el conjunto de oligonucleótidos mutagénicos para generar un conjunto de ácidos nucleicos variantes que codifican el sitio activo o codifican el sitio de unión al sustrato que codifican una variedad de variaciones de aminoácido en cada codón de aminoácido que fue mutagenizado, produciendo de ese modo una biblioteca de ácidos nucleicos que codifican una pluralidad de sitios activos de glucanasa modificados o sitios de unión al sustrato. En un aspecto, el método comprende mutagenizar el primer ácido nucleico de la etapa (a) por medio de un método que comprende un sistema de evolución directa optimizada, Mutagénesis por Saturación del Sitio de Genes (GSSM™), reensamblaje de ligación sintética (SLR), PCR propensa a error, barajado, mutagénesis dirigida a oligonucleótidos, PCR de ensamblaje, mutagénesis por PCR sexual, mutagénesis in vivo, mutagénesis por inserción de casetes, mutagénesis de ensamblaje recursivo, mutagénesis de conjunto exponencial, mutagénesis de sitio específico, reensamblaje de genes, reensamblaje de ligación sintética (SLR) y una de sus combinaciones. En otro aspecto, el método comprende mutagenizar el primer ácido nucleico de la etapa (a) o variantes por medio de un método que comprende recombinación, recombinación de secuencia recursiva, mutagénesis de ADN modificada por fosfotioato, mutagénesis de molde que contiene uracilo, mutagénesis de dúplex discontinuo, mutagénesis por reparación de apareamiento erróneo, mutagénesis de cepa anfitriona de reparación defectuosa, mutagénesis química, mutagénesis radiogénica, mutagénesis por supresión, mutagénesis por restricción-selección, mutagénesis por

restricción-purificación, síntesis artificial de genes, mutagénesis de ensamblaje, creación de multímeros de ácidos nucleicos quiméricos y una de sus combinaciones.

5 Los métodos para elaborar una molécula pequeña comprenden las siguientes etapas: (a) proporcionar una pluralidad de enzimas biosintéticas capaces de sintetizar o modificar una molécula pequeña, donde una de las enzimas comprende una enzima glucanasa codificada por un ácido nucleico de la invención; (b) proporcionar un sustrato para al menos una de las enzimas de la etapa (a); y (c) hacer reaccionar el sustrato de la etapa (b) con las enzimas en condiciones que faciliten una pluralidad de reacciones biocatalíticas para generar una molécula pequeña mediante una serie de reacciones biocatalíticas. La invención proporciona métodos para modificar una molécula pequeña que
 10 comprende las siguientes etapas: (a) proporcionar una enzima glucanasa, donde la enzima comprende un polipéptido de la invención, o, un polipéptido codificado por un ácido nucleico de la invención, o una subsecuencia del mismo; (b) proporcionar una molécula pequeña; y (c) hacer reaccionar la enzima de la etapa (a) con la molécula pequeña de la etapa (b) en condiciones que faciliten una reacción enzimática catalizada por la enzima glucanasa, modificando de ese modo una molécula pequeña por una reacción enzimática de glucanasa. En un aspecto, el método puede comprender una pluralidad de sustratos de molécula pequeña para la enzima de la etapa (a), generándose de ese modo una biblioteca de moléculas pequeñas modificadas producidas por al menos una reacción enzimática catalizada por la enzima glucanasa. En un aspecto, el método puede comprender una pluralidad de enzimas adicionales en condiciones que faciliten una pluralidad de reacciones biocatalíticas por las enzimas para formar una biblioteca de moléculas pequeñas modificadas producidas por la pluralidad de reacciones enzimáticas.
 20 En otro aspecto, el método puede comprender adicionalmente la etapa de someter a ensayo la biblioteca para determinar si una molécula pequeña modificada concreta que muestra una actividad deseada está presente en la biblioteca. La etapa de someter a ensayo la biblioteca puede comprender adicionalmente las etapas de eliminación sistemática de todas las reacciones biocatalíticas menos una utilizadas para producir una porción de la pluralidad de las moléculas pequeñas modificadas en la biblioteca sometiendo a ensayo la porción de la molécula pequeña modificada por la presencia o ausencia de la molécula pequeña modificada concreta con una actividad deseada, e
 25 identificación de al menos una reacción biocatalítica específica que produce la molécula pequeña modificada concreta de actividad deseada.

30 Los métodos para determinar un fragmento funcional de una enzima glucanasa comprenden las etapas de: (a) proporcionar una enzima glucanasa, donde la enzima comprende un polipéptido de la invención, o un polipéptido codificado por un ácido nucleico de la invención, o una subsecuencia del mismo; y (b) suprimir una pluralidad de residuos de aminoácidos de la secuencia de la etapa (a) y someter a ensayo la subsecuencia restante para determinar una actividad glucanasa, determinando de ese modo un fragmento funcional de una enzima glucanasa. En un aspecto, la actividad glucanasa se mide proporcionando un sustrato de glucanasa y detectando un descenso
 35 en la cantidad del sustrato o un incremento en la cantidad de un producto de reacción.

Los métodos para construir células completas de fenotipos nuevos o modificados utilizando análisis de flujo metabólico en tiempo real, comprenden las siguientes etapas: (a) elaborar una célula modificada modificando la composición genética de una célula, donde la composición genética es modificada mediante la adición a la célula de
 40 un ácido nucleico de la invención; (b) cultivar la célula modificada para generar una pluralidad de células modificadas; (c) medir al menos un parámetro metabólico de la célula controlando el cultivo celular de la etapa (b) en tiempo real; y, (d) analizar los datos de la etapa (c) para determinar si el parámetro medido difiere de una medición comparable en una célula no modificada en condiciones similares, identificando de ese modo un fenotipo construido en la célula utilizando análisis de flujo metabólico en tiempo real. En un aspecto, la composición genética de la célula se puede modificar por medio de un método que comprende la delección de una secuencia o modificación de una secuencia en la célula, o, la desactivación de la expresión de un gen. En un aspecto, el método puede comprender adicionalmente seleccionar una célula que comprende un fenotipo construido recientemente. En otro aspecto, el método puede comprender cultivar la célula seleccionada, generándose de ese modo un nueva cepa celular que comprende un fenotipo construido recientemente.
 50

Los métodos para aumentar la termotolerancia o la termoestabilidad de un polipéptido de glucanasa, comprenden la glicosilación de un polipéptido de glucanasa, donde el polipéptido comprende al menos treinta aminoácidos contiguos de un polipéptido de la invención; o un polipéptido codificado por una secuencia de ácido nucleico de la invención, incrementando de ese modo la termotolerancia o la termoestabilidad del polipéptido de glucanasa. En un
 55 aspecto, la actividad específica de la glucanasa puede ser termoestable o termotolerante a una temperatura en el intervalo de más de alrededor de 37°C a alrededor de 95°C.

Los métodos para expresar en exceso un polipéptido de glucanasa recombinante en una célula comprenden expresar un vector que comprende un ácido nucleico que comprende un ácido nucleico de la invención o una
 60 secuencia de ácido nucleico de la invención, donde las identidades de secuencia se determinan mediante análisis con un algoritmo de comparación de secuencia o mediante inspección visual, donde la expresión en exceso se efectúa mediante el uso de un promotor de alta actividad, un vector dicistrónico o mediante amplificación génica del vector.

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

Los métodos para expresar una secuencia de ácido nucleico heteróloga en una célula vegetal comprenden las siguientes etapas: (a) transformar la célula vegetal con una secuencia de ácido nucleico heteróloga conectada operablemente a un promotor, donde la secuencia de ácido nucleico heteróloga comprende un ácido nucleico de la invención; (b) cultivar la planta en condiciones en las que la secuencia de ácido nucleico heteróloga se exprese en la célula vegetal. Los métodos para expresar una secuencia de ácido nucleico heteróloga en una célula vegetal comprenden las siguientes etapas: (a) transformar la célula vegetal con una secuencia de ácido nucleico heteróloga conectada operablemente a un promotor, donde la secuencia de ácido nucleico heteróloga comprende una secuencia de la invención; (b) cultivar la planta en condiciones en las que la secuencia de ácido nucleico heteróloga se exprese en la célula vegetal.

La invención proporciona métodos para hidrolizar, romper o desorganizar una composición que comprende glucano que comprenden las siguientes etapas: (a) proporcionar un polipéptido de la invención que tiene una actividad glucanasa, o un polipéptido codificado por un ácido nucleico de la invención; (b) proporcionar una composición que comprende glucano; y (c) poner en contacto el polipéptido de la etapa (a) con la composición de la etapa (b) en condiciones en las que la glucanasa hidroliza, rompe o desorganiza la composición que comprende glucano. En un aspecto, la composición comprende una célula vegetal, una célula bacteriana, una célula de levadura, una célula de insecto, o una célula animal. De este modo, la composición puede comprender cualquier planta o parte de planta, cualquier alimento o pienso que contenga glucano (p. ej., un pienso para animales, tal como un pienso para pollos), un producto de desecho y similares. La invención proporciona métodos para licuar o separar una composición que comprende glucano que comprenden las siguientes etapas: (a) proporcionar un polipéptido de la invención que tiene una actividad glucanasa, o un polipéptido codificado por un ácido nucleico de la invención; (b) proporcionar una composición que comprende un glucano; y (c) poner en contacto el polipéptido de la etapa (a) con la composición de la etapa (b) en condiciones en las que la glucanasa separa, reblandece o licua la composición que comprende glucano.

La invención proporciona composiciones detergentes que comprenden un polipéptido de la invención, o un polipéptido codificado por un ácido nucleico de la invención, donde el polipéptido tiene una actividad glucanasa, p. ej. endoglucanasa. La glucanasa puede ser una glucanasa no tensioactiva o una glucanasa tensioactiva. La glucanasa se puede formular en una composición líquida no acuosa, un sólido moldeado, una forma granular, una forma particulada, una tableta comprimida, una forma de gel, una pasta o una forma en suspensión. La invención proporciona métodos para lavar un objeto que comprenden las siguientes etapas: (a) proporcionar una composición que comprende un polipéptido de la invención que tiene una actividad glucanasa, o un polipéptido codificado por un ácido nucleico de la invención; (b) proporcionar un objeto; y (c) poner en contacto el polipéptido de la etapa (a) y el objeto de la etapa (b) en condiciones en las que la composición pueda lavar el objeto.

La invención proporciona productos textiles o tejidos, que incluyen, p. ej., fibras, que comprenden un polipéptido de la invención, o un polipéptido codificado por un ácido nucleico de la invención. En un aspecto, los productos textiles o tejidos comprenden fibras que contienen glucano. La invención proporciona métodos para tratar un producto textil o tejido (p. ej., eliminando una mancha de una composición) que comprenden las siguientes etapas: (a) proporcionar una composición que comprende un polipéptido de la invención que tiene una actividad glucanasa, o un polipéptido codificado por un ácido nucleico de la invención; (b) proporcionar un producto textil o tejido que comprende un glucano; y (c) poner en contacto el polipéptido de la etapa (a) y la composición de la etapa (b) en condiciones en las que la glucanasa pueda tratar el producto textil o tejido (p. ej., eliminación de la mancha). La invención proporciona métodos para mejorar el acabado de un tejido que comprenden las siguientes etapas: (a) proporcionar una composición que comprende un polipéptido de la invención que tiene una actividad glucanasa, o un polipéptido codificado por un ácido nucleico de la invención; (b) proporcionar un tejido; y (c) poner en contacto el polipéptido de la etapa (a) y el tejido de la etapa (b) en condiciones en las que el polipéptido pueda tratar el tejido mejorando de ese modo el acabado del tejido. En un aspecto, el tejido es una lana o una seda.

La invención proporciona piensos (p. ej., un pienso para animales, tal como un pienso para pollos), o alimentos que comprenden un polipéptido de la invención, o un polipéptido codificado por un ácido nucleico de la invención. La invención proporciona métodos para hidrolizar un glucano u otro polisacárido en un pienso o un alimento antes de su consumo por un animal que comprenden las siguientes etapas: (a) obtener un material para pienso que comprende

una glucanasa, p. ej., endoglucanasa de la invención, o una glucanasa codificada por un ácido nucleico de la invención; y (b) añadir el polipéptido de la etapa (a) al material para pienso o alimento en una cantidad suficiente durante un período de tiempo suficiente para ocasionar la hidrólisis de un glucano u otro polisacárido y la formación de un alimento o pienso tratado, hidrolizando de ese modo un glucano u otro polisacárido en el alimento o el pienso antes de su consumo por el animal. En un aspecto, la invención proporciona métodos para hidrolizar un glucano u otro polisacárido en un pienso o un alimento después de su consumo por un animal que comprenden las siguientes etapas: (a) obtener un material para pienso que comprende una glucanasa de la invención, o una glucanasa codificada por un ácido nucleico de la invención; (b) añadir el polipéptido de la etapa (a) al material para pienso o alimento; y (c) administrar el material para pienso o alimento al animal, donde después de su consumo, la glucanasa causa la hidrólisis de un glucano u otro polisacárido en el pienso o alimento en el tracto digestivo del animal. El alimento o el pienso (p. ej., un pienso para animales, tal como un pienso para pollos), pueden ser, p. ej., un cereal, un grano, un maíz y similares.

En otro aspecto, la invención proporciona métodos para disminuir la viscosidad de los glucanos en una composición, p. ej., en un alimento o un pienso (p. ej., un pienso para animales, tal como un pienso para pollos), tratando la composición con una glucanasa de la invención, o, incluyendo una glucanasa de la invención en la composición. El alimento o pienso puede comprender cebada o trigo, p. ej., un alimento para pienso para una dieta con alto contenido de cebada o alto contenido de trigo, tal como una dieta para aves de corral. En un aspecto, la invención proporciona métodos para minimizar la disminución de la humedad en la alimentación de un animal (p. ej., un ave, tal como un ave de corral doméstica) en un alimento o pienso tratado con o que comprende una glucanasa de la invención. En un aspecto, la invención proporciona métodos para aumentar la tasa de crecimiento y/o la conversión de pienso en la alimentación de un animal (p. ej., un ave, tal como un ave de corral doméstica) en un alimento o pienso tratado con o que comprende una glucanasa de la invención. En un aspecto, la invención proporciona métodos para disminuir los excrementos en la alimentación de un animal (p. ej., un ave, tal como un ave de corral doméstica) en un alimento o pienso tratado con o que comprende una glucanasa de la invención.

La invención proporciona alimentos o suplementos nutricionales para un animal (p. ej., un ave de corral, tal como un pollo) que comprenden un polipéptido de la invención, p. ej., un polipéptido codificado por el ácido nucleico de la invención. En un aspecto, el polipéptido en el alimento o suplemento nutricional puede ser glicosilado. La invención proporciona matrices de liberación de enzimas comestibles que comprenden un polipéptido de la invención, p. ej., un polipéptido codificado por el ácido nucleico de la invención. En un aspecto, la matriz de liberación comprende un pelete que comprende una enzima de la invención, p. ej., un pelete que comprende una enzima termotolerante o termoestable de la invención). En un aspecto, el polipéptido puede ser glicosilado (que en un aspecto puede hacer a la enzima más termotolerante o termoestable). En un aspecto, la actividad glucanasa, p. ej., endoglucanasa es termotolerante. En otro aspecto, la actividad glucanasa es termoestable.

La invención proporciona un alimento, un pienso (p. ej., un pienso para animales, tal como un pienso para pollos), o un suplemento nutricional que comprende un polipéptido de la invención. La invención proporciona métodos para utilizar una glucanasa como suplemento nutricional en una dieta para animales, comprendiendo el método: preparar un suplemento nutricional que contiene una enzima glucanasa que comprende al menos treinta aminoácidos contiguos de un polipéptido de la invención; y administrar el suplemento nutricional a un animal para incrementar la utilización de un glucano contenido en un pienso o un alimento ingerido por el animal. El animal puede ser un ser humano, un rumiante o un animal monogástrico. Por ejemplo, el animal puede ser un ave, p. ej., un pollo. La enzima glucanasa se puede preparar mediante la expresión de un polinucleótido que codifica la glucanasa en un organismo tal como una bacteria, una levadura, una planta, un insecto, un hongo y un animal.

Los organismos ilustrativos para expresar los polipéptidos de la invención pueden ser *S. li*, *Streptomyces* sp., *Bacillus* sp. y *Lactobacillus* sp.

La invención proporciona matrices de liberación de enzimas comestibles que pueden comprender una enzima glucanasa recombinante termoestable, p. ej., un polipéptido de la invención. Los métodos para liberar un suplemento de glucanasa a un animal (un ser humano, un rumiante, un animal monogástrico, un ave, p. ej., un pollo), comprenden: preparar una matriz de liberación de enzimas comestible en forma de peletes que comprende un portador comestible granulado y una enzima glucanasa aislada, sintética o recombinante termoestable, donde los peletes dispersan fácilmente la enzima glucanasa contenida en los mismos en medios acuosos, y administrar la matriz de liberación de enzimas comestible al animal. La enzima glucanasa recombinante puede comprender un polipéptido de la invención. El portador comestible granulado puede comprender un portador seleccionado del grupo que consiste en un germen de grano, un germen de grano residuo de la extracción de aceite, un forraje, una alfalfa, un heno, una cáscara de soja, una harina de semilla de girasol y un salvado de trigo. El portador comestible puede comprender germen de grano residuo de la extracción de aceite. La enzima glucanasa puede ser glicosilada para proporcionar termoestabilidad en condiciones de peletización. La matriz de liberación se puede formar peletizando una mezcla que comprende un germen de grano y una glucanasa. Las condiciones de peletización pueden incluir la aplicación de vapor de agua. Las condiciones de peletización pueden comprender la aplicación de una temperatura

en exceso de alrededor de 80°C durante alrededor de 5 minutos y la enzima conserva una actividad específica de al menos 350 a alrededor de 900 unidades por miligramo de enzima.

5 La invención proporciona métodos para mejorar la textura y el sabor de un producto lácteo que comprenden las siguientes etapas: (a) proporcionar un polipéptido de la invención que tiene una actividad glucanasa, o una glucanasa codificada por un ácido nucleico de la invención; (b) proporcionar un producto lácteo; y (c) poner en contacto el polipéptido de la etapa (a) y el producto lácteo de la etapa (b) en condiciones en las que la glucanasa pueda mejorar la textura o el sabor del producto lácteo. En un aspecto, el producto lácteo comprende un queso o un yogurt. Los productos lácteos comprenden una glucanasa de la invención, o son codificados por un ácido nucleico de la invención.

15 La invención proporciona métodos para mejorar la extracción de aceite de un material vegetal rico en aceite que comprenden las siguientes etapas: (a) proporcionar un polipéptido de la invención que tiene una actividad glucanasa, o una glucanasa codificada por un ácido nucleico de la invención; (b) proporcionar un material vegetal rico en aceite; y (c) poner en contacto el polipéptido de la etapa (a) y el material vegetal rico en aceite. En un aspecto, el material vegetal rico en aceite comprende una semilla rica en aceite. El aceite puede ser un aceite de soja, un aceite de oliva, un aceite de colza (canola) o un aceite de girasol y similares.

20 En un aspecto, la invención proporciona métodos para utilizar una glucanasa de la invención para producir azúcares fermentables que pueden ser convertidos en etanol combustible. En un aspecto, la invención proporciona combustibles que comprenden un polipéptido de la invención que tiene actividad glucanasa, o una glucanasa codificada por un ácido nucleico de la invención. En un aspecto, se utiliza una enzima de la invención para catalizar la hidrólisis de celulosas, hemicelulosas y ligninas. La degradación de celulosa se puede utilizar para la conversión de biomasa vegetal en combustibles y productos químicos. Véase, p. ej., Kohlmann (1996) Adv. Space Res. 18:251-265; Perez (2002) Int Microbiol. 5:53-63.

30 La invención proporciona métodos para preparar un zumo, jarabe, puré o extracto de fruta u hortalizas que comprenden las siguientes etapas: (a) proporcionar un polipéptido de la invención que tiene una actividad glucanasa, o una glucanasa codificada por un ácido nucleico de la invención; (b) proporcionar una composición o un líquido que comprende un material de frutas u hortalizas; y (c) poner en contacto el polipéptido de la etapa (a) y la composición, preparando de ese modo el zumo, jarabe, puré o extracto de fruta u hortalizas.

35 La invención proporciona papeles o productos de papel o la pasta de papel que comprenden una glucanasa de la invención, o un polipéptido codificado por un ácido nucleico de la invención. La invención proporciona métodos para tratar un papel o una pasta de papel o madera que comprende las siguientes etapas: (a) proporcionar un polipéptido de la invención que tiene una actividad glucanasa, o una glucanasa codificada por un ácido nucleico de la invención; (b) proporcionar una composición que comprende un papel o una pasta de papel o madera; y (c) poner en contacto el polipéptido de la etapa (a) y la composición de la etapa (b) en condiciones en las que la glucanasa pueda tratar el papel o la pasta de papel o madera. En un aspecto, la composición farmacéutica actúa como coadyuvante digestivo o antimicrobiano (p. ej., contra *Salmonella*). En un aspecto, el tratamiento es profiláctico. En un aspecto, la invención proporciona productos para la higiene bucal que comprenden un polipéptido de la invención que tiene una actividad glucanasa, o una glucanasa codificada por un ácido nucleico de la invención. El producto para la higiene bucal puede comprender una pasta dentífrica, una crema dental, un gel o un polvo dentífrico, una preparación odóntica, un colutorio, una formulación de enjuagado pre- o postcepillado, un chicle, una gragea o un caramelo. La invención proporciona composiciones para limpieza de lentes de contacto que comprenden un polipéptido de la invención que tiene una actividad glucanasa, o una glucanasa codificada por un ácido nucleico de la invención.

50 En un aspecto, la invención proporciona métodos para eliminar o proteger animales de un microorganismo que comprende un glucano u otro polisacárido que comprenden administrar un polipéptido de la invención. El microorganismo puede ser una bacteria que comprende un glucano, p. ej., *Salmonella*.

55 Otro aspecto de la invención es un método para generar una variante que incluye obtener un ácido nucleico que tiene una secuencia de la invención, secuencias esencialmente idénticas a ésta, secuencias complementarias a una secuencia de la invención, fragmentos que comprenden al menos 30 nucleótidos consecutivos de las secuencias precedentes 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.

60 Un medio legible por ordenador puede tener almacenada una secuencia de ácido nucleico o polipéptido de la invención. Un sistema informático puede incluir un procesador y un dispositivo de almacenamiento de datos donde el dispositivo de almacenamiento de datos tiene almacenada una secuencia de ácido nucleico o polipéptido de la invención. Un método compara una primera secuencia con una secuencia de referencia donde la primera secuencia es una secuencia de ácido nucleico o polipéptido de la invención. El método incluye leer la primera secuencia y la secuencia de referencia a través del uso de un programa informático que compara secuencias; y determinar las diferencias entre la primera secuencia y la secuencia de referencia con el programa informático. Un método para

identificar una característica en una secuencia de ácido nucleico o polipéptido de la invención, incluye leer la secuencia a través del uso de un programa informático que identifica características en las secuencias; e identificar características en la secuencia con el programa informático.

5 Otro aspecto más de la invención es un método para catalizar la rotura de glucano o uno de sus derivados, que comprende la etapa de poner en contacto un glucano u otro polisacárido o derivado del mismo con un polipéptido de la invención en condiciones que faciliten la rotura de un glucano.

10 Un análisis identifica fragmentos o variantes de un polipéptido de la invención, que conservan la función enzimática de un polipéptido de la invención. Este análisis incluye poner en contacto un polipéptido de la invención con una molécula sustrato en condiciones que permitan que el fragmento de polipéptido o variante funcionen y detectar un 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 identificando de ese modo un fragmento o variante de tales secuencias.

15 Una sonda de ácido nucleico de un oligonucleótido tiene de alrededor de 10 a 50 nucleótidos de longitud y es un segmento de al menos 10 nucleótidos contiguos que es complementario al menos en 50% a una región diana de ácido nucleico de una secuencia de ácido nucleico de la invención; y que hibrida con la región diana de ácido nucleico en condiciones de moderada a altamente restrictivas para formar un dúplex de diana:sonda detectable.

20 Otro aspecto de la invención es una sonda de polinucleótido para el aislamiento o la identificación de genes de glucanasa que tienen una secuencia que es la misma que, o completamente complementaria a al menos una secuencia de ácido nucleico de la invención.

25 En otro aspecto más, la invención proporciona una preparación de proteína que comprende un polipéptido que tiene una secuencia de aminoácidos de la invención donde la preparación de proteína es un líquido. Otro aspecto más de la invención proporciona una preparación de proteína que comprende un polipéptido que tiene una secuencia de aminoácidos de la invención donde el polipéptido es un sólido.

30 Un método para modificar moléculas pequeñas comprende la etapa de mezclar al menos un polipéptido de la invención con al menos una molécula pequeña, para producir al menos una molécula pequeña modificada a través de al menos una reacción biocatalítica, donde el al menos un polipéptido tiene actividad glucanasa.

35 Otro aspecto de la invención es un vector de clonación de una secuencia que codifica un polipéptido de la invención que tiene actividad glucanasa. Otro aspecto de la invención es una célula anfitriona que comprende una secuencia que codifica un polipéptido de la invención. En otro aspecto adicional, la invención proporciona un vector de expresión susceptible de replicación en una célula anfitriona que comprende un ácido nucleico de la invención o un ácido nucleico que codifica un polinucleótido de la invención.

40 En otro aspecto, la invención proporciona un método de acondicionamiento de masas que comprende poner en contacto masa con al menos un polipéptido de la invención en condiciones suficientes para acondicionar la masa. Otro aspecto de la invención es un método de producción de bebidas que comprende la administración de al menos un polipéptido de la invención en condiciones suficientes para disminuir la viscosidad del mosto o la cerveza, o, aumentar la claridad (p. ej., clarificación) de la bebida.

45 Las glucanasas, p. ej., endoglucanasas de la invención se utilizan para romper los glucanos u otros polisacáridos de alto peso molecular en piensos para animales (p. ej., un pienso para un ser humano, rumiante, animal monogástrico, ave, p. ej., un pollo). La adición de las enzimas de la invención estimula las velocidades de crecimiento mejorando la digestibilidad, lo que también mejora la calidad de la deposición del animal. La glucanasa funciona a través del tracto gastrointestinal reduciendo la viscosidad intestinal e incrementando la difusión de enzimas pancreáticas. Adicionalmente, las enzimas de la invención se pueden utilizar en el tratamiento de las paredes de las células del endospermo de los cereales forrajeros y las proteínas vegetales. Las enzimas novedosas de la invención se pueden administrar a un animal con el fin de incrementar la utilización del glucano u otro polisacárido en el alimento. Esta actividad de las enzimas de la invención se puede utilizar para romper el material insoluble de las paredes celulares, liberando nutrientes de las paredes celulares, que a continuación se hacen asequibles para el animal. También cambia la hemicelulosa a azúcares nutritivos de manera que se liberan los nutrientes atrapados anteriormente en las paredes celulares. Las enzimas glucanasas de la invención pueden producir compuestos que pueden ser una fuente nutritiva para la microflora ruminal.

60 Un método para la utilización de glucanasa como suplemento nutricional en las dietas de los animales, comprende la preparación de un suplemento nutricional que contiene una enzima glucanasa recombinante que comprende al menos treinta aminoácidos contiguos de una secuencia de aminoácidos de la invención y la administración del suplemento nutricional a un animal para incrementar la utilización de un glucano u otro polisacárido contenido en el alimento ingerido por el animal.

Un método la liberación de un suplemento de glucanasa a un animal puede comprender la preparación de una matriz comestible de liberación de enzimas en forma de peletes que comprenden un portador comestible granulado y una glucanasa recombinante o sintética termoestable de la enzima de la invención, donde las partículas dispersan rápidamente la enzima glucanasa contenida en la misma en medios acuosos, y la administración de la matriz de liberación de enzimas comestible al animal. El portador comestible granulado puede comprender un portador seleccionado del grupo que consiste en germen de grano residuo de la extracción de aceite, forraje, alfalfa, heno, cáscara de soja, harina de semilla de girasol y salvado de trigo.

En otro aspecto, la invención proporciona un ácido nucleico aislado, sintético o recombinante que comprende una secuencia de la invención que codifica un polipéptido que tiene actividad glucanasa, donde la secuencia contiene una secuencia señal. La invención también proporciona un ácido nucleico aislado, sintético o recombinante que comprende una secuencia que codifica un polipéptido de la invención que tiene actividad glucanasa, y la secuencia contiene una secuencia señal de otra glucanasa, mananasa, o xilanasas. Adicionalmente, la invención proporciona un ácido nucleico aislado, sintético o recombinante que comprende una secuencia de la invención que codifica un polipéptido que tiene actividad glucanasa y la secuencia no contiene una secuencia señal.

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

Todas las publicaciones, patentes, solicitudes de patente, secuencias GenBank y depósitos ATCC, citados en la presente memoria y el disco compacto (presentados por cuadruplicado) que contienen una lista de secuencias se incorporan a la presente expresamente como referencia para todos los fines.

Breve Descripción de los Dibujos

Los siguientes dibujos son ilustrativos de aspectos de la invención y no limitan de ningún modo el alcance de la invención abarcada por las reivindicaciones.

La Figura 1 es un diagrama de bloques de un sistema informático.

La Figura 2 es un diagrama de flujo que ilustra un aspecto de un procedimiento para comparar una nueva secuencia de nucleótidos o proteínas con una base de datos de secuencias con el fin de determinar los niveles de homología entre la nueva secuencia y las secuencias en la base de datos.

La Figura 3 es un diagrama de flujo que ilustra un aspecto de un procedimiento en un ordenador para determinar si dos secuencias son homólogas.

La Figura 4 es un diagrama de flujo que ilustra un aspecto de un procedimiento identificador 300 para detectar la presencia de una característica en una secuencia.

La Figura 5 es una tabla que resume las actividades relativas de varias enzimas ilustrativas de la invención en diversas condiciones.

La Figura 6 es una ilustración en forma gráfica de un conjunto de datos ("datos de muestra") ilustrativos que se ilustra en forma de una "curva patrón", como se comenta en el Ejemplo 3.

La Figura 7 y la Figura 8 ilustran los resultados de los análisis de la actividad glucanasa que demuestran una mejora de la expresión en *Pichia pastoris* de una glucanasa de referencia que tiene una secuencia expuesta en el SEQ ID NO: 464, codificada por una versión de codón optimizado del SEQ ID NO: 5 (es decir, siendo la versión optimizada el SEQ ID NO: 463), como se comenta en el Ejemplo 4, más abajo.

La Figura 9 ilustra los resultados de los análisis de la actividad glucanasa que muestran el perfil de una glucanasa codificada por el SEQ ID NO: 6, como se comenta en el Ejemplo 5, más abajo.

La Figura 10 ilustra los resultados de los análisis de la actividad glucanasa que muestran la determinación de la vida media de una glucanasa codificada por el SEQ ID NO: 6, como se comenta en el Ejemplo 5, más abajo.

Símbolos de referencia parecidos en los diferentes dibujos indican elementos parecidos.

Descripción Detallada de la Invención

La invención proporciona polipéptidos y polinucleótidos que los codifican y a métodos para elaborarlos y utilizarlos. Los polipéptidos de la invención tienen una actividad glucanasa, por ejemplo, polipéptidos capaces de hidrolizar los enlaces glicosídicos presentes en un glucano, p. ej., catalizando la hidrólisis de los enlaces β -1,4-xilosídicos internos. Las enzimas de la invención se pueden utilizar para elaborar y/o procesar alimentos, piensos (p. ej., para un ser humano, un rumiante, un animal monogástrico, un ave, p. ej., un pollo), bebidas, suplementos nutricionales, productos textiles, detergentes y similares. Las enzimas de la invención se pueden utilizar en composiciones farmacéuticas y complementos dietéticos. Las glucanasas de la invención son particularmente útiles en el procesamiento de alimentos, panificación, piensos o alimentos para animales, bebidas, detergentes, procesamiento de pasta de papel y procedimientos con papel.

Definiciones

El término "anticuerpo" incluye un péptido o polipéptido derivado de, modelado después o codificado sustancialmente por un gen de inmunoglobulina o genes de inmunoglobulina, o fragmentos de los mismos, capaces de unirse específicamente a un antígeno o epítipo, véase, p. ej. *Fundamental Immunology*, Tercera Edición, W.E. Paul, ed., Raven Press, N.Y. (1993); Wilson (1994) *J. Immunol. Methods* 175:267-273; Yarmush (1992) *J. Biochem. Biophys. Methods* 25:85-97. El término anticuerpo incluye porciones de unión al antígeno, es decir, "sitios de unión al antígeno", (p. ej., fragmentos, subsecuencias, regiones determinantes de la complementariedad (CDR)) que conservan la capacidad de unirse al antígeno, incluyendo (i) un fragmento Fab, un fragmento monovalente que consiste en los dominios VL, VH, CL y CH1; (ii) un fragmento F(ab')₂, un fragmento bivalente que comprende dos fragmentos Fab unidos por un puente disulfuro en la región bisagra; (iii) un fragmento Fd que consiste en los dominios VH y CH1; (iv) un fragmento Fv que consiste en los dominios VL y VH de un solo brazo de un anticuerpo, (v) un fragmento dAb (Ward et al., (1989) *Nature* 341:544-546), que consiste en un dominio VH; y (vi) una región determinante de la complementariedad (CDR) aislada. Los anticuerpos de cadena sencilla también están incluidos como referencia en el término "anticuerpo".

Los términos "matriz" o "micromatriz" o "biochip" o "chip" según se utilizan en la presente memoria son una pluralidad de elementos diana, comprendiendo cada elemento diana una cantidad definida de uno o más polipéptidos (incluyendo anticuerpos) o ácidos nucleicos inmovilizados sobre una zona definida de una superficie sustrato, como se comenta con más detalle, más abajo.

Según se utiliza en la presente memoria, los términos "ordenador", "programa informático" y "procesador" se utilizan en sus contextos generales más amplios e incorporan todos estos dispositivos, como se describe detalladamente más abajo. Una "secuencia codificante de" o una "secuencia que codifica" un polipéptido o proteína concretos, es una secuencia de ácido nucleico que es transcrita y traducida a un polipéptido o proteína cuando se coloca bajo el control de secuencias reguladoras apropiadas.

Las frases "ácido nucleico" o "secuencia de ácido nucleico" según se utilizan en la presente memoria hacen referencia a un oligonucleótido, nucleótido, polinucleótido, o a un fragmento de cualquiera de estos, a ADN o ARN de origen genómico o sintético que pueden ser monocatenarios o bicatenarios y pueden representar una hebra efectora o antisentido, a ácido peptidonucleico (PNA), o a cualquier material de tipo ADN o de tipo ARN, de origen natural o sintético. Las frases "ácido nucleico" o "secuencia de ácido nucleico" incluye oligonucleótido, nucleótido, polinucleótido, o a un fragmento de cualquiera de estos, a ADN o ARN (p. ej., ARNm, ARNr, ARNt, ARNi) de origen genómico o sintético que puede ser monocatenario o bicatenario y pueden representar una cadena directa o antisentido, a ácido peptidonucleico (PNA), o a cualquier material de tipo ADN o de tipo ARN, de origen natural o sintético, incluyendo, p. ej., ARNi, ribonucleoproteínas (p. ej., p. ej., ARNi bicatenarios, p. ej., RNPI). El término abarca ácidos nucleicos, es decir, oligonucleótidos, que contienen análogos conocidos de nucleótidos naturales. El término también abarca estructuras de tipo ácido nucleico con esqueletos sintéticos, véanse p. ej., Mata (1997) *Toxicol. Appl. Pharmacol.* 144:189-197; Strauss-Soukup (1997) *Biochemistry* 36:8692-8698; Samstag (1996) *Antisense Nucleic Acid Drug Dev* 6:153-156. "Oligonucleótido" incluye un polidesoxinucleótido monocatenario o dos cadenas complementarias de polidesoxinucleótidos que se pueden sintetizar químicamente. Tales oligonucleótidos sintéticos no tienen fosfato 5' y por tanto no se ligan a otro oligonucleótido sin añadir un fosfato con un ATP en presencia de una quinasa. Un oligonucleótido sintético se puede ligar a un fragmento que no haya sido desfosforilado.

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

El término "gen" significa el segmento de ADN implicado en la producción de una cadena polipeptídica; incluye las regiones anteriores y posteriores a la región codificante (líder y trailer) así como, cuando sea aplicable, las secuencias interpuestas (intrones) entre segmentos codificantes individuales (exones). "Conectado operablemente" según se utilizan en la presente memoria hace referencia a una relación funcional entre dos o más segmentos de ácido nucleico (p. ej., ADN). Típicamente, hace referencia a la relación funcional de secuencias reguladoras transcripcionales con una secuencia transcrita. Por ejemplo, un promotor esta conectado operablemente a una secuencia codificante, tal como un ácido nucleico de la invención, si estimula o modula la transcripción de la secuencia codificante en una célula anfitriona u otro sistema de expresión apropiado. Generalmente, las secuencias reguladoras de la transcripción promotoras que están conectadas operablemente a una secuencia transcrita son contiguas físicamente a la secuencia transcrita, es decir, actúan en cis. Sin embargo, algunas secuencias reguladoras de la transcripción, tales como los intensificadores, no necesitan estar contiguos físicamente o localizados en íntima proximidad a las secuencias codificantes cuya transcripción intensifican.

El término "casete de expresión" según se utiliza en la presente memoria hace referencia a una secuencia de nucleótidos que es capaz de afectar a la expresión de un gen estructural (es decir, una secuencia de proteínas

- codificante, tal como una glucanasa de la invención) en un anfitrión compatible con tales secuencias. Los casetes de expresión incluyen al menos un promotor conectado operablemente con la secuencia codificante del polipéptido; y, opcionalmente, con otras secuencias, p. ej., señales de terminación de la transcripción. También se pueden utilizar factores necesarios o útiles para efectuar la expresión, p. ej., intensificadores. De este modo, los casetes de expresión también incluyen plásmidos, vectores de expresión, virus recombinantes, cualquier forma de vector de "ADN desnudo" recombinante, y similares. Un "vector" comprende un ácido nucleico que puede infectar, transfectar, transducir transitoriamente o permanentemente una célula. Se advertirá que un vector puede ser un ácido nucleico desnudo, o un ácido nucleico que forma complejo con proteínas o lípidos. El vector comprende opcionalmente ácidos nucleicos y/o proteínas, y/o membranas virales o bacterianas (p. ej., una membrana celular, una envoltura lipídica viral, etc.). Los vectores incluyen, pero no están limitados a replicones (p. ej., replicones de ARN, bacteriófagos) a los que se pueden anclar los fragmentos de ADN y replicarlos. Los vectores incluyen así, pero no están limitados a ARN, ADN o ARN circulares o lineales autoreplicantes autónomos (p. ej., plásmidos, virus, y similares, véase, p. ej., la Patente de los Estados Unidos Núm. 5.217.879), e incluyen plásmidos con expresión y sin expresión. Cuando se describe que un microorganismo recombinante o cultivo celular alberga un "vector de expresión" esto incluye tanto ADN circular y lineal extracromosómico como ADN que ha sido incorporado al cromosoma o a los cromosomas del anfitrión. Cuando un vector está siendo mantenido por una célula anfitriona, el vector puede ser replicado establemente por las células durante la mitosis en forma de una estructura autónoma, o es incorporado al genoma de anfitrión.
- Según se utiliza en la presente memoria, el término "promotor" incluye todas las secuencias capaces de dirigir la transcripción de una secuencia codificante en una célula, p. ej., una célula vegetal. De este modo, los promotores utilizados en los constructos de la invención incluyen elementos de control de la transcripción que actúan en cis y secuencias reguladoras que están implicadas en la regulación o la modulación de la duración y/o la velocidad de la transcripción de un gen. Por ejemplo, un promotor puede ser un elemento de control de la transcripción que actúa en cis, incluyendo un intensificador, un promotor, un terminador de la transcripción, un origen de replicación, una secuencia de integración cromosómica, regiones no traducidas 5' y 3', o una secuencia intrónica, que están implicados en la regulación de la transcripción. Estas secuencias que actúan en cis interactúan típicamente con proteínas u otras biomoléculas para llevar a cabo (activar/desactivar, regular, modular, etc.) la transcripción. Los promotores "constitutivos" son aquellos que dirigen la expresión continuamente en la mayoría de las condiciones ambientales y estados de desarrollo o diferenciación celular. Los promotores "inducibles" o "regulables" dirigen la expresión del ácido nucleico de la invención bajo la influencia de las condiciones ambientales o condiciones de desarrollo. Los ejemplos de las condiciones ambientales que puede afectar a la transcripción por promotores inducibles incluyen condiciones anaeróbicas, temperatura elevada, sequía, o la presencia de luz.
- Los promotores "específicos de tejido" son elementos de control de la transcripción que únicamente son activos en células o tejidos o órganos concretos, p. ej., en plantas o animales. La regulación específica del tejido se puede lograr por medio de ciertos factores intrínsecos que aseguran que los genes que codifican proteínas específicas para un tejido dado se expresen. Se sabe que tales factores existen en mamíferos y plantas de manera que permiten que se desarrollen tejidos específicos.
- El término "planta" incluye plantas completas, partes de plantas (p. ej., hojas, tallos, flores, raíces, etc.), protoplastos vegetales, semillas y células vegetales y sus progenies. La clase de plantas que se pueden utilizar en el método de la invención es generalmente tan amplio como la clase de plantas superiores susceptibles de técnicas de transformación, incluyendo angiospermas (plantas monocotiledóneas y dicotiledóneas), así como gimnospermas. Ésta incluye plantas de una variedad de niveles de ploidía, incluyendo los estados poliploide, diploide, haploide y hemizigoto. Según se utiliza en la presente memoria, el término "planta transgénica" incluye plantas o células vegetales en las que se ha insertado una secuencia de ácido nucleico heteróloga, p. ej., los ácidos nucleicos y diversos constructos recombinantes (p. ej., casetes de expresión) de la invención.
- Los "plásmidos" pueden ser asequibles comercialmente, accesibles al público en una base no restringida, o se pueden construir a partir de plásmidos asequibles de acuerdo con procedimientos publicados. Plásmidos equivalentes a los descritos en la presente memoria son conocidos en la técnica y resultarán evidentes para los expertos normales en la técnica.
- "Aminoácido" o "secuencia de aminoácidos" según se utilizan en la presente memoria hacen referencia a un oligopéptido, péptido, polipéptido, o secuencia de proteínas, o a un fragmento, porción, o subunidad de cualquiera de estos y a moléculas de origen natural o sintéticas.
- "Aminoácido" o "secuencia de aminoácidos" incluyen un oligopéptido, péptido, polipéptido, o secuencia de proteínas, o a un fragmento, porción, o subunidad de cualquiera de éstos, y a moléculas de origen natural o sintéticas. El término "polipéptido" según se utiliza de la presente memoria, hace referencia a aminoácidos unidos entre sí por enlaces peptídicos o enlaces peptídicos modificados, es decir, isoésteres peptídicos y pueden contener aminoácidos modificados distintos de los 20 aminoácidos codificados por genes. Los polipéptidos se pueden modificar por medio de procedimientos naturales, tales como procesamiento post-traducciona, o mediante técnicas de modificación

química que son bien conocidos en la técnica. Las modificaciones se pueden producir en cualquier parte del polipéptido, incluyendo el esqueleto peptídico, las cadenas laterales de aminoácidos y los extremos amino o carboxilo. Se apreciará que puede estar presente el mismo tipo de modificación en el mismo grado o en grados variables en diferentes sitios en un polipéptido dado. Asimismo un polipéptido dado puede tener muchos tipos de modificaciones. Las modificaciones incluyen acetilación, acilación, ribosilación de ADP, amidación, unión covalente de flavina, unión covalente de un radical hemo, unión covalente de un nucleótido o derivado de nucleótido, unión covalente de un lípido o derivado de lípido, unión covalente de un fosfatidilinositol, ciclación de entrecruzamiento, formación de enlace disulfuro, desmetilación, formación de entrecruzamientos covalentes, formación de cisteína, formación de piroglutamato, formilación, gamma-carboxilación, glicosilación, formación del ancla de GPI, hidroxilación, yodación, metilación, miristoilación, oxidación, pegilación, procesamiento de xilano hidrolasa, fosforilación, prenilación, racemización, selenilación, sulfatación y adición mediada por ARN de transferencia de aminoácidos a proteínas por ejemplo arginilación. (Véanse Creighton, T.E., *Proteins - Structure and Molecular Properties* 2ª Ed., W.H. Freeman y Company, New York (1993); *Posttranslational Covalent Modification of Proteins*, B.C. Johnson, Ed., Academic Press, New York, págs. 1-12 (1983)). Los péptidos y polipéptidos de la invención también incluyen todas las formas "miméticas" y "peptidomiméticas", como se describe con más detalle, más abajo.

Según se utiliza en la presente memoria, el término "aislado" significa que el material es separado de su entorno original (*p. ej.*, el entorno natural si es de origen natural). Por ejemplo, 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 ya que tal vector o composición no es parte de su entorno natural. Según se utiliza en la presente memoria, 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 uno 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.

Según se utiliza en la presente memoria, el término "recombinante" significa que el ácido nucleico es adyacente a un "esqueleto" del ácido nucleico al que no es 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 troncales de ácido nucleico. Las moléculas troncales de acuerdo con la invención incluyen ácidos nucleicos tales como vectores de expresión, ácidos nucleicos autoreplicantes, virus, ácidos nucleicos integrantes y otros vectores o ácidos nucleicos utilizados 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 troncales recombinantes. Más típicamente, los ácidos nucleicos enriquecidos representan 50% o más del número de insertos de ácido nucleico en la población de moléculas troncales recombinantes. En un aspecto, los ácidos nucleicos enriquecidos representan 90% o más del número de insertos de ácido nucleico en la población de moléculas troncales 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 utilizar 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 asequibles comercialmente (Cambridge Research Biochemicals). Tales kits de laboratorio asequibles 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 utiliza tal sistema, una placa de varillas o clavijas se invierte y se inserta en una segunda placa de pocillos o reservorios correspondientes, que contienen soluciones 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 soluciones apropiadas, los aminoácidos se arman para dar los péptidos deseados. Además, están disponibles varios sistemas de síntesis peptídica FMOC asequible. 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, mediante síntesis directa o mediante síntesis de una serie de fragmentos que se pueden acoplar utilizando otras técnicas conocidas.

Una secuencia promotora está "conectada operablemente a" una secuencia codificante cuando la ARN polimerasa que inicia la transcripción hacia el promotor transcribirá la secuencia codificante a ARNm.

5 Los "plásmidos" se nombran por una "p" minúscula precedida y/o seguida de letras mayúsculas y/o números. Los plásmidos de partida en la presente memoria son asequibles comercialmente, accesibles al público en una base no restringida, o pueden ser construidos a partir de plásmidos disponibles de acuerdo con procedimientos publicados. Además, los plásmidos equivalentes a los descritos en la presente memoria son conocidos en la técnica y resultarán evidentes para los expertos normales en la técnica.

10 "Digestión" de ADN hace referencia a la escisión catalítica del ADN con una enzima de restricción que actúa únicamente en ciertas secuencias en el ADN. Las diversas enzimas de restricción utilizadas en la presente memoria están disponibles en el mercado y sus condiciones de reacción, cofactores y otros requerimientos se utilizarían como conocen los expertos normales en la técnica. Con fines analíticos, típicamente se utiliza 1 µg de plásmido o fragmento de ADN con alrededor de 2 unidades de enzima en alrededor de 20 µl de solución tampón. Con el fin de
15 aislar fragmentos de ADN para la construcción del plásmido, e digieren típicamente de 5 a 50 µg de ADN con 20 a 250 unidades de enzima en un gran volumen. Los tampones y las cantidades de sustrato apropiados para las enzimas de restricción concretas son especificados por el fabricante. Normalmente se utilizan tiempos de incubación de alrededor de 1 hora a 37°C, pero pueden variar de acuerdo con las instrucciones del fabricante. Después de la digestión, se puede realizar la electroforesis en gel para aislar el fragmento deseado.

20 La frase "sustancialmente idéntico" en el contexto de dos ácidos nucleicos o polipéptidos, hace referencia a dos o más secuencias que tienen, p. ej., 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
25 más identidad de nucleótidos o residuos de aminoácidos (secuencia), cuando se comparan y alinean para una correspondencia máxima, según se mide utilizando uno de los algoritmo de comparación de secuencias conocidos o mediante inspección visual. Típicamente, existe identidad sustancial a lo largo de una región de al menos alrededor de 100 residuos y muy comúnmente las secuencias son sustancialmente idénticas a lo largo de al menos alrededor de 150-200 residuos. En algunos aspectos, las secuencias son sustancialmente idénticas a lo largo de la longitud
30 completa de las regiones codificantes.

Adicionalmente una secuencia de aminoácidos "sustancialmente idéntica" es una secuencia que difiere de una secuencia de referencia en una o más sustituciones, deleciones, o inserciones de aminoácidos conservativas o no conservativas, particularmente cuando tal sustitución se produce en un sitio que no es el sitio activo de la molécula y
35 siempre que el polipéptido conserve esencialmente sus propiedades funcionales. Una sustitución de aminoácido conservativa, por ejemplo, sustituye un aminoácido por otro de la misma clase (p. ej., sustitución de un aminoácido hidrófobo, tal como isoleucina, valina, leucina, o metionina, por otro, o sustitución de un aminoácido polar por otro, tal como la sustitución de arginina por lisina, ácido glutámico por ácido aspártico o glutamina por asparragina). Se pueden suprimir uno o más aminoácidos, por ejemplo, de un polipéptido de glucanasa, dando como resultado la
40 modificación de la estructura del polipéptido, sin alterar significativamente su actividad biológica. Por ejemplo, los aminoácidos amino- o carboxilo terminales que no son requeridos para la actividad biológica de la glucanasa se pueden eliminar. Las secuencias de polipéptidos modificadas de la invención se pueden someter a ensayo para determinar la actividad biológica de la glucanasa mediante cualquier número de métodos, incluyendo poner en
45 contacto la secuencia de polipéptidos modificada con un sustrato de glucanasa y determinar si el polipéptido modificado disminuye la cantidad de sustrato específico en el análisis o aumenta los bioproductos de la reacción enzimática de un polipéptido de glucanasa funcional con el sustrato.

Los "fragmentos" según se utiliza en la presente memoria son una porción de una proteína de origen natural que puede existir en al menos dos conformaciones diferentes. Los fragmentos pueden tener la misma o sustancialmente
50 la misma secuencia de aminoácidos que la proteína de origen natural. "Sustancialmente la misma" significa que una secuencia de aminoácidos en su mayor parte, pero no completamente, la misma, pero conserva al menos una actividad funcional de la secuencia con la que está relacionada. En general dos secuencias de aminoácidos son "sustancialmente las mismas" o "sustancialmente homólogas" si son idénticas en al menos alrededor de 85%. También están incluidos los fragmentos que tienen estructuras tridimensionales diferentes a la de la proteína de
55 origen natural. Un ejemplo de esto, es una molécula "pro-forma", tal como una proproteína de baja actividad que se puede modificar mediante escisión para producir una enzima madura con una actividad significativamente mayor.

"Hibridación" hace referencia al procedimiento mediante el cual una hebra de ácido nucleico se une a una hebra complementaria a través de emparejamiento de bases. Las reacciones de hibridación pueden ser sensibles y
60 selectivas de manera que una secuencia de interés concreta se puede identificar incluso en muestras en las que está presente a bajas concentraciones. Las condiciones adecuadamente restrictivas pueden estar definidas, por ejemplo, por las concentraciones de sal o formamida en las soluciones de prehibridación e hibridación, o por la temperatura de hibridación y son bien conocidas en la técnica. En particular, la restricción se puede incrementar reduciendo la concentración de sal, incrementando la concentración de formamida, o subiendo la temperatura de

hibridación. En aspectos alternativos, los ácidos nucleicos de la invención se definen por su capacidad de hibridar en diversas condiciones restrictivas (p. ej., altas, medias, y bajas), como se muestra en la presente memoria.

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

15 El término "variante" hace referencia a polinucleótidos o polipéptidos de la invención modificados en uno o más pares de bases, codones, intrones, exones, o residuos de aminoácidos (respectivamente) conservando aún la actividad biológica de una glucanasa de la invención. Las variantes se pueden producir por cualquier número de medios incluyendo métodos tales como, por ejemplo, PCR propensa a error, barajado, mutagénesis dirigida a oligonucleótidos, PCR de ensamblaje, mutagénesis por PCR sexual, mutagénesis *in vivo*, mutagenésis por inserción de casetes, mutagenésis de ensamblaje recursivo, mutagenésis de conjunto exponencial, mutagénesis de sitio específico, reensamblaje de genes, GSSM™ y cualquiera de sus combinaciones.

20 El término "Mutagénesis por Saturación" o "Mutagénesis por Saturación de Sitio de Genes™" o "GSSM™" incluye un método en el que se utilizan cebadores oligonucleotídicos degenerados para introducir mutaciones puntuales en un polinucleótido, como se describe en detalle, más abajo.

25 El término "sistema de evolución dirigida optimizada" o "evolución dirigida optimizada" incluye un método para reensamblar fragmentos de secuencias de ácido nucleico relacionadas, p. ej., genes relacionados, y se explica en detalle, más abajo.

30 El término "reensamblaje de ligación sintética" o "SLR" incluye un método para ligar fragmentos de oligonucleótidos de un modo no estocástico, y se explica en detalle, más abajo.

Generación y Manipulación de Ácidos Nucleicos

35 La invención proporciona ácidos nucleicos aislados, recombinantes y sintéticos (p. ej., un ácido nucleico ilustrativo de la invención, incluyendo el SEQ ID NO: 37, y secuencias que tienen una identidad de secuencia con el SEQ ID NO: 37; ácidos nucleicos que codifican polipéptidos de la invención, p. ej., las secuencias de aminoácidos mostradas en el SEQ ID NO: 38.

40 La invención también proporciona 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. Los métodos para descubrir nuevas secuencias de glucanasas utilizan los ácidos nucleicos de la invención. Los métodos para inhibir la expresión de genes, transcritos y polipéptidos de glucanasas utilizan los ácidos nucleicos de la invención. También se proporcionan métodos para modificar los ácidos nucleicos de la invención, p. ej., mediante el reensamblaje de ligación sintética, el sistema de evolución dirigida optimizada y/o la mutagénesis por saturación.

Los ácidos nucleicos de la invención se pueden elaborar, aislar y/o manipular, p. ej., mediante clonación y expresión de bibliotecas de ADNc, amplificación de mensaje o ADN genómico mediante PCR, y similares.

50 Las siguientes secuencias ilustrativas de la descripción se obtuvieron inicialmente de las siguientes fuentes, como se expone en la Tabla 1 más abajo:

Tabla 1

SEQ ID NO:	Fuente
291, 292	<i>Aquifex aeolicus</i>
161, 162	Archaea
175, 176	Archaea
367, 368	Archaea
479, 480	Archaea
495, 496	Archaea

ES 2 401 795 T3

<u>SEQ ID NO:</u>	<u>Fuente</u>
59, 60	Archaea
75, 76	Archaea
109, 110	Bacteria
229, 230	Bacteria
261, 262	Bacteria
263, 264	Bacteria
273, 274	Bacteria
277, 278	Bacteria
287, 288	Bacteria
293, 294	Bacteria
295, 296	Bacteria
331, 332	Bacteria
333, 334	Bacteria
363, 364	Bacteria
365, 366	Bacteria
369, 370	Bacteria
395, 396	Bacteria
397, 398	Bacteria
401, 402	Bacteria
427, 428	Bacteria
433, 434	Bacteria
435, 436	Bacteria
439, 440	Bacteria
447, 448	Bacteria
449, 450	Bacteria
455, 456	Bacteria
483, 484	Bacteria
485, 486	Bacteria
499, 500	Bacteria
5, 6	Bacteria
231, 232	Bacteria
67, 68	Bacteria
517, 518	Bacteria
399, 400	Thermotoga sp.
1, 2	Desconocida
101, 102	Desconocida
103, 104	Desconocida
105, 106	Desconocida
107, 108	Desconocida

ES 2 401 795 T3

<u>SEQ ID NO:</u>	<u>Fuente</u>
11, 12	Desconocida
111, 112	Desconocida
113, 114	Desconocida
115, 116	Desconocida
117, 118	Desconocida
119, 120	Desconocida
121, 122	Desconocida
123, 124	Desconocida
125, 126	Desconocida
127, 128	Desconocida
129, 130	Desconocida
13,14	Desconocida
131, 132	Desconocida
133, 134	Desconocida
135, 136	Desconocida
137, 138	Desconocida
139, 40	Desconocida
141, 142	Desconocida
143,144	Desconocida
145, 146	Desconocida
147, 148	Desconocida
149, 150	Desconocida
15, 16	Desconocida
151, 152	Desconocida
153, 154	Desconocida
155, 156	Desconocida
157, 158	Desconocida
159, 160	Desconocida
163, 164	Desconocida
165, 166	Desconocida
167, 168	Desconocida
169, 170	Desconocida
17, 18	Desconocida
171, 172	Desconocida
173, 174	Desconocida
177, 178	Desconocida
179, 180	Desconocida
181, 182	Desconocida
183, 184	Desconocida

ES 2 401 795 T3

<u>SEQ ID NO:</u>	<u>Fuente</u>
185, 186	Desconocida
187, 188	Desconocida
189, 190	Desconocida
19, 20	Desconocida
191, 192	Desconocida
193, 94	Desconocida
195, 196	Desconocida
197, 198	Desconocida
199, 200	Desconocida
201, 202	Desconocida
203, 204	Desconocida
205, 206	Desconocida
207, 208	Desconocida
209, 210	Desconocida
21, 22	Desconocida
211, 212	Desconocida
213, 214	Desconocida
215, 216	Desconocida
217, 218	Desconocida
219, 220	Desconocida
221, 222	Desconocida
223, 224	Desconocida
225, 226	Desconocida
227, 228	Desconocida
23, 24	Desconocida
233, 234	Desconocida
235, 236	Desconocida
237, 238	Desconocida
239, 240	Desconocida
241, 242	Desconocida
243, 244	Desconocida
245, 246	Desconocida
247, 248	Desconocida
249, 250	Desconocida
25, 26	Desconocida
251, 252	Desconocida
253, 254	Desconocida
255, 256	Desconocida
257, 258	Desconocida

ES 2 401 795 T3

<u>SEQ ID NO:</u>	<u>Fuente</u>
259, 260	Desconocida
265, 266	Desconocida
267, 268	Desconocida
269, 270	Desconocida
27, 28	Desconocida
271, 272	Desconocida
275, 276	Desconocida
279, 280	Desconocida
281, 282	Desconocida
283, 284	Desconocida
285, 286	Desconocida
289, 290	Desconocida
29, 30	Desconocida
297, 298	Desconocida
299, 300	Desconocida
3, 4	Desconocida
301, 302	Desconocida
303, 304	Desconocida
305, 306	Desconocida
307, 308	Desconocida
309, 310	Desconocida
31, 32	Desconocida
311, 312	Desconocida
313, 314	Desconocida
315, 316	Desconocida
317, 318	Desconocida
319, 320	Desconocida
321, 322	Desconocida
323, 324	Desconocida
325, 326	Desconocida
327, 328	Desconocida
329, 330	Desconocida
33, 34	Desconocida
335, 336	Desconocida
337, 338	Desconocida
339, 340	Desconocida
341, 342	Desconocida
343, 344	Desconocida
345, 346	Desconocida

ES 2 401 795 T3

<u>SEQ ID NO:</u>	<u>Fuente</u>
347, 348	Desconocida
349, 350	Desconocida
35, 36	Desconocida
351, 352	Desconocida
353, 354	Desconocida
355, 356	Desconocida
357, 358	Desconocida
359, 360	Desconocida
361, 362	Desconocida
37, 38	Desconocida
371, 372	Desconocida
373, 374	Desconocida
375, 376	Desconocida
377, 378	Desconocida
379, 380	Desconocida
381, 382	Desconocida
383, 384	Desconocida
385, 386	Desconocida
387, 388	Desconocida
389, 390	Desconocida
39, 40	Desconocida
391, 392	Desconocida
393, 394	Desconocida
403, 404	Desconocida
405, 406	Desconocida
407, 408	Desconocida
409, 410	Desconocida
41, 42	Desconocida
411, 412	Desconocida
413,414	Desconocida
415, 416	Desconocida
417,418	Desconocida
419,420	Desconocida
421, 422	Desconocida
423, 424	Desconocida
425, 426	Desconocida
429, 430	Desconocida
43, 44	Desconocida
431, 432	Desconocida

ES 2 401 795 T3

<u>SEQ ID NO:</u>	<u>Fuente</u>
437, 438	Desconocida
441, 442	Desconocida
443, 444	Desconocida
445, 446	Desconocida
45, 46	Desconocida
451, 452	Desconocida
453, 454	Desconocida
457, 458	Desconocida
459, 460	Desconocida
461, 462	Desconocida
463, 464	Artificial
465, 466	Desconocida
467, 468	Desconocida
469,470	Desconocida
47, 48	Desconocida
471, 472	Desconocida
473, 474	Desconocida
475, 476	Desconocida
477, 478	Desconocida
481, 482	Desconocida
487, 488	Desconocida
489, 490	Desconocida
49, 50	Desconocida
491, 492	Desconocida
493, 494	Desconocida
497, 498	Desconocida
501, 502	Desconocida
503, 504	Desconocida
505, 506	Desconocida
507, 508	Desconocida
509, 510	Desconocida
51, 52	Desconocida
511, 512	Desconocida
513, 514	Desconocida
515, 516	Desconocida
53, 54	Desconocida
55, 56	Desconocida
57, 58	Desconocida
61, 62	Desconocida

<u>SEQ ID NO:</u>	<u>Fuente</u>
63, 64	Desconocida
65, 66	Desconocida
69, 70	Desconocida
7, 8	Desconocida
71, 72	Desconocida
73, 74	Desconocida
77, 78	Desconocida
79, 80	Desconocida
81, 82	Desconocida
83, 84	Desconocida
85, 86	Desconocida
87, 88	Desconocida
89, 90	Desconocida
9, 10	Desconocida
91, 92	Desconocida
93, 94	Desconocida
95, 96	Desconocida
97, 98	Desconocida
99, 100	Desconocida

5 Al poner en práctica los métodos de la invención, los genes homólogos se puede modificar mediante manipulación de un ácido nucleico molde, como se describe en la presente memoria. La invención se puede poner en práctica junto con cualquier método o protocolo o dispositivo conocido en la técnica, que son bien descritos en la bibliografía científica y de patentes.

10 Un aspecto de la invención es un ácido nucleico aislado que comprende una de las secuencias de las secuencias de 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.

15 Los ácidos nucleicos aislados de la invención, se pueden utilizar para preparar 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 bases consecutivas de uno de los polipéptidos de la invención.

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

30 El ácido nucleico aislado que codifica uno de los polipéptidos de la invención, puede incluir, pero no está limitado a: solo la secuencia codificante 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, según se utiliza de la presente memoria, el término

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

5 Alternativamente, las secuencias de ácido nucleico de la invención, se pueden mutagenizar utilizando 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. Según se utiliza 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 con el fin de incrementar el nivel del polipéptido producido por las células anfitrionas que contienen un vector que codifica el polipéptido mediante la introducción de codones o pares de codones que aparecen frecuentemente en el organismo anfitrión.

15 La invención también se refiere a polinucleótidos que tienen cambios de nucleótido que dan como resultado sustituciones, adiciones, deleciones, fusiones y truncamientos de aminoácido en los polipéptidos de la invención. Tales cambios de nucleótido se puede introducir utilizando técnicas tales como mutagénesis dirigida al sitio, mutagénesis química al azar, deleción mediante exonucleasa III y otras técnicas de ADN recombinante. Alternativamente, tales cambios de nucleótido pueden ser variantes alélicas de origen natural que son aisladas identificando los ácidos nucleicos que hibridan específicamente con sondas que comprenden al menos 10, 15, 20, 25, 30, 35, 40, 50, 75, 100, 150, 200, 300, 400, o 500 bases consecutivas de una de las secuencias de la invención (o las secuencias complementarias a éstas) en condiciones de restricción alta, moderada, o baja como se proporciona en la presente memoria.

Técnicas Generales

25 Los ácidos nucleicos utilizados 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 (p. ej., glucanasas) generados a partir de estos ácidos nucleicos se pueden aislar o clonar individualmente y someter a ensayo para determinar una actividad deseada. Se puede utilizar cualquier sistema de expresión recombinante, incluyendo sistemas de expresión bacterianos, de mamíferos, de levadura, de insecto o de células vegetales.

35 Alternativamente, estos ácidos nucleicos se pueden sintetizar *in vitro* por medio de técnica de síntesis química bien conocidas, como describen, p. ej., 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 de los Estados Unidos Núm. 4.458.066.

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

45 Otro medio útil para obtener y manipular los ácidos nucleicos utilizados para poner en práctica los métodos de la invención es la clonación en muestras genómicas, y, si se desea, escrutar y re-clonar los insertos aislados o amplificados en, p. ej., clones genómicos o clones de ADNc. Las fuentes de ácido nucleico utilizadas en los métodos de la invención incluyen bibliotecas genómicas o de ADNc contenidas, p. ej., en cromosomas artificiales de mamífero (MAC), véanse, p. ej., las Patentes de los Estados Unidos Núms. 5.721.118; 6.025.155; cromosomas artificiales humanos, véase, p. ej., Rosenfeld (1997) Nat. Genet. 15:333-335; cromosomas artificiales de levadura (YAC); cromosomas artificiales bacterianos (BAC); cromosomas artificiales P1, véase, p. ej., Woon (1998) Genomics 50:306-316; vectores derivados de P1 (PAC), véase, p. ej., Kern (1997) Biotechniques 23:120-124; cósmidos, virus recombinantes, fagos o plásmidos.

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

60 La invención proporciona proteínas de fusión y ácidos nucleicos que las codifican. Un polipéptido de la invención se puede fusionar a un péptido o polipéptido heterólogo, tal como péptidos de identificación N-terminal que confieren las características deseadas, tales como aumento de estabilidad o purificación simplificada. Los péptidos y polipéptidos de la invención se pueden sintetizar y expresar también en forma de proteínas de fusión con uno o más dominios adicionales conectados a los mismos para, p. ej., producir un péptido más inmunogénico, para aislar más fácilmente un péptido sintetizado recombinantemente, para identificar y aislar anticuerpos y células B que expresan

anticuerpos, y similares. Los dominios que facilitan la detección y purificación incluyen, p. ej., péptidos queladores de metal tales como tractos de polihistidina, y módulos de histidina-triptófano que permiten la purificación en metales inmovilizados, dominios de proteína A que permiten la purificación en inmunoglobulinas inmovilizadas, y el dominio utilizado el sistema de purificación extensión/afinidad FLAGS (Immunex Corp, Seattle WA). La inclusión de un conector de secuencias que se puede dividir tal como el Factor Xa o enteroquinasa (Invitrogen, San Diego CA) entre un dominio de purificación y el péptido que contiene el motivo o el polipéptido para facilitar la purificación. Por ejemplo, un vector de expresión puede incluir una secuencia de ácidos nucleicos que codifique el epítipo conectado a seis residuos de histidina seguidos por una tiorredoxina y un sitio de escisión para la enteroquinasa (ver p.ej., Williams (1995) *Biochemistry* 34:1787-1797; Dobeli (1998) *Protein Expr. Purif.* 12:404-14). Los residuos de histidina facilitan la detección y la purificación mientras que el sitio de escisión de la enteroquinasa proporciona una manera para purificar el epítipo del resto de proteínas de fusión. La tecnología relativa a los vectores que codifican las proteínas de fusión y la aplicación de proteínas de fusión están bien descritas en la bibliografía científica y de patentes, véase p. ej., Kroll (1993) *ADN Cell. Biol.*, 12:441-53.

15 *Secuencias de Control de la Transcripción y la Traducción*

La invención proporciona secuencias de ácido nucleico (p. ej., ADN) de la invención conectadas operativamente a una o varias secuencias de control de la expresión (p. ej., transcripcional o traduccional), p. ej., promotores o intensificadores, 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 ilustrativos incluyen *lacI*, *lacZ*, T3, T7, *gpt*, *lambda PR*, *PL* y *trp*. Los promotores eucarióticos ilustrativos incluyen temprano inmediato de CMV, timidina quinasa de HSV, temprano y tardío de SV40, LTR de retrovirus, y metalotioneína I de ratón.

Los promotores adecuados para expresar un polipéptido en bacterias incluyen los promotores *lac* o *trp* de *E. coli*, el promotor *lacI*, el promotor *lacZ*, el promotor T3, el promotor T7, el promotor *gpt*, el promotor *PR* de *lambda*, el promotor *PL* de *lambda*, promotores de operones que codifican enzimas glicolíticas tales como 3-fosfoglicerato quinasa (PGK), y el promotor de la fosfatasa ácida. Los promotores eucarióticos incluyen el promotor temprano inmediato de CMV, el promotor de la timidina quinasa de HSV, promotores de choque térmico, el promotor temprano y tardío de SV40, las LTR de retrovirus, y el promotor de la metalotioneína I de ratón. También se pueden utilizar otros promotores conocidos por controlar la expresión de genes en células procarióticas o eucarióticas o sus virus. Los promotores adecuados para expresar el polipéptido o fragmento del mismo en bacterias incluyen los promotores *lac* o *trp* de *E. coli*, el promotor *lacI*, el promotor *lacZ*, el promotor T3, el promotor T7, el promotor *gpt*, el promotor *P_R* de *Lambda*, el promotor *P_L* de *lambda*, promotores de operones que codifican enzimas glicolíticas tales como 3-fosfoglicerato quinasa (PGK) y el promotor de la fosfatasa ácida. Los promotores fúngicos incluyen el promotor del factor α . Los promotores eucarióticos incluyen el promotor temprano inmediato de CMV, el promotor de la timidina quinasa de HSV, promotores de choque térmico, el promotor temprano y tardío de SV40, las LTR de retrovirus, y el promotor de la metalotioneína I de ratón. También se pueden utilizar otros promotores conocidos por controlar la expresión de genes en células procarióticas o eucarióticas o sus virus.

40 *Promotores Vegetales Específicos de Tejidos*

La invención proporciona casetes de expresión que se pueden expresar de una manera específica del tejido, p. ej., que puede expresar una glucanasa de la invención de una manera específica del tejido. La invención también proporciona plantas o semillas que expresan una glucanasa de la invención de una manera específica del tejido. La especificidad del tejido puede ser específico de la semilla, específico del tallo, específico de la hoja, específico de la raíz, específico del fruto y similares.

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

La invención utiliza promotores específicos de tejido o constitutivos derivados de virus que pueden incluir, p. ej., el promotor subgenómico de tobamovirus (Kumagai (1995) *Proc. Natl. Acad. Sci. USA* 92:1679-1683; el virus

baciliforme tungro del arroz (RTBV), que replica solo en células de floema en plantas de arroz infectadas, con su promotor que dirige la expresión del gen informador específico del floema fuerte; el promotor del virus del mosaico de la nervadura de la yuca (CVMV), con la actividad más alta en los elementos vasculares, en células del mesófilo de las hojas, y en las puntas de las raíces (Verdaguer (1996) *Plant Mol. Biol.* 31:1129-1139).

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

Los promotores específicos de tejido pueden promover la transcripción solo en un cierto marco temporal del estado de desarrollo de ese tejido. Véase, p. ej., Blázquez (1998) *Plant Cell* 10:791-800, que caracteriza el promotor del gen LEAFY de *Arabidopsis*. Véase también Cardon (1997) *Plant J* 12:367-77, que describe el factor de transcripción SPL3, que reconoce un motivo de secuencia conservado en la región promotora del gen AP1 de identidad del meristemo floral de *A. thaliana*; y Mandel (1995) *Plant Molecular Biology*, Vol. 29, págs. 995-1004, que describe el promotor del meristemo eIF4. Se pueden utilizar los promotores específicos de tejidos que son activos a lo largo del ciclo de vida de un tejido concreto. En un aspecto, los ácidos nucleicos de la invención están conectados operablemente a un promotor activo principalmente solo en células de fibra de algodón. En un aspecto, los ácidos nucleicos de la invención están conectados operablemente a un promotor activo principalmente durante las fases de elongación de las células de la fibras de algodón, p. ej., como describe Rinehart (1996) *supra*. Los ácidos nucleicos pueden estar conectados operablemente al promotor del gen Fb12A para ser expresados preferentemente en células de fibra de algodón (Ídem). Véase también, John (1997) *Proc. Natl. Acad. Sci. USA* 89:5769-5773; John, et al., Patentes de los Estados Unidos Núms. 5.608.148 y 5.602.321, que describen promotores específicos de fibra de algodón y métodos para la construcción de plantas de algodón transgénicas. También se pueden utilizar promotores específicos de raíces para expresar los ácidos nucleicos de la invención. Los ejemplos de los promotores específicos de raíces incluyen el promotor del gen de la alcohol deshidrogenasa (DeLisle (1990) *Int. Rev. Cytol.* 123:39-60). Otros promotores que se pueden utilizar para expresar los ácidos nucleicos de la invención incluyen, p. ej., promotores específicos de ovulo, específicos de embrión, específicos de endospermo, específicos de integumento, específicos del revestimiento de la semilla, o alguna de sus combinaciones; un promotor específico de hoja (véase, p. ej., Busk (1997) *Plant J.* 11:1285-1295, que describe un promotor específico de hoja en el maíz); el promotor ORF13 de *Agrobacterium rhizogenes* (que muestra alta actividad en raíces, véase, p. ej., Hansen (1997) *supra*); un promotor específico de polen de maíz (véase, p. ej., Guerrero (1990) *Mol. Gen. Genet.* 224:161-168); se puede utilizar un promotor de tomate activo durante la maduración del fruto, la senescencia y la abscisión de las hojas y, en menor grado, de las flores (véase, p. ej., Blume (1997) *Plant J.* 12:731-746); un promotor específico del pistilo del gen SK2 de la patata (véase, p. ej., Ficker (1997) *Plant Mol. Biol.* 35:425-431); el gen Blec4 del guisante, que es activo en tejido epidérmico de ápices de brotes vegetativos y florales de alfalfa transgénica que lo hacen una herramienta útil para dirigir la expresión de genes foráneos a la capa epidérmica de brotes o fibras que crecen activamente; el gen BEL1 específico de óvulos (véase, p. ej., Reiser (1995) *Cell* 83:735-742, GenBank Núm. U39944); y/o, el promotor de Klee, Patente de los Estados Unidos Núm. 5.589.583, que describe una región promotora vegetal que es capaz de convertir elevados niveles de transcripción en tejido meristemático y/o células que se dividen rápidamente.

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

Los ácidos nucleicos de la invención se pueden conectar también operablemente a promotores vegetales que son inducibles después de la exposición a reactivos químicos que se pueden aplicar a la planta, tales como herbicidas o antibióticos. Por ejemplo, se puede utilizar el promotor In2-2 de maíz, activado por protectores contra herbicidas o bencenosulfonamida, (De Veylder (1997) *Plant Cell Physiol.* 38:568-577); la aplicación de diferentes protectores contra herbicidas induce distintos patrones de expresión génica, incluyendo expresión en la raíz, hidatodos, y el meristemo apical de los brotes. La secuencia codificante puede estar bajo el control de, p. ej., un promotor inducible de tetraciclina, p. ej., como se ha descrito con las plantas de tabaco transgénicas que contienen el gen de la arginina descarboxilasa de *Avena sativa* L. (avena) (Masgrau (1997) *Plant J.* 11:465-473); o, un elemento sensible al ácido salicílico (Stange (1997) *Plant J.* 11:1315-1324). Utilizando promotores inducidos químicamente (p. ej., hormona o plaguicida), es decir, un promotor sensible a un agente químico que se puede aplicar a la planta transgénica en el

campo, la expresión de un polipéptido de la invención se puede inducir en una fase concreta de desarrollo de la planta. De este modo, la invención también proporciona plantas transgénicas que contienen un gen inducible codificante de polipéptidos de la invención cuya gama de anfitriones está limitada a especies de plantas objetivo, tales como maíz, arroz, cebada, trigo, patata u otros cultivos, inducible en cualquier estado de desarrollo del cultivo.

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

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

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

Vectores de expresión y vehículos de clonación

La invención proporciona vectores de expresión y vehículos de clonación que comprenden ácidos nucleicos de la invención, p. ej., secuencias que codifican las glucanasas de la invención. Los vectores de expresión y los vehículos de clonación de la invención pueden comprender partículas virales, baculovirus, fagos, plásmidos, fagémidos, cósmidos, fósmidos, cromosomas artificiales bacterianos, ADN viral (p. ej., vaccinia, adenovirus, virus de la viruela aviar, pseudorrabia y derivados de SV40), cromosomas artificiales basados en P1, plásmidos de levadura, cromosomas artificiales de levadura, y otros vectores cualesquiera específicos para anfitriones de interés específicos (tales como bacillus, Aspergillus y levadura). Los vectores de la invención pueden incluir secuencias de ADN cromosómico, no cromosómico y sintético. Es conocido un gran número de vectores adecuados por los expertos en la técnica, y están disponibles en el mercado. Los vectores ilustrativos incluyen: vectores bacterianos: pQE (Qiagen), plásmidos pBluescript, vectores pNH, (vectores lambda-ZAP (Stratagene); ptrc99a, pKK223-3, pDR540, pRIT2T (Pharmacia); eucarióticos: pXT1, pSG5 (Stratagene), pSVK3, pBPV, pMSG, pSVLSV40 (Pharmacia). Sin embargo, se puede utilizar cualquier otro plásmido u otro vector con tal que sean replicables y viables en el anfitrión. Se pueden emplear vectores con bajo número de copias o alto número de copias con la presente invención.

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

En un aspecto, la vectores de expresión contienen uno o más genes marcadores seleccionables para permitir la selección de células anfitrionas que contienen el vector. Tales marcadores seleccionables incluyen genes que codifican la dihidrofolato reductasa o genes que confieren resistencia a la neomicina para el cultivo de células eucarióticas, genes que confieren resistencia a tetraciclina o ampicilina en *E. coli*, y el gen TRP1 de *S. cerevisiae*. Las regiones promotoras se pueden seleccionar a partir de cualquier gen deseado utilizando vectores de cloranfenicol transferasa (CAT) u otros vectores con marcadores seleccionables.

Los vectores para expresar el polipéptido o fragmento del mismo en células eucarióticas pueden contener también intensificadores para incrementar los niveles de expresión. Los intensificadores son elementos que actúan en cis de ADN, usualmente de alrededor de 10 a alrededor de 300 pb de longitud que actúan sobre un promotor para incrementar su transcripción. Los ejemplos incluyen intensificador de SV40 en el lado tardío del origen de replicación

pb 100 a 270, el intensificador del promotor temprano de citomegalovirus, el intensificador del polioma en el lado tardío del origen de replicación, y los intensificadores de adenovirus.

Una secuencia de ácido nucleico se puede insertar en un vector por medio de una variedad de procedimientos. En general, la secuencia se liga a la posición deseada en el vector después de la digestión del inserto y el vector con endonucleasas de restricción apropiadas. Alternativamente, los extremos romos en el inserto y en el vector se pueden ligar. Una variedad de mecanismos de clonación con conocidas en la técnica, p. ej., como describen Ausubel y Sambrook. Se considera que tales procedimientos y otros se encuentran al alcance de los expertos en la técnica.

El vector puede estar en forma de un plásmido, una partícula viral, o un fago. Otros vectores incluyen secuencias de ADN cromosómico, no cromosómico y sintético, derivados de SV40; plásmidos bacterianos, ADN de fagos, baculovirus, plásmidos de levadura, vectores derivados de combinaciones de plásmidos y ADN de fago, ADN viral tal como vaccinia, adenovirus, virus de la viruela aviar, y pseudorrabia. Una variedad de vectores de clonación y expresión para su uso con anfitriones procarióticos y eucarióticos son descritos, p. ej., por Sambrook.

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

Los ácidos nucleicos de la invención se pueden expresar en casetes de expresión, vectores o virus y expresar transitoriamente o establemente en células vegetales y semillas. Un sistema de expresión transitorio ilustrativo utiliza sistemas de expresión episómicos, p. ej., ARNB viral del virus del mosaico de la coliflor (CaMV) generado en el núcleo mediante transcripción de un mini-cromosoma episómico que contiene ADN superenrollado, véase, p. ej., Covey (1990) Proc. Natl. Acad. Sci. USA 87:1633-1637. Alternativamente, se pueden insertar secuencias codificantes, es decir, todas o sub-fragmentos de secuencias de la invención en un genoma de una célula anfitriona vegetal correspondiente a una parte integrante del ADN cromosómico del anfitrión. Se pueden expresar de esta manera transcritos directos o antisentido. Un vector que comprende las secuencias (p. ej., regiones promotoras o codificantes) de ácidos nucleicos de la invención puede comprender un gen marcador que confiere un fenotipo seleccionable en una célula vegetal o una semilla. Por ejemplo, el marcador puede codificar resistencia a biocidas, particularmente resistencia a antibióticos, tal como resistencia a kanamicina, G418, bleomicina, higromicina, o resistencia a herbicidas, tal como resistencia a clorosulfurón o Basta.

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

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

Los vectores de expresión de la invención pueden incluir también un gen marcador seleccionable para permitir la selección de cepas bacterianas que han sido transformadas, p. ej., genes que vuelven la bacteria resistente a fármacos tales como ampicilina, cloramfenicol, eritromicina, kanamicina, neomicina y tetraciclina. Los marcadores seleccionables pueden incluir también genes biosintéticos, tales como los de las rutas biosintéticas de la histidina, el triptófano y la leucina.

La secuencia de ADN en el vector de expresión está conectada operativamente a una o varias 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* de *lambda* y *trp*. Los promotores eucarióticos incluyen el temprano

inmediato de CMV, timidina quinasa de HSV, temprano y tardío de ate SV40, las LTR de retrovirus y metalotioneína I de ratón. La selección del vector y el promotor apropiados está al nivel de conocimiento práctico 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 puede incluir también secuencias apropiadas para amplificar la expresión. Las regiones promotoras se pueden seleccionar entre cualquier gen deseado utilizando vectores de la cloranfenicol transferasa (CAT) u otros vectores con marcadores seleccionables. Además, los vectores de expresión contienen preferiblemente uno o más genes marcadores seleccionables para proporcionar un rasgo fenotípico para la selección de células anfitrionas transformadas tal como dihidrofolato reductasa o resistencia a la neomicina para un cultivo celular eucariótico, o tal como resistencia a tetraciclina o ampicilina en *E. coli*.

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

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

Además, los vectores de expresión contienen típicamente uno o más genes marcadores seleccionables para permitir la selección de células anfitrionas 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 eucarióticas, genes que confieren resistencia a tetraciclina o ampicilina en *E. coli* y el gen *TRP1* de *S. cerevisiae*.

En algunos aspectos, el ácido nucleico que codifica uno de los polipéptidos de la 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 el aumento de estabilidad o la simplificación de la purificación.

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 del inserto y del 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* 2^a Ed., Cold Spring Harbor Laboratory Press (1989). Se considera que tales procedimientos y otros están al alcance de de los expertos en la técnica.

El vector puede estar, por ejemplo, en forma de un plásmido, una partícula viral, o un fago. Otros vectores incluyen secuencias de ADN cromosómico, no cromosómico y sintético, derivados de SV40; plásmidos bacterianos, ADN de fagos, baculovirus, plásmidos de levadura, vectores derivados de combinaciones de plásmidos y ADN de fagos, ADN viral tal como vaccinia, adenovirus, virus de la viruela aviar y pseudorrabia. Una variedad de vectores de clonación y expresión para su uso con anfitriones procarióticos y eucarióticos son descritos por Sambrook, et al., *Molecular Cloning: A Laboratory Manual*, 2a Ed., Cold Spring Harbor, N.Y., (1989).

Células anfitrionas y célula transformada

La invención también proporciona una célula transformada que comprende una secuencia de ácido nucleico de la invención, p. ej., una secuencia que codifica una glucanasa de la invención, o un vector de la invención. La célula anfitriona puede ser cualquiera de las células anfitrionas familiares para los expertos en la técnica, incluyendo células procarióticas, células eucarióticas, 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*, *Streptomyces*, *Bacillus subtilis*, *Salmonella typhimurium* y diversas especies dentro de los géneros *Pseudomonas*, *Streptomyces*, y *Staphylococcus*. Las células de insecto ilustrativas incluyen S2 de *Drosophila* y Sf9 de *Spodoptera*. Las células animales ilustrativas incluyen CHO, COS o melanoma de Bowes o cualquier línea celular de ratón o humana. La selección de un anfitrión 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. Véase, p. ej., Weising (1988) Ann. Rev. Genet. 22:421-477; Patente de los Estados Unidos Núm. 5.750.870.

- 5 El vector se puede introducir en las células anfitrionas utilizando cualquiera de una variedad de técnicas, incluyendo transformación, transfección, transducción, infección viral, 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)).
- 10 En un aspecto, los ácidos nucleicos o vectores de la invención 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 diana. Los métodos ilustrativos incluyen precipitación con CaPO_4 , fusión de liposomas, lipofección (p. ej., LIPOFECTIN™), electroporación, infección viral, etc. Los ácidos nucleicos candidato se pueden integrar establemente en el genoma de la célula anfitriona (por ejemplo, con introducción retroviral) o pueden existir transitoriamente 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, vectores retrovirales capaces de transfectar tales dianas.
- 20 Cuando sea apropiado, las células anfitrionas construidas se pueden cultivar en medios nutrientes convencionales según sea apropiado para activar los promotores, seleccionar los transformantes o amplificar los genes de la invención. Después de la transformación de una cepa anfitriona adecuada y el crecimiento de la cepa anfitriona hasta una densidad celular apropiada, se puede introducir el promotor seleccionado mediante métodos apropiados (p. ej., 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.
- 25 Las células se pueden cosechar mediante centrifugación, romper mediante métodos físicos o químicos, y el extracto bruto resultante conservar para su purificación adicional. Las células microbianas empleadas para la expresión de las proteínas se pueden romper mediante cualquier método conveniente, incluyendo ciclos de congelación-descongelación, sonicación, rotura mecánica, o uso de agentes de lisis celular. Tales métodos son bien conocidos por los expertos en la técnica. El polipéptido expresado o fragmentos del mismo se pueden recuperar y purificar a partir de cultivos celulares recombinantes mediante métodos que incluyen precipitación con sulfato de amonio o etanol, extracción con ácido, cromatografía de intercambio aniónico o catiónico, cromatografía sobre fosfocelulosa, cromatografía de interacción hidrófoba, cromatografía de afinidad, cromatografía en hidroxilapatita y cromatografía en lectinas. Se pueden utilizar etapas de replegamiento de proteínas, según sea necesario, al completar la configuración del polipéptido. Si se desea, se puede emplear cromatografía líquida de alta resolución (HPLC) para las etapas de purificación finales.
- 30 Los constructos en las células anfitrionas se pueden utilizar de una manera convencional para producir el producto génico codificado por la secuencia recombinante. Dependiendo del anfitrión empleado en un procedimiento de producción recombinante, los polipéptidos producidos por las células anfitrionas que contienen el vector pueden estar glicosilados o pueden estar no glicosilados. Los polipéptidos de la invención pueden también incluir o no incluir un residuo de aminoácido de metionina inicial.
- 40 También se pueden emplear sistemas de traducción sin células para producir un polipéptido de la invención. Los sistemas de traducción sin células pueden utilizar ARNm transcritos a partir de un constructo de ADN que comprende un promotor conectado operablemente a un ácido nucleico que codifica el polipéptido o fragmento del mismo. En algunos aspectos, el constructo de ADN se puede linealizar antes de llevar a cabo una reacción de transcripción *in vitro*. El ARNm transcrito se incuba a continuación con un extracto de traducción sin células apropiado, tal como un extracto de reticulocito de conejo, para producir el polipéptido deseado o un fragmento del mismo.
- 45 Los vectores de expresión pueden contener uno o más genes marcadores seleccionables para proporcionar un rasgo fenotípico para la selección de células anfitrionas transformadas tal como dihidrofolato reductasa o resistencia a la neomicina para un cultivo celular eucariótico, o tal como resistencia a tetraciclina o ampicilina en *E. coli*.
- 50 Las células anfitrionas que contienen los polinucleótidos de interés, p. ej., los 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 utilizado previamente con la célula anfitriona 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 de polinucleótidos que codifica una enzima que tiene la actividad intensificada.
- 60

La invención proporciona un método para expresar en exceso una glucanasa recombinante en una célula que comprende expresar un vector que comprende un ácido nucleico de la invención o, un ácido nucleico que hibrida en condiciones restrictivas con una secuencia de ácido nucleico de la invención. La expresión en exceso se puede efectuar por medio de cualquier método, p. ej., uso de un promotor con alta actividad, un vector dicistrónico o mediante amplificación génica del vector.

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

Los detalles adicionales concernientes a este enfoque se encuentran en la bibliografía pública y/o son conocidos por los expertos en la técnica. En una ilustración no limitante concreta, tal bibliografía disponible al público incluye el documento EP 0659215 (WO 9403612 A1) (Nevalainen et al.); Lapidot, A., Mechaly, A., Shoham, Y., "Overexpression and single-step purification of a thermostable xylanase from *Bacillus stearothermophilus* T-6", J. Biotechnol. Nov 51:259-64 (1996); Lüthi, E., Jasmal, N.B., Bergquist, P.L., "Xylanase from the extremely thermophilic bacterium *Caldocellum saccharolyticum*: overexpression of the gen in *Escherichia coli* and characterization of the gen product", Appl. Environ. Microbiol. Sep 56:2677-83 (1990); y Sung, W.L., Luk, C.K., Zahab, D.M., Wakarchuk, W., "Overexpression of the *Bacillus subtilis* and *circulans* xylanases in *Escherichia coli*", Protein Expr. Purif. Jun 4:200-6 (1993), aunque estas referencias no ilustran las enzimas de la invención de la presente solicitud.

La célula anfitriona puede ser cualquiera de las células anfitrionas familiares para los expertos en la técnica, incluyendo células procarióticas, células eucarióticas, células de mamífero, células de insecto, o células vegetales. Como ejemplos representativos de los anfitriones apropiados, se pueden mencionar: células bacterianas, tales como *E. coli*, *Streptomyces*, *Bacillus subtilis*, *Salmonella typhimurium* y diversas especies de los géneros *Pseudomonas*, *Streptomyces* y *Staphylococcus*, célula fúngica, 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 anfitrión apropiado se encuentra dentro de las capacidades de los expertos en la técnica.

El vector se puede introducir en las células anfitrionas utilizando cualquiera de una variedad de técnicas, incluyendo transformación, transfección, transducción, infección viral, 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)).

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

Las células se pueden cosechar mediante centrifugación, romper mediante métodos físicos o químicos, y el extracto bruto resultante conservar para su purificación adicional. Las células microbianas empleadas para la expresión de las proteínas se pueden romper mediante cualquier método conveniente, incluyendo ciclos de congelación-descongelación, sonicación, rotura mecánica, o uso de agentes de lisis celular. Tales métodos son bien conocidos por los expertos en la técnica. El polipéptido expresado o fragmentos del mismo se pueden recuperar y purificar a partir de cultivos celulares recombinantes mediante métodos que incluyen precipitación con sulfato de amonio o etanol, extracción con ácido, cromatografía de intercambio aniónico o catiónico, cromatografía sobre fosfocelulosa, cromatografía de interacción hidrófoba, cromatografía de afinidad, cromatografía en hidroxilapatita y cromatografía en lectinas. Se pueden utilizar etapas de replegamiento de proteínas, según sea necesario, al completar la configuración del polipéptido. Si se desea, se puede emplear cromatografía líquida de alta resolución (HPLC) para las etapas de purificación finales.

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

Los constructos en las células anfitrionas se pueden utilizar de una manera convencional para producir el producto génico codificado por la secuencia recombinante. Dependiendo del anfitrión empleado en un procedimiento de producción recombinante, los polipéptidos producidos por las células anfitrionas que contienen el vector pueden estar glicosilados o pueden estar no glicosilados. Los polipéptidos de la invención pueden también incluir o no incluir un residuo de aminoácido de metionina inicial.

Alternativamente, 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 de los mismos se pueden producir sintéticamente por medio de sintetizadores peptídicos convencionales. En otros aspectos, se pueden emplear fragmentos o porciones de los polipéptidos para producir el polipéptido completo correspondiente mediante síntesis peptídica; por lo tanto, los fragmentos se pueden emplear como intermedios para la producción de polipéptidos completos.

También se pueden emplear sistemas de traducción sin células para producir 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 de los mismos utilizando los ARNm transcritos a partir de un constructo de ADN que comprende un promotor conectado operablemente a un ácido nucleico que codifica el polipéptido o fragmento del mismo. En algunos aspectos, el constructo de ADN se puede linealizar antes de llevar a cabo una reacción de transcripción in vitro. El ARNm transcrito se incuba a continuación con un extracto de traducción sin células apropiado, tal como un extracto de reticulocito de conejo, para producir el polipéptido deseado o un fragmento del mismo.

Amplificación de Ácidos Nucleicos

Al poner en práctica la invención, los ácidos nucleicos de la invención y los ácidos nucleicos que codifican las glucanasas de la invención, o los ácidos nucleicos modificados de 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 cebadoras de amplificación para amplificar ácidos nucleicos de la invención. Un experto en la técnica puede diseñar pares de secuencias cebadoras de amplificación para cualquier parte de o para toda la longitud de estas secuencias.

Un ácido nucleico puede ser amplificado con un par de cebadores, p. ej., un par de cebadores como se muestra mediante alrededor de los primeros (los 5') 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, o 25 residuos de un ácido nucleico de la invención, y alrededor de los primeros (los 5') 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, o 25 residuos de la hebra complementaria.

Un par de secuencias cebadoras de amplificación para amplificar un ácido nucleico que codifica un polipéptido que tiene una actividad glucanasa, puede ser capaz de amplificar un ácido nucleico que comprende una secuencia de la invención, o fragmentos o subsecuencias del mismo. Uno o cada miembro del par de secuencias cebadoras de amplificación pueden comprender un oligonucleótido que comprende al menos alrededor de 10 a 50 bases consecutivas de la secuencia, o alrededor de 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, o 25 bases consecutivas de la secuencia. Los pares de cebadores de amplificación pueden comprender un primer miembro que tiene una secuencia como la mostrada mediante alrededor de los primeros (los 5') 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, o 25 residuos de un ácido nucleico de la invención, y un segundo miembro que tiene una secuencia como la mostrada mediante alrededor de los primeros (los 5') 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, o 25 residuos de la hebra complementaria del primer miembro. La invención proporciona glucanasas generadas mediante amplificación, p. ej., reacción en cadena de la polimerasa (PCR), utilizando un par de cebadores de amplificación de la invención. Los métodos para elaborar glucanasas mediante amplificación, p. ej., reacción en cadena de la polimerasa (PCR), utilizan tal par de cebadores de amplificación. El par de cebadores de amplificación amplifica un ácido nucleico de una biblioteca, p. ej., una genoteca, tal como una biblioteca ambiental.

Las reacciones de amplificación se pueden utilizar también para cuantificar la cantidad de ácido nucleico en una muestra (tal como la cantidad de mensaje en una muestra celular), marcar el ácido nucleico (p. ej., aplicarlo a una matriz o una transferencia), detectar el ácido nucleico, o cuantificar la cantidad de un ácido nucleico específico en una muestra. El mensaje aislado de una célula o una biblioteca de ADNc se puede amplificar.

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, p. ej., reacción en cadena de la polimerasa, PCR (véase, p. ej., 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, p. ej., Wu (1989) Genomics 4:560; Landegren (1988) Science 241:1077; Barringer (1990) Gene 89:117); amplificación basada en la transcripción (véase, p. ej., Kwok (1989) Proc. Natl. Acad. Sci. USA 86:1173); y, replicación de secuencias autosostenida (véase, p. ej., Guatelli (1990) Proc. Natl. Acad. Sci. USA 87:1874); amplificación de la replicasa Q Beta (véase, p. ej., Smith (1997) J. Clin. Microbiol. 35:1477-1491), análisis de amplificación de la replicasa Q Beta automático (véase, p. ej., Burg (1996) Mol. Cell. Probes 10:257-271) y otras técnicas mediadas por ARN polimerasa (p. ej., NASBA, Cangene, Mississauga, Ontario); véase también Berger

(1987) *Methods Enzymol.* 152:307-316; Sambrook; Ausubel; Patentes de los Estados Unidos Núms. 4.683.195 y 4.683.202; Sooknanan (1995) *Biotechnology* 13:563-564.

Determinación del grado de identidad de secuencia

5 La invención proporciona ácidos nucleicos que comprenden las secuencias que tienen al menos alrededor de 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, o más, o identidad de secuencia completa (100%) con el SEQ ID NO: 37.

10 La invención proporciona polipéptidos que comprenden las secuencias que tienen al menos 90%, 91 %, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, o más, o identidad de secuencia completa (100%) con un polipéptido ilustrativo de la invención. El grado de identidad de secuencia (homología) se puede determinar utilizando cualquier programa informático y parámetros asociados, incluyendo los descritos de la presente memoria, tales como BLAST 2.2.2. o FASTA versión 3.0t78, con los parámetros por defecto.

15 Las secuencias de ácido nucleico de 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 ilustrativa de la invención y secuencias sustancialmente idénticas a éstas. Las secuencias homólogas y los fragmentos de las secuencias de ácido nucleico de la invención pueden hacer 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 de la presente memoria, incluyendo FASTA versión 3.0t78 con los parámetros por defecto. Las secuencias homólogas también incluyen secuencias de ARN en las que las uridinas remplazan a las timinas en las secuencias de ácido nucleico de la invención. Las secuencias homólogas se pueden obtener utilizando 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*, 3^a Ed., W. H Freeman & Co., New York.) o en cualquier otro formato que registre la identidad de los nucleótidos en una secuencia.

20 Se contemplan particularmente para su uso en este aspecto diversos programas de comparación de secuencias identificados en otra parte en esta solicitud de patente. Las homologías de las secuencias de proteínas y/o ácidos nucleicos se pueden evaluar utilizando 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).

25 La homología o la identidad se miden con frecuencia utilizando programas de análisis de secuencias (p. ej., Sequence Analysis Software Package de the 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 delecciones, sustituciones u otras modificaciones. Los términos "homología" e "identidad" en el contexto de dos o más ácidos nucleicos o secuencias de polipéptidos, hacen referencia a dos o más secuencias o subsecuencias que son la misma o tienen un porcentaje especificado de residuos 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 utilizando cualquier número de algoritmos de comparación de secuencias o mediante alineamiento manual e inspección visual.

30 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 utiliza 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 secuencia. Se pueden utilizar parámetros del programa por defecto, o se pueden diseñar parámetros alternativos. El algoritmo de comparación de secuencia 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.

35 Una "ventana de comparación", según se utiliza de la presente memoria, incluye la referencia a un segmento de una cualquiera de las varias posiciones contiguas seleccionadas del grupo que consiste en 20 a 600, usualmente alrededor de 50 a alrededor de 200, más usualmente 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 su comparación son bien conocidos en la técnica. El alineamiento óptimo de secuencias para su comparación se puede llevar a cabo, p. ej., por medio del algoritmo de homología local de Smith & Waterman, *Adv. Appl. Math.* 2:482, 1981, mediante el algoritmo de alineamiento de homología de Needleman & Wunsch, *J. Mol. Biol.* 48:443, 1970, por medio

de la búsqueda para el método de similitud de Pearson & Lipman, Proc. Nat'l. Acad. Sci. USA 85:2444, 1988, por medio de implementaciones computarizadas de estos algoritmos (GAP, BESTFIT, FASTA y TFASTA en The 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 at the National Center for Biological Information "Herramienta de Búsqueda de Alineamiento Local Básico"), ALIGN, AMAS (Analysis of Multiply Aligned Sequences "Análisis de Secuencias Alineadas Multiplicadas"), AMSP (Protein Multiple Sequence Alignment "Alineamiento Múltiple de Secuencias 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 Multisequencia de Patrón Inducido"), SAGA (Sequence Alignment by Genetic Algorithm "Alineamiento de Secuencia mediante Algoritmo Genético") y WHAT-IF. Tales programas de alineamiento se pueden utilizar también para escrutar bases de datos genómicas para identificar secuencias de polinucleótidos que tienen secuencias sustancialmente idénticas. Están disponibles varias bases de datos genómicas, por ejemplo, una porción sustancial del genoma humano está disponibles como parte del Proyecto de Secuenciación del Genoma Humano "Human Genome Sequencing Project" (J. Roach, http://weber.u.washington.edu/~roach/human_genome_progress_2.html) (Gibbs, 1995). Ya se han secuenciado al menos veintinueve 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 realizado progresos significativos en la secuenciación de genomas de organismos modelo, tales como ratón, *C. elegans* y *Arabidopsis sp.* Diversas bases de datos que contienen información genómica anotada con cierta información funcional son mantenidas por diferentes organizaciones y son accesibles a través de Internet.

Un ejemplo de un algoritmo útil son los algoritmos BLAST y BLAST 2.0, que son descritos 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 realizar los análisis BLAST están disponibles al público a través del National Center for Biotechnology Information. Este algoritmo implica identificar en primer lugar pares de secuencia 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.*, *supra*). Estos éxitos con la palabra vecina inicial actúan como semillas para iniciar búsquedas para encontrar HSP más largos que las contienen. Los éxitos con las palabras se prolongan en ambas direcciones a lo largo de cada secuencia tanto como aumente la puntuación de alineamiento cumulativa. Las puntuaciones cumulativas se calculan utilizando, para las secuencias de nucleótidos, los parámetros M (la puntuación de recompensa para un par de residuos emparejados; siempre >0). Para las secuencias de aminoácidos, se utiliza una matriz de puntuación para calcular la puntuación cumulativa. La prolongación de los éxitos con las palabras en cada dirección se detiene cuando: la puntuación de alineamiento cumulativa disminuye en la cantidad X de su valor máximo alcanzado; la puntuación cumulativa tiende a cero o menos, debido a la acumulación de uno o más alineamientos de residuos 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 de nucleótidos) utiliza 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 utiliza como parámetros por defecto una longitud de palabra de 3 y expectativas (E) de 10 y la matriz de puntuación BLOSUM62 (véase Henikoff & Henikoff, Proc. Natl. Acad. Sci. USA 89:10915, 1992) 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, p. ej., Karlin & Altschul, Proc. Natl. Acad. Sci. USA 90:5873, 1993). Una medida de similitud proporcionada por el algoritmo BLAST es la probabilidad de la suma mínima ($P(N)$), que proporciona una indicación de la probabilidad por la cual ocurriría por casualidad un emparejamiento entre dos secuencias de nucleótidos o de aminoácidos. Por ejemplo, un ácido nucleico se considera similar a una secuencia de referencia si la probabilidad de la suma mínima en una comparación del ácido nucleico de ensayo con respecto al ácido nucleico de referencia es menor de alrededor de 0,2, más preferiblemente menor de alrededor de 0,01 y muy preferiblemente menor de alrededor de 0,001.

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

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

151. Los programas BLAST identifican secuencias homólogas identificando segmentos similares, que son referidos en la presente memoria 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 ácido nucleico. Los pares de segmentos de alta puntuación se identifican preferiblemente (*es decir*, se alinean) por medio de una matriz de puntuación, muchas de las cuales son conocidas en la técnica. Preferiblemente, la matriz de puntuación utilizada es la matriz BLOSUM62 (Gonnet et al., Science 256:1443-1445, 1992; Henikoff y Henikoff, Proteins 17:49-61, 1993). Menos preferiblemente, también se pueden utilizar las matrices PAM o PAM250 (véase, *p. ej.*, 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 utilizados con los algoritmos anteriores se pueden adaptar dependiendo de la longitud de la secuencia y el grado de homología estudiado. En algunos aspectos, los parámetros pueden ser los parámetros por defecto utilizados por los algoritmos en ausencia de instrucciones para el usuario.

30 **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 polipéptido de la invención se puede almacenar, registrar, y manipular sobre cualquier medio que pueda ser leído y al que se pueda acceder mediante un ordenador.

Los ordenadores, los sistemas informáticos, los medios legibles por ordenador, los productos de programas informáticos y similares han registrado o almacenado el ácido nucleico y las secuencias de polipéptidos de la invención. Según se utiliza en la presente memoria, 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 ácido nucleico y/o de polipéptidos de la invención.

Los polipéptidos de la invención incluyen las secuencias de polipéptidos de la invención, *p. ej.*, las secuencias ilustrativas de la invención, y secuencias sustancialmente idénticas a las mismas, y fragmentos de cualquiera de las secuencias precedentes. Las secuencias de polipéptidos sustancialmente idénticas, u homólogas, hacen referencia a una secuencia de polipéptidos que tiene al menos una identidad de secuencia 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 completa (100%) con una secuencia ilustrativa de la invención.

La homología se puede determinar utilizando cualquiera de los programas informáticos y los parámetros descritos de la presente memoria, incluyendo FASTA versión 3.0t78 con los parámetros por defecto o con parámetros modificados cualesquiera. Las secuencias homólogas se pueden obtener utilizando 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 del Grupo B y secuencias esencialmente idénticas a éstas, 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., New York.) o en cualquier otro formato que refiera la identidad de los polipéptidos en una secuencia.

Una secuencia de ácido nucleico o polipéptido de la invención se puede almacenar, registrar, y manipular sobre cualquier medio que pueda ser leído y al que se pueda acceder mediante un ordenador. Según se utiliza en la

5 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 ácido nucleico de la invención, una o más de las secuencias de polipéptidos de la invención. Un medio legible por ordenador puede tener registradas al menos 2, 5, 10, 15, o 20 o más secuencias de ácido nucleico de la invención.

10 Un medio legible por ordenador pueden tener registradas una o más de las secuencias de ácido nucleico como se muestra en la invención. Un medio legible por ordenador pueden tener registradas una o más de las secuencias de polipéptidos como se muestra en las secuencias de aminoácidos del Grupo B y secuencias esencialmente idénticas a éstas. Un medio legible por ordenador puede tener registradas al menos 2, 5, 10, 15, o 20 o más de las secuencias como se ha mostrado anteriormente.

15 Los medios legibles por ordenador incluyen medios legibles magnéticamente, medios legibles ópticamente, medios legibles electrónicamente y medios magnéticos/ópticos. Por ejemplo, el 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.

20 Los sistemas (*p. ej.*, sistemas basados en Internet) incluyen sistemas informáticos que almacenan y manipulan la información de la secuencia descrita en la presente memoria. Un ejemplo de un sistema informático 100 se ilustra en forma de diagrama de bloques en la Figura 1. Según se utiliza en la presente memoria, "un sistema informático" hace referencia a los componentes del equipo físico, los componentes de soporte lógico y los componentes de almacenamiento de datos utilizados para analizar una secuencia de nucleótidos de una secuencia de ácido nucleico de la invención, o una secuencia de polipéptidos de la invención. El sistema informático 100 incluye típicamente un procesador para procesar, acceder y manipular los datos de secuencia. El procesador 105 puede ser cualquier 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.

30 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 son adecuados cualquiera de los sistemas informáticos disponibles en la actualidad.

35 El sistema informático 100 puede incluir un procesador 105 conectado a un bus que está conectado a una memoria principal 115 (en un aspecto implementada como RAM) y uno o más dispositivos de almacenamiento de datos internos 110, tales como un disco duro y/u otros medios legibles por ordenador que tienen datos registrados en el mismo. El sistema informático 100 puede incluir adicionalmente uno o más dispositivos de recuperación de datos 40 118 para leer los datos almacenados en los dispositivos de almacenamiento de datos internos 110.

45 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 (*p. ej.*, vía Internet) etc. El dispositivo de almacenamiento de datos interno 110 puede ser un medio legible por ordenador extraíble tal como un disco flexible, un disco compacto, una cinta magnética, etc. que contiene registrados lógica de control y/o datos. El sistema informático 100 puede incluir ventajosamente o 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.

50 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.

55 El soporte lógico para acceder a, y procesar, las secuencias de nucleótidos de una secuencia de ácido nucleico de la invención, (tales como las herramientas de búsqueda, las herramientas de comparación y las herramientas de modelado etc.) puede residir en la memoria principal 115 durante la ejecución.

60 En algunos aspectos, el sistema informático 100 puede comprender adicionalmente un algoritmo de comparación de secuencia para comparar una secuencia de ácido nucleico de la invención, o una secuencia de polipéptidos de la invención, almacenado en un medio legible por ordenador con una o varias secuencias de nucleótidos o polipéptidos de referencia almacenadas en un medio legible por ordenador. Un "algoritmo de comparación de secuencia" hace referencia a uno o más programas que son implementados (localmente o remotamente) en el sistema informático 100 para comparar una secuencia de nucleótidos con otras secuencias de nucleótidos y/o compuestos almacenados en los medios de almacenamiento de datos. Por ejemplo, el algoritmo de comparación de secuencia puede

comparar las secuencias de nucleótidos de una secuencia de ácido nucleico como de la invención, o una secuencia de polipéptidos 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.

5 La Figura 2 es un diagrama de flujo que ilustra un aspecto de un procedimiento 200 para comparar un nuevo nucleótido o secuencia de proteínas con una base de datos de secuencias con el fin de determinar los niveles de homología entre la nueva secuencia y las secuencias de la base de datos. La base de datos de secuencias puede ser una base de datos privada almacenada en el sistema informático 100, o una base de datos pública GENBANK que está disponible a través de Internet.

10 El procedimiento 200 comienza en un estado de inicio 201 y a continuación se mueve a un estado 202 donde la nueva secuencia que se va a comparar se almacena en una memoria en un sistema informático 100. Como se ha comentado anteriormente, la memoria podría ser cualquier tipo de memoria, incluyendo RAM o un dispositivo de almacenamiento interno.

15 El procedimiento 200 se mueve a continuación a un estado 204 donde una base de datos de secuencias se abre para el análisis y la comparación. El procedimiento 200 se mueve a continuación a un estado 206 donde la primera secuencia almacenada en la base de datos se lee en una memoria en el ordenador. A continuación se realiza 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 la realización de una comparación exacta entre la nueva secuencia y la primera secuencia en la base de datos. Los métodos bien conocidos son conocidos por los expertos en la técnica para comparar dos secuencias de nucleótidos o proteínas, incluso si no son idénticas. Por ejemplo, se pueden introducir espacios en una secuencia con el fin de aumentar el nivel de homología entre las dos secuencias sometidas a ensayo. Los parámetros que controlan si se introducen espacios u otras características en una secuencia durante la comparación son insertados normalmente por el usuario del sistema informático.

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

30 Si se realiza una determinación de que las dos secuencias son iguales, el procedimiento 200 se mueve a un estado 214 donde 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 habían introducido. Una vez que el nombre de la secuencia almacenada es presentado al usuario, el procedimiento 200 se mueve a un estado de decisión 218 donde se toma una determinación si existen más secuencias en la base de datos. Si no existen más secuencias en la base de datos, el procedimiento 200 termina en un estado final 220. Sin embargo, si existen más secuencias en la base de datos, el procedimiento 200 se mueve a un estado 224 donde se mueve un puntero hasta la siguiente secuencia en la base de datos de manera que ésta se puede comparar con la nueva secuencia. De esta manera, la nueva secuencia se alinea y compara con cada secuencia de la base de datos.

40 Se debe observar que si se ha tomado la determinación en el estado de decisión 212 de que las secuencias no eran homólogas, el procedimiento 200 se movería inmediatamente al estado de decisión 218 con el fin de determinar si estarían disponibles otras secuencias cualesquiera en la base de datos para su comparación.

45 Por lo tanto, un sistema informático puede comprender un procesador, un dispositivo de almacenamiento de datos que tiene almacenada una secuencia de ácido nucleico de la invención, o una secuencia de polipéptidos de la invención, un dispositivo de almacenamiento de datos que tiene almacenadas recuperablemente secuencias de nucleótidos o secuencias de polipéptidos de referencia que se van a comparar con una secuencia de ácido nucleico de la invención, o una secuencia de polipéptidos de la invención 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 identificar motivos estructurales en el código de ácido nucleico descrito anteriormente una secuencia de ácido nucleico de la invención, o una secuencia de polipéptido, o puede identificar motivos estructurales en las secuencias que son comparadas con estos códigos de ácidos nucleicos y códigos de polipéptidos. El dispositivo de almacenamiento de datos pueden tener almacenadas las secuencias de al menos 2, 5, 10, 15, 20, 25, 30 o 40 o más de las secuencias de ácido nucleico de la invención, o las secuencias de polipéptidos de la invención.

50 Un método para determinar el nivel de homología entre una secuencia de ácido nucleico de la invención, o una secuencia de polipéptido de la invención, y una secuencia de nucleótidos de referencia puede incluir leer el código de ácido nucleico o el código de polipéptido y la secuencia de nucleótidos o polipéptidos 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 del polipéptido y la secuencia de nucleótidos o polipéptidos de referencia con el programa informático. El programa informático puede ser cualquiera de varios programas informáticos para determinar niveles de homología, incluyendo los enumerados específicamente de la presente memoria, (p. ej.,

BLAST2N con los parámetros por defecto o con parámetros modificados cualesquiera). El método se puede implementar utilizando los sistemas informáticos descritos anteriormente. El método se puede realizar también leyendo al menos 2, 5, 10, 15, 20, 25, 30 o 40 o más de la secuencias de ácido nucleico de la invención descritas anteriormente, o las secuencias de polipéptidos de la invención a través del uso del programa informático y determinar homología entre el código de ácido nucleico o el código de polipéptidos y las secuencias de nucleótidos o las secuencias de polipéptidos de referencia.

La Figura 3 es un diagrama de flujo que ilustra un aspecto de un procedimiento 250 en un ordenador para determinar si dos secuencias son homólogas. El procedimiento 250 comienza en un estado de inicio 252 y se mueve a continuación a un estado 254 donde una primera secuencia que se va a comparar se almacena en una memoria. La segunda secuencia que se va a comparar se almacena a continuación en una memoria en un estado 256. El procedimiento 250 se mueve a continuación a un estado 260 donde el primer carácter en la primera secuencia se lee y a continuación a un estado 262 donde se lee el primer carácter de la segunda secuencia. Se debe entender que si la secuencia es una secuencia de nucleótidos, el carácter sería normalmente A, T, C, G o U. Si la secuencia es una secuencia de proteínas, ésta está preferiblemente en el código de aminoácidos de una sola letra de manera que se pueden comprar fácilmente la primera y la segunda secuencias.

A continuación se realiza una determinación en un estado de decisión 264 si los dos caracteres son iguales. Si son iguales, el procedimiento 250 se mueve a un estado 268 donde se leen los siguientes caracteres en la primera y segunda secuencias. A continuación se realiza una determinación de si los siguientes caracteres son iguales. Si lo son, el procedimiento 250 continúa este bucle hasta que dos caracteres no son iguales. Si se toma la determinación de que los dos caracteres siguientes no son iguales, el procedimiento 250 se mueve a un estado de decisión 274 para determinar si hay algún otro carácter que cualquier secuencia lea.

Si no hay ningún carácter más para leer, el procedimiento 250 se mueve a un estado 276 donde el nivel de homología entre la primera y la segunda secuencias se presenta al usuario. El nivel de homología se determina calculando la proporción de caracteres entre las secuencias que fueron iguales del número total de secuencias en la primera secuencia. De este modo, si cada carácter en una primera secuencia de 100 nucleótido se alineó con cada carácter en una segunda secuencia, el nivel de homología sería 100%.

Alternativamente, el programa informático puede ser un programa informático que compara las secuencias de nucleótidos de una secuencia de ácido nucleico como se muestra de la invención, con una o más secuencias de nucleótidos 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 a una secuencia del 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 como se muestra en la invención, contiene un polimorfismo de un solo nucleótido (SNP) con respecto a una secuencia de nucleótidos de referencia.

Un método para determinar si una secuencia de ácido nucleico de la invención, difiere en uno o más nucleótidos de una secuencia de nucleótidos de referencia comprende las etapas de leer el código de ácido nucleico y la secuencia de nucleótidos de referencia a través del uso de un programa informático que identifica diferencias entre secuencias de ácido nucleico e identificar diferencias entre el código de ácido nucleico y la secuencia de nucleótidos de referencia con el programa informático. En algunos aspectos, el programa informático es un programa que identifica polimorfismos de un solo nucleótido. El método se puede implementar mediante los sistemas informáticos descritos anteriormente y el método ilustrado en la Figura 3. El método se puede realizar también leyendo al menos 2, 5, 10, 15, 20, 25, 30, o 40 o más de las secuencias de ácido nucleico de la invención y las secuencias de nucleótidos de referencia a través del uso del programa informático e identificar diferencias entre el código de ácidos nucleicos y las secuencias de nucleótidos de referencia con el programa informático.

El sistema informático puede comprender adicionalmente un identificador para identificar características dentro de una secuencia de ácido nucleico de la invención o una secuencia de polipéptidos de la invención.

Un "identificador" hace referencia a uno o más programas que identifican ciertas características dentro de una secuencia de ácido nucleico de la invención, o una secuencia de polipéptidos de la invención. El identificador puede comprender un programa que identifica un marco de lectura abierto en una secuencia de ácido nucleico de la invención.

La Figura 4 es un diagrama de flujo que ilustra un aspecto de un procedimiento identificador 300 para detectar la presencia de una característica en una secuencia. El procedimiento 300 comienza en un estado de inicio 302 y se mueve a continuación a un estado 304 donde una primera secuencia que se va a verificar para determinar sus características se almacena en una memoria 115 en el sistema informático 100. El procedimiento 300 se mueve a continuación a un estado 306 donde se abre una base de datos de características de secuencia. Tal base de datos incluiría una lista de cada uno de los atributos de las características junto con el nombre de la característica. Por

ejemplo, un nombre de característica podría ser "Codón de Inicio" y el atributo sería "ATG". Otro ejemplo sería el nombre de la característica "Caja TAATAA" y el atributo de la característica sería "TAATAA". Un ejemplo de tal base de datos es producido por el University of Wisconsin Genetics Computer Group. Alternativamente, las características pueden ser motivos polipeptídicos estructurales tales como hélices alfa, láminas beta, o motivos polipeptídicos funcionales tales como sitios activos enzimáticos, motivos hélice-giro-hélice o otros motivos conocidos por los expertos en la técnica.

Una vez que la base de datos de características se abre en el estado 306, el procedimiento 300 se mueve a un estado 308 donde la primera característica se lee a partir de la base de datos. A continuación se realiza una comparación del atributo de la primera característica con la primera secuencia en un estado 310. Después se realiza una determinación en un estado de decisión 316 si se encontró el atributo de la característica en la primera secuencia. Si se encontró el atributo, el procedimiento 300 se mueve a un estado 318 donde se presenta al usuario el nombre de la característica encontrada.

El procedimiento 300 se mueve a continuación a un estado de decisión 320 donde se realiza una determinación de si existen características de desplazamiento en la base de datos. Si no existen más características, el procedimiento 300 termina en un estado final 324. Sin embargo, si existen más características en la base de datos, el procedimiento 300 lee la siguiente característica de la secuencia en un estado 326 y vuelve al estado 310 donde el atributo de la siguiente característica se compara frente a la primera secuencia. Se debe observar que si no se encuentra el atributo de la característica en la primera secuencia en el estado de decisión 316, el procedimiento 300 se mueve directamente al estado de decisión 320 con el fin de determinar si existe cualquier otra característica en la base de datos.

Un método de identificación de una característica dentro de una secuencia de ácido nucleico de la invención, o una secuencia de polipéptidos de la invención, comprende leer el código o los códigos del ácido o los ácidos nucleicos o el código o los códigos del polipéptido o los polipéptidos a través del uso de un programa informático que identifica características e identificar características en el código o los códigos del ácido o los ácidos nucleicos con el programa informático. El programa informático puede comprender un programa informático que identifica marcos de lectura abiertos. El método se puede realizar 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 de polipéptidos de la invención, a través del uso del programa informático e identificar características dentro de los código de ácido nucleico o los código de polipéptidos con el programa informático.

Una secuencia de ácido nucleico de la invención, o una secuencia de polipéptidos de la invención, se puede almacenar y manipular en una variedad de programas de tratamiento de datos en una variedad de formatos. Por ejemplo, una secuencia de ácido nucleico de la invención, o una secuencia de polipéptidos de la invención, se pueden almacenar en forma de texto en un archivo de tratamiento de texto, tal como Microsoft WORD™ o WORDPERFECT™ o en forma de un archivo ASCII en una variedad de programas de base de datos familiares para los expertos en la técnica, tales como DB2™, SYBASE™, o ORACLE™. Además, muchos programas informáticos y bases de datos se pueden utilizar como algoritmo de comparación de secuencias, identificadores, o fuentes de secuencias de nucleótidos o secuencias de polipéptidos de referencia que se van a comparar con una secuencia de ácido nucleico de la invención, o una secuencia de polipéptidos de la invención. Se pretende que la siguiente lista no limite la invención pero proporcione apoyo a los programas y bases de datos que son útiles con las secuencias de ácido nucleico de la invención, o las secuencias de polipéptidos de la invención.

Los programas y bases de datos que se pueden utilizar incluyen, pero no están limitados a: MacPattem (EMBL), DiscoveryBase (Molecular Applications Group), GeneMine (Molecular Applications Group), Look (Molecular Applications Group), MacLook (Molecular Applications Group), BLAST y BLAST2 (NCBI), BLASTN y BLASTX (Altschul et al, J. Mol. Biol. 215: 403, 1990), FASTA (Pearson y Lipman, Proc. Natl. Acad. Sci. USA, 85: 2444, 1988), FASTDB (Brutlag et al. Comp. Apágs. Biosci. 6:237-245, 1990), Catalyst (Molecular Simulations Inc.), Catalyst/SHAPE (Molecular Simulations Inc.), Cerius². DBAccess (Molecular Simulations Inc.), HypoGen (Molecular Simulations Inc.), Insight II, (Molecular Simulations Inc.), Discover (Molecular Simulations Inc.), CHARMM (Molecular Simulations Inc.), Felix (Molecular Simulations Inc.), DelPhi, (Molecular Simulations Inc.), QuanteMM, (Molecular Simulations Inc.), Homology (Molecular Simulations Inc.), Modeler (Molecular Simulations Inc.), ISIS (Molecular Simulations Inc.), Quanta/Protein Design (Molecular Simulations Inc.), WebLab (Molecular Simulations Inc.), WebLab Diversity Explorer (Molecular Simulations Inc.), Gene Explorer (Molecular Simulations Inc.), SeqFold (Molecular Simulations Inc.), la base de datos MDL Available Chemicals Directory, la base de datos MDL Drug Data Report, la base de datos Comprehensive Medicinal Chemistry, la base de datos Derwent's World Drug Index, la base de datos BioByteMasterFile, la base de datos Genbank y la base de datos Genseq. Muchos otros programas y bases de datos resultarían evidentes para los expertos en la técnica dada la presente descripción.

Los motivos que se pueden detectar utilizando los programas anteriores incluyen las secuencias que codifican cremalleras de leucina, motivos hélice-giro-hélice, sitios de glicosilación, 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 secuencias homeobox, tramos ácidos, sitios activos enzimáticos, sitios de unión al sustrato y sitios de escisión enzimática.

Hibridación de ácidos nucleicos

5 La invención proporciona ácidos nucleicos aislados, sintéticos o recombinantes que hibridan en condiciones restrictivas con el SEQ ID NO: 37.

10 Las condiciones restrictivas pueden ser condiciones altamente restrictivas, condiciones restrictivas intermedias y/o condiciones poco restrictivas, incluyendo las condiciones altamente restrictivas y poco restrictivas descritas en la presente memoria. En un aspecto, es el rigor de las condiciones de lavado el que establece las condiciones que determinan si un ácido nucleico se encuentra dentro del alcance de la invención, como se comenta más abajo.

15 En aspectos alternativos, los ácidos nucleicos de la invención definidos por su capacidad para hibridar en condiciones restrictivas pueden estar entre alrededor de cinco residuos y la longitud completa del ácido nucleico de la invención; p. ej., 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, residuos de longitud. También se incluyen los ácidos nucleicos más cortos que la longitud completa. Estos ácidos nucleicos pueden ser útiles, p. ej., como sondas de hibridación, sondas de marcaje, sondas oligonucleotídicas para PCR, ARNi (de hebra sencilla o doble), secuencias antisentido o secuencias que codifican péptidos de unión a anticuerpos (epítomos), motivos, sitios activos (dominios catalíticos (CD)) y similares.

20 En un aspecto, los ácidos nucleicos de la invención se definen por su capacidad para hibridar en condiciones altamente restrictivas que comprenden condiciones en formamida de alrededor de 50% a alrededor de 37°C a 42°C. En un aspecto, los ácidos nucleicos de la invención se definen por su capacidad para hibridar en condiciones poco restrictivas que comprenden condiciones en formamida de alrededor de 35% a 25% a alrededor de 30°C a 35°C.

30 Alternativamente, ácidos nucleicos de la invención se definen por su capacidad para hibridar en condiciones altamente restrictivas que comprenden condiciones a 42°C en formamida al 50%, 5X SSPE, SDS al 0,3%, y una secuencia repetitiva que bloquea el ácido nucleico, tal como cot-1 o ADN de esperma de salmón (p. ej., 200 n/ml ADN de esperma de salmón sometido a cizalladura y desnaturalizado). En un aspecto, ácidos nucleicos de la invención se definen por su capacidad para hibridar en condiciones poco restrictivas que comprenden formamida al 35% a una temperatura reducida de 35°C.

35 En las reacciones de hibridación de ácido nucleico, las condiciones utilizadas para lograr un nivel de restricción concreto variarán, dependiendo de la naturaleza de los ácidos nucleicos que estén hibridando. Por ejemplo, la longitud, el grado de complementariedad, la composición de la secuencia de nucleótidos (p. ej., contenido de GC vs. AT) y el tipo de ácido nucleico (p. ej., ARN vs. ADN) de las regiones hibridantes de los ácidos nucleicos se pueden considerar al seleccionar las condiciones de hibridación. Una consideración adicional es si uno de los ácidos nucleicos está inmovilizado, por ejemplo, sobre un filtro.

40 La hibridación se puede llevar a cabo en condiciones poco restrictivas, moderadamente restrictivas o altamente restrictivas. Como ejemplo de una hibridación de ácido nucleico, una membrana polimérica que contiene ácidos nucleicos desnaturalizados inmovilizados se prehibrida en primer lugar durante 30 minutos a 45°C en una solución que consiste en NaCl 0,9 M, NaH₂PO₄ 50 mM, pH 7,0, Na₂EDTA 5,0 mM, SDS al 0,5%, 10X Denhardt y ácido polirriboadenílico de 0,5 mg/ml. Después se añaden a la solución aproximadamente 2X10⁷ cpm (actividad específica 4-9 X10⁸ cpm/ug) de sonda oligonucleotídica marcada en el extremo con P³². Al cabo de 12-16 horas de incubación, la membrana se lava durante 30 minutos a temperatura ambiente en 1X SET (NaCl 150 mM, Tris-hidrocloreuro 20 mM, pH 7,8, Na₂EDTA 1 mM) que contiene SDS al 0,5%, seguido de un lavado de 30 minutos en 1X SET de nueva aportación a T_m-10°C para la sonda oligonucleotídica. La membrana se expone a continuación a una película autorradiográfica para la detección de las señales de hibridación.

Se podría considerar que todas las hibridaciones precedentes fueron en condiciones altamente restrictivas.

55 Después de la hibridación, se puede lavar un filtro para eliminar cualquier sonda detectable no unida específicamente. Las condiciones restrictivas utilizadas para lavar los filtros se pueden variar dependiendo de la naturaleza de los ácidos nucleicos que estén siendo hibridados, la longitud de los ácidos nucleicos que estén siendo hibridados, el grado de complementariedad, la composición de la secuencia de nucleótidos (p. ej., contenido de GC vs. AT) y el tipo de ácido nucleico (p. ej., ARN vs. DNA). Los ejemplos de los lavados con condiciones restrictivas progresivamente altas son los siguientes: 2X SSC, SDS al 0,1% a temperatura ambiente durante 15 minutos (condiciones poco restrictivas); 0,1X SSC, SDS al 0,5% a temperatura ambiente durante 30 minutos a 1 hora (condiciones moderadamente restrictivas); 0,1X SSC, SDS al 0,5% durante 15 a 30 minutos entre la temperatura de hibridación y 68°C (condiciones altamente restrictivas); y NaCl 0,15 M durante 15 minutos a 72°C (condiciones muy restrictivas). Se puede llevar a cabo un lavado final en condiciones restrictivas en 0,1X SSC a temperatura ambiente.

Los Ejemplos anteriores son meramente ilustrativos de un conjunto de condiciones que se pueden utilizar para lavar filtros. Un experto en la técnica conocerá que existen numerosas fórmulas para los lavados en diferentes condiciones restrictivas. Algunos otros Ejemplos se proporcionan más abajo.

- 5 Los ácidos nucleicos que han hibridado con la sonda se identifican mediante autorradiografía u otras técnicas convencionales.

10 El procedimiento anterior se puede modificar para identificar ácidos nucleicos que tienen niveles decrecientes de homología con la secuencia de la sonda. Por ejemplo, para obtener ácidos nucleicos de homología decreciente con la sonda detectable, se pueden utilizar condiciones menos restrictivas. Por ejemplo, la temperatura de hibridación puede disminuir en incrementos de 5°C de 68°C a 42°C en un tampón de hibridación que tiene una concentración de Na⁺ de aproximadamente 1M. Después de la hibridación, el filtro se puede lavar con 2X SSC, SDS al 0,5% a la temperatura de hibridación. Se considera que estas condiciones son condiciones "moderadas" por encima de 50°C y condiciones "bajas" por debajo de 50°C. Un ejemplo específico de condiciones de hibridación "moderadas" se produce cuando la hibridación anterior se lleva a cabo a 55°C. Un ejemplo específico de condiciones de hibridación "poco restrictivas" se produce cuando la hibridación anterior se lleva a cabo a 45°C.

20 Alternativamente, la hibridación se puede llevar a cabo en tampones, tales como 6X SSC, que contienen formamida a una temperatura de 42°C. En este caso, la concentración de formamida en el tampón de hibridación se puede reducir en incrementos de 5% de 50% a 0% para identificar los clones que tienen niveles decrecientes de homología con la sonda. Después de la hibridación, el filtro se puede lavar con 6X SSC, SDS al 0,5% a 50°C. Se considera que estas condiciones son condiciones "moderadas" por encima de formamida al 25% y condiciones "bajas" por debajo de formamida al 25%. Un ejemplo específico de condiciones de hibridación "moderadas" se produce cuando la hibridación anterior se lleva a cabo en formamida al 30%. Un ejemplo específico de condiciones de hibridación "poco restrictivas" se produce cuando la hibridación anterior se lleva a cabo en formamida al 10%.

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

40 Estos métodos se pueden utilizar para aislar los ácidos nucleicos de la invención. Por ejemplo, los métodos precedentes se pueden utilizar para aislar ácidos nucleicos que tienen una secuencia con una homología de al menos alrededor de 97%, al menos 95%, al menos 90%, 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 éstas. La homología se puede medir utilizando 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 en la presente memoria. Tales variantes alélicas pueden tener una sustitución, delección o adición de uno o más nucleótidos cuando se comparan con los ácidos nucleicos de la invención.

50 Adicionalmente, los procedimientos anteriores se pueden utilizar para aislar ácidos nucleicos que codifican polipéptidos que tienen una homología de al menos alrededor de 99%, 95%, al menos 90% 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 las mismas según se determina utilizando un algoritmo de alineamiento de secuencias (p. ej., tal como el algoritmo FASTA versión 3.0t78 con los parámetros por defecto).

55 **Sondas Oligonucleotídicas y métodos para su utilización**

60 La invención también proporciona sondas de ácidos nucleicos que se pueden usar, p. ej., para identificar ácidos nucleicos que codifican un polipéptido con una actividad glucanasa o fragmentos del mismo o para identificar genes de glucanasa. En un aspecto, la sonda comprende al menos 10 bases consecutivas de un ácido nucleico de la invención. Alternativamente, una sonda de la invención 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. Las sondas se pueden utilizar en matrices de la invención, véase la discusión de más abajo, incluyendo, p. ej., matrices capilares. Las sondas de la invención también se pueden utilizar para aislar otros ácidos nucleicos o polipéptidos.

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

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

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

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

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

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

60 Los ácidos nucleicos aislados de la invención, las secuencias complementarias a éstas, o un fragmento que comprende al menos 10, 15, 20, 25, 30, 35, 40, 50, 75, 100, 150, 200, 300, 400, o 500 bases consecutivas de una de las secuencias de la invención, o las secuencias complementarias a éstas se pueden utilizar como sondas para identificar y aislar ácidos nucleicos relacionados. En algunos aspectos, los ácidos nucleicos relacionados pueden ser ADNc o ADN genómicos de organismos distintos del que se aislado el ácido nucleico. Por ejemplo, los otros organismos pueden ser organismos relacionados. En tales procedimientos, un ácido nucleico muestra se pone en contacto con la sonda en condiciones que permitan a la sonda hibridar específicamente con secuencias

relacionadas. La hibridación de la sonda con ácidos nucleicos del organismo relacionado se detecta a continuación utilizando cualquiera de los métodos descritos anteriormente.

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

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

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

20 La prehibridación se puede llevar a cabo en 6X SSC, 5X reactivo de Denhardt, SDS al 0,5%, 100 µg de ADN de esperma de salmón fragmentado desnaturalizado o 6X SSC, 5X reactivo de Denhardt, SDS al 0,5%, 100 µg de ADN de esperma de salmón fragmentado desnaturalizado, formamida al 50%. Las fórmulas para las soluciones de SSC y Denhardt son enumeradas por Sambrook *et al.*, *supra*.

25 La hibridación se lleva a cabo añadiendo la sonda detectable a las soluciones de prehibridación enumeradas anteriormente. Cuando la sonda comprende ADN de doble hebra, este se desnaturaliza antes de su adición a la solución de hibridación. El filtro se pone en contacto con la solución de hibridación durante un período de tiempo suficiente para permitir que la sonda hibride con los ADNc o los ADN genómicos que contienen secuencias complementarias a estas y homólogas a estas. Para sondas de más de 200 nucleótidos de longitud, la hibridación se puede llevar a cabo a 15-25°C por debajo de la T_m . Para sondas más cortas, tales como las sondas oligonucleotídicas, la hibridación se puede llevar a cabo a 5-10°C por debajo de la T_m . Típicamente, para las hibridaciones en 6X SSC, la hibridación se lleva a cabo a aproximadamente 68°C. Usualmente, para las hibridaciones en soluciones que contienen formamida al 50%, la hibridación se lleva a cabo a aproximadamente 42°C.

35 **Inhibición de la Expresión de Enzimas (Glucanasas)**

La invención proporciona ácidos nucleicos complementarios a (p. ej., secuencias antisentido a) los ácidos nucleicos de la invención, p. ej., ácidos nucleicos que codifican endoglucanasas. Las secuencias antisentido son capaces de inhibir el transporte, el empalme o la transcripción de genes que codifican glucanasa, que codifican endoglucanasas. La inhibición se puede efectuar a través de la elección como diana de ADN genómico o ARN mensajero. La transcripción o la función del ácido nucleico elegido como diana se puede inhibir, por ejemplo, mediante hibridación y/o escisión. Un grupo particularmente útil de inhibidores proporcionado por la presente invención incluye oligonucleótidos que son capaces de unirse al gen de la glucanasa o a su mensaje, evitando o inhibiendo en cualquier caso la producción o la función de la glucanasa. La asociación puede ser a través de hibridación específica de la secuencia. Otra clase útil de inhibidores incluye oligonucleótidos que ocasionan la inactivación o la escisión del mensaje de la glucanasa. El oligonucleótido puede tener una actividad enzimática que causa tal escisión, tal como ribozimas. El oligonucleótido puede ser modificado químicamente o conjugado con una enzima o composición capaz de escindir el ácido nucleico complementario. Se puede escrutar una asociación de muchos de estos oligonucleótidos diferentes para determinar los que tienen la actividad deseada. De este modo, la invención proporciona diversas composiciones para la inhibición de la expresión de la glucanasa en un nivel de ácido nucleico y/o proteína, p. ej., secuencias antisentido, ARNi y ribozimas que comprenden secuencias de glucanasa de la invención y los anticuerpos anti-glucanasa de la invención.

55 La inhibición de la expresión de la glucanasa puede tener una variedad de aplicaciones industriales. Por ejemplo, la inhibición de expresión de la glucanasa puede ralentizar o evitar la descomposición. La descomposición puede ocurrir cuando se degradan enzimáticamente polisacáridos, p. ej., polisacáridos estructurales. Esto puede conducir al deterioro, o podredumbre, de frutos y hortalizas. En un aspecto, las composiciones de la invención que inhiben la expresión y/o la actividad de las glucanasas, p. ej., anticuerpos, oligonucleótidos antisentido, ribozimas y ARNi, se utilizan para ralentizar o evitar la descomposición. La aplicación sobre una planta o producto vegetal (p. ej., un cereal, un grano, un fruto, semilla, raíz, hoja, etc.) de anticuerpos, oligonucleótidos antisentido, ribozimas y ARNi de la invención puede ralentizar o evitar la descomposición. Estas composiciones también puede ser expresadas por la planta (p. ej., una planta transgénica) u otro organismo (p. ej., una bacteria u otro microorganismo transformado con un gen de glucanasa de la invención).

Las composiciones de la invención para la inhibición de la expresión de la glucanasa (p. ej., secuencias antisentido, ARNi, ribozimas, anticuerpos) se pueden utilizar como composiciones farmacéuticas, p. ej., como agentes antipatógenos o en otras terapias, p. ej., como anti-microbianos para, p. ej., *Salmonella*.

Oligonucleótidos Antisentido

Los oligonucleótidos antisentido son capaces de unirse al mensaje de la glucanasa, que pueden inhibir un gen diana o mensaje para, p. ej., inhibir una actividad glucano hidrolasa (p. ej., catalizando la hidrólisis de enlaces β -1,4-xilosídicos internos) eligiendo como diana el ARNm. Las estrategias para el diseño de oligonucleótidos antisentido están bien descritas en la bibliografía científica y de patentes, y el experto en la técnica puede diseñar tales oligonucleótidos de glucanasa utilizando los reactivos de la invención novedosos. Por ejemplo, los protocolos de paseo cromosómico/cartografiado de ARN para escrutar los oligonucleótidos antisentido eficaces son bien conocidos en la técnica, véase, p. ej., Ho (2000) *Methods Enzymol.* 314:168-183, que describe un análisis de cartografiado de ARN, que está basado en técnicas moleculares convencionales para proporcionar un método fácil y fiable para la selección potente de secuencias antisentido. Véase también Smith (2000) *Eur. J. Pharm. Sci.* 11:191-198.

Los ácidos nucleicos de origen natural se utilizan como oligonucleótidos antisentido. Los oligonucleótidos antisentido pueden ser de cualquier longitud; por ejemplo, en aspectos alternativos, los oligonucleótidos antisentido tienen entre alrededor de 5 a 100, alrededor de 10 a 80, alrededor de 15 a 60, alrededor de 18 a 40. La longitud óptima se puede determinar mediante escrutinio rutinario. Los oligonucleótidos antisentido pueden estar presentes a cualquier concentración. La concentración óptima se puede determinar mediante escrutinio rutinario. Se conoce una amplia gama de análogos de nucleótidos y ácidos nucleicos de origen no natural, sintéticos que pueden abordar este problema potencial. Por ejemplo, se pueden utilizar ácidos peptidonucleicos (PNA) que contiene cadenas principales no iónicas, tales como unidades de N-(2-aminoetil)glicina. También se pueden utilizar oligonucleótidos antisentido que tienen conexiones 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 la cadena principal de ADN sintéticos proporcionados por la invención también pueden incluir fosforo-ditioato, metilfosfonato, fosforamidato, alquilfosfotriésteres, sulfamato, 3'-tioacetal, metileno(metilimino), 3'-N-carbamato, y ácidos morfolinocarbamato-nucleicos, como se ha descrito anteriormente.

Se puede utilizar metodología química combinatoria para crear un enorme número de oligonucleótidos que se pueden escrutar rápidamente para determinar los oligonucleótidos específicos que tienen afinidades y especificidades de unión apropiadas hacia cualquier diana, tal como the secuencias de glucanasa directa y antisentido de la invención (véase, p. ej., Gold (1995) *J. de Biol. Chem.* 270:13581-13584).

Ribozimas Inhibidoras

Las ribozimas son capaces de unirse al mensaje o a los genes de glucanasa. Estas ribozimas pueden inhibir la actividad glucanasa, p. ej., eligiendo como diana el ARNm. Las estrategias diseñar ribozimas y seleccionar la secuencia antisentido específica de glucanasa para elegirla como diana están bien descritas en la bibliografía científica y de patentes, y el experto en la técnica puede diseñar tales ribozimas utilizando los reactivos novedosos de la invención. Las ribozimas actúan uniéndose a un ARN diana a través de la porción de unión al ARN diana de una ribozima que es mantenida en íntima proximidad con una porción enzimática del ARN que escinde el ARN diana. De este modo, la ribozima reconoce y se une a una diana de ARN a través de los pares de bases complementarios, y una vez unida en el sitio correcto, actúa enzimáticamente para escindir e inactivar el ARN diana. La escisión de un ARN diana de esta manera destruirá su capacidad para dirigir la síntesis de una proteína codificada si la escisión ocurre en la secuencia codificante. Después de que se ha unido una ribozima y escindido su diana de ARN, ésta se puede liberar de ese ARN para unirse y escindir nuevas dianas repetidamente.

En ciertas circunstancias, la naturaleza enzimática de una ribozima puede ser ventajosa sobre otras tecnologías, tales como la tecnología antisentido (donde una molécula de ácido nucleico se une simplemente a un ácido nucleico diana para bloquear su transcripción, translación o asociación con otra molécula) puesto que la concentración eficaz de ribozima necesaria para efectuar un tratamiento terapéutico puede ser inferior que la de un oligonucleótido antisentido. Esta ventaja potencial refleja la capacidad de la ribozima para actuar enzimáticamente. De este modo, una sola molécula de ribozima es capaz de escindir muchas moléculas de ARN diana. Además, una ribozima es típicamente un inhibidor altamente específico, dependiendo la especificidad de inhibición no solo de mecanismo de emparejamiento de bases, sino también del mecanismo por medio del cual la molécula inhibe la expresión del ARN al que se une. Esto es, la inhibición está causada por la escisión de la diana de ARN y de ese modo la especificidad se define como la proporción de la velocidad de escisión del ARN elegido como diana sobre la velocidad de escisión del ARN no elegido como diana. Este mecanismo de escisión depende de factores adicionales a aquellos implicados en el emparejamiento de bases. De este modo, la especificidad de acción de una ribozima puede ser mayor que la de un oligonucleótido antisentido que se une al mismo sitio del ARN.

La ribozima, p. ej., una molécula de ARN ribozima enzimática, se puede formar en un motivo cabeza de martillo, un motivo en horquilla, como un motivo del virus delta de la hepatitis, un motivo de intrón del grupo I y/o un ARN de tipo RNasaP asociado con una secuencia guía de ARN. Los ejemplos de los motivos cabeza de martillo son descritos, p. ej., por Rossi (1992) *Aids Research and Human Retrovirus* 8:183; los motivos en horquilla por Hampel (1989) *Biochemistry* 28:4929, y Hampel (1990) *Nuc. Acids Res.* 18:299; el motivo del virus delta de la hepatitis por Perrotta (1992) *Biochemistry* 31:16; el motivo RNasaP por Guerrier-Takada (1983) *Cell* 35:849; y el intrón del grupo I por Cech Patente de los Estados Unidos Núm. 4.987.071. No se pretende que la enumeración de estos motivos específicos sea limitante. Los expertos en la técnica advertirán que una ribozima de la invención, p. ej., una molécula de ARN enzimática de esta invención, puede tener un sitio de unión al sustrato específico complementario a uno o más de las regiones de ARN del gen diana. Una ribozima de la invención puede tener una secuencia de nucleótidos en o rodeando a ese sitio de unión al sustrato que confiere una actividad de escisión de ARN a la molécula.

Interferencia por ARN (ARNi)

Una molécula inhibidora de ARN, denominada moléculas de "ARNi" puede comprender una secuencia de glucanasa de la invención. La molécula de ARNi comprende una molécula de ARN de doble hebra (ARNdh). El ARNi puede inhibir la expresión de un gen de glucanasa. El ARNi puede tener alrededor de 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25 o más nucleótidos dúplex de longitud. Si bien que la invención no está limitada por cualquier mecanismo de acción concreto, el ARNi pueden entrar en una célula y causar la degradación de un ARN de hebra sencilla (ARNhs) de secuencias similares o idénticos, incluyendo ARNm endógenos. Cuando una célula se expone a ARN de doble hebra (ARNdh), el ARNm del gen homólogo es degradado selectivamente por un procedimiento denominado interferencia por ARN (RNAi). Un posible mecanismo básico detrás del ARNi es la rotura de un ARN de doble hebra (ARNdh) emparejando una secuencia de genes específica en piezas cortas denominado ARN interferente corto, que desencadena la degradación del ARNm que empareja su secuencia. Los ARNi se pueden utilizar en agentes terapéuticos silenciadores de genes, véase, p. ej., Shuey (2002) *Drug Discov. Today* 7:1040-1046. Los métodos para degradar selectivamente ARN hacen uso de los ARNi de la invención. El procedimiento se puede poner en práctica *in vitro*, *ex vivo* o *in vivo*. En un aspecto, las moléculas de ARNi se pueden utilizar para generar una mutación de pérdida de función en una célula, un órgano o un animal. Los métodos para elaborar y utilizar las moléculas ARNi para degradar selectivamente ARN son bien conocidos en la técnica, véase, p. ej., las Patentes de los Estados Unidos Núms. 6.506.559; 6.511.824; 6.515.109; 6.489.127.

Modificación de Ácidos Nucleicos

La invención proporciona métodos para generar variantes de los ácidos nucleicos de la invención, p. ej., los que codifican una glucanasa. Estos métodos se pueden repetir o utilizar en diversas combinaciones para generar glucanasas que tienen una actividad alterada o diferente o una estabilidad alterada o diferente de los de una glucanasa codificada por el ácido nucleico molde. Estos métodos también se pueden repetir o utilizar en diversas combinaciones, p. ej., para generar variaciones en la expresión del gen/mensaje, la traducción del mensaje o la estabilidad del mensaje. En otro aspecto, la composición genética de una célula se altera, p. ej., mediante modificación de un gen homólogo *ex vivo*, seguido de su inserción en la célula.

Un ácido nucleico de la invención se puede alterar mediante cualquier método. Por ejemplo, métodos aleatorios o estocásticos, o métodos no estocásticos, o "evolución dirigida", véase, p. ej., la Patente de los Estados Unidos Núm. 6.361.974. Los métodos para la mutación de genes aleatoria son bien conocidos en la técnica, véase, p. ej., la Patente de los Estados Unidos Núm. 5.830.696. Por ejemplo, los mutágenos se pueden utilizar para mutar aleatoriamente un gen. Los mutágenos incluyen, p. ej., irradiación con luz ultravioleta o gamma, o un mutágeno químico, p. ej., mitomicina, ácido nitroso, psoralenos fotoactivados, solos o combinados, para inducir roturas de 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 de nucleótidos, p. ej., nitrosoguanidina, 5-bromouracilo, 2-aminopurina, o acridina. Estos agentes se pueden añadir a una reacción PCR en lugar del precursor de nucleótido mutando de este modo la secuencia. También se pueden utilizar agentes intercalantes tales como proflavina, acriflavina, quinacrina y similares.

Se puede utilizar cualquier técnica de biología molecular, p. ej., mutagénesis aleatoria por PCR, véase, p. ej., Rice (1992) *Proc. Natl. Acad. Sci. USA* 89:5467-5471; o, mutagénesis múltiple combinatoria por inserción de casetes, véase, p. ej., Cramer (1995) *Biotécnicas* 18:194-196. Alternativamente, los ácidos nucleicos, p. ej., los genes, se pueden reensamblar después de la fragmentación aleatoria, "estocástica", véanse, p. ej., las Patentes de los Estados Unidos Núms. 6.291.242; 6.287.862; 6.287.861; 5.955.358; 5.830.721; 5.824.514; 5.811.238; 5.605.793. En aspectos alternativos, se introducen modificaciones, adiciones o deleciones mediante PCR propensa a error, barajado, mutagénesis dirigida a oligonucleótidos, PCR de ensamblaje, mutagénesis por PCR sexual, mutagénesis *in vivo*, mutagénesis por inserción de casetes, mutagénesis de ensamblaje recursivo, mutagénesis de conjunto exponencial, mutagénesis de sitio específico, reensamblaje de genes, Mutagénesis de Saturación de Sitios de Genes (GSSM™), reensamblaje por ligación sintética (SLR), recombinación, recombinación de secuencia recursiva,

mutagénesis de ADN modificada por fosfotioato, mutagénesis del molde que contiene uracilo, mutagénesis del dúplex con huecos, mutagénesis de reparación puntual de emparejamientos erróneos, mutagénesis de la cepa anfitriona de reparación deficiente, mutagénesis química, mutagénesis radiogénica, mutagénesis por delección, mutagénesis de restricción-selección, mutagénesis de restricción-purificación, síntesis de genes artificiales, mutagénesis de ensamblaje, creación de multímeros de ácido nucleico quimérico, y/o una combinación de estos y otros métodos.

Las siguientes publicaciones describen una variedad de procedimientos de recombinación recursiva y/o métodos que se pueden incorporar a los métodos de la invención: Stemmer (1999) "Molecular breeding of virus 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 fosforilation using DNA family shuffling" *Nature Biotechnology* 17:259-264; Cramer (1998) "DNA shuffling of a family of genes from diverse species accelerates directed evolution" *Nature* 391:288-291; Cramer (1997) "Molecular evolution of an arsenate detoxification pathway by ADN shuffling", *Nature Biotechnology* 15:436-438; Zhang (1997) "Directed evolution of an effective fucosidase from a galactosidase by ADN shuffling and screening" *Proc. Natl. Acad. Sci. USA* 94:4504-4509; Patten et al. (1997) "Applications of ADN Shuffling to Pharmaceuticals and Vaccines" *Current Opinion in Biotechnology* 8:724-733; Cramer et al. (1996) "Construction and evolution of antibody-phage libraries by ADN 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" In: *The Encyclopedia of Molecular Biology*. VCH Publishers, New York. págs. 447-457; Cramer y Stemmer (1995) "Combinatorial multiple mutagenesis by insertion of cassettes creates all the permutations of mutant and wildtype cassettes" *BioTechniques* 18:194-195; Stemmer et al. (1995) "Single-step assembly of a gen and entire plasmid form large numbers of oligodeoxyribonucleotides" *Gene*, 164:49-53; Stemmer (1995) "The Evolution of Molecular Computation" *Science* 270: 1510; Stemmer (1995) "Searching Sequence Space" *Bio/Technology* 13:549-553; Stemmer (1994) "Rapid evolution of a protein in vitro by ADN 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 ADN 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 & 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" en *Nucleic Acids & Molecular Biology* (Eckstein, F. y Lilley, D. M. J. eds., Springer Verlag, Berlin)); mutagenesis utilizando 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 en Enzymol.* 154, 367-382; y Bass et al. (1988) "Mutant Trp repressors with new DNA-binding specificities" *Science* 242:240-245); mutagenesis dirigida a oligonucleótidos (Methods in Enzymol. 100: 468-500 (1983); Methods in Enzymol. 154: 329-350 (1987); Zoller (1982) "Oligonucleotide-directed mutagenesis using M13-derived vectors: an efficient and general procedure for the production of point mutations in any ADN fragment" *Nucleic Acids Res.* 10:6487-6500; Zoller & Smith (1983) "Oligonucleotide-directed mutagenesis of fragments of ADN cloned into M13 vectors" *Methods en Enzymol.* 100:468-500; y Zoller (1987) Oligonucleotide-directed mutagenesis: a simple method using two oligonucleotide primers and a single-stranded ADN template" *Methods en Enzymol.* 154:329-350); mutagenesis de ADN modificada con fosfotioato (Taylor (1985) "The use of phosphorothioate-modified ADN in restriction enzyme reactions to prepare nicked DNA" *Nucl. Acids Res.* 13: 8749-8764; Taylor (1985) "The rapid generation of oligonucleotide-directed mutations at high frequency using phosphorothioate-modified DNA" *Nucl. Acids Res.* 13: 8765-8787 (1985); Nakamaye (1986) "Inhibition of restriction endonuclease Nci I cleavage by phosphorothioate groups and its application to mutagenesis directed to oligonucleotides" *Nucl. Acids Res.* 14: 9679-9698; Sayers (1988) "Y-T Exonucleases in phosphorothioate-based mutagenesis directed to oligonucleotides" *Nucl. Acids Res.* 16:791-802; y Sayers et al. (1988) "Strand specific cleavage of phosphorothioate-containing ADN by reaction with restriction endonucleases in presence of ethidium bromide" *Nucl. Acids Res.* 16: 803-814); mutagenesis utilizando ADN dúplex con huecos (Kramer et al. (1984) "The gapped duplex ADN approach to oligonucleotide-directed mutation construction" *Nucl. Acids Res.* 12: 9441-9456; Kramer & Fritz (1987) *Methods in Enzymol.* "Oligonucleotide-directed construction of mutations via gapped duplex DNA" 154:350-367; Kramer (1988) "Improved enzymatic in vitro reactions in the gapped duplex ADN approach to oligonucleotide-directed construction of mutations" *Nucl. Acids Res.* 16: 7207; y Fritz (1988) "Oligonucleotide-directed construction of mutations: a gapped duplex ADN procedure without enzymatic reactions in vitro" *Nucl. Acids Res.* 16: 6987-6999).

Los protocolos adicionales que se pueden utilizar para poner en práctica la invención incluyen reparación puntual de emparejamientos erróneos (Kramer (1984) "Point Mismatch Repair" *Cell* 38:879-887), mutagénesis utilizando cepas anfitrionas de reparación deficiente (Carter et al. (1985) "Improved oligonucleotide site-directed mutagenesis using

M13 vectors" Nucl. Acids Res. 13: 4431-4443; y Carter (1987) "Improved mutagenesis directed to oligonucleotides using M13 vectores" Methods in Enzymol. 154: 382-403), mutagénesis por delección (Eghtedarzadeh (1986) "Use of oligonucleotides to generate large deletions" Nucl. Acids Res. 14: 5115), 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 de genes total (Nambiar et al. (1984) "Total synthesis and cloning of a gene coding for the ribonuclease S protein" Science 223: 1299-1301; Sakamar y Khorana (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), double-strand break repair (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" Proc. Natl. Acad. Sci. USA, 83:7177-7181). Se pueden encontrar detalles adicionales sobre muchos de los métodos anteriores en Methods in Enzymology Volumen 154, que también describe controles útiles para la solución de problemas con diversos métodos de mutagénesis.

Los protocolos que se pueden utilizar para poner en práctica la invención se describen, p. ej., en la Patente de los Estados Unidos Núms. 5.605.793 de Stemmer (Feb. 25, 1997), "Methods for In Vitro Recombination;" Patente de los Estados Unidos Núm. 5.811.238 de Stemmer et al. (Sep. 22, 1998) "Methods for Generating Polynucleotides having Desired Characteristics by Iterative Selection and Recombination;" Patente de los Estados Unidos Núm. 5.830.721 de Stemmer et al. (Nov. 3, 1998), "DNA Mutagenesis by Random Fragmentation and Reassembly;" Patente de los Estados Unidos Núm. 5.834.252 de Stemmer, et al. (Nov. 10, 1998) "End-Complementary Polymerase Reaction;" Patente de los Estados Unidos Núm. 5.837.458 de Minshull, et al. (Nov. 17, 1998), "Methods and Compositions for Cellular and Metabolic Engineering;" documento WO 95/22625, Stemmer and Cramer, "Mutagenesis by Random Fragmentation and Reassembly;" documento WO 96/33207 de Stemmer and Lipschutz "End Complementary Polymerase Chain Reaction;" documento WO 97/20078 de Stemmer and Cramer "Methods for Generating Polynucleotides having Desired Characteristics by Iterative Selection and Recombination;" documento WO 97/35966 de Minshull and Stemmer, "Methods and Compositions for Cellular and Metabolic Engineering;" documento WO 99/41402 de Punnonen et al. "Targeting of Genetic Vaccine Vectors;" documento WO 99/41383 de Punnonen et al. "Antigen Library Immunization;" documento WO 99/41369 de Punnonen et al. "Genetic Vaccine Vector Engineering;" documento WO 99/41368 de Punnonen et al. "Optimization of Immunomodulatory Properties of Genetic Vaccines;" EP 752008 de Stemmer and Cramer, "DNA Mutagenesis by Random Fragmentation and Reassembly;" EP 0932670 de Stemmer "Evolving Cellular DNA Uptake by Recursive Sequence Recombination;" documento WO 99/23107 de Stemmer et al., "Modification of Virus Tropism and Host Range by Viral Genome Shuffling;" documento WO 99/21979 de Apt et al., "Human Papillomavirus Vectors;" documento WO 98/31837 de Cardayre et al. "Evolution of Whole Cells and Organisms by Recursive Sequence Recombination;" documento WO 98/27230 de Patten and Stemmer, "Methods and Compositions for Polypeptide Engineering;" documento WO 98/27230 de 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 de Arnold et al., "Recombination of Polynucleotide Sequences Using Random or Defined Primers", documento WO 99/29902 de Arnold et al., "Method for Creating Polynucleotide and Polypeptide Sequences", documento WO 98/41653 de Vind, "An in Vitro Method for Construction of a DNA Library", documento WO 98/41622 de Borchert et al., "Method for Constructing a Library Using DNA Shuffling", y documento WO 98/42727 de Pati and Zarling, "Sequence Alterations using Homologous Recombination".

Los protocolos que se pueden utilizar para poner en práctica los métodos (proporcionando detalles concernientes a diversos métodos generadores de diversidad) se describen, p. ej., en la Solicitud de Patente de los Estados Unidos con el Núm. de Serie (USSN) 09/407.800, "SHUFFLING OF CODON ALTERED GENES" de Patten et al. presentada el 28 de Sep. de 1999; "EVOLUTION OF WHOLE CELLS AND ORGANISMS BY RECURSIVE SEQUENCE RECOMBINATION" de Cardayre et al., Patente de los Estados Unidos Núm. 6.379.964; "OLIGONUCLEOTIDE MEDIATED NUCLEIC ACID RECOMBINATION" de Cramer et al., Patentes de los Estados Unidos Núms. 6.319.714; 6.368.861; 6.376.246; 6.423.542; 6.426.224 y documento PCT/US00/01203; "USE OF CODON-VARIED OLIGONUCLEOTIDE SYNTHESIS FOR SYNTHETIC SHUFFLING" de Welch et al., Patente de los Estados Unidos Núm. 6.436.675; "METHODS FOR MAKING CHARACTER STRINGS, POLYNUCLEOTIDES & POLYPEPTIDES HAVING DESIRED CHARACTERISTICS" de Selifonov et al., presentada el 18 de Enero de 2000, (PCT/US00/01202) y, p. ej. "METHODS FOR MAKING CHARACTER STRINGS, POLYNUCLEOTIDES & POLYPEPTIDES HAVING DESIRED CHARACTERISTICS" de Selifonov et al., presentada el 18 de Julio de 2000 (Documento U.S. Núm. Ser. 09/618,579); "METHODS OF POPULATING DATA STRUCTURES FOR USE IN EVOLUTIONARY STIMULATIONS" de Selifonov y Stemmer, presentada el 18 de Enero de 2000 (PCT/US00/01138); y "SINGLE-STRANDED NUCLEIC ACID TEMPLATE-MEDIATED RECOMBINATION AND NUCLEIC ACID FRAGMENT ISOLATION" de Affholter, presentada el 6 de Sep. de 2000 (Documento U.S. Núm. Ser. 09/656,549); y Patentes de los Estados Unidos Núms. 6.177.263; 6.153.410.

Los métodos no estocásticos, o de "evolución dirigida" incluyen, p. ej., la Mutagénesis de Saturación de Sitios de Genes (GSSM™), reensamblaje por ligación sintética (SLR), o una de sus combinaciones se utilizan para modificar los ácidos nucleicos de la invención para generar glucanasas con propiedades nuevas o alteradas (p. ej., actividad bajo condiciones altamente ácidas o alcalinas, temperaturas altas o bajas, y similares). Los polipéptidos codificados por los ácidos nucleicos modificados se pueden escrutar para determinar una actividad antes de someter a ensayo la hidrólisis del glucano u otro polisacárido u otra actividad. Se puede utilizar cualquier modalidad o protocolo de ensayo, p. ej., utilizando una plataforma de matriz capilar. Véanse, p. ej., las Patentes de los Estados Unidos Núms. 6.361.974; 6.280.926; 5.939.250.

10 **Mutagénesis por saturación, o, GSSM™**

En un aspecto, se utilizan cebadores codónicos que contienen una secuencia N,N,G/T degenerada para introducir mutaciones puntuales en un polinucleótido, p. ej., una glucanasa o un anticuerpo de la invención, para generar un conjunto de polipéptidos progenie en el que en cada posición de aminoácidos está representada en cada posición de aminoácido, p. ej., un residuo de aminoácido en un sitio activo enzimático (dominios catalíticos (CD)) o un sitio de unión al ligando elegido como diana que se va a modificar. Estos oligonucleótidos pueden comprender una primera secuencia homóloga contigua, una secuencia N,N,G/T degenerada, y, opcionalmente, una segunda secuencia homóloga. Los productos traducionales de la progenie aguas abajo a partir del uso de tales oligonucleótidos incluyen todos los cambios de aminoácido posibles en cada sitio de aminoácido a lo largo del polipéptido, debido a que la degeneración de la secuencia N,N,G/T incluye codones para los 20 aminoácidos. En un aspecto, uno de tales oligonucleótidos degenerados (que comprende, p. ej., un casete N,N,G/T degenerado) se utiliza para someter cada codón original en un molde polinucleotídico parental a una gama completa de sustituciones codónicas. En otro aspecto, se utilizan al menos dos casetes degenerados - en el mismo oligonucleótido o no, para someter al menos dos codones originales en un molde polinucleotídico parental a una gama completa de sustituciones codónicas. Por ejemplo, en un oligonucleótido puede estar contenida más de una secuencia N,N,G/T para introducir mutaciones de aminoácido en más de un sitio. Esta pluralidad de secuencias N,N,G/T pueden estar directamente contigua, o separada by una o más secuencias de nucleótidos adicionales. En otro aspecto, se pueden usar oligonucleótidos prácticos para introducir adiciones y deleciones solos o combinados con los codones que contienen una secuencia N,N,G/T, para introducir cualquier combinación o permutación de adiciones, deleciones, y/o sustituciones de aminoácidos.

En un aspecto, la mutagénesis simultánea de dos o más posiciones de aminoácido contiguas se realiza utilizando un oligonucleótido que contiene tripletes N,N,G/T contiguos, es decir una secuencia (N,N,G/T)_n degenerada. En otro aspecto, se utilizan casetes degenerados que tienen menos degeneración que la secuencia N,N,G/T. Por ejemplo, puede ser deseable en algunos casos (p. ej. en un oligonucleótido) una secuencia de triplete degenerado que comprende solo un N, donde dicho N puede estar en la segunda o tercera posición del triplete. Se pueden utilizar otras bases cualesquiera incluyendo cualquiera de sus combinaciones y permutaciones en las dos posiciones restantes del triplete. Alternativamente, puede ser deseable en algunos casos (p. ej. en un oligo) utilizar una secuencia de triplete N,N,N degenerado.

En un aspecto, el uso de tripletes degenerados (p. ej., tripletes N,N,G/T) permite la generación sistemática y fácil de una gama completa de aminoácidos naturales posibles (para un total de 20 aminoácidos) en todas y cada una de las posiciones de aminoácidos en un polipéptido (en aspectos alternativos, los métodos también incluyen la generación de menos de todas las sustituciones posibles por posición de residuo de aminoácido, o codón,). Por ejemplo, para un polipéptido de 100 aminoácidos, se pueden generar 2000 especies distintas (es decir 20 aminoácidos posibles por posición X 100 posiciones de aminoácidos). A través del uso de un oligonucleótido o conjunto de oligonucleótidos que contienen un triplete N,N,G/T degenerado, 32 secuencias individuales pueden codificar los 20 aminoácidos naturales posibles. De este modo, en un recipiente de reacción en el que una secuencia parental de polinucleótidos se somete a mutagénesis por saturación utilizando al menos uno de tales oligonucleótidos, se generan 32 polinucleótidos progenie distintos que codifican 20 polipéptidos distintos. En contraste, el uso de un oligonucleótido no degenerado en la mutagénesis dirigida al sitio conduce solo a un producto polipeptídico progenie por recipiente de reacción. Los oligonucleótidos no degenerados se pueden utilizar opcionalmente combinados con los cebadores degenerados descritos; por ejemplo, los oligonucleótidos no degenerados se pueden utilizar para generar mutaciones puntuales específicas en un polinucleótido de trabajo. Esto proporciona un medio para generar mutaciones puntuales silenciosas específicas, mutaciones puntuales que conducen a cambios de aminoácido correspondientes, y mutaciones puntuales que ocasionan la generación de codones de parada y la correspondiente expresión de fragmentos polipeptídicos.

En un aspecto, cada recipiente de reacción de mutagénesis por saturación contiene polinucleótidos que codifican al menos 20 moléculas de polipéptido progenie (p. ej., glucanasas, mananasas, o xilanasas) de manera que los 20 aminoácidos naturales están representados en una posición de aminoácido específica correspondiente a la posición del codón mutagenizada en el polinucleótido parental (otros aspectos utilizan menos de los 20 combinaciones naturales). La progenie de polipéptidos degenerada 32 veces generada a partir de cada recipiente de reacción de mutagénesis por saturación se puede someter a amplificación clonal (p. ej. clonada en un anfitrión adecuado, p. ej.,

anfitrión *E. coli*, utilizando, p. ej., un vector de expresión) y someter a escrutinio de la expresión. Cuando una progenie de polipéptido individual se identifica para que presente un cambio de propiedades favorable (cuando se compara con el polipéptido parental, tal como el aumento de la actividad de hidrólisis del glucano en condiciones alcalinas o ácidas), ésta se puede secuenciar para identificar la sustitución de aminoácido correspondientemente favorable contenida allí.

En un aspecto, después de mutagenizar todas y cada una de las posiciones de aminoácido en un polipéptido parental utilizando mutagénesis por saturación como se describe en la presente memoria, se pueden identificar cambios de aminoácidos favorables en más de una posiciones de aminoácido. Se pueden generar una o más moléculas progenie que contiene una combinación todas o parte de estas sustituciones de aminoácido favorables. Por ejemplo, si se identifican 2 cambios de aminoácido favorables específicos en cada una de 3 posiciones de aminoácidos en un polipéptido, las permutaciones incluyen 3 posibilidades en cada posición (sin cambio del aminoácido original, y cada uno de dos cambios favorables) y 3 posiciones. De este modo, hay $3 \times 3 \times 3$ o 27 posibilidades totales, incluyendo 7 que se examinaron previamente - 6 mutaciones puntuales individuales (es decir 2 en cada una de las tres posiciones) y no hay cambio en ninguna posición.

En otro aspecto más, se puede utilizar mutagénesis de saturación de sitio junto con barajado, quimerización, recombinación y otros procedimientos mutagenizantes, junto con escrutinio. Esta invención proporciona el uso de uno o varios procedimientos de mutagenización cualesquiera, incluyendo mutagénesis por saturación, de una manera iterativa. En una ejemplificación, se utiliza el uso iterativo de uno o varios procedimientos de mutagenización cualesquiera combinados con escrutinio.

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

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

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

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

Se aprecia, no obstante, que el uso de un triplete degenerado (tal como N,N,G/T o una secuencia triplete N,N,G/C) como se describe en la presente memoria es ventajoso por diversas razones. En un aspecto, esta invención proporciona un medio para generar sistemáticamente y bastante fácilmente la sustitución de la gama completa de aminoácidos posibles (para un total de 20 aminoácidos) en todas y cada una de las posiciones de aminoácidos en un polipéptido. De este modo, para un polipéptido de 100 aminoácidos, la invención proporciona a una manera de generar sistemáticamente y bastante fácilmente 2000 especies distintas (es decir, 20 aminoácidos posibles por posición en 100 posiciones de aminoácidos). Se aprecia que se proporcionan, a través del uso de un oligo que contiene una secuencia triplete N,N,G/T o N,N,G/C degenerada, 32 secuencias individuales que codifican 20 aminoácidos posibles. De este modo, en un recipiente de reacción en el que una secuencia parental de polinucleótidos se somete a mutagénesis por saturación utilizando uno de tales oligos, se generan 32 polinucleótidos

progenie distintos que codifican 20 polipéptidos distintos. En contraste, el uso de un oligo no degenerado en la mutagénesis dirigida al sitio conduce a solo un producto progenie de polipéptidos por recipiente de reacción.

Esta invención también proporciona el uso de oligos no degenerados, que se pueden utilizar opcionalmente combinados con los cebadores degenerados descritos. Se aprecia que en algunas situaciones, es ventajoso utilizar oligos no degenerados para generar mutaciones puntuales específicas en un polinucleótido de trabajo. Esto proporciona un medio para generar mutaciones puntuales silenciosas específicas, mutaciones puntuales que conducen a cambios aminoácido y mutaciones puntuales que causan la generación de codones de parada y la expresión correspondiente de fragmentos peptídicos.

De este modo, en un aspecto de esta invención, cada recipiente de reacción de mutagénesis por saturación puede contener polinucleótidos que codifican al menos moléculas progenie de 20 polipéptidos de manera que los 20 aminoácidos están representados en una posición de aminoácido específica correspondiente a la posición codónica mutagenizada en el polinucleótido parental. La progenie de polipéptidos degenerada 32 veces generada a partir de cada recipiente de reacción de la mutagénesis por saturación se puede someter a amplificación clonal (*p. ej.*, clonar en un anfitrión *E. coli* adecuado utilizando un vector de expresión) y someter a escrutinio de expresión. Cuando se identifica una progenie de polipéptido individual mediante escrutinio para presentar un cambio de propiedades favorable (cuando se compara con el polipéptido parental), ésta se puede secuenciar para identificar sustitución de aminoácido correspondientemente favorable contenida allí.

Se aprecia que después de mutagenizar todas y cada una de las posiciones de aminoácido en un polipéptido parental utilizando mutagénesis por saturación como se describe en la presente memoria, se pueden identificar los cambios de aminoácido favorables en más de una posición de aminoácido. Se pueden generar una o más moléculas progenie nuevas que contienen una combinación de todas o parte de estas sustituciones de aminoácido favorables. Por ejemplo, si se identifican 2 cambios aminoácido favorables específicos en cada una de las 3 posiciones de aminoácidos en un polipéptido, las permutaciones incluyen 3 posibilidades en cada posición (sin cambio del aminoácido original y cada uno de dos los cambios favorables) y 3 posiciones. De este modo, hay $3 \times 3 \times 3$ o 27 posibilidades totales, incluyendo 7 que se examinaron previamente - 6 mutaciones puntuales individuales (es decir, 2 en cada una de las tres posiciones) y sin cambio en ninguna posición.

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

Además para realizar la mutagénesis a lo largo de la secuencia completa de un gen, la presente invención proporciona que la mutagénesis se puede utilizar para reemplazar cada una de cualquier número de bases en una secuencia de polinucleótidos, donde el número de bases que se va a mutagenizar es preferiblemente cada número entero de 15 a 100.000. De este modo, en lugar de mutagenizar cada posición a lo largo de una molécula, se puede someter cada base o un número discreto de bases (totalizando un subgrupo preferiblemente de 15 a 100.000) a mutagénesis. En un aspecto, se utiliza un nucleótido separado para mutagenizar cada posición o grupo de posiciones a lo largo de una secuencia de polinucleótidos. Un grupo de 3 posiciones que se vayan a mutagenizar puede ser un codón. Las mutaciones se introducen preferiblemente utilizando un cebador mutagénico, que contiene un casete heterólogo, también referido como casete mutagénico. Los casetes ilustrativos pueden tener de 1 a 500 bases. Cada posición de nucleótido en tales casetes heterólogos pueden ser N, A, C, G, T, A/C, A/G, A/T, C/G, C/T, G/T, C/G/T, A/G/T, A/C/T, A/C/G, o E, donde E es cualquier base que no sea A, C, G, o T (E se puede referir como oligo diseñador).

En un sentido general, la mutagénesis por saturación comprende la mutagenización de un conjunto completo de casetes mutagénicos (donde cada casete tiene preferiblemente alrededor de 1-500 bases de longitud) en secuencias de polinucleótidos definidas que se van a mutagenizar (donde la secuencia que se va a mutagenizar tiene preferiblemente de alrededor de 15 a 100.000 bases de longitud). De este modo, se introduce un grupo de mutaciones (que oscila de 1 a 100 mutaciones) en cada casete que se vaya a mutagenizar. Un agrupamiento de mutaciones que se va a introducir en un casete puede ser diferente o igual que un segundo agrupamiento de mutaciones que se va a introducir en un segundo casete durante la aplicación de una ronda de mutagénesis por saturación. Tales agrupamientos se ilustran mediante deleciones, adiciones, agrupamientos de codones concretos y agrupamientos de casetes de nucleótidos concretos.

Las secuencias definidas que se van a mutagenizar incluyen un gen completo, una ruta, un ADNc, un marco de lectura abierto (ORF) completo y un promotor completo, un intensificador, un represor/transactivador, un origen de replicación, un intrón, un operador, o cualquier grupo funcional polinucleotídico. Generalmente, las "secuencias definidas" para este fin pueden ser polinucleótidos cualesquiera que tienen una secuencia de polinucleótidos de 15 bases y secuencias de polinucleótidos de entre 15 bases y 15.000 bases de longitud (esta invención

específicamente nombra cada número entero intermedio). Las consideraciones en la elección de los agrupamientos de codones incluyen los tipos de aminoácidos codificados por un casete mutagénico degenerado.

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

10 **Reensamblaje por ligación sintética (SLR)**

La invención proporciona un sistemas de modificación de genes no estocástico denominado "reensamblaje por ligación sintética", o simplemente "SLR", un "proceso de evolución dirigida", para generar polipéptidos, glucanasas de la invención, con propiedades nuevas o alteradas.

15 El SLR es un método para ligar fragmentos de oligonucleótidos entre sí no estocásticamente. Este método difiere de barajado de oligonucleótidos estocástico en que los elementos fundamentales del ácido nucleico no se barajan, concatenan o quimerizan aleatoriamente, sino que en lugar de eso se ensamblan no estocásticamente. Véase, p. ej., la Solicitud de Patente de los Estados Unidos con el Núm. de Serie (USSN)09/332,835 titulada "Synthetic Ligation Reensamblaje in Directed Evolution" y presentada el 14 de Juno de 1999 ("USSN 09/332,835"). En un aspecto, el SLR comprende las siguientes etapas: (a) proporcionar un polinucleótido molde, donde el polinucleótido molde comprende una secuencia que codifica un gen homólogo; (b) proporcionar una pluralidad de polinucleótidos unidades fundamentales, donde los polinucleótidos unidades fundamentales se diseñan para su reensamblaje mediante entrecruzamiento con el polinucleótido molde en una secuencia predeterminada, y un polinucleótido unidad fundamental comprende una secuencia que es una variante del gen homólogo y una secuencia homóloga al polinucleotídico molde que flanquea la secuencia variante; (c) combinar un polinucleótido unidad fundamental con un polinucleótido molde de manera que el polinucleótido unidad fundamental se reensamble mediante entrecruzamiento con el polinucleótido molde para generar polinucleótidos que comprenden variaciones de secuencia de genes homólogos.

30 El SLR no depende de la presencia de altos niveles de homología entre los polinucleótidos que se van a reordenar. De este modo, este método se puede utilizar para generar no estocásticamente bibliotecas (o conjuntos) de moléculas progenie que comprenden más de 10^{100} quimeras diferentes. El SLR se puede utilizar para generar bibliotecas que comprenden más de 10^{1000} quimeras progenie diferentes. De este modo, los aspectos de la presente invención incluyen métodos no estocásticos para producir un conjunto de molécula de ácido nucleico quiméricas finalizadas que tienen un orden de ensamblaje global que se elige mediante diseño. Este método incluye las etapas de generar mediante diseño una pluralidad de elementos fundamentales de ácido nucleico específicos que tienen extremos ligables mutuamente compatibles prácticos, y ensamblar estos elementos fundamentales de ácido nucleico, de manera que se logre un orden de ensamblaje global diseñado.

40 Se considera que los extremos ligables mutuamente compatibles de los elementos fundamentales de ácido nucleico que se van a ensamblar son "prácticos" para este tipo de ensamblaje ordenado si posibilitan el acoplamiento de los elementos fundamentales en los órdenes predeterminados. De este modo, el orden de ensamblaje global en el que el elementos fundamentales de ácido nucleico pueden estar acoplados se especifica mediante el diseño de los extremos ligables. Si se va a utilizar más de una etapa de ensamblaje, el orden de ensamblaje global en el que pueden estar acoplados los elementos fundamentales de ácido nucleico también está especificado por el orden sucesivo de la etapa o las etapas de ensamblaje. En un aspecto, las piezas fundamentales recocidas se tratan con una enzima, tal como una ligasa (p. ej. ADN ligasa de T4), para lograr la unión covalente de las piezas fundamentales.

50 El diseño de los elementos fundamentales del oligonucleótido se obtiene analizando un conjunto de secuencias progenitoras de moldes de ácido nucleico que sirven como una base para la producción de un conjunto progenie de polinucleótidos quiméricos finalizados. Estos moldes de oligonucleótido parentales sirven así como fuente de información de secuencia que ayuda al diseño de los elementos fundamentales de ácido nucleico que se van a mutagenizar, p. ej., quimerizar o barajar. Las secuencias de una pluralidad de moldes de ácido nucleico parentales se alinean con el fin de seleccionar uno o más puntos de demarcación. Los puntos de demarcación pueden estar localizados en una zona de homología, y comprenden uno o más nucleótidos. Estos puntos de demarcación con compartidos preferiblemente por al menos dos de los moldes progenitores. Los puntos de demarcación se pueden utilizar de ese modo para delinear los límites de elementos fundamentales de los oligonucleótidos que se van a generar con el fin de reordenar los polinucleótidos parentales. Los puntos de demarcación identificados y seleccionados en las moléculas progenitoras sirven como puntos de quimerización potenciales en el ensamblaje de las moléculas progenie quiméricas finales. Un punto de demarcación puede ser una zona de homología (que comprende al menos un base nucleotídica homóloga) compartida por al menos dos secuencias de polinucleótidos parentales. Alternativamente, un punto de demarcación puede ser una zona de homología que es compartida por al menos la mitad de las secuencias de polinucleótidos parentales, o, puede ser una zona de homología que es

compartida por al menos dos tercios de las secuencias de polinucleótidos parentales. Incluso más preferiblemente los puntos de demarcación prácticos son una zona de homología que es compartida por al menos tres cuartas partes de las secuencias de polinucleótidos parentales, o, puede ser compartida por casi todas las secuencias de polinucleótidos parentales. En un aspecto, un punto de demarcación es una zona de homología que es compartida por todas las secuencias de polinucleótidos parentales.

Se puede realizar exhaustivamente un procedimiento de reensamblaje por ligación con el fin de generar una biblioteca exhaustiva de polinucleótidos quiméricos progenie. En otras palabras, todas las combinaciones ordenadas posibles de los elementos fundamentales de ácido nucleico están representadas en el conjunto de moléculas de ácido nucleico quiméricas finalizadas. Al mismo tiempo, en otro aspecto, el orden de ensamblaje (es decir el orden de ensamblaje de cada elemento fundamental en secuencia 5' a 3' de cada ácido nucleico quimérico finalizado) en cada combinación es mediante diseño (o no estocástico) como se ha descrito anteriormente. Debido a la naturaleza no estocástica de esta invención, se reduce enormemente la posibilidad de productos secundarios no deseados.

El método de reensamblaje por ligación se puede realizar sistemáticamente. Por ejemplo, el método se realiza con el fin de generar una biblioteca compartimentalizada sistemáticamente de moléculas progenie, con compartimentos que pueden ser escrutados sistemáticamente, p. ej. de uno en uno. En otras palabras esta invención proporciona que, a través del uso selectivo y juicioso de elementos específicos fundamentales de ácido nucleico, acoplado al uso selectivo y juicioso de reacciones de ensamblaje recorridas sucesivamente, se puede lograr un diseño en el que se elaboran conjuntos específicos de productos progenie en cada uno de los diversos recipientes de reacción. Esto permite la realización de un procedimiento sistemático de examen y escrutinio. De este modo, estos métodos permiten examinar sistemáticamente un número potencialmente muy grande de moléculas progenie en grupos más pequeños. Debido a su capacidad para realizar quimerizaciones de una manera que es altamente flexible aunque también exhaustiva y sistemática, particularmente cuando hay un bajo nivel de homología entre las moléculas progenitoras, estos métodos proporcionan la generación de una biblioteca (o conjunto) que comprende un gran número de moléculas progenie. Debido a la naturaleza no estocástica del reensamblaje por ligación de la presente invención, las moléculas progenie generadas comprenden preferiblemente una biblioteca de moléculas de ácido nucleico quiméricas finalizadas que tienen un orden de ensamblaje global que se elige mediante diseño. La mutagénesis por saturación y los métodos de evolución dirigida optimizados también se pueden utilizar para generar especies moleculares progenie diferentes. Se aprecia que esto proporciona libertad de elección y control con respecto a la selección de los puntos de demarcación, el tamaño y el número de los elementos fundamentales de ácido nucleico, y el tamaño y el diseño de los acoplamientos. Se aprecia además, que el requerimiento de homología intermolecular se encuentra muy relajado por la operabilidad. De hecho, los puntos de demarcación se pueden elegir incluso en zonas de poca o ninguna homología intermolecular. Por ejemplo, debido al balanceo codónico, es decir la degeneración de los codones, se pueden introducir sustituciones de nucleótidos en los elementos fundamentales de ácido nucleico sin alterar el aminoácido codificado originalmente en el molde progenitor correspondiente. Alternativamente, se puede alterar un codón de manera que se altere la codificación para un aminoácido original. Esta invención proporciona que tales sustituciones se pueden introducir en el elemento fundamental del ácido nucleico con el fin de incrementar la incidencia de puntos de demarcación homólogos intermoleculares y permitir así un aumento del número de acoplamientos que se van a lograr entre los elementos fundamentales, lo que a su vez permite la generación de un mayor número de moléculas quiméricas progenie.

Un método no estocástico se denomina reensamblaje sintético de genes que está algo relacionado con el barajado estocástico, salvo que los elementos fundamentales del ácido nucleico no son barajados o concatenados o quimerizados aleatoriamente, sino que en vez de eso se ensamblan no estocásticamente.

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

Un método no estocástico produce un conjunto de moléculas de ácido nucleico quiméricas finalizadas que tienen un orden de ensamblaje global que se elige mediante diseño, cuyo método comprende las etapas de generación mediante diseño de una pluralidad de elementos específicos fundamentales de ácidos nucleicos que tienen extremos ligables mutuamente compatibles prácticos y ensamblaje de estos elementos fundamentales de ácido nucleico, de manera que se logre un orden de ensamblaje global diseñado.

Se considera que los extremos ligables mutuamente compatibles de los elementos fundamentales de ácido nucleico que se van a ensamblar son "prácticos" para este tipo de ensamblaje ordenado si posibilitan que los elementos fundamentales se acoplen en los órdenes predeterminados. De este modo, en un aspecto, el orden de ensamblaje global en el que los elementos fundamentales de ácido nucleico pueden estar acoplados se especifica mediante el diseño de los extremos ligables y, si se va a utilizar más de una etapa de ensamblaje, el orden de ensamblaje global en el que los elementos fundamentales de ácido nucleico pueden estar acoplados es especificado también por el

orden sucesivo de la etapa o etapas de ensamblaje. Las piezas fundamentales recocidas se pueden tratar con una enzima, tal como una ligasa (p. ej., ADN ligasa de T4) para lograr la unión covalente de las piezas fundamentales.

5 Alternativamente, se obtiene el diseño de elementos fundamentales de ácido nucleico después del análisis de las secuencias de un conjunto de moldes de ácido nucleico progenitores que sirven como base para la producción de un conjunto progenie de moléculas de ácido nucleico quiméricas finalizadas. Estos moldes de ácido nucleico progenitores sirven de este modo como fuente de información de secuencia que ayuda al diseño del elementos fundamentales de ácido nucleico que se van a mutagenizar, es decir quimerizar o barajar.

10 En una ilustración, la descripción proporciona la quimerización de una familia de genes relacionados y su familia codificada productos relacionados. En una ilustración concreta, los productos codificados son enzimas. Las glucanasas de la presente invención se pueden mutagenizar de acuerdo con los métodos descritos en la presente memoria.

15 De este modo, las secuencias de una pluralidad de moldes de ácido nucleico progenitores (p. ej., polinucleótidos de la invención) se alinean con el fin de seleccionar uno o más puntos de demarcación, cuyos puntos de demarcación pueden estar localizados en una zona de homología. Los puntos de demarcación se pueden utilizar para delinear los límites de los elementos fundamentales del ácido nucleico que se vaya a generar. De este modo, los puntos de demarcación identificados y seleccionados en las moléculas progenitoras sirven como puntos de quimerización potenciales en el ensamblaje de las moléculas progenie.

20 Típicamente un punto de demarcación práctico es una zona de homología (que comprende al menos una base nucleotídica homóloga) compartida por al menos dos moldes progenitores, pero el punto de demarcación puede ser una zona de homología que es compartida por al menos la mitad de los moldes progenitores, al menos dos terceras partes de los moldes progenitores, al menos tres cuartas partes de los moldes progenitores y preferiblemente casi todos los moldes progenitores. Incluso más preferiblemente un punto de demarcación aún más práctico es una zona de homología que es compartida por todos los moldes progenitores.

25 El proceso de reensamblaje de genes se realiza exhaustivamente con el fin de generar una biblioteca exhaustiva. En otras palabras, todas las combinaciones ordenadas posibles de elementos fundamentales de ácido nucleico están representadas en el conjunto de moléculas de ácido nucleico quiméricas finalizadas. Al mismo tiempo, el orden de ensamblaje (es decir el orden de ensamblaje de cada elemento fundamental en la secuencia 5' a 3' de cada ácido nucleico quimérico finalizado) en cada combinación es mediante diseño (o no estocástico). Debido a la naturaleza no estocástica del método, se reduce enormemente la posibilidad de productos secundarios no deseados.

30 El método puede proporcionar que el proceso de reensamblaje de genes se realice sistemáticamente, por ejemplo para generar una biblioteca compartimentalizada sistemáticamente, con compartimentos que pueden ser escrutados sistemáticamente, p. ej., de uno en uno. En otras palabras, esto proporciona que a través del uso selectivo y juicioso de elementos específicos fundamentales de ácido nucleico, acoplado al uso selectivo y juicioso de reacciones de ensamblaje recorridas sucesivamente, se pueda lograr un diseño experimental en el que se elaboran conjuntos específicos de productos progenie en cada uno de los diversos recipientes de reacción. Esto permite la realización de un procedimiento sistemático de examen y escrutinio. De este modo, estos métodos permiten examinar sistemáticamente un número potencialmente muy grande de moléculas progenie en grupos más pequeños.

35 Debido a su capacidad para realizar quimerizaciones de una manera que altamente flexible aunque también exhaustiva y sistemática, particularmente cuando hay un bajo nivel de homología entre las moléculas progenitoras es posible la generación de una biblioteca (o conjunto) que comprende un gran número de moléculas progenie. Debido a la naturaleza no estocástica del presente reensamblaje de genes, las moléculas progenie generadas comprenden preferiblemente una biblioteca de moléculas de ácido nucleico quiméricas finalizadas que tienen un orden de ensamblaje global que se elige mediante diseño. En un aspecto concreto, tal biblioteca generada comprende más de 10^3 a más de 10^{1000} especies moleculares progenie diferentes.

40 Un conjunto de moléculas de ácido nucleico quiméricas finalizadas, producidas como se ha descrito, puede comprender un polinucleótido que codifica un polipéptido. Este polinucleótido puede ser gen, que puede ser un gen elaborado por el hombre. Este polinucleótido forma parte de una ruta génica, que puede ser una ruta génica elaborada por el hombre. Uno o más genes elaborados por el hombre generados de este modo se pueden incorporar a una ruta génica elaborada por el hombre, tal como una ruta génica operable en un organismo eucariótico (incluyendo una planta).

45 En otra ilustración, la naturaleza sintética de la etapa en la que se generan los elementos fundamentales permite el diseño e introducción de nucleótidos (p. ej., uno o más nucleótidos, que pueden ser, por ejemplo, codones o intrones o secuencias reguladoras) que se pueden eliminar después opcionalmente en un procedimiento *in vitro* (p. ej., mediante mutagénesis) o en un procedimiento *in vivo* (p. ej., utilizando la capacidad de empalme del gen de un organismo anfitrión). Se aprecia que en muchos casos puede ser también deseable la introducción de estos

nucleótidos por otras muchas razones además del beneficio potencial de la creación de un punto de demarcación práctico.

5 De este modo, esto proporciona que se pueda utilizar un elemento fundamental de ácido nucleico para introducir un intrón. Los intrones funcionales se pueden introducir en un gen elaborado por el hombre de la invención. Los intrones funcionales se pueden introducir en una ruta génica elaborada por el hombre. Esto genera un polinucleótido quimérico que es un gen elaborado por el hombre que contiene uno (o más) intrones introducidos artificialmente

10 Por lo tanto, esto también proporciona la generación de un polinucleótido quimérico que es una ruta génica elaborada por el hombre que contiene uno (o más) intrones introducidos artificialmente. Preferiblemente, el intrón o los intrones introducidos artificialmente pueden ser funcionales en una o más células anfitrionas para el empalme de genes al igual que los intrones de origen natural sirven funcionalmente para el empalme de genes. Un procedimiento produce polinucleótidos que contienen intrones elaborados por el hombre para introducirlos en organismos anfitriones para la recombinación y/o el empalme.

15 Un gen elaborado por el hombre producido de este modo puede servir también como sustrato para la recombinación con otros ácido nucleico. Asimismo, una ruta génica elaborada por el hombre producida de este modo puede servir también como sustrato para la recombinación con otro ácido nucleico. La recombinación se puede facilitar mediante, o se produce en, zonas de homología entre el gene que contiene intrones elaborado por el hombre, y un ácido nucleico, que sirve como pareja de recombinación. En un aspecto, la pareja de recombinación puede ser también un ácido nucleico generado mediante estos métodos, incluyendo un gen elaborado por el hombre o una ruta génica elaborada por el hombre. La recombinación puede ser facilitada por, o se puede producir en zonas de homología que existen en uno (o más) intrones introducidos artificialmente en el gen elaborado por el hombre.

20 El método de reensamblaje sintético de genes utiliza una pluralidad de elementos fundamentales de ácido nucleico, cada uno de los cuales tiene preferiblemente dos extremos ligables. Los dos extremos ligables de cada elemento fundamental del ácido nucleico puede tener dos extremos romos (es decir, cada uno tiene un saliente de cero nucleótidos), o preferiblemente un extremo romo y un saliente, o aún más preferiblemente dos salientes.

25 Un saliente útil para este fin puede ser un saliente 3' o un saliente 5'. De este modo, un elemento fundamental del ácido nucleico puede tener un saliente 3' o alternativamente un saliente 5' o alternativamente dos salientes 3' o alternativamente dos salientes 5'. El orden general en el que se ensamblan los elementos fundamentales de ácido nucleico para formar una molécula de ácido nucleico quimérico finalizada se determina mediante diseño experimental intencionado y no es aleatorio.

30 En un aspecto, un elemento fundamental del ácido nucleico se genera mediante síntesis química de dos ácidos nucleicos de hebra sencilla (también referidos como oligos de hebra sencilla) y poniéndolos en contacto de manera que se les permita recocerse para formar un elemento fundamental de ácido nucleico de doble hebra.

35 Un elemento fundamental de ácido nucleico de doble hebra puede tener un tamaño variable. Los tamaños de estos elementos fundamentales pueden ser pequeños o grandes. Los tamaños ilustrativos para el elemento fundamental oscilan de 1 par de bases (sin incluir ningún saliente) a 100.000 pares de bases (sin incluir ningún saliente). Se proporcionan también otros intervalos de tamaño ilustrativos, que tienen límites inferiores de 1 pb a 10.000 pb (incluyendo cada valor entero intermedio) y límites superiores de 2 pb a 100.000 pb (incluyendo cada valor entero intermedio).

40 Existen muchos métodos por medio de los cuales se puede generar un elemento fundamental de ácido nucleico de doble hebra que sea práctico; y éstos son conocidos en la técnica y pueden ser realizados fácilmente por el experto en la técnica.

45 Un elemento fundamental de ácido nucleico de doble hebra se puede generar generando en primer lugar dos ácidos nucleicos de hebra sencilla y permitiéndoles recocerse para formar un elemento fundamental de ácido nucleico de doble hebra. Las dos hebras de un elemento fundamental de ácido nucleico de doble hebra pueden ser complementarias a cada nucleótido aparte de cualquiera de los que forman un saliente; sin contener de este modo emparejamientos erróneos, aparte de cualquier saliente o salientes.

50 Las dos hebras de un elemento fundamental de ácido nucleico de doble hebra son complementarias en no todos los nucleótidos aparte de cualquiera de los que forman un saliente. De este modo, se puede utilizar un elemento fundamental de ácido nucleico de doble hebra para introducir degeneración codónica. La degeneración codónica se puede introducir utilizando la mutagénesis de saturación de sitio descrita en la presente memoria, utilizando uno o más casetes N,N,G/T o alternativamente utilizando uno o más casetes N,N,N.

60

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

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

15 En un aspecto la descripción descrita en la presente memoria se refiere al uso de ciclos repetidos de reordenamiento reductivo, recombinación y selección que permite la evolución molecular dirigida de secuencias lineales altamente complejas, tales como ADN, ARN o proteínas a través de recombinación.

Sistema de Evolución Dirigida Optimizada

20 Un sistema de modificación de genes no estocástico denominado "sistema de evolución dirigida optimizada" genera polipéptidos, p. ej., glucanasas o anticuerpos de la invención, con propiedades nuevas o alteradas. La evolución directa optimizada está dirigida al uso de ciclos repetidos de reorganización reductiva, recombinación y selección que permiten la evolución molecular dirigida de los ácidos nucleicos a través de recombinación. La evolución dirigida optimizada permite la generación de una gran población de secuencias quiméricas evolucionadas, donde la
25 población generada está enriquecida significativamente en secuencias que tienen un número predeterminado de eventos de entrecruzamiento.

Un evento de entrecruzamiento es un punto en una secuencia quimérica en el que se produce un desplazamiento en la secuencia de una variante parental a otra variante parental. Tal punto se encuentra normalmente en la conexión
30 en la que los oligonucleótidos de los dos orígenes se ligan entre sí para formar una sola secuencia. Este método permite el cálculo de las concentraciones correctas de las secuencias de oligonucleótidos de manera que la población final de secuencias quiméricas está enriquecida con el número elegido de eventos de entrecruzamiento. Esto proporciona más control sobre las variantes quiméricas de elección que tienen un número predeterminado de eventos de entrecruzamiento.

35 Además, este método proporciona un medio conveniente para explorar una cantidad enorme del posible espacio variable de proteína en comparación con otros sistemas. Previamente, si se generaran, por ejemplo, 10^{13} moléculas quiméricas durante un reacción, sería extremadamente difícil someter a ensayo este alto número de variantes quiméricas para una actividad concreta. Por otra parte, una porción significativa de la población progenie tendría un
40 número muy alto de eventos de entrecruzamiento que darían como resultado proteínas que tendrían menos probabilidad de tener aumentos de los niveles de una actividad concreta. Utilizando estos métodos, la población de moléculas quiméricas puede estar enriquecida en esas variantes que tienen un número concreto de eventos de entrecruzamiento. De este modo, aunque todavía se pudieran generar 10^{13} moléculas quiméricas durante una
45 reacción, cada una de las moléculas elegidas para un análisis adicional tendría muy probablemente, por ejemplo, solo tres eventos de entrecruzamiento. Puesto que la población progenie resultante se podría desviar para que tuviera un número predeterminado de eventos de entrecruzamiento, los límites de variedad funcional entre las moléculas quiméricas son reducidos. Esto proporciona un número de variables más controlable cuando se calcula qué oligonucleótido de los polinucleótidos parentales originales sería responsable de la influencia en un rasgo concreto .

50 Un método para crear una secuencia de polinucleótidos progenie quimérica es crear oligonucleótidos correspondientes a fragmentos o porciones de cada secuencia parental. Cada oligonucleótido incluye preferiblemente una región de solapamiento única de manera que la mezcla de oligonucleótidos entre sí da como resultado una nueva variante que tiene fragmentos de oligonucleótidos ensamblados en el orden correcto. También
55 se puede encontrar información adicional, p. ej., en el documento USSN 09/332.835; Patente de los Estados Unidos Núm. 6.361.974.

El número de oligonucleótidos generados para cada variante parental produce una relación con respecto al número total de entrecruzamientos resultantes en la molécula quimérica que se ha creado por último. Por ejemplo, se
60 podrían proporcionar tres variantes de secuencias de nucleótidos parentales para que experimentaran una reacción de ligación con el fin de encontrar una variante quimérica que tuviera, por ejemplo, mayor actividad a alta temperatura. Como ejemplo, se puede generar un conjunto de 50 secuencias de oligonucleótidos correspondientes a cada una de las porciones de cada variante parental. Por lo tanto, durante el procedimiento de reensamblaje por ligación podría haber hasta 50 eventos de entrecruzamiento en cada una de las secuencias quiméricas. La

probabilidad de que cada uno de los polinucleótidos quiméricos generados contenga oligonucleótidos de cada variante parental en orden alterno es muy baja. Si cada fragmento de oligonucleótido está presente en la reacción de ligación en la misma cantidad molar es probable que en algunas posiciones se ligen oligonucleótidos del mismo polinucleótido parental con otro contiguo y de este modo no se produzca un evento de entrecruzamiento. Si la concentración de cada oligonucleótido de cada progenitor se mantiene constante durante cualquier etapa de ligación en este Ejemplo, existe un tercio de probabilidades (suponiendo 3 progenitores) de que un oligonucleótido de la misma variante parental se ligue en la secuencia quimérica y no produzca entrecruzamiento.

Por lo tanto, se puede determinar la función de densidad de probabilidad (FDP) para pronosticar la población de eventos de entrecruzamiento que es probable que se produzcan durante cada etapa en una reacción de ligación dado un número establecido de variantes parentales, un número de oligonucleótidos correspondiente a cada variante, y las concentraciones de cada variante durante cada etapa en la reacción de ligación. Las estadísticas y matemáticas subyacentes a la determinación de la FDP se describen más abajo. Mediante la utilización de estos métodos, se puede calcular tal función de densidad de probabilidad, y de este modo enriquecer la población progeñie quimérica para un número predeterminado de eventos de entrecruzamiento resultantes de una reacción de ligación concreta. Por otra parte, puede estar predeterminado un número objetivo de eventos de entrecruzamiento, y el sistema programado a continuación para calcular las cantidades de partida de cada oligonucleótido parental durante cada etapa en la reacción de ligación para dar como resultado una función de densidad de probabilidad que tiene como eje central el número predeterminado de eventos de entrecruzamiento. Estos métodos están dirigidos al uso de ciclos repetidos de reordenamiento reductivo, recombinación y selección que permite la evolución molecular dirigida de un ácido nucleico que codifica un polipéptido a través de recombinación. Este sistema permite la generación de una gran población de secuencias quiméricas evolucionadas, donde la población generada está enriquecida significativamente en secuencias que tienen un número predeterminado de eventos de entrecruzamiento. Un evento de entrecruzamiento es un punto en una secuencia quimérica en el que se produce un desplazamiento en la secuencia de una variante parental a otra variante parental. Tal punto se encuentra normalmente en la conexión en la que los oligonucleótidos de los dos orígenes se ligan entre sí para formar una sola secuencia. Este método permite el cálculo de las concentraciones correctas de las secuencias de oligonucleótidos de manera que la población final de secuencias quiméricas está enriquecida con el número elegido de eventos de entrecruzamiento. Esto proporciona más control sobre las variantes quiméricas de elección que tienen un número predeterminado de eventos de entrecruzamiento.

Además, estos métodos proporcionan un medio conveniente para explorar una cantidad enorme del posible espacio variable de proteína en comparación con otros sistemas. Utilizando los métodos descritos de la presente memoria, la población de moléculas quiméricas puede estar enriquecida para esas variantes que tienen un número concreto de eventos de entrecruzamiento. De este modo, aunque todavía se pudieran generar 10^{13} moléculas quiméricas durante una reacción, cada una de las moléculas elegidas para un análisis adicional tendría muy probablemente, por ejemplo, solo tres eventos de entrecruzamiento. Puesto que la población progeñie resultante se podría desviar para que tuviera un número predeterminado de eventos de entrecruzamiento, los límites de variedad funcional entre las moléculas quiméricas son reducidos. Esto proporciona un número de variables más controlable cuando se calcula qué oligonucleótido de los polinucleótidos parentales originales sería responsable de la influencia en un rasgo concreto .

En un aspecto, el método crea una secuencia de polinucleótidos progeñie quimérica creando oligonucleótidos correspondiente a fragmentos o porciones de cada secuencia parental. Cada oligonucleótido incluye en un aspecto una única región de solapamiento de manera que la mezcla de oligonucleótidos entre sí da como resultado una nueva variante que tiene cada fragmento de oligonucleótido ensamblado en el orden correcto. Véase también el documento USSN 09/332.835.

Determinación de los Eventos de Entrecruzamiento

Los aspectos incluyen un sistema y soporte lógico reciben una función de densidad de probabilidad de entrecruzamiento (FDE), el número de genes parentales que se van a reensamblar, y el número de fragmentos en el reensamblaje como entradas. La salida de este programa es un "fragmento FDE" que se puede que se puede utilizar para determinar una receta para producir genes reensamblados, y la FDE de entrecruzamiento estimada de esos genes. El procedimiento descrito en la presente memoria se realiza preferiblemente en MATLAB™ (The Mathworks, Natick, Massachusetts) un lenguaje de programación y entorno de desarrollo para cálculo técnico.

Procedimientos Iterativos

Estos procedimientos se pueden repetir iterativamente. Por ejemplo, un ácido nucleico (o, el ácido nucleico) responsable de un fenotipo de xilanasa nuevo o alterado se identifica, re-aísla, se modifica de nuevo, se vuelve a someter a ensayo para determinar su actividad. Este procedimiento se puede repetir iterativamente hasta que se construya un fenotipo deseado. Por ejemplo, se puede construir una ruta bioquímica anabólica o catabólica completa en una célula, incluyendo, p. ej., actividad glucanasa.

De un modo similar, si se determina que un oligonucleótido concreto no afecta en absoluto al rasgo deseado (p. ej., un nuevo fenotipo de glucanasa), éste se puede eliminar como una variable sintetizando oligonucleótidos parentales más grandes que incluyen la secuencia que se va a eliminar. Puesto que la incorporación de la secuencia en una secuencia más grande evita cualquier evento de entrecruzamiento, no habrá ninguna variación de esta secuencia en los polinucleótidos progenie. Esta práctica iterativa para determinar cuáles oligonucleótidos están más relacionados con el rasgo deseado, y cuáles no están relacionados, permite una exploración más eficaz de las variantes de proteínas posibles podrían proporcionar un rasgo o actividad concretos.

10 **Barajado *in vivo***

El barajado *in vivo* de moléculas se utiliza en métodos que proporcionan variantes de polipéptidos de la invención, p. ej., anticuerpos, glucanasas, y similares. El barajado *in vivo* se puede realizar utilizando la propiedad natural de las células para recombinar múltimeros. Si bien la recombinación *in vivo* ha proporcionado la ruta natural principal para la diversidad molecular, la recombinación genética sigue siendo un procedimiento relativamente complejo que implica 1) el reconocimiento de homologías; 2) escisión de la hebra, invasión de la hebra, y etapas metabólicas que conducen a la producción de un quiasma recombinante; y finalmente 3) la resolución del quiasma a moléculas recombinadas discretas. La formación del quiasma requiere el reconocimiento de secuencias homólogas.

En otro aspecto, la invención incluye un método para producir un polinucleótido híbrido a partir de al menos un primer polinucleótido y un segundo polinucleótido. La invención se puede utilizar para producir un polinucleótido híbrido introduciendo al menos un primer polinucleótido y un segundo polinucleótido que comparten al menos una región de homología de secuencia parcial en una célula anfitriona adecuada. Las regiones de homología de secuencia parcial promueven procedimientos que dan como resultado reorganización de secuencia produciendo un polinucleótido híbrido. El término "polinucleótido híbrido", según se utiliza de la presente memoria, es cualquier secuencia de nucleótidos que resulta de este método y contiene secuencias de al menos dos secuencias de polinucleótidos. Tales polinucleótidos híbridos pueden resultar de eventos de recombinación intermoleculares que promueven la integración de secuencias entre moléculas de ADN. Además, tales polinucleótidos híbridos pueden resultar de procedimientos de reordenamiento reductivo intramolecular que utilizan secuencias repetidas para alterar una secuencia de nucleótidos en una molécula de ADN.

El reordenamiento *in vivo* está centrado en procedimientos "inter-moleculares" referidos colectivamente como "recombinación" que en bacterias, es considerado generalmente como un fenómeno "dependiente de RecA". La invención puede basarse en procedimientos de recombinación de una célula anfitriona para recombinar y reordenar secuencias, o la capacidad de las células para mediar procedimientos reductivos para reducir la complejidad de las secuencias cuasi repetidas en la célula mediante delección. Este procedimiento de "reordenamiento reductivo" se produce mediante un procedimiento "intra-molecular", independiente de RecA.

Por lo tanto, se pueden generar polinucleótidos novedosos mediante el procedimiento de reordenamiento reductivo. El método implica la generación de constructos que contienen secuencias consecutivas (secuencias codificantes originales), su inserción en un vector apropiado y su posterior introducción en una célula anfitriona apropiada. El reordenamiento de las entidades moleculares individuales se produce mediante procedimientos combinatorios entre las secuencias consecutivas en el constructo que posee regiones de homología, o entre unidades cuasi repetidas. El procedimiento de reordenamiento recombina y/o reduce la complejidad y la extensión de las secuencias repetidas y da como resultado la producción de especies moleculares novedosas. Se pueden aplicar diversos tratamientos para aumentar la velocidad de reordenamiento. Estos podrían incluir tratamiento con luz ultravioleta, o agentes químicos que dañan el ADN y/o el uso de líneas de células anfitrionas que despliegan niveles incrementados de "inestabilidad genética". De este modo el procedimiento de reordenamiento puede implicar recombinación homóloga o la propiedad natural de las secuencias cuasi repetidas para dirigir su propia evolución.

Las secuencias repetidas o "cuasi repetidas" juegan un papel en la inestabilidad genética. Las "cuasi-repeticiones" son repeticiones que no están restringidas a su estructura unitaria original. Las unidades cuasi repetidas se pueden presentar como una matriz de secuencias en un constructo; unidades consecutivas de secuencias similares. Una vez ligadas, las conexiones entre las secuencias consecutivas se hacen esencialmente invisibles y la naturaleza cuasi repetitiva del constructo resultante es ahora continua a nivel molecular. El procedimiento de delección que realiza la célula para reducir la complejidad del constructo resultante funciona entre las secuencias cuasi repetidas. Las unidades cuasi repetidas proporcionan un repertorio prácticamente ilimitado sobre los cuales pueden ocurrir eventos de deslizamiento. Los constructos que contienen las cuasi repeticiones proporcionan de este modo eficazmente suficiente elasticidad molecular como para que puedan ocurrir eventos de delección (y potencialmente inserción) virtualmente en cualquier parte en las unidades cuasi repetitivas.

Cuando las secuencias cuasi repetidas están ligadas todas en la misma orientación, por ejemplo cabeza a cola o viceversa, la célula no puede distinguir las unidades individuales. Por consiguiente, el procedimiento reductivo puede ocurrir por todas las secuencias. En contraste, cuando por ejemplo, las unidades se presentan cabeza a cabeza, en lugar de cabeza a cola, la inversión perfila los extremos de la unidad adyacente de manera que la formación de la

deleción favorecerá la pérdida de unidades discretas. De este modo, es preferible con el presente método que las secuencias se encuentren en la misma orientación. La orientación aleatoria de secuencias cuasi repetidas dará como resultado la pérdida de eficacia de reordenamiento, mientras que la orientación uniforme de las secuencias ofrecerá la mayor eficacia. Sin embargo, si bien la posesión de menos secuencias contiguas en la misma orientación reduce la eficacia, aún puede proporcionar suficiente elasticidad para la recuperación eficaz de las moléculas novedosas. Los constructos se pueden elaborar con las secuencias cuasi repetidas en la misma orientación para permitir una mayor eficacia.

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

- a) Se pueden utilizar cebadores que incluyen una cabeza poli-A y una cola poli-T que cuando se elaboran de hebra sencilla proporcionan orientación. Esto se logra elaborando a partir de ARN las primeras pocas bases de los cebadores y por tanto eliminándolas fácilmente con ARNasa H.
- b) Se pueden utilizar cebadores que incluyen sitios de escisión por restricción únicos. Se requerirían sitios múltiples, una batería de secuencias únicas y etapas de ligación y síntesis repetidas.
- c) Las pocas bases internas del cebador podrían estar tioladas y se utilizaría una exonucleasa para producir las moléculas apropiadamente dotadas de cola.

La recuperación de las secuencias se basa en la identificación de vectores de clonación con un índice repetitivo reducido (IR). Las secuencias codificantes reordenadas se pueden recuperar de este modo mediante amplificación. Los productos se vuelven a clonar y a expresar. La recuperación de los vectores de clonación con IR reducido se puede efectuar mediante:

- 1) El uso de vectores solo mantenidos establemente cuando el constructo se reduce en complejidad.
- 2) La recuperación física de los vectores acortados mediante procedimientos físicos. En este caso, el vector de clonación sería recuperado utilizando procedimientos de aislamiento de plásmidos convencionales y se fraccionaría su tamaño en gel de agarosa, o una columna con un bajo peso molecular de corte utilizando procedimientos convencionales.
- 3) La recuperación de vectores que contienen genes interrumpidos se puede seleccionar cuando el tamaño del inserto disminuye.
- 4) El uso de técnicas de selección directa con un vector de expresión y la selección apropiada.

Las secuencias codificantes (por ejemplo, genes) de organismos relacionados pueden mostrar un alto grado de homología y codifican productos proteicos bastante diversos. Estos tipos de secuencias son particularmente útiles en la presente invención en forma de cuasi repeticiones. Sin embargo, mientras que los Ejemplos ilustrados más abajo demuestran el reordenamiento de secuencias codificantes originales casi idénticas (cuasi repeticiones), este procedimiento no está limitado a tales repeticiones idénticas.

El siguiente ejemplo muestra un método de la invención. Se describen las secuencias codificantes de ácido nucleico (cuasi repeticiones) derivadas de tres (3) especies únicas. Cada secuencia codifica una proteína con un conjunto de propiedades distintas. Cada una de las secuencias difiere en uno o unos pocos pares de bases en una única posición de la secuencia. Las secuencias cuasi repetidas se amplifican y se ligan separadamente o colectivamente en ensamblajes aleatorios de manera que estén disponibles todas las permutaciones y combinaciones posibles en la población de moléculas ligadas. El número de unidades cuasi repetidas puede estar controlado por las condiciones de ensamblaje. El número medio de unidades cuasi repetidas en un constructo se define como el índice repetitivo (IR).

Una vez formados, los constructos pueden ser fraccionados, o no, por tamaño en un gel de agarosa de acuerdo con los protocolos publicados, insertados en un vector de clonación y transfectados en una célula anfitriona apropiada. Las células se propagan a continuación y se efectúa el "reordenamiento reductivo". La velocidad del procedimiento de reordenamiento reductivo se puede estimular mediante la introducción de desperfectos en el ADN si se desea. El que la reducción del IR esté mediada por la formación de deleción entre secuencias repetidas por medio de un mecanismo "intra-molecular", o mediada por eventos de tipo recombinación a través de mecanismos "intermoleculares" es irrelevante. El resultado final es un reordenamiento de las moléculas en todas las combinaciones posibles.

Opcionalmente, el método comprende la etapa adicional de escrutinio de los miembros de la biblioteca de la reserva barajada para identificar miembros de la biblioteca barajados individuales que tienen la capacidad de unirse o interactuar de otro modo, o catalizar una reacción concreta (p. ej., tal como el dominio catalítico de una enzima) con una macromolécula predeterminada, tal como por ejemplo un receptor proteínico, un oligosacárido, virión, u otro compuesto o estructura predeterminados.

Los polipéptidos que se identifican a partir de tales bibliotecas se pueden utilizar para fines terapéuticos, diagnósticos, de investigación y propósitos relacionados (p. ej., catalizadores, solutos para incrementar la osmolaridad de una solución acuosa y similares) y/o se pueden someter a uno o más ciclos de barajado y/o selección adicionales.

En otro aspecto, se prevé que antes de o durante la recombinación o el reordenamiento, los polinucleótidos generados por estos métodos se puedan someter a agentes o procedimientos que promuevan la introducción de mutaciones en los polinucleótidos originales. La introducción de tales mutaciones aumentaría la diversidad de los híbridos polinucleotídico resultantes y los polipéptidos codificados de allí. Los agentes o procedimientos que promueven la mutagénesis pueden incluir, pero no están limitados a: (+)-CC-1065, o un análogo sintético tal como (+)-CC-1065-(N3-Adenina) (Véase Sun y Hurley, (1992)); un aducto de 4'-fluro-4-aminobifenilo N-acetilado o desacetilado capaz de inhibir la síntesis de ADN (Véase, por ejemplo, van de Poll *et al.* (1992)); o un aducto de 4-aminobifenilo N-acetilado o desacetilado capaz de inhibir la síntesis de ADN (Véase también, van de Poll *et al.* (1992), págs. 751-758); cromo trivalente, una sal de cromo trivalente, un hidrocarburo aromático policíclico (HAP), un aducto de ADN capaz de inhibir la replicación de ADN, tal como 7-bromometilbenz[a]antraceno ("BMA"), tris(2,3-dibromopropil)fosfato ("Tris-BP"), 1,2-dibromo-3-cloropropano ("DBCP"), 2-bromoacroleína (2BA), benzo[a]pireno-7,8-dihidrodiol-9-10-epóxido ("BPDE"), una sal halogenada de platino(II), N-hidroxi-2-amino-3-metilimidazo[4,5-f]-quinolina ("N-hidroxi-IQ") y N-hidroxi-2-amino-1-metil-6-fenilimidazo[4,5-f]-piridina ("N-hidroxi-PhIP"). Los medios ilustrativos para ralentizar o detener la amplificación mediante PCR consisten en luz UV (+)-CC-1065 y (+)-CC-1065-(N3-Adenina). Los medios incluidos particularmente son los aductos de ADN o polinucleótidos que comprenden los aductos de ADN de los polinucleótidos o reservas de polinucleótidos, que se pueden liberar o eliminar mediante un procedimiento que incluye calentar la solución que comprende los polinucleótidos antes de su procesamiento adicional.

Un método para producir proteínas recombinantes que tienen actividad biológica es mediante tratamiento de una muestra que comprende polinucleótidos molde de doble hebra que codifican una proteína de tipo salvaje en condiciones de acuerdo con la invención que proporcionan la producción de polinucleótidos híbridos o reordenados.

Producción de secuencias variantes

La invención también proporciona métodos adicionales para elaborar variantes de secuencia de las secuencias de ácido nucleico (p. ej., glucanasa) de la invención. La invención también proporciona métodos adicionales para el aislamiento de glucanasas utilizando los ácidos nucleicos y polipéptidos de la invención. En un aspecto, la invención proporciona variantes de una secuencia codificante de glucanasa (p. ej., un gen, ADNc o mensaje) de la invención, que se pueden alterar mediante cualquier método, incluyendo, p. ej., métodos aleatorios o estocásticos, o métodos no estocásticos, o de "evolución dirigida", como se ha descrito anteriormente.

Las variantes aisladas pueden ser origen natural. También se pueden crear variantes *in vitro*. Las variantes se pueden crear utilizando técnicas de ingeniería genética tales como mutagénesis dirigida al sitio, mutagénesis química al azar, procedimientos de delección con Exonucleasa III, y técnicas de clonación convencionales. Alternativamente, tales variantes, fragmentos, análogos, o derivados se pueden crear utilizando procedimientos de síntesis o modificación química. Otros métodos para elaborar variantes son también familiares para los expertos en la técnica. Estos incluyen procedimientos en los que se modifican secuencias de ácido nucleico obtenidas de productos aislados naturales para generar ácidos nucleicos que codifican polipéptidos que tienen características que potencian su valor en aplicaciones industriales o de laboratorio. En tales procedimientos, se generan y caracterizan un gran número de secuencias variantes que tienen una o más diferencias de nucleótidos con respecto a la secuencia obtenida a partir del producto aislado natural. Estas diferencias de nucleótidos pueden dar como resultado cambios de aminoácidos con respecto a los polipéptidos codificados por los ácidos nucleicos de los productos aislados naturales.

Por ejemplo, se pueden crear variantes utilizando PCT propensa a error. En la PCR propensa a error, la PCR se realiza en condiciones en las que la fidelidad de la copia de la ADN polimerasa es baja, de manera que se obtiene un alta tasa de mutaciones puntuales a lo largo de toda la longitud del producto de PCR. La PCR propensa a error es descrita, p. ej., por Leung, D.W., et al., en *Technique*, 1:11-15, 1989) y Caldwell, R. C. & Joyce G.F., *PCR Methods Applic.*, 2:28-33, 1992. En resumen, tales procedimientos, los ácidos nucleicos que se van a mutagenizar se mezclan con cebadores para PCR, tampón de reacción, MgCl₂, MnCl₂, polimerasa Taq y una concentración apropiada de dNTPs para lograr una alta la velocidad de mutación puntual a lo largo de la longitud completa del producto de PCR. Por ejemplo, la reacción se puede realizar utilizando 20 fmoles del ácido nucleico que se va a mutagenizar, 30 pmoles de cada cebador de PCR, un tampón de reacción que comprende KCl 50 mM, Tris HCl 10 mM (pH 8.3) y gelatina al 0,01%, MgCl₂ 7 mM, MnCl₂ 0,5 mM, 5 unidades de polimerasa Taq, dGTP 0,2 mM, dATP 0,2 mM, dCTP 1 mM, y dTTP 1 mM. La PCR se puede realizar 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 pueden variarse 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 utilizando mutagénesis dirigida a oligonucleótidos para generar mutaciones de sitio específico en cualquier ADN clonado de interés. La mutagénesis de oligonucleótidos la describen, p. ej., Reidhaar-Olson (1988) en *Science* 241:53-57. En resumen, en tales procedimientos se sintetiza una pluralidad de oligonucleótidos de doble hebra que portan una o más mutaciones que se van a introducir en el ADN clonado y se inserta en el ADN clonado que se va 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 la 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 producen en paralelo un gran número de reacciones PCR diferentes en el mismo vial, primando los productos de una reacción sobre los productos de otra reacción. La PCR de ensamblaje se describe, p. ej., en la Patente de los Estados Unidos Núm. 5.965.408.

Otro método más para generar variantes es la mutagénesis por PCR sexual. En la mutagénesis por PCR sexual, se produce la recombinación homóloga forzada entre moléculas de ADN de secuencias de ADN diferentes pero íntimamente relacionadas *in vitro*, como resultado de la fragmentación al azar de la molécula de ADN basada en la homología de la secuencia, seguido de fijación del entrecruzamiento por extensión de cebador en una reacción PCR. La mutagénesis por PCR sexual la describe, p. ej., Stemmer (1994) en *Proc. Natl. Acad. Sci. USA* 91:10747-10751. En resumen, en tales procedimientos se digiere con ADNasa una pluralidad de ácidos nucleicos que se van a recombinar para generar fragmentos que tienen un tamaño medio de 50-200 nucleótidos. Los fragmentos de tamaño medio deseado se purifican y resuspenden una mezcla para PCR. La PCR se lleva a cabo en condiciones que faciliten la recombinación entre los fragmentos de ácido nucleico. Por ejemplo, la PCR se puede realizar suspendiendo los fragmentos purificados a una concentración de 10-30 ng/μl en una solución 0,2mM de cada dNTP, MgCl₂ 2,2 mM, KCl 50 mM, Tris HCl 10 mM, pH 9,0, y Triton X-100 al 0,1%. Se añaden 2,5 unidades de polimerasa Taq por 100:1 de mezcla de reacción y la PCR se realiza utilizando del siguiente régimen: 94°C durante 60 segundos, 94°C durante 30 segundos, 50-55°C durante 30 segundos, 72°C durante 30 segundos (30-45 veces) y 72°C durante 5 minutos. Sin embargo, se apreciará que estos parámetros se pueden variar según sea apropiado. En algunos aspectos, se pueden incluir oligonucleótidos en las reacciones PCR. En otros aspectos, se puede utilizar el fragmento de Klenow de ADN polimerasa I en un primer grupo de reacciones PCR y se puede utilizar polimerasa Taq en un grupo de reacciones posterior. Las secuencias recombinantes se aíslan y se evalúan las actividades de los polipéptidos que codifican.

También se pueden crear variantes mediante mutagénesis *in vivo*. En algunos aspectos, se generan mutaciones al azar en una secuencia de interés propagando la secuencia de interés en una cepa bacteriana, tal como una cepa de *E. coli*, que porta mutaciones en uno o más de las rutas de reparación del ADN. Tales cepas "mutadoras" tienen una tasa de mutación aleatoria más alta que un progenitor de tipo salvaje. La propagación del ADN en una de estas cepas generará finalmente mutaciones al azar en el ADN. Las cepas mutadoras adecuadas para su uso para la mutagénesis *in vivo* se describen en la Publicación PCT Núm. WO 91/16427, publicada el 31 de Octubre de 1991, titulada "Methods for Phenotype Creation from Multiple Gene Populations".

También se pueden generar variantes utilizando mutagénesis por inserción de casetes. En la mutagénesis por inserción de casetes una pequeña región de una molécula de ADN de doble hebra se reemplaza por un "casete" de oligonucleótido sintético que difiere de la secuencia nativa. El oligonucleótido contiene con frecuencia una secuencia nativa completamente y/o parcialmente aleatorizada.

También se puede utilizar mutagénesis recursiva de conjunto para generar variantes. La mutagénesis recursiva de conjunto es un algoritmo para construir proteínas (mutagénesis de proteínas) desarrollado para producir poblaciones diversas de mutantes relacionados fenotípicamente cuyos miembros difieren en la secuencia de aminoácidos. Este método utiliza un mecanismo de retroalimentación para controlar las rondas sucesivas de mutagénesis combinatoria por inserción de casetes. La mutagénesis recursiva de conjunto es descrita por Arkin, A.P. y Youvan, D.C., *PNAS*, USA, 89:7811-7815, 1992.

En algunos aspectos, la variantes se crean utilizando mutagénesis de conjunto exponencial. La mutagénesis de conjunto exponencial es un procedimiento para generar bibliotecas combinatorias con un alto porcentaje de mutantes únicos y funcionales, donde se asignan al azar en paralelo grupos de residuos para identificar, en cada posición alterada, los aminoácidos que conducen a proteínas funcionales. La mutagénesis de conjunto exponencial es descrita por Delegrave, S. y Youvan, D.C., en *Biotechnology Research*, 11:1548-1552, 1993. La mutagénesis al azar y dirigida al sitio son descritas por Arnold, F.H., en *Current Opinion in Biotechnology*, 4:450-455, 1993.

En algunos aspectos, las variantes se crean utilizando procedimientos de barajado donde porciones de una pluralidad de ácidos nucleicos que codifican polipéptidos distintos se fusionan entre sí para crear secuencias de ácidos nucleicos quiméricas que codifican polipéptidos quiméricos como se describe en La Patente de los Estados Unidos Núm. 5.965.408, presentada el 9 de Julio de 1996, titulada, "Method of ADN Reassembly by Interrupting

Synthesis" y la Patente de los Estados Unidos Núm. 5.939.250, presentada el 22 de Mayo de 1996, titulada, "Producción of Enzymes Having Desired Activities by Mutagenesis".

5 Las variantes de los polipéptidos de la invención pueden ser variantes en las que uno o más de los residuos de aminoácidos de los polipéptidos de la invención son sustituidos por residuo de aminoácido conservado o no conservado (en un aspecto un residuo de aminoácido conservado) y tal residuo de aminoácido sustituido puede ser uno codificado o no por el código genético.

10 La invención 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 ha comentado en la presente memoria (p. ej., las sustituciones de aminoácidos conservativas son aquellas que sustituyen un aminoácido dado en un polipéptido por otro aminoácido de características parecidas). La invención proporciona polipéptidos (y los ácidos nucleicos que los codifican) donde cualquiera, alguno o todos los residuos de aminoácidos están sustituidos por otro aminoácido de características parecidas, p. ej., una sustitución de aminoácido conservativa.

15 Las sustituciones conservativas son aquellas que sustituyen un aminoácido dado en un polipéptido por otro aminoácido de características parecidas. Las sustituciones consideradas generalmente como conservativas son los siguientes remplazos: remplazos de un aminoácido alifático tal como Alanina, Valina, Leucina y Isoleucina por otro aminoácido alifático; remplazo de una Serina por una Treonina o vice versa; remplazo de un residuo ácido tal como ácido Aspártico y ácido Glutámico por otro residuo ácido; remplazo de un residuo que porta un grupo amida, tal como Asparragina y Glutamina, por otro residuo que porta un grupo amida; intercambio de un residuo alcalino tal como Lisina y Arginina por otro residuo alcalino; y remplazo de un residuo aromático tal como Fenilalanina, Tirosina por otro residuo aromático. En aspectos alternativos, estas sustituciones conservativas pueden ser también equivalentes sintéticos de estos aminoácidos.

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

30 Otras variantes más son aquellas en las que el polipéptido está asociado con otro compuesto, tal como un compuesto para incrementar la vida media del polipéptido (por ejemplo, polietilenglicol).

Son variantes adicionales aquellas en las que se fusionan al polipéptido aminoácidos adicionales, tal como una secuencia líder, una secuencia secretora, una secuencia de proproteína o una secuencia que facilita la purificación, el enriquecimiento, o la estabilización del polipéptido.

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

40 **Optimización de codones para lograr altos niveles de expresión de proteínas en células anfitrionas**

45 La invención proporciona métodos para modificar los ácidos nucleicos que codifican glucanasas para modificar el uso de los codones. En un aspecto, la invención proporciona métodos para modificar codones en un ácido nucleico que codifica una glucanasa para incrementar o disminuir su expresión en una célula anfitriona. La invención también proporciona ácidos nucleicos que codifican una glucanasa, modificada para incrementar su expresión en una célula anfitriona, la glucanasa modificada de este modo, y los métodos para elaborar la glucanasa modificada. El método comprende identificar un codón "no preferido" o "menos preferido" en el ácido nucleico codificante de la glucanasa y reemplazar uno o más de esos codones no preferidos o menos preferidos por un "codón preferido" que codifica el mismo aminoácido que el codón reemplazado y reemplazar al menos un codón no preferido o menos preferido en el ácido nucleico por un codón preferido que codifica el mismo aminoácido. Un codón preferido es un codón representado en exceso en secuencias codificantes en genes en la célula anfitriona y un codón no preferido o menos preferido es un codón infrarrepresentado en secuencias codificantes en genes en la célula anfitriona.

55 Las células anfitrionas para expresar los ácidos nucleicos, casetes de expresión y vectores de la invención incluyen bacterias, levadura, 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 de codón alterado y polipéptidos elaborados con los ácidos nucleicos de codón alterado. Las células anfitrionas ilustrativas incluyen bacterias gram negativas, tales como *Escherichia coli*; bacterias gram positivas, tales como *Streptomyces*, *Lactobacillus gasseri*, *Lactococcus lactis*, *Lactococcus cremoris*, *Bacillus sp.*, *Bacillus subtilis*, *Bacillus cereus*. Las células anfitrionas ilustrativas también incluyen organismos eucarióticos, p. ej., diversas levaduras, tales como *Saccharomyces sp.*, incluyendo *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, *Pichia pastoris*, y *Kluyveromyces lactis*, *Hansenula polymorpha*, *Aspergillus niger*, y células y líneas celulares de mamífero y células y líneas celulares de insecto. De este modo, la invención también incluye ácidos nucleicos y polipéptidos optimizados

en cuanto a su expresión en estos organismos y especies, p. ej., los ácidos nucleicos de la invención tienen codones optimizados para su expresión en una célula anfitriona, p. ej., *Pichia* sp., p. ej., *P. pastoris*, *Saccharomyces* sp., o *Bacillus* sp., *Streptomyces* sp., y similares.

5 Por ejemplo, los codones de un ácido nucleico que codifica un polipéptido de la invención, p. ej., una glucanasa, o una enzima similar aislados de una célula bacteriana, se modifican de manera que el ácido nucleico (que codifica la enzima) sea expresado óptimamente en una célula bacteriana diferente de la que derivó la enzima (p. ej., glucanasa), 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, véase, p. ej., la Patente de los Estados Unidos
10 Núm. 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 codones de optimización en sistemas de ratón; Outchkourov (2002) Protein Expr. Purif. 24:18-24, que describe codones de optimización en levaduras; Feng (2000) Biochemistry 39:15399-15409, que describe codones de optimización en *E. coli*; Humphreys (2000) Protein Expr. Purif. 20:252-264, que describe el uso de un codón de optimización 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 de codones de ADN basado en internet.

Por ejemplo, como se comenta en el Ejemplo 4, más abajo, el ácido nucleico que codifica el polipéptido que tiene una secuencia mostrada en el SEQ ID NO: 6 (p. ej., SEQ ID NO: 5) se sometió a optimización de codones para determinar la expresión óptima en *Pichia pastoris*; el ácido nucleico que codifica la enzima de codón optimizado de *Pichia pastoris* es el SEQ ID NO: 463. El polipéptido que tiene una secuencia mostrada como el SEQ ID NO: 464 en la posición 91 es alanina (SEQ ID NO: 464), y en un aspecto alternativo, valina (como en el SEQ ID NO: 6). De un modo similar, el ácido nucleico ilustrativo que codifica el SEQ ID NO: 6 (es decir, el SEQ ID NO: 5) puede codificar o bien alanina o bien valina (u otra sustitución conservativa) en la posición 91. De hecho, la invención 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 comenta en la presente memoria (p. ej., las sustituciones de aminoácidos conservativas son aquellas que sustituyen un aminoácido dado en un polipéptido por otro aminoácido de características similares), como se comenta en la presente memoria.

30 Animales no humanos transgénicos

La invención proporciona animales no humanos transgénicos que comprenden un ácido nucleico, un polipéptido (p. ej., una glucanasa), un casete de expresión o vector o una célula transfectada o transformada de la invención. La invención también proporciona métodos para elaborar y utilizar estos animales no humanos transgénicos.

Los animales no humanos transgénicos pueden ser, p. ej., cabras, conejos, ovejas, cerdos, vacas, ratas y ratones, que comprenden los ácidos nucleicos de la invención. Estos animales se pueden utilizar, p. ej., como modelos *in vivo* para estudiar la actividad glucanasa, o, como modelos para escrutar agentes que cambian la actividad glucanasa *in vivo*. Las secuencias codificantes de los polipéptidos que se van a expresar en los animales no humanos transgénicos se pueden diseñar para que sean constitutivas, o, estén bajo el control de factores reguladores transcripcionales específicos de tejido, específicos del desarrollo o inducibles. Los animales no humanos transgénicos se pueden diseñar y generar utilizando cualquier método conocido en la técnica; véanse, p. ej., las Patentes de los Estados Unidos Núms. 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 elaboran y utilizan células transformadas y huevos y ratones, ratas, conejos, ovejas, cerdos y vacas transgénicos. Véase también, p. ej., Pollock (1999) J. Immunol. Methods 231:147-157, que describe la producción de proteínas recombinantes en la leche de animales productores de leche transgénicos; Baguisi (1999) Nat. Biotechnol. 17:456-461, que demuestra la producción de cabras transgénicas. La Patente de los Estados Unidos Núm. 6.211.428, describe la realización y utilización de mamíferos no humanos transgénicos que expresan en sus cerebros un constructo de ácido nucleico que comprende una secuencia de ADN. La Patente de los Estados Unidos Núm. 5.387.742, describe la inyección de secuencias ADN recombinante o sintético clonadas en huevos de ratón fertilizados, la implantación de los huevos inyectados en hembras pseudo-preñadas, y el crecimiento a largo plazo de ratones transgénicos cuyas células expresan proteínas relacionadas con la patología de la enfermedad de Alzheimer. La Patente de los Estados Unidos Núm. 6.187.992, describe la elaboración y utilización de un ratón transgénico cuyo genoma comprende una interrupción del gen codificante de la proteína precursora del amiloide (PPA).

También se pueden utilizar "animales que carecen de la expresión específica de un gen (knockout)" para poner en práctica los métodos de la invención. Por ejemplo, en un aspecto, los animales transgénicos o modificados de la invención comprenden un "animal knockout", p. ej., un "ratón knockout", construido para no expresar un gen endógeno, que es remplazado por un gen que expresa una glucanasa de la invención, o, una proteína de fusión que comprende una glucanasa de la invención.

Plantas y Semillas Transgénicas

La invención proporciona planta y semillas transgénicas que comprenden un ácido nucleico, un polipéptido (p. ej., una glucanasa), un casete de expresión o vector o una célula transfectada o transformada de la invención. La invención también proporciona productos vegetales, p. ej., aceites, semillas, hojas, extractos y similares, que comprenden un ácido nucleico y/o un polipéptido (p. ej., una glucanasa) de la invención. La planta transgénica puede ser dicotiledónea (dos cotiledones) o monocotiledónea (un cotiledón). La invención también proporciona métodos para elaborar y utilizar estas plantas y semillas transgénicas. La planta o célula vegetal transgénica que expresan un polipéptido de la presente invención se pueden construir de acuerdo con cualquier método conocido en la técnica. Véase, por ejemplo, la Patente de los Estados Unidos Núm. 6.309.872.

Los ácidos nucleicos y constructos de expresión de la invención se pueden introducir en una célula vegetal por medio de cualquier método. Por ejemplo, los ácidos nucleicos o constructos de expresión se pueden introducir en el genoma de una planta anfitriona deseada, o, los ácidos nucleicos o constructos de expresión pueden ser episomas. La introducción en el genoma de una planta deseada puede ser de manera que la producción de glucanasa por el anfitrión esté regulada por elementos de control de la transcripción o la traducción endógenos. La invención también proporciona "plantas knockout" donde la inserción de la secuencia génica, p. ej., mediante recombinación homóloga, ha interrumpido la expresión del gen endógeno. Los medios para generar plantas "knockout" son bien conocidos en la técnica, véase, p. ej., 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 utilizar para conferir rasgo deseados a esencialmente cualquier planta, p. ej., a plantas que producen almidón, tales como la patata, el trigo, el arroz, la cebada, y similares. Los ácidos nucleicos de la invención se pueden utilizar para manipular las rutas metabólicas de una planta con el fin de optimizar o alterar la expresión en el anfitrión de la glucanasa. Estos pueden cambiar la actividad glucanasa en una planta. Alternativamente, una glucanasa de la invención se puede utilizar 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 producto novedoso.

En un aspecto, la primera etapa en la producción de una planta transgénica implica elaborar un constructo de expresión para su expresión en una célula vegetal. Estos mecanismos son bien conocidos en la técnica. Pueden incluir seleccionar y clonar un promotor, una secuencia codificante para facilitar la unión eficaz de los ribosomas al ARNm y seleccionar las secuencias terminadoras de genes apropiadas. Un promotor constitutivo ilustrativo es CaMV35S, del virus del mosaico de la coliflor, que generalmente da como resultado un alto grado de expresión en plantas. Otros promotores son más específicos y responden a indicaciones en el entorno interno o externo de la planta. Un promotor inducible por luz ilustrativo es el promotor del gen cab, codificante de la proteína de unión a clorofila a/b grande.

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

Se pueden añadir genes con marcadores seleccionables al constructo génico con el fin de identificar células vegetales o tejidos que tengan integrado el transgen satisfactoriamente. Esto puede ser necesario puesto que lograr la incorporación y expresión de genes en células vegetales es un evento raro, que ocurre solo en un pequeño porcentaje de tejidos o células diana. Los genes marcadores seleccionables codifican proteínas que proporcionan resistencia a agentes que son normalmente tóxicos para las plantas, tales como antibióticos o herbicidas. Únicamente las células vegetales que tienen integrado el gen marcador seleccionable sobrevivirán cuando crezcan sobre un medio que contiene el antibiótico o herbicida apropiado. Como para los otros genes insertados, los genes marcadores también requieren secuencias promotoras y de terminación para su funcionamiento apropiado.

En un aspecto, la elaboración de plantas o semillas transgénicas comprende la incorporación de secuencias de la invención y, opcionalmente, genes marcadores a un constructo de expresión diana (p. ej., un plásmido), junto con la colocación de las secuencias promotoras y terminadoras. Esto puede implicar la transferencia del gen modificado a la planta a través de un método adecuado. Por ejemplo, se puede introducir un constructo directamente en el ADN genómico de la célula vegetal utilizando técnicas tales como electroporación y microinyección de protoplastos de células vegetales, o los constructos se pueden introducir directamente en tejidos vegetales utilizando métodos balísticos, tales como bombardeo de partículas de ADN. Por ejemplo, véase, p. ej., 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 discute el uso de bombardeo de partículas para introducir transgenes en trigo; y Adam (1997) supra, para el uso de bombardeo de partículas para introducir YAC en células vegetales. Por ejemplo,

Rinehart (1997) supra, utilizó bombardeo de partículas para generar plantas de algodón transgénicas. El aparato para acelerar partículas se describe en la Patente de los Estados Unidos Núm. 5.015.580; y, aparato de aceleración de partículas BioRad (Biolistics) PDS-2000 asequible comercialmente; véanse también, John, Patente de los Estados Unidos Núm. 5.608.148; y Ellis, Patente de los Estados Unidos Núm. 5.681.730, que describe transformación mediada por partículas de gimnospermas.

En un aspecto, los protoplastos se pueden inmovilizar y se les pueden inyectar ácidos nucleicos, p. ej., un constructo de expresión. Aunque la regeneración de plantas a partir de protoplastos no es fácil con cereales, es posible la regeneración de plantas en legumbres utilizando embriogénesis somática de callos derivados de protoplastos. Los tejidos organizados se pueden transformar con ADN desnudo utilizando la técnica de la pistola génica, en la que el ADN se aplica como recubrimiento sobre microproyectiles de tungsteno, disparo 1/100^a del tamaño de las células, que transporta el ADN profundamente a las células u organelos. Se induce a continuación la regeneración del tejido transformado, usualmente mediante embriogénesis somática. Esta técnica ha resultado satisfactoria en diversas especies de cereales incluyendo maíz y arroz.

También se pueden introducir ácidos nucleicos, p. ej., constructos de expresión, en células vegetales utilizando virus recombinantes. Se pueden transformar células vegetales utilizando vectores virales, tales como, p. ej., vectores derivados del virus del mosaico del tabaco (Rouwendal (1997) Plant Mol. Biol. 33:989-999), véase Porta (1996) "Use of viral replicons for the expresion of genes in plants", Mol. Biotechnol. 5:209-221.

Alternativamente, se pueden combinar ácidos nucleicos, p. ej., un constructo de expresión, con regiones flanqueantes de ADN-T adecuadas e introducir en un vector anfitrión de *Agrobacterium tumefaciens* convencional. Las funciones de virulencia del anfitrión *Agrobacterium tumefaciens* dirigirá la inserción del constructo y el marcador adyacente en el ADN de la célula vegetal cuando la célula es infectada por la bacteria. Las transformaciones técnicas mediadas por *Agrobacterium tumefaciens*, incluyendo el desarme y uso de vectores binarios, están bien descritas en la bibliografía científica. Véase, p. ej., Horsch (1984) Science 233:496-498; Fraley (1983) Proc. Natl. Acad. Sci. USA 80:4803 (1983); Gene Transfer to Plants, Potrykus, ed. (Springer-Verlag, Berlin 1995). El ADN en una célula de *A. tumefaciens* están contenido en el cromosoma bacteriano así como en otra estructura conocida como plásmido Ti (inductor de tumor). El plásmido Ti contiene un tramo de ADN denominado ADN T (~20 kb de longitud) que es transferido a célula vegetal en el procedimiento de infección y una serie de genes vir (virulencia) que dirige el procedimiento de infección. *A. tumefaciens* puede infectar solo una planta a través de heridas: cuando una raíz o tallo de una planta se hiere produce ciertas señales químicas, en respuesta a la cual, los genes vir de *A. tumefaciens* se activan y dirigen una serie de eventos necesarios para la transferencia del ADN T desde el plásmido Ti al cromosoma de la planta. El ADN T entra a continuación en la célula vegetal a través de la herida. Se especula que el ADN T espera hasta que el ADN de la planta se está replicando o transcribiendo, después se autoinserta en el ADN de la planta expuesto. Con el fin de utilizar *A. tumefaciens* como vector transgen, se tiene que eliminar la sección del ADN T que induce tumor, mientras que se conservan las regiones limítrofes del ADN T y los genes vir. El transgen se inserta a continuación entre las regiones limítrofes del ADN T, donde se transfiere a la célula vegetal y se integra en los cromosomas de las plantas.

La invención proporciona la transformación de plantas monocotiledóneas utilizando los ácidos nucleicos de la invención, incluyendo cereales importantes, véase Hiei (1997) Plant Mol. Biol. 35:205-218. Véanse también, p. ej., Horsch, Science (1984) 233:496; Fraley (1983) Proc. Natl. Acad. Sci USA 80:4803; Thykjaer (1997) supra; Park (1996) Plant Mol. Biol. 32:1135-1148, que comentan la integración de ADN T en ADN genómico. Véase también D'Halluin, Patente de los Estados Unidos Núm. 5.712.135, que describe un procedimiento para la integración estable de un ADN que comprende un gen que es funcional en una célula de un cereal, u otra planta monocotiledónea.

En un aspecto, la tercera etapa puede implicar la selección y regeneración de plantas completas capaces de transmitir el gen diana incorporado a la siguiente generación. Tales técnicas de regeneración se basan en la manipulación de ciertas fitohormonas en un medio de crecimiento de cultivo de tejido, basándose típicamente en un marcador biocida y/o herbicida que ha sido introducido junto con la secuencia de nucleótidos deseada. La regeneración de plantas a partir de protoplastos cultivados es descrita por Evans et al., en Protoplasts Isolation and Culture, Handbook of Plant Cell Culture, págs. 124-176, MacMillilan Publishing Company, New York, 1983; y Binding, Regeneration of Plants, Plant Protoplasts, págs. 21-73, CRC Press, Boca Raton, 1985. La regeneración se puede obtener también a partir de callos, explantes, órganos, o partes de los mismos de plantas. Tales técnicas de regeneración son descritas generalmente por Klee (1987) en Ann. Rev. of Plant Phys. 38:467-486. Para obtener plantas completas a partir de tejidos transgénicos tales como embriones inmaduros, estos se pueden cultivar en condiciones ambientales controladas en una serie de medios que contienen nutrientes y hormonas, un procedimiento conocido como cultivo de tejidos. Una vez que se han generado plantas completas y producen semilla, comienza la evaluación de la progenie.

Después de incorporar establemente el casete de expresión en las plantas transgénicas, éste se puede introducir en otras plantas mediante cruzamiento sexual. Se puede utilizar cualquiera de las diversas técnicas de cría convencionales, dependiendo de las especies que se vayan 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 recombinante de la invención se pueden cruzar sexualmente con una segunda planta para obtener un producto final. De este modo, la semilla de la invención se puede obtener 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 (p. ej., la expresión de los polipéptidos de la invención para producir una planta en la que se altera el comportamiento de floración) se pueden potenciar cuando ambas plantas progenitoras expresan los polipéptidos (p. ej., una glucanasa) de la invención. Los efectos deseados se pueden pasar a las generaciones de plantas futuras mediante métodos de propagación convencionales.

Los ácidos nucleicos y polipéptidos de la invención son expresados o insertados en cualquier planta o semilla. Las plantas transgénicas de la invención pueden ser dicotiledóneas o monocotiledóneas. Los ejemplos de las plantas transgénicas de un solo cotiledón de la invención son céspedes, tales como gramíneas pratenses (poa de los prados, *Poa*), hierbas forrajeras tales como festuca, lolium, hierbas de zonas templadas, tales como *Agrostis*, y cereales, p. ej., trigo, avenas, centeno, cebada, arroz, sorgo, y maíz. Los ejemplos de plantas transgénicas de dos cotiledones de la invención son tabaco, legumbres, tal como altramuces, patata, remolacha azucarera, guisante, judías y soja, y plantas crucíferas (familia *Brassicaceae*), tales como coliflor, semilla de colza, y el organismo modelo íntimamente relacionado *Arabidopsis thaliana*. De este modo, las plantas y semillas transgénicas de la invención incluyen una amplia gama de plantas, incluyendo, pero no limitadas a, especies de los géneros *Anacardium*, *Arachis*, *Asparagus*, *Atropa*, *Avena*, *Brassica*, *Citrus*, *Citrullus*, *Capsicum*, *Carthamus*, *Cocos*, *Coffea*, *Cucumis*, *Cucurbita*, *Daucus*, *Elaeis*, *Fragaria*, *Glycine*, *Gossypium*, *Helianthus*, *Heterocallis*, *Hordeum*, *Hyoscyamus*, *Lactuca*, *Linum*, *Lolium*, *Lupinus*, *Lycopersicon*, *Malus*, *Manihot*, *Majorana*, *Medicago*, *Nicotiana*, *Olea*, *Oryza*, *Panicum*, *Pennisetum*, *Persea*, *Phaseolus*, *Pistachia*, *Pisum*, *Pyrus*, *Prunus*, *Raphanus*, *Ricinus*, *Secale*, *Senecio*, *Sinapis*, *Solanum*, *Sorghum*, *Theobromus*, *Trigonella*, *Triticum*, *Vicia*, *Vitis*, *Vigna*, y *Zea*.

En realizaciones alternativas, los ácidos nucleicos de la invención son expresados en plantas que contienen células de fibras, incluyendo, p. ej., algodón, árbol del algodón (Kapok, Ceiba pentandra), sauce del desierto, arbusto de creosota, salsola, balsa, ramio, kenaf, hemp, rosa de Jamaica, yute, sisal, abacá y linaza. 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 invención también proporciona plantas transgénicas que se van a utilizar para producir grandes cantidades de los polipéptidos (p. ej., una glucanasa o un anticuerpo) de la invención. Por ejemplo, véase Palmgren (1997) Trends Genet. 13:348; Chong (1997) Transgenic Res. 6:289-296 (que producen proteína láctea humana beta-caseína en plantas de patata transgénicas utilizando un promotor de manopina sintasa (mas1',2') inducible por auxina, bidireccional con métodos de transformación del disco de la hoja mediados por *Agrobacterium tumefaciens*).

Utilizando procedimientos conocidos, un experto en la técnica puede escrutar plantas de la invención detectando el aumento o disminución del ARNm transgen o la proteína en plantas transgénicas. Los medios para detectar y cuantificar los ARNm o proteínas son bien conocidos en la técnica.

Polipéptidos y péptidos

En un aspecto, la invención proporciona polipéptidos aislados, sintéticos o recombinantes que tienen una identidad de secuencia (p. ej., al menos aproximadamente 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, o más, o identidad de secuencia completa (100%)) con el SEQ ID NO: 38.

El polipéptido tiene una actividad glucanasa, p. ej., puede hidrolizar un enlace glicosídico en un polisacárido, p. ej., un glucano. En un aspecto, el péptido tiene una actividad glucanasa que comprende la catálisis de enlaces 1,4-beta-D-glicosídicos o enlaces β -1,3-glucosídicos. En un aspecto, la actividad endoglucanasa comprende una actividad endo-1,4-beta-endoglucanasa. En un aspecto, la actividad endoglucanasa comprende la hidrólisis de un glucan para producir un glucano u oligómero de glucano de peso molecular más pequeño. En un aspecto, el glucano comprende un beta-glucano, tal como un beta-glucano soluble en agua.

La Figura 5 es una tabla que resume las actividades relativas de varias enzimas en diversas condiciones, p. ej., pH y temperatura variables. En la Figura 5: NO = no determinado; * pH o temperatura óptima no determinados pero las actividades enzimáticas se midieron a pH y/o temperatura indicados; 1 estabilidad térmica, tiempo que la enzima conservó una actividad significativa (aprox. > 50%) o tiempo en el que la enzima ha perdido 50% de su actividad (t1/2) a la temperatura indicada; 2 RA = actividad relativa a pH 2,6, 4,0, 5,5, 7,0, 8,0, 9,0 o a 25°C, 37°C, 50°C, 65°C, 75°C, 85°C con respecto a la actividad al pH y temperatura óptima respectivamente; 3 RA a pH 3,75, 5, 5,3, 6,25, 7 para pH opt. y 40, 55, 70, 90°C para temp opt; 4 BBG = beta-glucano de cebada, CMC = carboximetilcelulosa. El agrupamiento en familias de las glucanasas se comentan más abajo.

La invención también proporciona polipéptidos quiméricos (y los ácidos nucleicos que los codifican) que comprenden al menos dos enzimas de la invención o subsecuencias de las mismas, p. ej., sitios activos, o dominios catalíticos (CD). Una proteína quimérica de la invención (p. ej., una proteína de fusión, u, otro heterodímero, p. ej., dos dominios enlazados por otros medios, p. ej., un conector, o, electrostáticamente) puede comprender un polipéptido (p. ej., sitio activo o péptido de dominio catalítico) de la invención y otro polipéptido (p. ej., sitio activo o péptido de dominio catalítico) de la invención u otro polipéptido. Por ejemplo, una proteína quimérica de la invención puede tener actividad mananasa y glucanasa, etc. En un aspecto la proteína quimérica de la invención comprende una fusión de dominios, p. ej., un dominio sencillo puede exhibir actividad glucanasa o cualquier combinación de actividades glucanasa/xilanasa/mananasa.

En primer lugar se clasificaron en familias las glicosidasa hidrolasas en 1991, véase, p. ej., Henrissat (1991) *Biochem. J.* 280:309-316. Desde entonces, la clasificación se ha actualizado continuamente, véanse, p. ej., Henrissat (1993) *Biochem. J.* 293:781-788; Henrissat (1996) *Biochem. J.* 316:695-696 Henrissat; (2000) *Plant Physiology* 124:1515-1519. Existen aproximadamente 87 familias de glicosidasa hidrolasas identificadas. Las glucanasas de la descripción están categorizadas como familias, p. ej., las familias 3, 5, 6, 8, 9, 12, y 16, como se muestran más abajo en la Tabla 2.

Tabla 2

SEQ ID NO:	Familia
1, 2	5
101, 102	16
103, 104	5
105, 106	5
107, 108	5
109, 110	5
11, 12	8
111, 112	5
113, 114	16
115, 116	5
117, 118	5
119, 120	16
121, 122	12
123, 124	8
125, 126	16
127, 128	5
129, 130	5
13, 14	8
131, 132	9
133, 134	8
135, 136	5+CBD
137, 138	5
139, 140	8
141, 142	9
143, 144	5
145, 146	5+CBD+SLH
147, 148	5+CBD+SLH
149, 150	5

ES 2 401 795 T3

SEQ ID NO:	Familia
15, 16	3
151, 152	16
153, 154	5
155, 156	9+CBD
157, 158	16
159, 160	16
161, 162	16
163, 164	9
165, 166	5
167, 168	5
169, 170	5
17, 18	9 Y 1
171, 172	16
173, 174	16
175, 176	5
177, 78	16
179, 180	5+CBD
181, 182	16
183, 184	8
185, 186	NA
187, 188	8 Y 1
189, 190	5
19, 20	5
191, 192	16
193, 194	5
195, 196	16
197,198	16
199,200	16
201,202	5
203, 204	3
205, 206	5
207, 208	5
209,210	16
21,22	12
211,212	8
213,214	16
215,216	6
217, 218	16
219,220	5

ES 2 401 795 T3

SEQ ID NO:	Familia
221, 222	5
223, 224	5
225, 226	8
227, 228	5+CBD
229, 230	5
23, 24	12
231,232	5
233,234	5
235, 236	5
237,238	6
239, 240	NA+CBD
241,242	5
245, 246	9
247, 248	8
249, 250	5
25, 26	8
251, 252	9+CBD
253, 254	5
255,256	5
257, 258	9
259, 260	5
261,262	5
263, 264	1+CBD
265, 266	NA
267,268	5
269, 270	9
27, 28	8
271, 272	9
273, 274	48+CBD
275, 276	8
277, 278	3
279, 280	5
281,282	9
283, 284	5
285, 286	5
287, 288	6
289, 290	8
29, 30	9
291,292	8

ES 2 401 795 T3

SEQ ID NO:	Familia
293, 294	6
295, 296	9+DOCR
297, 298	9
299, 300	5
3, 4	8
301, 302	5
303, 304	9+CBD
305,306	5
307, 308	5
309,310	10
31, 32	5
311,312	5
313, 314	5
315,316	5
317,318	5
319, 320	5
321, 322	5
323, 324	5
325, 326	5
327, 328	5
329, 330	9+CBD
33, 34	8
333, 334	5
335,336	6
337, 338	5
339, 340	6
341,342	5
343,344	5
345, 346	6
347, 348	5
349, 350	5
35,36	12
351, 352	5+CBD
353, 354	12
355, 356	5+CBD
357, 358	5
359,360	5
361,362	5
363, 364	CBD

ES 2 401 795 T3

SEQ ID NO:	Familia
365, 366	5
367,368	5
369,370	5
37, 38	5 y/o 6
371,372	5
373,374	5
375,376	9
377, 378	5
379, 380	3
381,382	9
383, 384	5
385, 386	8
387, 388	5
389,390	9
39,40	5
391,392	9
395, 396	8
397, 398	3
399, 400	5
401, 402	5 o 6+CDB
403,404	5
405, 406	5
407,408	5
409,410	5
41,42	12
411,412	5
413, 414	6
415, 416	9+CDB
417, 418	5
419, 420	5
421,422	5
423,424	9
425, 426	44
427,428	5
429, 430	3
43, 44	16
431,432	9
433, 434	6
435, 436	5

ES 2 401 795 T3

SEQ ID NO:	Familia
437, 438	5
439, 440	5
441,442	9
443, 444	NA
445, 446	NA
447, 448	26
449, 450	5+DOCR
45, 46	9
451,452	5
453, 454	5 y 26
455,456	1
457, 458	5
459, 460	9
461,462	5
463, 464	5
465, 468	5
467, 468	10
469, 470	5
47, 48	8
471,472	16
473, 474	5
475, 476	5
477, 478	11
481,482	5
483, 484	16
485,486	16
487, 488	12
489, 490	5
49, 50	5
491,492	11
493, 494	16
495, 496	5
497,498	16
499, 500	16
5, 6	5
501,502	1
503,504	5
505,506	5
507, 508	1

SEQ ID NO:	Familia
509, 510	5
51,52	5
511,512	26
513,514	26
515,516	5
517,518	3
53,54	5
55, 56	5
57, 58	9
59, 60	16
61,62	12
63, 64	16
65, 66	16
67, 68	9
69, 70	5
7, 8	9
71, 72	16
73, 74	5
75, 76	12
77, 78	5
79, 80	CBD
81, 82	16
83, 84	5
87, 88	16
89, 90	16
9, 10	5
91,92	3
93, 94	6
95, 96	16
97, 98	5
99, 100	16

Los polipéptidos de la invención incluyen glucanasas en una forma activa o inactiva. Por ejemplo, los polipéptidos de la invención incluyen proproteínas antes de la "maduración" o el procesamiento de las secuencias prepro, p. ej., por medio de una enzima de procesamiento de proproteínas, tal como una proproteína convertasa para generar una proteína mutante "activa". Los polipéptidos de la invención incluyen glucanasas inactivas por otras razones, p. ej., antes de la "activación" mediante un evento de procesamiento post-traducciona, p. ej., una acción de endo- o exo-peptidasa o proteinasa, un evento de fosforilación, una amidación, una glicosilación o una sulfuración, un evento de dimerización, y similares. Los polipéptidos de la invención incluyen todas las formas activas, incluyendo subsecuencias activas, p. ej., dominios catalíticos o sitios activos, de la glucanasa.

5

10

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.

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

10 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 residuos. Los polipéptidos de la invención puede ser también más cortos que los péptidos ilustrativos completos. Los polipéptidos (péptidos, fragmentos) pueden tener un tamaño que oscila entre alrededor de 5 y la longitud total de un polipéptido, p. ej., una enzima, tal como una glucanasa; siendo los tamaños ilustrativos 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 residuos, p. ej., residuos contiguos de una glucanasa ilustrativa de la invención.

20 Los péptidos de la invención (p. ej., una subsecuencia de un polipéptido ilustrativo de la invención) pueden ser útiles, p. ej., como sondas de marcaje, antígenos, tolerágenos, motivos, sitios activos de glucanasas (p. ej., "dominios catalíticos"), secuencias señal y/o dominios prepro.

25 Los polipéptidos y péptidos de la invención se pueden aislar a partir de fuentes naturales, ser sintéticos, o ser polipéptidos generados recombinantemente. Los péptidos y proteínas se pueden expresar recombinantemente *in vitro* o *in vivo*. Los péptidos y polipéptidos de la invención se pueden elaborar y aislar utilizando cualquier método conocido en la técnica. Los polipéptidos y péptidos de la invención se pueden sintetizar también, completamente o en parte, utilizando métodos químicos bien conocidos en la técnica. Véanse p. ej., 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 y Delivery Systems* (1995) Technomic Publishing Co., Lancaster, PA. Por ejemplo, la síntesis peptídica se puede realizar utilizando diversas técnicas en fase sólida (véanse p. ej., Roberge (1995) *Science* 269:202; Merrifield (1997) *Methods Enzymol.* 289:3-13) y la síntesis automática se puede lograr, p. ej., utilizando el Sintetizador de Péptidos ABI 431A (Perkin Elmer) de acuerdo con las instrucciones proporcionadas por el fabricante.

35 Los péptidos y polipéptidos de la invención pueden estar también glicosilados. La glicosilación se puede añadir post-traduccionalmente ya sea químicamente o mediante mecanismos biosintéticos celulares, donde el último incorpora el uso de motivos de glicosilación conocidos, que pueden ser nativos para la secuencia o se pueden añadir en forma de un péptido o añadir a la secuencia de ácido nucleico codificante. La glicosilación puede estar conectada a O o conectada a N.

40 Los péptidos y polipéptidos de la invención, como se ha definido anteriormente, incluyen todas las formas "miméticas" y "peptidomiméticas". Los términos "mimético" y "peptidomimético" hacen referencia a un compuesto químico sintético que tiene esencialmente las mismas características estructurales y/o funcionales de los polipéptidos de la invención. El mimético puede estar compuesto completamente de análogos de aminoácidos no naturales, sintéticos, o, es una molécula quimérica parcialmente de aminoácidos de péptidos naturales y parcialmente de análogos de aminoácidos no naturales. El mimético también puede incorporar cualquier cantidad de sustituciones conservativas de aminoácidos naturales con tal que tales sustituciones no alteren tampoco sustancialmente la estructura y/o actividad del mimético. Como con los polipéptidos de la invención que son variantes conservativas, la experimentación rutinaria determinará cuándo un mimético se encuentra dentro del alcance de la invención, es decir, que su estructura y/o función no está sustancialmente alterada. De este modo, en un aspecto, una composición mimética se encuentra dentro del alcance de la invención si tiene una actividad glucanasa.

55 Las composiciones miméticas de polipéptidos de la invención pueden contener cualquier combinación de componentes estructurales no naturales. En un aspecto alternativo, las composiciones miméticas de la invención incluyen uno o todos los siguientes tres grupos estructurales: a) grupos de unión a residuos distintos de las uniones mediante enlace amida natural ("enlace peptídico"); b) residuos no naturales en lugar de residuos de aminoácido de origen natural; o c) residuos que inducen mimetismo estructural secundario, es decir, que inducen o estabilizan una estructura secundaria, p. ej., un giro beta, un giro gamma, una lámina beta, una conformación en hélice alfa, y similares. Por ejemplo, un polipéptido de la invención puede ser caracterizado como mimético cuando todos o algunos de sus residuos se conectan mediante medios químicos distintos de los enlaces peptídicos naturales. Los residuos de peptidomiméticos individuales se pueden conectar mediante enlaces peptídicos, otros enlaces o medios de acoplamiento químicos, tales como, p. ej., glutaraldehído, ésteres de N-hidroxisuccinimida, maleimidas bifuncionales, N,N'-d ciclohexilcarbodiimida (DCC) o N,N'-diisopropilcarbodiimida (DIC). Los grupos conectores que

pueden ser una alternativa a las uniones mediante enlace amida tradicional ("enlace peptídico") incluyen, p. ej., cetometileno (p. ej., $-C(=O)-CH_2-$ para $-C(=O)-NH-$), aminometileno (CH_2-NH), etileno, olefina ($CH=CH$), éter (CH_2-O), tioéter (CH_2-S), tetrazol (CN_4-), tiazol, retroamida, tioamida, o éster (véase, p. ej., Spatola (1983) en *Chemistry and Biochemistry of Amino Acids, Peptides and Proteins*, Vol. 7, págs. 267-357, "Peptide Backbone Modifications", Marcell Dekker, NY).

Un polipéptido de la invención se puede caracterizar también como mimético por contener todos o algunos residuos no naturales en lugar de residuos de aminoácido de origen natural. Los residuos no naturales están bien descritos en la bibliografía científica y de patentes; unas pocas composiciones no naturales ilustrativas útiles como miméticos de residuos de aminoácidos naturales y sus líneas de actuación se describen más abajo. Se pueden generar miméticos de aminoácidos aromáticos reemplazando por, p. ej., D- o L- naftilalanina; D- o L- fenilglicina; D- o L-2 tienoilalanina; D- o L-1, -2, 3-, o 4-pirenoilalanina; D- o L-3-tienoilalanina; D- o L-(2-piridinil)-alanina; D- o L-(3-piridinil)-alanina; D- o L-(2-pirazinil)-alanina; D- o L-(4-isopropil)-fenilglicina; D-(trifluorometil)-fenilglicina; D-(trifluorometil)-fenilalanina; D-p-fluoro-fenilalanina; D- o L-p-bifenilfenilalanina; D- o L-p-metoxi-bifenilfenilalanina; D- o L-2-indol(alquil)alaninas; y, D- o L-alquilalaninas, donde el alquilo puede ser metilo, etilo, propilo, hexilo, butilo, pentilo, isopropilo, iso-butilo, sec-isotilo, iso-pentilo sustituidos o no sustituidos, o aminoácidos no ácidos. Los anillos aromáticos de un aminoácido no natural incluyen, p. ej., anillos aromáticos de tiazolilo, tiofenilo, pirazolilo, benzimidazolilo, naftilo, furanilo, pirrolilo, y piridilo.

Los miméticos de aminoácidos ácidos se pueden generar sustituyendo por, p. ej., aminoácidos no carboxilados mientras se mantenga una carga negativa; (fosfona)alanina; treonina sulfatada. Los grupos laterales carboxilados (p. ej., aspartilo o glutamilo) se pueden modificar también selectivamente mediante reacción con carbodiimidas ($R'-N-C-N-R'$) tales como, p. ej., 1-ciclohexil-3(2-morfolinil-(4-etil)carbodiimida o 1-etil-3(4-azonio-4,4-dimetilpentil)carbodiimida. Aspartilo o glutamilo se pueden convertir también en residuos de asparraginilo y glutaminilo mediante reacción con iones amonio. Los miméticos de aminoácidos alcalinos se pueden generar mediante sustitución con, p. ej., (además de lisina y arginina) los aminoácidos ornitina, citrulina, o ácido (guanidino)-acético, o ácido (guanidino)alquil-acético, donde alquilo se ha definido anteriormente. Los derivados de nitrilo (p. ej., que contienen el radical CN en lugar de COOH) se pueden sustituir por asparragina o glutamina. Los residuos de asparraginilo y glutaminilo pueden ser desaminados a los residuos de aspartilo o glutamilo correspondientes. Se pueden generar miméticos de residuos de arginina haciendo reaccionar arginilo, p. ej., con uno o más reactivos convencionales, incluyendo, p. ej., fenilglicoxal, 2,3-butanodiona, 1,2-ciclo-hexanodiona, o ninhidrina, preferiblemente en condiciones alcalinas. Los miméticos de residuos de tirosina se pueden generar haciendo reaccionar tirosilo, p. ej., con compuestos de diazonio aromáticos o tetranitrometano. Se pueden utilizar N-acetilimidazol y tetranitrometano para formar especies de O-acetiltirosilo y 3-nitroderivados, respectivamente. Los miméticos de residuos de cisteína se pueden generar haciendo reaccionar residuos de cisteinilo, p. ej., con alfa-haloacetatos tales como ácido 2-cloroacético o cloroacetamida y las aminas correspondientes; para producir derivados de carboximetilo o carboxamidometilo. Los residuos de miméticos de cisteína se pueden generar también haciendo reaccionar residuos de cisteinilo, p. ej., con bromo-trifluoroacetona, ácido alfa-bromo-beta-(5-imidazoil)propiónico; fosfato de cloroacetilo, N-alquilmaleimidias, disulfuro de 3-nitro-2-piridilo; disulfuro de metil-2-piridilo; p-cloromercuribenzoato; 2-cloromercuri-nitrofenol; o, cloro-7-nitrobenzo-oxa-1,3-diazol. Los miméticos de lisina se pueden generar (y los residuos amino terminales pueden ser alterados) haciendo reaccionar lisinilo, p. ej., con anhídrido de ácido succínico u otros anhídridos de ácidos carboxílicos. La lisina y otros miméticos de residuos que contienen alfa-amino se pueden generar también mediante reacción con imidoésteres, tales como picolinimidato de metilo, fosfato de piridoxal, piridoxal, cloroborohidruro, ácido trinitro-bencenosulfónico, O-metilisourea, 2,4-pentanodiona, y reacciones catalizadas por transamidasa con glioxilato. Los miméticos de metionina se pueden generar mediante reacción, p. ej., con metioninsulfóxido. Los miméticos de prolina incluyen, p. ej., ácido piperídico, ácido tiazolidinocarboxílico, 3- o 4-hidroxiprolina, deshidroprolina, 3- o 4-metilprolina, o 3,3-dimetilprolina. Los miméticos de residuos de histidina se pueden generar haciendo reaccionar histidilo, p. ej., con dietilprocarbonato o bromuro de para-bromofenacilo. Otros miméticos incluyen, p. ej., los generados mediante hidroxilación de prolina y lisina; fosforilación de los grupos hidroxilo de residuos de serilo o treonilo; metilación de los grupos alfa-amino de lisina, arginina y histidina; acetilación de la amina N-terminal; metilación de residuos amida de la cadena principal o sustitución con N-metil aminoácidos; o amidación de grupos carboxilo C-terminales.

Un residuo, p. ej., u aminoácido, de un polipéptido de la invención se puede reemplazar también por un aminoácido (o residuo de peptidomimético) de quiralidad opuesta. De este modo, cualquier aminoácido de origen natural en configuración L (que puede ser referida también como R o S, dependiendo de la estructura de la entidad química) se puede reemplazar por el aminoácido del mismo tipo estructural químico o un peptidomimético, pero de quiralidad opuesta, referido como D-aminoácido, pero también se puede referir como forma R o S.

Los métodos pueden modificar los polipéptidos de la invención mediante procedimientos naturales, tales como procesamiento post-traduccional (p. ej., fosforilación, acilación, etc.), o mediante técnicas de modificación químicas, dando como resultado los polipéptidos modificados. Las modificaciones se pueden producir en cualquier parte en el polipéptido, incluyendo el esqueleto peptídico, las cadenas laterales de aminoácidos y los extremos amino o carboxilo. Se apreciará que el mismo tipo de modificación puede estar presente en el mismo o en grados variables

en diversos sitios en un polipéptido dado. Asimismo un polipéptido dado puede tener muchos tipos de modificaciones. Las modificaciones incluyen acetilación, acilación, ribosilación del ADP, amidación, unión covalente de flavina, unión covalente de un radical hemo, unión covalente de un nucleótido o derivado de nucleótido, unión covalente de un lípido o derivado de lípido, unión covalente de un fosfatidilinositol, ciclación de entrecruzamiento, formación de enlaces disulfuro, desmetilación, formación de entrecruzamientos covalentes, formación de cisteína, formación de piroglutamato, formilación, gamma-carboxilación, glicosilación, formación de ancla de GPI, hidroxilación, yodación, metilación, miristilación, oxidación, pegilación, procesamiento proteolítico, fosforilación, prenilación, racemización, selenoilación, sulfación, y adición de aminoácidos mediada por ARN de transferencia a proteínas tal como arginilación. Véase, p. ej., Creighton, T.E., *Proteins - Structure and Molecular Properties* 2a Ed., W.H. Freeman y Company, New York (1993); *Posttranslational Covalent Modification of Proteins*, B.C. Johnson, Ed., Academic Press, New York, págs. 1-12 (1983).

También se pueden utilizar 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 la década 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 se han empleado recientemente en diseños de péptidos en laboratorio asequibles comercialmente y kits de síntesis (Cambridge Research Biochemicals). Tales kits de laboratorio asequibles 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 se conectan a una sola placa. Cuando se utiliza tal sistema, una placa de varillas o clavijas se invierte y se inserta en una segunda placa de pocillos o reservorios correspondientes, que contienen soluciones 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 soluciones apropiadas, los aminoácidos se arman para dar los péptidos deseados. Además, están disponibles varios sistemas de síntesis peptídica Fmoc asequible. 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, mediante síntesis directa o mediante síntesis de una serie de fragmentos que se pueden acoplar utilizando otras técnicas conocidas.

La invención incluye glucanasas de la invención con y sin señal. El polipéptido que comprende una secuencia señal de la invención puede ser una glucanasa de la invención u otra glucanasa u otra enzima u otro polipéptido.

La descripción incluye glucanasas inmovilizadas, anticuerpos anti-glucanasa y fragmentos de los mismos. Los métodos para inhibir la actividad glucanasa pueden utilizar mutantes negativos dominantes o anticuerpos anti-glucanasa de la invención. La invención incluye heterocomplejos, p. ej., proteínas de fusión, heterodímeros, etc., que comprenden las glucanasas de la invención.

Los polipéptidos de la invención pueden tener actividad glucanasa en diversas condiciones, p. ej., extremas en pH y/o temperatura, agentes oxidantes, y similares. La invención proporciona métodos que conducen a preparaciones de glucanasa alternativas con diferentes eficacias y estabildades catalíticas, p. ej., frente a temperatura, agentes oxidantes y condiciones de lavado variables. En un aspecto, las variantes de glucanasa se pueden producir utilizando técnicas de mutagénesis dirigida al sitio y/o mutagénesis al azar. En un aspecto, se puede utilizar la evolución dirigida para producir una gran variedad de variantes de glucanasa con especificidades y estabilidad alternativas.

Las proteínas de la invención también son útiles como reactivos de investigación para identificar moduladores de glucanasa, p. ej., activadores o inhibidores de la actividad glucanasa. En resumen, se añaden muestras de ensayo (compuestos, caldos, extractos, y similares) a análisis de glucanasa para determinar su capacidad para inhibir la escisión del sustrato. Los inhibidores identificados de este modo se pueden utilizar en industria e investigación para reducir o evitar la proteólisis no deseada. Los inhibidores de glucanasas, se pueden combinar para incrementar el espectro de actividad.

Las enzimas de la invención también son útiles como reactivos de investigación para digerir proteínas o para secuenciar proteínas. Por ejemplo, una glucanasa se puede utilizar para romper polipéptidos en fragmentos más pequeños para su secuenciación, p. ej. utilizando un secuenciador automático.

La invención también proporciona métodos para descubrir nuevas glucanasas utilizando los ácidos nucleicos, polipéptidos y anticuerpos de la invención. En un aspecto, se escrutan bibliotecas de fagémidos para el descubrimiento de glucanasas basado en su expresión. En otro aspecto, se escrutan bibliotecas de fagos lambda para el descubrimiento de glucanasas basado en su expresión. El escrutinio de las bibliotecas de fagos o fagémidos puede permitir la detección de clones tóxicos; el acceso mejorado al sustrato; la reducción de la necesidad de construir un anfitrión, evitando el potencial de cualquier tendencia resultante de la escisión masiva de la biblioteca; y, un crecimiento más rápido a bajas densidades de clones. El escrutinio de bibliotecas de fagos o fagémidos puede ser en fase líquida o en fase sólida. En un aspecto, la invención proporciona el escrutinio en fase líquida. Esto

proporciona una mayor flexibilidad en condiciones de análisis; flexibilidad de sustrato adicional; mayor sensibilidad para clones débiles; y facilidad de automatización a lo largo del escrutinio en fase sólida.

Los métodos de escrutinio utilizan las proteínas y ácidos nucleicos de la invención y la automatización robótica para posibilitar la ejecución de muchos miles de reacciones biocatalíticas y análisis de escrutinio en un corto período de tiempo, p. ej., por día, así como la garantía de un alto nivel de exactitud y reproducibilidad (véase la discusión de matrices, más abajo). Como resultado, se puede producir una biblioteca de compuestos derivados en cuestión de semanas. Para enseñanzas adicionales sobre modificación de moléculas, incluyendo moléculas pequeñas, véase PCT/US94/09174.

Un polipéptido aislado o purificado comprende la secuencia de uno de 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. Como se ha comentado anteriormente, tales polímeros se pueden obtener insertando un ácido nucleico que codifica el polipéptido en un vector de manera que la secuencia codificante esté conectada operablemente a una secuencia capaz de dirigir la expresión del polipéptido codificado en una célula anfitriona adecuada. Por ejemplo, el vector de expresión puede comprender un promotor, un sitio de unión al ribosoma para el inicio de la traducción y un terminador de la transcripción. El vector puede incluir también secuencias apropiadas para amplificar la expresión.

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

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 mediante aislamiento de los ácidos nucleicos que los codifican utilizando 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 digestión con glucano hidrolasa, electroforesis en gel y/o microsecuenciación. La secuencia de polipéptido o fragmento homólogos 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 utilizando cualquiera de los programas descritos anteriormente.

Un análisis identifica fragmentos o variantes de la invención, que conservan la función enzimática de los polipéptidos de la invención. Por ejemplo los fragmentos o variantes de dichos polipéptidos, se pueden utilizar para catalizar reacciones bioquímicas, que indican que el fragmento o variante conserva la actividad enzimática de un polipéptido de la invención.

El análisis para determinar si los fragmentos o variantes conservan la actividad enzimática de los polipéptidos de la invención incluye las etapas de poner en contacto el fragmento de polipéptido o variante con una molécula sustrato en condiciones que permitan que el fragmento de polipéptido o variante funcione y detecte cualquier disminución del nivel de sustrato o un incremento en el nivel de producto de reacción específico de la reacción entre el polipéptido y el sustrato.

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 de los mismos se pueden utilizar en una variedad de aplicaciones. Por ejemplo, los polipéptidos o fragmentos de los mismos se pueden utilizar para catalizar reacciones bioquímicas. De acuerdo con un aspecto de la invención, se proporciona un procedimiento para la utilización de los polipéptidos de la invención o polinucleótidos que codifican tales polipéptidos para hidrolizar los enlaces glicosídicos. En tales procedimientos, una sustancia que contiene un enlace glicosídico (p. ej., un almidón) se pone en contacto con uno de los polipéptidos de las secuencias de aminoácidos del Grupo B, o secuencias esencialmente idénticas a éstas en condiciones que faciliten la hidrólisis del enlace glicosídico.

La presente invención explota las propiedades catalíticas únicas de las enzimas. Mientras que el uso de biocatalizadores (es decir, enzimas purificadas o brutas, células no vivas o vivas) en transformaciones químicas

requiere normalmente la identificación de un biocatalizador concreto que reacciona con un compuesto de partida específico, la presente invención utiliza biocatalizadores seleccionados y condiciones de reacción que son específicas para grupos funcionales que están presentes en muchos compuestos de partida, tales como moléculas pequeñas. Cada biocatalizador es específico para un grupo funcional, o varios grupos funcionales relacionados y puede reaccionar con muchos compuestos de partida que contienen este grupo funcional.

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

Las enzimas reaccionan en sitios específicos de un compuesto de partida sin afectar al resto de la molécula, un procedimiento que es muy difícil de lograr utilizando métodos químicos tradicionales. Este alto grado de especificidad biocatalítica proporciona los medios para identificar un solo compuesto activo dentro de la biblioteca. La biblioteca está caracterizada por la serie de reacciones biocatalíticas para producirla, una denominada "historia biosintética". El escrutinio de la biblioteca para determinar las actividades biológicas y el trazado de la historia biosintética identifica la secuencia de reacción específica que produce el compuesto activo. La secuencia de reacción se repite y se determina la estructura del compuesto sintetizado. Este modo de identificación, a diferencia de otros enfoques de síntesis y escrutinio, no requiere tecnologías de inmovilización y los compuestos se pueden sintetizar y someter a ensayo libres en solución utilizando virtualmente cualquier tipo de análisis de escrutinio. Resulta importante observar, que el alto grado de especificidad de las reacciones enzimáticas sobre grupos funcionales permite el "seguimiento" de reacciones enzimáticas específicas que constituyen la biblioteca producida biocatalíticamente.

Muchas de las etapas procedimentales se realizan utilizando automatización robótica que posibilita la ejecución de muchos miles de reacciones biocatalíticas y análisis de escrutinio por día así como la garantía de un alto nivel de exactitud y reproducibilidad. Como resultado, se puede producir una biblioteca de compuestos derivados en cuestión de semanas que podría tardar años en producirse utilizando los métodos químicos actuales.

Un método para modificar moléculas pequeñas comprende poner en contacto un polipéptido codificado por un polinucleótido descrito en la presente memoria o fragmentos enzimáticamente activos del mismo con una molécula pequeña para producir una molécula pequeña modificada. Se somete a ensayo una biblioteca de moléculas pequeñas modificadas para determinar si una molécula pequeña modificada está presente en la biblioteca que exhibe una actividad deseada. Se identifica una reacción biocatalítica específica que produce la molécula pequeña modificada de actividad deseada eliminando sistemáticamente cada una de las reacciones biocatalíticas utilizadas para producir una porción de la biblioteca y después sometiendo a ensayo las moléculas pequeñas producidas en la porción de la biblioteca para determinar la presencia o ausencia de la molécula pequeña modificada con la actividad deseada.

Las reacciones biocatalíticas específicas que producen la molécula pequeña modificada de actividad deseada se repiten opcionalmente. Las reacciones biocatalíticas se llevan a cabo con un grupo de biocatalizadores que reaccionan con radicales estructurales distintos encontrados en la estructura de una molécula pequeña, cada biocatalizador es específico para un radical estructural o un grupo de radicales estructurales relacionados; y cada biocatalizador reacciona con muchas moléculas pequeñas diferentes que contienen el radical estructural distinto.

Secuencias señal, dominios prepro y catalíticos

Las secuencias señal de glucanasa (p. ej., péptidos señal (SP)), dominios prepro y dominios catalíticos (CD) (p. ej., sitios activos) pueden ser péptidos aislados o recombinantes o puede ser parte de una proteína de fusión, p. ej., en forma de un dominio heterólogo en una proteína quimérica. Los ácidos nucleicos codifican estos dominios catalíticos (CD), dominios prepro y secuencias señal (SP, p. ej., un péptido que tiene una secuencia que comprende/consiste en residuos amino terminales de un polipéptido de la invención). Una secuencia señal comprende un péptido que comprende/consiste en una secuencia como la mostrada en los residuos 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 de un polipéptido de la invención.

En un aspecto, la invención también proporciona polipéptidos quiméricos (y los ácidos nucleicos que los codifican) que comprenden al menos dos enzimas de la invención o subsecuencias de las mismas, p. ej., dominios catalíticos (CD) o sitios activos. Por ejemplo, una proteína quimérica de la invención puede tener actividad mananasa y glucanasa, etc. En un aspecto, la proteína quimérica de la invención comprende una fusión de dominios, p. ej., un dominio sencillo puede exhibir actividad glucanasa o cualquier combinación de actividades glucanasa/xilanasa/mananasa (p.ej., en forma de proteína quimérica recombinante).

5 Las secuencias señal aisladas, sintéticas o recombinantes pueden comprender/consistir en una secuencia señal, p. ej., como se muestra en la Tabla 3, de más abajo, y los polipéptidos pueden comprender estas secuencias señal. El polipéptido puede ser otra glucanasa de la invención, otra glucanasa o mananasa, o xilanasa, o glicosidasa o hidrolasa, u otro tipo de enzima o polipéptido. Por ejemplo, leyendo la Tabla 3, una secuencia señal aislada, sintética o recombinante se muestra mediante los residuos 1 a 21 del SEQ ID NO: 102, que está codificado por, p. ej., una subsecuencia del SEQ ID NO: 101; o una secuencia señal aislada, sintética o recombinante se muestra mediante los residuos 1 a 30 del SEQ ID NO: 104, que está codificado por una subsecuencia del SEQ ID NO: 103, etc.

Tabla 3

<u>SEQ ID NO:</u>	<u>Señal (AA)</u>	<u>Secuencia Señal Pronosticada</u>
101, 102	1-21	
103, 104	1-30	
105, 106	1-33	
107, 108	1-18	
109, 110	1-25	
11, 12	1-25	
111, 112	1-39	
113, 114	1-24	
115, 116	1-21	
117, 118	1-29	
121, 122	1-24	
123, 124	1-30	
125, 126	1-19	
127, 128	1-22	
129, 130	1-40	
13, 14	1-22	
133, 134	1-34	
135, 136	1-53	
137, 138	1-37	
139, 140	1-22	
141, 142	1-32	
143, 144	1-29	
145, 146	1-29	
147, 148	1-29	
149, 150	1-30	
15, 16	1-44	
151, 152	1-22	
153, 154	1-24	
155, 156	1-21	
157, 158	1-38	
159, 160	1-34	
161, 162	1-19	
163, 164	1-19	
165, 166	1-51	

ES 2 401 795 T3

SEQ ID NO:	Señal (AA)	Secuencia Señal Pronosticada
169, 170	1-22	
171, 172	1-22	
173, 174	1-19	
175, 176	1-19	
177, 178	1-26	
179,180	1-71	
181,182	1-35	
183,184	1-22	
185, 186	1-17	
187, 188	1-22	
19,20	1-68	
191, 192	1-20	
193, 194	1-29	
199, 200	1-22	MKTKLISTLVAGLIVISPATYA
201, 202	1-32	
203,204	1-27	
205,206	1-34	
207, 208	1-28	
209, 210	1-22	
21,22	1-34	
211,212	1-22	
213, 214	1-22	
215,216	1-42	
217,218	1-23	
219,220	1-27	
221, 222	1-29	
223,224	1-31	
225, 226	1-29	
227, 228	1-22	MTSKHFFKITLMSILLFTTTLA
229, 230	1-25	MKRRNWNYYLLIILLVISAFTLISAQ
23,24	1-22	
235, 236	1-19	MKSVLALALIVSINLVLLA
237, 238	1-29	MTRRSIVRSSSNKWLVLAGAALLACTALG
241,242	1-39	MSSFKASAINPRMAGALTRSLYAAGFSLAVSTLSTQAYA
243, 244	1-26	MKKLLKLSMLSTSVLALGIMASSGAIA
247, 248	1-21	MNVLRSGIVTMLLLAAFSVQA
25, 26	1-21	
251,252	1-23	MLKKLALAAGIAAATLAASGSHG

ES 2 401 795 T3

SEQ ID NO:	Señal (AA)	Secuencia Señal Pronosticada
255, 256	1-28	MKRTGWTLKLLLAALLLLPATLGLHNGA
257, 258	1-30	MYRLFFRSLKRSGILLPVLLYFFILPSATA
269, 270	1-22	MKFTLTPLLCGFALLLGCAVQA
27, 28	1-19	
271,272	1-18	MVSMLLLTVGAVSVSAVS
273,274	1-37	MPRLRARTRPRRQLTALAAALSLPLGLTAVGATTAQA
275, 276	1-27	MQNLFKRVFFHLLLALLAGCAGPSPV
277,278	1-28	MSGRSAGRGPWARLWALAAVGALVAGA
279,280	1-38	MQNKIINTKIKLRKFMSQLIKITYIFIIIFCMQRTYA
281,282	1-18	MKKLILTLFSLWAISAYA
285, 286	1-27	MRKSIRSFSILLAITFIIALLSFPAMG
287,288	1-30	MNPRSLRRRTTAALAALAACAALLATQAQA
289, 290	1-29	MFPRLSPSRFRQVTLTLLTLGLVSLTGCA
29, 30	1-19	
291,292	1-16	MKFFTLLFFLSFVFS
293,294	1-27	MRRRIRALVAALSALPLALVVAPSAHA
3, 4	1-25	
301,302	1-17	MAIGISATMLLAMPQQA
303, 304	1-30	MSCRTLMSRRVWGLLLLWGGLFLRTGSVTG
305, 306	1-26	MNPKYIYRITFLLISILSMTALQSFS
307, 308	1-52	MVWTPARSTLAGSSEIPLMTMNIFPNRKDSRMSLWIKGILCMM AGTVMVHG
31, 32	1-39	
311,312	1-24	MKRREFMLGGAGVAALASTLGVSA
313,314	1-20	MLIIGLLVLLGFSSCGRQA
315,316	1-32	MDKTITAKDSGKITALILILLVLPYAGYVVA
319, 320	1-29	MREIILKSGALLMVVILIVSILQILTVFA
323,324	1-32	MFQSLKMRTLSFLLLALLASFLALPTDVAHA
325,326	1-29	MKKIILKSGILLWILIVSILQILPVFA
327,328	1-46	MLVYRVSQKHLASLTVLVSLLLILAGCSSSSDSIAPVSSSSVSSA
33, 34	1-35	
331,332	1-28	MNNPTNGARRGRHRRRWSATALLGVPA
333, 334	1-28	MNRTRVLSAATLLALVATLASVPVTAQA
337, 338	1-29	MRNHLNVPFYFIFFLIASIFTVCSSSTA
339, 340	1-23	MNNPRILTYLLIGIWAVLIVFA
343,344	1-31	MRKIVKQINYLTSPVLGLLVLSLFFQVPTQA

ES 2 401 795 T3

SEQ ID NO:	Señal (AA)	Secuencia Señal Pronosticada
345, 346	1-33	MKRTRYGVRSPRSAPRFGVLFGAAAAGVLMTGA
349, 350	1-38	MNSSPVSVKKPCPVDRPNPLWAAGFSLALATLSTQTQA
351,352	1-28	MKKVSNARVLSFLLILVLIFGNLASVFA
357,358	1-32	MEKQICSNVFSTMLIIGLLVLLGFSSCGRQA
359,360	1-22	MRLITILATAVAILSTTSCS
361,362	1-20	MSRGILILVMSLVLSGAALA
363, 364	1-30	MRRTRSLLAGLALTAGLLTGAGAGAPPATA
365,366	1-50	MLGAPSPHFPMRRGMTKSQRRTWLTAVGSAIAGIAGLLLPVFAT AGAAQA
367, 368	1-42	MPHPKLLTNGGSYVSSKQKTVAIFVLFVVLGAVAGSIPASYA
37, 38	1-33	
373, 374	1-23	MNKILKLFSSLLLFAGICPALQA
377,378	1-30	MKMLTTLKKPLLKKTALALLTSAMVAPAF
381,382	1-26	MRAIRLSLSIAAGAVLLLAGCTTKPA
383, 384	1-28	MTMHRKLRHSIAAGALSAIFFVGLQAGA
387,388	1-25	MSIIKKVPLIFLCLLMFATSLFIFK
39, 40	1-24	
393, 394	1-29	MSKFLSLSNFFSLLVVCVLLGACSGGSSS
401,402	1-30	MGTSLMIKSTLTGMITAVAAVFTTSAafa
403, 404	1-32	MGKISKYFAMFLAFLMVFSSLFVNFQPRNVQA
407, 408	1-28	MRKNILMLAVAMIAAMCVTTSCGNKAQK
409, 410	1-24	MTRNWLKGKILAALLLAGCAIPAPA
411,412	1-24	MPYVRLALVCAWTVLACTGAPIA
413, 414	1-28	MSRHLISLGLLVWALGAMLWISSRDVA
419, 420	1-21	MLRKLILFCVLLSMSWVALA
421,422	1-26	MKTKSIYSIALSIALFFFTTAQTFS
423, 424	1-25	MWSQDVRKVWLVGFLLLVAGMPALA
425, 426	1-20	MRIRLATLALCAALSPVTFA
427, 428	1-43	METPMTSARSARPRRLRRYGIAGTALGALLGLATLPPTATA
43, 44	1-48	
431,432	1-22	MKFTLMPLLCGFALLGCAVQA
441,442	1-26	MRAIRLSLSIAAGAVLLLAGCTTKPA
445, 446	1-29	MRRFRWFLGLFVFFGIVIASQYGQTAAA
449, 450	1-25	MKKIVSLVCVLVMLVSILGSFSVVA
45, 46	1-16	
457, 458	1-39	MKSKTSTAAPSAGPLRNYKKTACIAVASTALLAGSASA

SEQ ID NO:	Señal (AA)	Secuencia Señal Pronosticada
459, 460	1-18	MKKLILTLFSLW AISAYA
461,462	1-28	MFKHLLHVLKIGFLPLLATLLLAGHAHG
465, 466	1-21	MLRKLIVSVFGFVMLTSA AAA
469, 470	1-25	MKYKAIFIYLVILFYSINIYANA
47, 48	1-23	
471, 472	1-22	MKTKLISTLVAGLIVISPATYA
473, 474	1-28	MKRKRVIHSLIVFFLMIGSFTSCGSVA
475, 476	1-25	MNLLAQYFSGLFLIFLISIFFVSSA
477,478	1-27	MKSIRSRSLATAVLALAGLVAAAGAQA
481,482	1-21	MKLLKLLIFLLITVIFSDVSA
483, 484	1-26	MYKRLLSSVLIIMLLLSAWSPISVQA
489, 490	1-23	MKYIFSYYIIMMILIGFIPVYGF
49, 50	1-42	
491,492	1-26	MSMFLSLKRVAALVCVAGFGISAANA
493, 494	1-28	MPYLKRVLLLLVTGLFMSLFAVTSTASA
495, 496	1-29	MSSKQKTVAIFVLFVALAGVAGSIPASYA
499, 500	1-17	MKKLVLVLLFPVFILA
505, 506	1-24	MKSKVKMFFAAAIVWSACSSTGYA
509, 510	1-20	MPKLLASFIALFFAANAAA
51, 52	1-38	
511,512	1-21	MKKLHILLALTAMTAFASCS

Dichas secuencias señal (SP) y/o secuencias prepro pueden ser péptidos aislados, o, secuencias conectadas a otro polipéptido de glucanasa, mananasa o xilanasa o no glucanasa, mananasa, o xilanasa, p. ej., en forma de una proteína de fusión (quimérica). Los polipéptidos pueden comprender secuencias señal de glucanasa, mananasa, o xilanasa. Los polipéptidos que comprenden secuencias señal SP de glucanasa, mananasa, o xilanasa y/o prepro pueden comprender secuencias heterólogas a una glucanasa de la invención (p. ej., una proteína de fusión que comprende SP y/o prepro de la invención y secuencias de otra glucanasa o una proteína no glucanasa). En un aspecto, la invención proporciona una glucanasa de la invención con SP y/o secuencias prepro heterólogas, p. ej., secuencias con una secuencia señal levadura. Una glucanasa de la invención puede comprender una SP y/o prepro heterólogas en un vector, p. ej., un vector de la serie pPIC (Invitrogen, Carlsbad, CA).

En un aspecto, las secuencias SP y/o prepro de la invención se identifican después de la identificación de los polipéptidos de glucanasa novedosos. Las rutas por medio de las cuales las proteínas se clasifican y transportan a su localización celular apropiada son referidas con frecuencia como rutas de direccionamiento de proteínas. Uno de los elementos más importantes en todos estos sistemas de direccionamiento es una secuencia de aminoácidos corta en el extremo amino de un polipéptido recién sintetizado denominado secuencia señal. Esta secuencia señal dirige una proteína a su localización apropiada en la célula y es eliminada durante el transporte o cuando la proteína alcanza su destino final. La mayoría de las proteínas lisosomales, de membrana, o secretadas tienen una secuencia señal amino terminal que las marca para la traslocación al lumen del retículo endoplásmico. Se han determinado más de 100 secuencias señal para proteínas en este grupo. Las secuencias señal pueden variar en longitud de 13 a 36 residuos de aminoácidos. Los expertos en la técnica conocen diversos métodos de reconocimiento de secuencias señal. Por ejemplo, en un aspecto, se identifican péptidos señal de glucanasa novedosos por medio de un método referido como SignalP. SignalP utiliza una red neural combinada que reconoce ambos péptidos señal y sus sitios de escisión. (Nielsen, et al., "Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites" Protein Engineering, vol. 10, núm. 1, págs. 1-6 (1997).

Se debe entender que en algunos aspectos las glucanasas de la invención pueden no tener secuencias o "dominios" SP y/o prepro. En un aspecto, la invención proporciona las glucanasas de la invención que carecen de todo o parte de un dominio SP y/o uno prepro. Una secuencia de ácido nucleico puede codificar una secuencia señal (SP) y/o

prepro de una glucanasa, mananasa, o xilanasa conectada operablemente a una secuencia de ácido nucleico de una glucanasa diferente u, opcionalmente, se puede desear una secuencia señal (SP) y/o dominio prepro de una proteína no glucanasa, mananasa, o xilanasa.

- 5 La descripción también proporciona polipéptidos aislados o recombinantes que comprenden las secuencias señal (SP), dominios prepro y/o dominios catalíticos (CD) de la invención y secuencias heterólogas. Las secuencias heterólogas son secuencias asociadas no naturalmente (p. ej., a una glucanasa, mananasa, o xilanasa) con una SP, dominio prepro y/o CD. La secuencia a la que están asociados no naturalmente la SP, el dominio prepro y/o el CD puede estar en el extremo amino terminal, el extremo carboxi terminal de la SP, el dominio prepro y/o el CD, y/o en ambos extremos de la SP y/o el CD. Un polipéptido aislado o recombinante puede comprender (o consistir en) un polipéptido que comprende una secuencia señal (SP), un dominio prepro y/o un dominio catalítico (CD) de la invención con la condición de que no esté asociado con cualquier secuencia a la que está asociado naturalmente (p. ej., una secuencia de glucanasa, mananasa o xilanasa). De un modo similar los ácidos nucleicos aislados o recombinantes codifican estos polipéptidos. De este modo, el ácido nucleico aislado o recombinante de la invención comprende una secuencia codificante para una secuencia señal (SP), dominio prepro y/o dominio catalítico (CD) y una secuencia heteróloga (es decir, una secuencia asociada no naturalmente con la secuencia señal (SP), el dominio prepro y/o el dominio catalítico (CD) de la invención). La secuencia heteróloga puede estar en el extremo 3' terminal, el extremo 5'-terminal, y/o en ambos extremos de la secuencia codificante de SP, dominio prepro y/o CD.

20 **Glucanasas, mananasas o xilanasas híbridas (quiméricas) y bibliotecas de péptidos**

En un aspecto, la invención proporciona glucanasas híbridas y proteínas de fusión, incluyendo bibliotecas de péptidos, que comprenden las secuencias de la invención. Las bibliotecas de péptidos se pueden utilizar para aislar moduladores peptídicos (p. ej., activadores o inhibidores) de dianas, tales como sustratos de glucanasa, receptores, 25 enzimas. Las bibliotecas de péptidos de la invención se pueden utilizar para identificar compañeros de unión formales de las dianas, tales como ligandos, p. ej., citoquinas, hormonas y similares. En un aspecto, la invención proporciona proteínas quiméricas que comprenden una secuencia señal (SP), un dominio prepro y/o un dominio catalítico (CD) de la descripción o una de sus combinaciones y una secuencia heteróloga (véase más arriba).

30 En un aspecto, las proteínas de fusión de la invención (p. ej., el radical peptídico) son estabilizadas conformacionalmente (con respecto a los péptidos lineales) para permitir una mayor afinidad de unión a las dianas. La invención proporciona fusiones de una glucanasa, mananasa, o xilanasa de la invención y otros péptidos, incluyendo péptidos conocidos y al azar. Se pueden fusionar de tal manera que la estructura de las glucanasas no resulte perturbada significativamente y el péptido sea estabilizado conformacionalmente metabólicamente o 35 estructuralmente. Esto permite la creación de una biblioteca de péptidos cuya presencia y cantidad en el interior de las células es verificada fácilmente.

Las variantes de la secuencia de aminoácidos de la invención se pueden caracterizar por medio de una naturaleza predeterminada de la variación, una característica que las aparta de una forma de origen natural, p. ej., una variación alélica o interespecie de una secuencia de glucanasa. En un aspecto, las variantes de la invención exhiben 40 la misma actividad biológica cualitativa que el análogo de origen natural. Alternativamente, las variantes se pueden seleccionar por tener características modificadas. En un aspecto, mientras que el sitio o región para introducir una variación de la secuencia de aminoácidos está predeterminado, la mutación per se no necesita estar predeterminada. Por ejemplo, con el fin de optimizar en funcionamiento de una mutación en un sitio dado, se puede llevar a cabo la mutagénesis al azar en el codón o región diana y escrutar las variantes de glucanasa expresadas 45 para determinar la combinación óptima de actividad deseada. Las técnicas para elaborar las mutaciones de sustitución en sitios predeterminados en el ADN que tiene una secuencia conocida son bien conocidas, como se comenta en la presente memoria por ejemplo, mutagénesis con cebador M13 y mutagénesis por PCR. Se puede realizar el escrutinio de los mutantes utilizando, p. ej., análisis de hidrólisis de glucano. En aspectos alternativos, las sustituciones de aminoácido pueden ser residuos individuales; las inserciones pueden estar en el orden de 50 alrededor de 1 a 20 aminoácidos, aunque se pueden realizar inserciones considerablemente mayores. Las deleciones pueden oscilar de alrededor de 1 a alrededor de 20, 30, 40, 50, 60, 70 residuos o más. Para obtener un derivado final con las propiedades óptimas, se pueden utilizar sustituciones, deleciones, inserciones o cualquier combinación de las mismas. Generalmente, estos cambios se realizan en unos pocos aminoácidos para minimizar la alteración de la molécula. Sin embargo, se pueden tolerar cambios mayores en ciertas circunstancias.

55 La invención proporciona una glucanasa en la que la estructura del esqueleto polipeptídico, la estructura secundaria y terciaria, p. ej., una estructura en hélice alfa o lámina beta, se ha modificado. En un aspecto, se han modificado la carga o el carácter hidrófobo. En un aspecto, se ha modificado el volumen de una cadena lateral. Se realizan cambios sustanciales en la función o la identidad inmunológica seleccionando sustituciones que son menos 60 conservativas. Por ejemplo, se pueden realizar sustituciones que afectan más significativamente a: la estructura de la cadena polipeptídica en la zona de alteración, por ejemplo una estructura en hélice alfa o lámina beta; una carga o sitio hidrófobo de la molécula, que puede ser un sitio activo; o una cadena lateral. La invención proporciona sustituciones en polipéptidos de la invención donde (a) un residuo hidrófilo, p. ej. serilo o treonilo, es sustituido por (o mediante) un residuo hidrófobo, p. ej. leucilo, isoleucilo, fenilalanilo, valilo o alanilo; (b) una cisteína o prolina es

5 sustituida por (o mediante) cualquier otro residuo; (c) un residuo que tiene una cadena lateral electropositiva, p. ej. lisilo, arginilo, o histidilo, es sustituido por (o mediante) un residuo electronegativo, p. ej. glutamilo o aspartilo; o (d) un residuo que tiene una cadena lateral voluminosa, p. ej. fenilalanina, es sustituido por (o mediante) uno que no tenga una cadena lateral, p. ej. glicina. Las variantes pueden mostrar la misma actividad biológica cualitativa (es decir actividad endoglucanasa) aunque se pueden seleccionar variantes para modificar las características de la glucanasa según se necesite.

10 En un aspecto, las glucanasas de la invención comprenden epítomos o etiquetas de purificación, secuencias señal u otras secuencias de fusión, etc. En un aspecto, la glucanasa de la invención se puede fusionar a péptido aleatorio para formar un polipéptido de fusión. Mediante "fusionado" o "conectado operablemente" en la presente memoria se quiere significar que el péptido aleatorio y la glucanasa se conectan entre sí, de tal manera que se minimiza la alteración de la estabilidad de la estructura de la glucanasa, p. ej., conserva la actividad glucanasa. El polipéptido de fusión (o polinucleótido de fusión que codifica el polipéptido de fusión) puede comprender también componentes adicionales, incluyendo múltiples péptidos en múltiples bucles.

15 En un aspecto, los péptidos y ácidos nucleicos que los codifican se aleatorizan, o bien se aleatorizan completamente o su aleatorización es parcial, p. ej. en la frecuencia del nucleótido/residuo generalmente o por posición. "Aleatorizado" significa que cada ácido nucleico y péptido consiste esencialmente en nucleótidos y aminoácidos al azar, respectivamente. En un aspecto, los ácidos nucleicos que dan lugar a los péptidos se pueden sintetizar químicamente, y de este modo pueden incorporar cualquier nucleótido en cualquier posición. De este modo, cuando los ácidos nucleicos se expresan para formar péptidos, se puede incorporar cualquier residuo de aminoácido a cualquier posición. El procedimiento sintético se puede diseñar para generar ácidos nucleicos aleatorizados, para permitir la formación de todas o la mayoría de las combinaciones posibles a lo largo de la longitud del ácido nucleico, formando de este modo una biblioteca de ácidos nucleicos aleatorizados. La biblioteca puede proporcionar una población suficientemente diversa estructuralmente productos de expresión aleatorizados que afecte a un intervalo suficiente probabilísticamente de respuestas celulares para proporcionar una o más células que exhiben una respuesta deseada. De este modo, la invención proporciona una biblioteca interacción suficientemente larga de manera que al menos uno de sus miembros tenga una estructura que proporcione su afinidad por cierta molécula, proteína, u otro factor.

30 Las endoglucanasas son enzimas multidominio que consisten opcionalmente en un péptido señal, un módulo de unión a carbohidrato, un dominio catalítico de glucanasa, un conector y/u otro dominio catalítico.

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

50 Un polipéptido híbrido resultante del método de la invención puede exhibir actividad enzimática especializada no desplegada por las enzimas originales. Por ejemplo, después de la recombinación y/o el reordenamiento reductivo de los polinucleótidos que codifican las actividades hidrolasa, el polipéptido híbrido resultante codificado por un polinucleótido híbrido se puede escrutar para determinar las actividades hidrolasa especializadas obtenidas de cada una de las enzimas originales, es decir el tipo de enlace sobre el que actúa la hidrolasa y la temperatura a la que funciona la hidrolasa. De este modo, por ejemplo, la hidrolasa se puede escrutar para averiguar las funcionalidades químicas que distinguen la hidrolasa híbrida de las hidrolasas originales, tales como: (a) amida (enlaces peptídicos), es decir, endoglucanasas; (b) enlaces éster, es decir, esterases y lipasas; (c) acetales, es decir, glicosidasas y, por ejemplo, la temperatura, pH o concentración de sal a la cual funciona el polipéptido híbrido.

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

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

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

Adicionalmente, la subclonación se puede realizar para aislar adicionalmente secuencias de interés. En la subclonación, una porción de ADN se amplifica, se digiere, generalmente mediante enzimas de restricción, para separar mediante corte la secuencia deseada, la secuencia deseada se liga en un vector receptor y se amplifica. En cada etapa de la subclonación, la porción se examina para determinar la actividad de interés, con el fin de asegurarse de que no se ha excluido el ADN que codifica la proteína estructural. El inserto se puede purificar en cualquier etapa de la subclonación, por ejemplo, mediante electroforesis en gel antes de la ligación en un vector o donde células que contienen el vector receptor y células que no contienen el vector receptor se colocan en medios selectivos que contienen, por ejemplo, un antibiótico, que destruirá las células que no contienen el vector receptor. Los métodos específicos de subclonación de insertos de ADNc en vectores son bien conocidos en la técnica (Sambrook et al., *Molecular Cloning: A Laboratory Manual*, 2ª Ed., Cold Spring Harbor Laboratory Press (1989)). En otro aspecto, las enzimas de la invención son subclones. Tales subclones pueden diferir de los clones parentales, por ejemplo, en su longitud, una mutación, una etiqueta o una marca.

En un aspecto, las secuencias señal de la invención se identifican después de la identificación de los polipéptidos de glucanasa novedosos. Las rutas mediante las cuales las proteínas se clasifican y se transportan a su localización celular apropiada son referidas con frecuencia como rutas de redireccionamiento de proteínas. Uno de los elementos más importantes en todos estos sistemas de redireccionamiento es una secuencia de aminoácidos corta en el extremo amino de un polipéptido recién sintetizado denominado secuencia señal. Esta secuencia señal dirige una proteína a su localización apropiada en la célula y es eliminada durante el transporte o cuando la proteína alcanza su destino final. La mayoría de las proteínas lisosomales, de membrana, o secretadas tienen una secuencia señal amino terminal que las marca para la traslocación al lumen del retículo endoplásmico. Se han determinado más de 100 secuencias señal para proteínas en este grupo. Las secuencias señal pueden variar en longitud de 13 a 36 residuos de aminoácidos. Los expertos en la técnica conocen diversos métodos de reconocimiento de secuencias señal. En un aspecto, los péptidos se identifican por medio de un método referido como SignalP. SignalP utiliza una red neural combinada que reconoce ambos péptidos señal y sus sitios de escisión. Véase, p. ej. Nielsen (1997) "Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites" *Protein Engineering*, vol. 10, núm. 1, págs. 1-6. Se debe entender que algunas de las glucanasas de la invención pueden contener o no secuencias señal. Puede ser deseable que incluyan una secuencia de ácido nucleico que codifica una secuencia señal de una glucanasa conectada operablemente a una secuencia de ácido nucleico de una glucanasa diferente u, opcionalmente, una secuencia señal de una proteína no glucanasa.

Los microorganismos a partir de los cuales se puede preparar el polinucleótido incluyen microorganismos procarióticos, tales como *Eubacteria* y *Archaeobacteria* y microorganismos eucarióticos inferiores tales como hongos, ciertas algas y protozoos. Los polinucleótidos se pueden aislar a partir de muestras ambientales en cuyo caso el ácido nucleico se puede recuperar sin cultivar de un organismo o recuperar de uno o más organismos cultivados. En un aspecto, tales microorganismos pueden ser extremófilos, tal como hipertermófilos, psicrófilos, psicrótrofos, halófilos, barófilos y acidófilos. Se pueden utilizar polinucleótidos que codifican enzimas aisladas de microorganismos extremófilos. Tales enzimas pueden funcionar a temperaturas por encima de alrededor de 100°C en fuentes de aguas termales terrestres y aberturas térmicas del fondo del mar, a temperaturas por debajo de 0°C en aguas árticas, en el entorno salino saturado del Mar Muerto, a valores de pH de alrededor de 0 en depósitos de hulla y en fuentes geotérmicas ricas en azufre, o a valores de pH mayores de 11 en lodos de aguas residuales. Por ejemplo, diversas esterasas y lipasas clonadas y expresadas en organismos extremófilos muestran alta actividad a lo largo de un amplio intervalo de temperaturas y pH.

Los polinucleótidos seleccionados y aislados como se ha descrito antes se introducen en una célula anfitriona adecuada. Una célula anfitriona adecuada es cualquier célula que es capaz de promover la recombinación y/o el reordenamiento reductivo. Los polinucleótidos seleccionados están ya preferiblemente en un vector que incluye las

secuencias de control apropiadas. La célula anfitriona puede ser una célula eucariótica superior, tal como una célula de mamífero, o una célula eucariótica inferior, tal como una célula de levadura, o preferiblemente, la célula anfitriona puede ser una célula procariótica, tal como una célula bacteriana. La introducción del constructo en la célula anfitriona se puede efectuar mediante transfección con fosfato de calcio, transfección mediada por DEAE-Dextrano, o electroporación (Davis *et al.*, 1986).

Como Ejemplos representativos de los anfitriones apropiados, se pueden mencionar: células bacterianas; tales como *E. coli*, *Streptomyces*, *Salmonella typhimurium*; célula fúngicas, tales como levaduras; células de insecto tales como S2 de *Drosophila* y Sf9 de *Spodoptera*; células animales tales como CHO, COS o melanoma de Bowes; adenovirus; y células vegetales. Se considera que la selección de un anfitrión apropiado se encuentra al alcance de los expertos en la técnica a partir de las enseñanzas de la presente memoria.

Remitiéndose concretamente a diversos sistemas de cultivo de células de mamíferos que se pueden emplear para proteínas recombinantes, los ejemplos de los sistemas de expresión de mamíferos incluyen las líneas de fibroblastos de riñón de mono COS-7, descritas en "SV40-transformed simian cells support the replication of early SV40 mutants" (Gluzman, 1981) y otras líneas celulares capaces de expresar un vector compatible, por ejemplo, las líneas celulares C127, 3T3, CHO, HeLa y BHK. Los vectores de expresión de mamífero comprenderán un origen de replicación, un promotor y un intensificador adecuados y también sitios de unión al ribosoma necesarios cualesquiera, un sitio de poliadenilación, un donador de empalme y sitios aceptores, secuencias de terminación de la transcripción y secuencias no transcritas flanqueantes 5'. Las secuencias de ADN derivadas de los sitios de poliadenilación y empalme de SV40 se pueden utilizar para proporcionar los elementos genéticos no transcritos requeridos.

En otro aspecto, se prevé que el método de la presente invención se pueda utilizar para generar polinucleótidos novedosos que codifican rutas bioquímicas de uno o más operones o complejos génicos o porciones de los mismos. Por ejemplo, las bacterias y muchos eucariotas tienen un mecanismo coordinado para la regulación génica cuyos productos están implicados en procedimientos relacionados. Estos genes se agrupan, en estructuras referidas como "complejos génicos", en un solo cromosoma y se transcriben juntos bajo el control de una sola secuencia reguladora, incluyendo un solo promotor que inicia la transcripción del complejo completo. De este modo, un complejo génico es un grupo de genes adyacentes que son idénticos o están relacionados, usualmente como su función. Un ejemplo de una ruta bioquímica codificada por complejos génicos son los policétidos.

El ADN del complejo génico se puede aislar de diferentes organismos y ligar en vectores, particularmente vectores que contienen secuencias reguladoras de la expresión que pueden controlar y regular la producción de una proteína detectable o la actividad de la matriz relacionada con la proteína de los complejos génicos ligados. Los vectores que tienen una capacidad excepcionalmente grande para la introducción de ADN exógeno son particularmente apropiados para su uso con tales complejos génicos y se describen a modo de ejemplo en la presente memoria por incluir el factor f (o factor de fertilidad) de *E. coli*. Este factor f *E. coli* es un plásmido que afecta a su propia transferencia de alta frecuencia durante la conjugación y es ideal para lograr y propagar establemente grandes fragmentos de ADN, tales como complejos génicos de muestras microbiana mixtas. Un aspecto de la invención es la utilización de vectores de clonación, referidos como "fósmidos" o vectores de cromosomas bacterianos artificiales (BAC). Estos están derivados del factor f de *E. coli* que es capaz de integrar establemente grandes segmentos de ADN genómico. Cuando se integran con ADN de una muestra ambiental no cultivada mixta, esto hace posible alcanzar grandes fragmentos genómicos en forma de una "biblioteca de ADN ambiental" estable. Otro tipo de vector para su uso la presente invención es un vector cosmídico. Los vectores cosmídicos fueron diseñados originalmente para clonar y propagar grandes segmentos de ADN genómico. La clonación en vectores cosmídicos la describen con detalle Sambrook *et al.*, en *Molecular Cloning: A Laboratory Manual*, 2^a Ed., Cold Spring Harbor Laboratory Press (1989). Una vez ligados en un vector apropiado, se pueden introducir dos o más vectores que contienen diferentes complejos génicos de policétido sintasa en una célula anfitriona adecuada. Las regiones de homología de secuencia parcial compartidas por los complejos génicos promoverán procedimientos tienen como consecuencia la reorganización de la secuencia dando como resultado un complejo génico híbrido. El complejo génico híbrido novedoso se puede escrutar a continuación para determinar las actividades potenciadas no encontradas en los complejos génicos originales.

Un método para producir un polipéptido híbrido biológicamente activo y rastrear tal polipéptido para determinar la actividad potenciada puede ser:

- 1) introduciendo al menos un primer polinucleótido en una conexión operable y un segundo polinucleótido en una conexión operable, compartiendo el al menos primer polinucleótido y segundo polinucleótido al menos una región de homología de secuencia parcial, en una célula anfitriona adecuada;
- 2) cultivando la célula anfitriona en condiciones que promuevan la reorganización de secuencia dando como resultado a polinucleótido híbrido en una conexión operable;
- 3) expresando un polipéptido híbrido codificado por el polinucleótido híbrido;
- 4) escrutando el polipéptido híbrido en condiciones que promuevan la identificación de la actividad biológica potenciada; y

5) aislando el polinucleótido que codifica el polipéptido híbrido.

Los métodos para escrutar diversas actividades enzimáticas son conocidos por los expertos en la técnica y se comentan a lo largo de la presente memoria. Tales métodos se pueden emplear cuando se aislen los polipéptidos y polinucleótidos de la invención.

Metodologías de Escrutinio y Dispositivos de Verificación "En línea (On-line)"

Al poner en practica los métodos de la invención, se pueden utilizar una variedad de aparatos y metodologías junto con los polipéptidos y ácidos nucleicos de la invención, p. ej., para escrutar polipéptidos para determinar su actividad glucanasa (p. ej., análisis tales como hidrólisis de caseína en zimogramas, la liberación de fluorescencia a partir de la gelatina, o la liberación p-nitroanilida a partir de diversos sustratos peptídicos pequeños), para escrutar compuestos como moduladores potenciales, p. ej., activadores o inhibidores, de una actividad glucanasa, para determinar los anticuerpos que se unen a un polipéptido de la invención, para determinar los ácidos nucleicos que hibridan con un ácido nucleico de la invención, para escrutar las células que expresan un polipéptido de la invención y similares. Además de los formatos de matrices descritos en detalle más abajo para escrutar muestras, también se puede utilizar formatos alternativos para poner en práctica los métodos de la invención. Tales formatos incluyen, por ejemplo, espectrómetros de masas, cromatógrafos, p. ej., HPLC de alto rendimiento y otras formas de cromatografía líquida, y formatos más pequeños, tales como placas de 1536 pocillos, placas de 384 pocillos y etcétera. Los aparatos de escrutinio de alto rendimiento se pueden adaptar y utilizar para poner en práctica los métodos de la invención, véase, p. ej., la Solicitud de Patente de los Estados Unidos Núm. 20020001809.

Matrices Capilares

Los ácidos nucleicos o polipéptidos de la invención se pueden inmovilizar en o aplicar a una matriz. Las matrices se pueden utilizar para escrutar o verificar bibliotecas de composiciones (p. ej., moléculas pequeñas, anticuerpos, ácidos nucleicos, etc.) para determinar su capacidad para unirse a, o modular la actividad de un ácido nucleico o un polipéptido de la invención. Las matrices capilares, tales como GIGAMATRIX™, Diversa Corporation, San Diego, CA; y las matrices descritas, p. ej., en Solicitud de Patente de los Estados Unidos Núm. 20020080350 A1; el documento WO 0231203 A; el documento WO 0244336 A, proporcionan un aparato alternativo para albergar y escrutar muestras. En un aspecto, la matriz capilar incluye una pluralidad de capilares formados en una matriz de capilares adyacentes, donde cada capilar comprende al menos una pared que define un lumen para retener una muestra. El lumen puede ser cilíndrico, cuadrado, hexagonal o de cualquier otra forma geométrica con tal que las paredes formen un lumen 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 planar. Los capilares se pueden unir entre sí, fusionándolos (p. ej., 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 planar sólido que contiene una pluralidad de orificios.

Se puede formar una matriz capilar de cualquier número de capilares individuales, por ejemplo, un intervalo de 100 a 4.000.000 de capilares. Adicionalmente, se puede formar una matriz capilar que tenga alrededor de 100.000 capilares individuales o más en el tamaño y forma convencionales de una placa Microtiter® para su ajuste en equipos de laboratorio convencionales. Los lúmenes se cargan manualmente o automáticamente utilizando cualquier acción o microinyección capilar utilizando una aguja delgada. Las muestras de interés se pueden eliminar posteriormente de los capilares individuales para su análisis o caracterización adicionales. Por ejemplo, se coloca una sonda de tipo aguja, delgada en comunicación fluida con un capilar seleccionado para añadir o retirar material del lumen.

En un análisis de escrutinio de un solo recipiente, los componentes del análisis se mezclan produciendo una solución de interés, antes de su inserción en la matriz capilar. El lumen se carga mediante acción capilar cuando al menos una porción de la matriz se sumerge en una solución de interés. Las reacciones químicas o biológicas y/o la actividad en cada capilar son verificadas para determinar los eventos detectables. Un evento detectable es referido con frecuencia como un "éxito", que usualmente se puede distinguir de los "no éxitos" que producen los capilares mediante detección óptica. De este modo, las matrices capilares permiten masivamente la detección paralela de "éxitos".

En un análisis de escrutinio de múltiples recipientes, un polipéptido o ácido nucleico, p. ej., 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, donde 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 controla la matriz capilar para determinar un evento detectable resultante de la reacción o no reacción de los dos componentes.

En un análisis de escrutinio 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, donde el lumen del capilar es recubierto con un material de unión para unir la partícula detectable al lumen. El primer líquido se puede eliminar a continuación del tubo capilar, donde la partícula detectable unida se mantiene dentro del capilar, y se puede introducir un segundo líquido en el tubo capilar. El capilar se controla después para determinar un evento detectable resultante de la reacción o no reacción de la partícula con el segundo líquido.

Matrices, o "Biochips"

Los ácidos nucleicos o polipéptidos de la invención se pueden inmovilizar en o aplicar a una matriz. Las matrices se pueden utilizar para escrutar o controlar bibliotecas de composiciones (p. ej., moléculas pequeñas, anticuerpos, ácidos nucleicos, etc.) para determinar su capacidad para unirse a o modular la actividad de un ácido nucleico o un polipéptido de la invención. Por ejemplo, en un aspecto de la invención, un parámetro controlado es la expresión de un transcrito de un gen de glucanasa. 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 con ácidos nucleicos inmovilizados sobre una matriz, o "biochip". Mediante la utilización de una "matriz" de ácidos nucleicos sobre un microchip, algunos o todos los transcritos de una célula se pueden cuantificar simultáneamente. Alternativamente, también se pueden utilizar matrices que comprenden ácido nucleico genómico para determinar el genotipo de una cepa construida mediante los métodos de la invención. Las matrices de polipéptidos también se pueden utilizar para cuantificar simultáneamente una pluralidad de proteínas. La presente invención se puede poner en práctica con cualquier "matriz" conocida también referida como "micromatriz" o "matriz de ácido nucleico" o "matriz de polipéptido" o "matriz de anticuerpo" o "biochip", o cualquier 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, p. ej., oligonucleótidos, inmovilizados sobre una zona definida de una superficie sustrato para su unión específica a una molécula muestra, p. ej., transcritos de ARNm.

Al poner en práctica los métodos de la invención, se puede incorporar en su totalidad o en parte cualquier matriz conocida y/o método para elaborar y utilizar matrices, o variaciones de los mismos, como se describe, por ejemplo, en las Patentes de los Estados Unidos Núms. 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, p. ej., los documentos WO 99/51773; WO 99/09217; WO 97/46313; WO 96/17958; véanse también, p. ej., 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 Patente de los Estados Unidos Publicadas Núms. 20010018642; 20010019827; 20010016322; 20010014449; 20010014448; 20010012537; 20010008765.

Anticuerpos y métodos de escrutinio basados en anticuerpos

La descripción proporciona anticuerpos aislados o recombinantes que se unen específicamente a una glucanasa de la invención. Estos anticuerpos se pueden utilizar para aislar, identificar o cuantificar una glucanasa de la invención o polipéptidos relacionados. Estos anticuerpos se pueden utilizar para aislar otros polipéptidos dentro del alcance de la invención u otras glucanasas relacionadas. Los anticuerpos se pueden diseñar para unirse a un sitio activo de una glucanasa. De este modo, la invención proporciona métodos para inhibir glucanasas utilizando los anticuerpos de la invención (véase la discusión anterior concerniente a las aplicaciones de las composiciones anti-glucanasa de la invención).

Los fragmentos de las enzimas de la invención pueden incluir fragmentos inmunogénicos de un polipéptido de la invención. La invención proporciona composiciones que comprenden un polipéptido o péptido de la invención y coadyuvantes o portadores y similares.

Los anticuerpos se pueden utilizar en inmunoprecipitación, tinción, columnas de inmovilización, y similares. Si se desea, las secuencias de ácido nucleico que codifican antígenos específicos se pueden generar mediante inmunización seguida de aislamiento del polipéptido o ácido nucleico, amplificación o clonación e inmovilización del polipéptido sobre una matriz de la invención. Alternativamente, los métodos de la invención se puede utilizar para modificar la estructura de un anticuerpo producido por una célula que se vaya a modificar, p. ej., se puede aumentar o disminuir la afinidad de un anticuerpo. Además, la capacidad para elaborar o modificar anticuerpos puede ser un fenotipo construido en una célula mediante los métodos de la invención.

Los métodos para la inmunización, la producción y el aislamiento de anticuerpos (policlonales y monoclonales) son conocidos por los expertos en la técnica y descritos en la bibliografía científica y de patentes, véase, p. ej., Coligan, CURRENT PROTOCOLS IN IMMUNOLOGY, Wiley/Greene, NY (1991); Stites (eds.) BASIC AND CLINICAL IMMUNOLOGY (7th ed.) Lange Medical Publications, Los Altos, CA ("Stites"); Goding, MONOCLONAL

ANTIBODIES: PRINCIPLES AND PRACTICE (2^a ed.) Academic Press, New York, NY (1986); Kohler (1975) Nature 256:495; Harlow (1988) ANTIBODIES, A LABORATORY MANUAL, Cold Spring Harbor Publicaciones, New York. También se pueden generar anticuerpos *in vitro*, p. ej., utilizando bibliotecas de presentación de fagos que expresan el sitio de unión al anticuerpo recombinante, además de los métodos *in vivo* tradicionales utilizando animales. Véanse, p. ej., Hoogenboom (1997) Trends Biotechnol. 15:62-70; Katz (1997) Annu. Rev. Biophys. Biomol. Struct. 26:27-45.

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 de los mismos, también se pueden utilizar para generar anticuerpos que se unen específicamente a los polipéptidos o fragmentos. Los anticuerpos resultantes se pueden utilizar en procedimientos de cromatografía de inmunoafinidad para aislar o purificar el polipéptido o para determinar si el polipéptido está presente en una muestra biológica. En tales procedimientos, una preparación de proteína, tal como un extracto, o una muestra biológica se pone en contacto con un anticuerpo capaz de unirse específicamente a uno de los polipéptidos de la 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.

En los procedimientos de inmunoafinidad, el anticuerpo se ancla a un soporte sólido, tal como una cuenta u otra matriz de la columna. La preparación de proteína se coloca en contacto con el anticuerpo en condiciones en las que el anticuerpo se une específicamente a uno de los polipéptidos de la invención, o un fragmento del mismo. Después de un lavado para eliminar las proteínas unidas no específicamente, se hacen eluir los polipéptidos unidos específicamente.

La capacidad de las proteínas en una muestra biológica para unirse al anticuerpo se puede determinar utilizando cualquiera de una variedad de procedimientos familiares para los expertos en la técnica. Por ejemplo, la unión se puede determinar marcando el anticuerpo con una marca detectable tal como un agente fluorescente, una marca enzimática, o un radioisótopo. Alternativamente, la unión del anticuerpo a la muestra se puede detectar utilizando un anticuerpo secundario que tiene en el mismo tal marca detectable. Los análisis concretos incluyen análisis ELISA, análisis sándwich, radioinmunoanálisis y Transferencias Western.

Los anticuerpos policlonales generados contra 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 de los mismos se pueden obtener mediante inyección directa de los polipéptidos en un animal o administrando los polipéptidos a un animal, por ejemplo, un animal no humano. El anticuerpo obtenido de este modo se unirá a continuación al propio polipéptido. De esta manera, se puede utilizar incluso una secuencia que codifica solo un fragmento del polipéptido para generar anticuerpos que se pueden unir al polipéptido nativo completo. Tales anticuerpos se pueden utilizar a continuación para aislar el polipéptido a partir de células que expresan ese polipéptido.

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

Las técnicas descritas para la producción de anticuerpos de cadena sencilla (Patente de los Estados Unidos Núm. 4.946.778) se pueden adaptar para producir anticuerpos de cadena sencilla para 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 de los mismos. Alternativamente, se pueden utilizar ratones transgénicos para expresar anticuerpos humanizados para estos polipéptidos o fragmentos de los mismos.

Los anticuerpos generados contra 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 de los mismos se pueden utilizar en el escrutinio de polipéptidos similares de otros organismos y muestras. En tales técnicas, los polipéptidos del organismo se ponen en contacto con el anticuerpo y se detectan los polipéptidos que se unen específicamente al anticuerpo. Cualquiera de los procedimientos descritos anteriormente se puede utilizar para detectar la unión al anticuerpo. Uno de tales análisis de escrutinio se describe en "Methods for Measuring Cellulase Activities", Methods in Enzymology, Vol 160, págs. 87-116.

Kits

Los kits pueden comprender las composiciones, p. ej., ácidos nucleicos, casetes de expresión, vectores, células, semillas transgénicas o plantas o partes de plantas, polipéptidos (p. ej., endoglucanasas) y/o anticuerpos de la invención. Los kits también pueden contener material básico para ilustrar las metodologías y usos industriales de la invención, como se describe en la presente memoria.

Construcción de la célula completa y parámetros metabólicos de medición

5 Los métodos de la invención proporcionan la evolución de la célula completa, o la construcción de la célula completa, de una célula para desarrollar una nueva cepa celular que tenga un nuevo fenotipo, p. ej., una actividad glucanasa nueva o modificada, modificando la composición genética de la célula. La composición genética se puede modificar mediante la adición a la célula de un ácido nucleico de la invención, p. ej., una secuencia codificante para una enzima de la invención. Véanse, p. ej., los documentos WO 0229032; WO 0196551.

10 Para detectar el nuevo fenotipo, se controla al menos un parámetro metabólico de una célula modificada en la célula en un marco temporal a "tiempo real" o "en línea". En un aspecto, se controla una pluralidad de células, tal como un cultivo celular, en "tiempo real" o "en línea". En un aspecto, se controla una pluralidad de parámetros metabólicos en "tiempo real" o "en línea". Los parámetros metabólicos se pueden controlar utilizando las glucanasas de la invención.

15 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:

20 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, p. ej. interacciones alostéricas, interacciones enzima-enzima etc,
 25 la compartimentalización intracelular de las enzimas o cualquier otra organización supramolecular de las enzimas, y,
 la presencia de gradientes de cualquier concentración de metabolitos, enzimas o moléculas efectoras o barreras a la difusión para su movimiento.

30 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 los cambios de la red metabólica completa en una célula. El fenotipo metabólico se basa en el cambio de la ruta de utilización con respecto a las condiciones ambientales, la regulación genética, el estado de desarrollo y el genotipo, etc. Después del cálculo del AFM en línea, el comportamiento dinámico de las células, su fenotipo y otras propiedades se pueden analizar investigando la utilización de la ruta. Por ejemplo, si aumenta el aporte de glucosa y el oxígeno disminuye 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 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 practica los métodos de la invención, también se pueden comparar los resultados del AFM con datos del transcriptoma y proteoma para diseñar experimentos y protocolos de ingeniería metabólica o barajado de genes, etc.

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

Control de la expresión de un transcrito de ARNm

50 El fenotipo ideado comprende el aumento o la disminución de la expresión de un transcrito de ARNm (p. ej., un mensaje de glucanasa) o generando nuevos transcritos (p. ej., glucanasa) en una célula. Esta expresión aumentada o reducida se puede rastrear comprobando la presencia de una glucanasa de la invención o mediante análisis de la actividad glucanasa. Los transcritos de ARNm, o mensajes, también se pueden detectar y cuantificar por medio de cualquier método conocido, incluyendo, p. ej., transferencias Northern, reacciones de amplificación cuantitativas, hibridación a matrices, y similares. Las reacciones de amplificación cuantitativas incluyen, p. ej., PCR cuantitativa, incluyendo, p. ej., reacción en cadena de la polimerasa mediante transcripción inversa cuantitativa, o RT-PCR; RT-PCR en tiempo real cuantitativa, o " RT-PCR cinética en tiempo real" (véanse, p. ej., Kreuzer (2001) Br. J. Haematol. 114:313-318; Xia (2001) Transplantation 72:907-914).

60 El fenotipo ideado se puede generar desactivando la expresión de un gen homólogo. La secuencia codificante del gen o uno o más elementos de control de la transcripción pueden estar desactivados, p. ej., promotores o intensificadores. De este modo, la expresión de un transcrito puede estar anulada completamente o solo mermada.

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

Control de la expresión de polipéptidos, péptidos y aminoácidos

El fenotipo ideado puede comprender el aumento o la disminución de la expresión de un polipéptido (p. ej., una glucanasa) o la generación de nuevos polipéptidos en una célula. Esta expresión aumentada o disminuida se puede rastrear determinando la cantidad de glucanasa presente o mediante análisis de la actividad glucanasa. Los polipéptidos, péptidos y aminoácidos también se pueden detectar y cuantificar por medio de cualquier método conocido en la técnica, incluyendo, p. ej., resonancia magnética nuclear (RMN), espectrofotometría, radiografía (radiomarcaje de proteínas), electroforesis, electroforesis capilar, cromatografía líquida de alta resolución (HPLC), cromatografía en capa fina (TLC), cromatografía de hiperdifusión, diversos métodos inmunológicos, p. ej. inmunoprecipitación, inmunodifusión, inmunolectroforesis, radioinmunoanálisis (RIA), análisis inmunoabsorbentes con enzima ligada (ELISA), análisis inmunofluorescentes, electroforesis en gel (p. ej., SDS-PAGE), tinción con anticuerpos, clasificador de células activadas por fluorescencia (FACS), pirólisis-espectrometría masas, Espectrometría de Infrarrojos por Transformada de Fourier, espectrometría Raman, GC-MS, y LC-Electropulverización y cap-LC-tándem-espectrometrías de masas por electropulverización, y similares. También se pueden escrutar bioactividades novedosas utilizando métodos, o variaciones de los mismos, descritos en la Patente de los Estados Unidos Núm. 6.057.103. Además, como se comenta con detalle más abajo, se pueden medir uno o más, o, todos los polipéptidos de una célula utilizando una matriz de proteínas.

Aplicaciones Industriales

Las enzimas glucanasas de la invención pueden ser catalizadores altamente selectivos. Pueden catalizar reacciones con estereo-, regio- y quimio-selectividades exquisitas sin precedente en la química sintética convencional. Por otra parte, las enzimas son notablemente versátiles. Las enzimas de la invención se pueden adaptar a funcionar en disolventes orgánicos, funcionar a pH extremos (por ejemplo, pH elevados y pH bajos) temperaturas extremas (por ejemplo, temperaturas elevadas y temperaturas bajas), niveles de salinidad extremos (por ejemplo, alta salinidad y baja salinidad) y catalizar reacciones con compuestos que no están relacionados estructuralmente con sus sustratos fisiológicos, naturales.

Composiciones Detergentes

La invención proporciona composiciones detergentes que comprenden uno o más polipéptidos (p. ej., endoglucanasas) de la invención, y métodos para elaborar y utilizar estas composiciones. La invención incorpora todos los métodos para elaborar y utilizar composiciones detergentes, véanse, por ejemplo las Patentes de los Estados Unidos Núms. 6.413.928; 6.399.561; 6.365.561; 6.380.147. Las composiciones detergentes pueden ser una composición acuosa de una o dos partes, una composición líquida no acuosa, un sólido moldeado, una forma granular, una forma particulada, una pastilla comprimida, un gel y/o una pasta y una forma en suspensión. Las glucanasas de la invención se pueden utilizar también como un producto aditivo detergente en forma sólida o líquida. Tales productos aditivos están destinados a funcionar como suplemento o refuerzo de composiciones detergentes convencionales y se pueden añadir en cualquier fase del procedimiento de limpieza.

El contenido de enzima activa real depende del método de fabricación de la composición detergente y no es crítico, suponiendo que la solución detergente tiene la actividad enzimática deseada. En un aspecto, la cantidad de glucanasa presente en la solución final oscila de alrededor de 0,001 mg a 0,5 mg por gramo de la composición detergente. La enzima concreta elegida para su uso en el procedimiento y los productos depende de las condiciones de utilidad final, incluyendo la forma física del producto, el pH de uso, la temperatura de uso, y los tipos de suciedad que vayan a ser degradados o alterados. La enzima se puede elegir para proporcionar una actividad y estabilidad óptimas para cualquier conjunto dado de condiciones de utilidad. En un aspecto, las glucanasas de la presente invención son activas en el pH que oscila de alrededor de 4 a alrededor de 12 y en el intervalo de temperatura de alrededor de 20°C a alrededor de 95°C. Los detergentes de la invención pueden comprender tensioactivos catiónicos, no iónicos semi-polares o zwitteriónicos; o, mezclas de los mismos.

Las glucanasas de la invención se pueden formular en detergentes en polvo y líquidos que tienen pH entre 4,0 y 12,0 a niveles de alrededor de 0,01 a alrededor de 5% (preferiblemente de 0,1% a 0,5%) en peso. Estas composiciones detergentes pueden incluir también otras enzimas tales como glucanasas, mananasas, o xilanasas, o celulasas, endoglicosidasas, endo-beta-1,4-glucanasas, beta-glucanasas, endo-beta-1,3(4)-glucanasas, catalasas, cutinasas, peroxidadasas, lacasas, lipasas, amilasas, glucoamilasas, pectinasas, reductasas, oxidadasas, fenoloxidasas, ligninasas, pululanadas, arabinanasas, hemicelulasas, mananasas, xiloglucanasas, pectín acetil esteradasas, ramnogalacturonano acetil esteradasas, poligalacturonadasas, ramnogalacturonadasas, galactanasas, proteasas, pectato

liasas, pectín metilesterasas, celobiohidrolasas y/o transglutaminasas. Estas composiciones detergentes pueden incluir también reforzantes y estabilizadores. Estas composiciones detergentes pueden incluir también reforzantes y estabilizadores.

5 La adición de una glucanasa de la invención a composiciones de limpieza convencionales no crea ninguna limitación de uso especial. En otras palabras, cualquier temperatura y pH adecuados para el detergente también son adecuados para las composiciones con tal que la enzima esté activa o sea tolerante al pH y/o temperatura de uso pretendidos. Además, una glucanasa de la invención se puede utilizar en una composición de limpieza sin detergentes, de nuevo sola o combinada con reforzantes y estabilizadores.

10 La presente invención proporciona composiciones de limpieza que incluyen composiciones detergentes para limpiar superficies duras, composiciones detergentes para limpiar tejidos, composiciones lavavajillas, composiciones de limpieza orales, composiciones de limpieza de dentaduras, o soluciones de limpieza de lentes de contacto.

15 Un método para lavar un objeto comprende poner en contacto el objeto con un polipéptido de la invención en condiciones suficiente para su lavado. Se puede incluir una glucanasa de la invención como aditivo detergente. La composición detergente de la invención se puede formular, por ejemplo, en forma de una composición detergente para al colada a mano o a máquina que comprende un polipéptido de la invención. Un aditivo para la colada adecuado para el pre-tratamiento de tejidos manchados puede comprender un polipéptido de la invención. Una composición suavizante de tejidos puede comprender una glucanasa de la invención. Alternativamente, una glucanasa de la invención se puede formular en forma de una composición detergente para uso en operaciones generales de limpieza de superficies duras domésticas. En aspectos alternativos, los aditivos para detergentes y las composiciones detergentes pueden comprender uno o más de otras enzimas tales como otras glucanasas, mananasas, o xilanasas, o una xilanasas, una lipasa, una cutinasa, otra glucanasa, una carbohidrasa, una celulasa, una pectinasa, una arabinasa, a galactanasa, una oxidasa, p. ej., una lactasa, y/o una peroxidasa (véase también, lo antedicho). Las propiedades de la enzima o las enzimas de la invención se eligen para que sean compatibles con el detergente seleccionado (es decir pH óptimo, compatibilidad con otros ingredientes enzimáticos y no enzimáticos, etc.) y la enzima o las enzimas están presentes en cantidades eficaces. En un aspecto, las enzimas glucanasas de la invención se utilizan para eliminar materiales malolientes de los tejidos. Las diversas composiciones detergentes y métodos para elaborarlas que se puede utilizar para poner en práctica la invención se describen, p. ej., en las Patentes de los Estados Unidos Núms. 6.387.690; 6.333.301; 6.329.333; 6.326.341; 6.297.038; 6.309.871; 6.204.232; 6.197.070; 5.856.164.

35 Cuando se formulan en forma de composiciones adecuadas para su uso en un método de lavado a máquina para la colada, las enzimas de la invención pueden comprender un compuesto tensioactivo y un reforzante. Estas pueden comprender adicionalmente uno o más componentes detergentes, p. ej., compuestos poliméricos orgánicos, agentes blanqueantes, enzimas adicionales, supresores del agua de lavado, dispersantes, dispersantes de jabón de cal, agentes suspensores de manchas y anti-redepósito e inhibidores de la corrosión. Las composiciones para la colada pueden contener también agentes suavizantes, como componentes detergentes adicionales. Tales composiciones que contienen carbohidrasa pueden proporcionar limpieza de tejidos, eliminación de la suciedad, mantenimiento de la blancura, suavizamiento, apariencia de color, inhibición de la transferencia de colorantes e higienización cuando se formulan en forma de composiciones detergentes para la colada.

45 La densidad de las composiciones detergentes para la colada puede oscilar de alrededor de 200 a 1500 g/litros, o, de alrededor de 400 a 1200 g/litro, o, de alrededor de 500 a 950 g/litro, o, 600 a 800 g/litro, de composición; ésta se puede medir a alrededor de 20°C.

50 La forma "compacta" de las composiciones detergentes para la colada se refleja mejor por su densidad y, en términos de composición, por la cantidad de sal de carga inorgánica. Las sales de cargas inorgánica son ingredientes convencionales de composiciones detergentes en forma de polvo. En composiciones detergentes convencionales, las sales de carga están presentes en cantidades sustanciales, típicamente de 17% a 35% en peso de la composición total. En un aspecto de las composiciones compactas, la sal de carga está presente en cantidades que no excedan de 15% de las composición, o, no exceda de 10%, o, no exceda de 5% en peso de la composición. Las sales de carga inorgánica se pueden seleccionar entre las sales de metales alcalinos y alcalinotérreos de sulfatos y cloruros, p. ej., sulfato de sodio.

55 Las composiciones detergentes líquidas pueden estar también en una "forma concentrada". En un aspecto, las composiciones detergentes líquidas pueden contener una cantidad inferior de agua, en comparación con los detergentes líquidos convencionales. En aspectos alternativos, el contenido de agua del detergente líquido concentrado es menor de 40%, o, menor de 30%, o, menor de 20% en peso de la composición detergente. Los compuestos detergentes pueden comprender formulaciones como las descritas en el documento WO 97/01629.

Las enzimas de la invención pueden ser útiles para formular diversas composiciones de limpieza. Varios compuestos conocidos son tensioactivos adecuados incluyendo detergentes no iónicos, aniónicos, catiónicos, o zwitteriónicos, se

pueden utilizar, p. ej., como se describe en las Patentes de los Estados Unidos Núms. 4.404.128; 4.261.868; 5.204.015. Además, las glucanasas de la invención se pueden utilizar, por ejemplo, en aplicaciones de jabón líquido o en pastilla, formulaciones para el cuidado de la vajilla, soluciones o productos de limpieza de lentes de contacto, hidrólisis de péptidos, tratamiento de residuos, aplicaciones textiles, como enzimas fusión-escisión en la producción de proteínas, y similares. Las glucanasas de la invención pueden proporcionar un aumento de rendimiento en una composición detergente en comparación con otra glucanasa detergente, esto es, el grupo enzimático puede aumentar la limpieza de ciertas manchas sensibles a las enzimas tales como hierba o sangre, según se determina mediante la evaluación habitual después de un ciclo de lavado convencional. Las glucanasas de la invención se pueden formular en detergentes en polvo y líquidos que tienen un pH entre 6,5 y 12,0 a niveles de alrededor de 0,01 a alrededor de 5% (por ejemplo, alrededor de 0,1% a 0,5%) en peso. Estas composiciones de limpieza detergentes pueden incluir también otras enzimas tales como glucanasas, mananasas, xilanasas, amilasas, celulasas, lipasas o endoglicosidasas conocidas, así como reforzantes y estabilizadores.

En un aspecto, la invención proporciona composiciones detergentes que tienen actividad glucanasa (una glucanasa de la invención) para su uso con compuestos de frutas, hortalizas y/o barro y arcilla (véase, por ejemplo, la Patente de los Estados Unidos Núm. 5.786.316).

Tratamiento de fibras y artículos textiles

Los métodos para tratar fibras y tejidos pueden utilizar una o más glucanasas de la invención. Las enzimas de la invención se pueden utilizar en cualquier método de tratamiento de fibras o tejidos, que son bien conocidos en la técnica, véase, p. ej., la Patente de los Estados Unidos Núm. 6.261.828; 6.077.316; 6.024.766; 6.021.536; 6.017.751; 5.980.581; la Publicación de Patente de los Estados Unidos Núm. 20020142438 A1. Por ejemplo, las enzimas de la invención se pueden utilizar en el desaprestado de fibras y/o tejidos. En un aspecto, se mejora el tacto y la apariencia de un tejido por medio de un método que comprende poner en contacto el tejido con una enzima de la invención en una solución. En un aspecto, el tejido se trata con la solución a presión. Por ejemplo, las enzimas de la invención se pueden utilizar en la eliminación de manchas.

En un aspecto, las enzimas de la invención se aplican durante o después del entretejido de los artículos textiles, o durante la fase de desaprestado, o durante una o más etapas de procesamiento de tejidos adicionales. Durante el entretejido de artículos textiles, los hebras se exponen a un esfuerzo mecánico considerable. Antes del entretejido en un telar mecánico, los hilos de urdimbre se recubren a menudo con almidón o derivados de almidón de apresto con el fin de aumentar la resistencia a la tracción y para evitar la rotura. Una vez que los artículos textiles se han tejido, se puede proceder a la fase de desaprestado del tejido. Esto puede estar seguido de una o más etapas de procesamiento del tejido adicionales. El desaprestado es el acto de eliminar el "apresto" de los artículos textiles. Después del entretejido, el recubrimiento de apresto se debe eliminar antes del procesamiento adicional del tejido con el fin de asegurar un resultado homogéneo y a prueba de lavado.

Las enzimas de la invención se pueden utilizar para tratar cualquier material celulósico, incluyendo fibras (p. ej., fibras de algodón, cáñamo, linaza o lino), tejidos cosidos y no cosidos, p. ej., géneros tricotados, géneros tejidos, tela vaquera, hilaturas, y tejidos de rizo, elaborados de algodón, mezclas de algodón o productos celulósicos naturales o fabricados por el hombre (p. ej. originados a partir de fibras de celulosa que contienen glucano por ejemplo de pasta de madera) o combinaciones de los mismos. Los ejemplos de las combinaciones son combinaciones de algodón o rayón/viscosa con uno o más materiales acompañantes tales como lana, fibras sintéticas (p. ej. fibras de poliamida, fibras acrílicas, fibras de poliéster, fibras de poli(alcohol vinílico), fibras de poli(cloruro de vinilo), fibras de poli(cloruro de vinilideno), fibras de poliuretano, fibras de poliurea, fibras de aramida), y fibras que contienen celulosa (p. ej. rayón/viscosa, ramio, cáñamo, linaza/lino, yute, fibras de acetato de celulosa, liocel).

Las enzimas de la invención se pueden utilizar para tratar tejidos o cualquier material que comprende glucano, o celulosa, incluyendo tejidos que contienen algodón, como aditivos detergentes, p. ej., en composiciones acuosas. Para la fabricación de telas, el tejido se puede cortar y coser en telas o prendas de vestir. Estas se pueden finalizar antes o después del tratamiento. En particular, para la fabricación de pantalones vaqueros, se han desarrollado diferentes métodos de acabado enzimático. El acabado de las prendas de vestir de tela vaquera se inicia normalmente con una etapa de desaprestado enzimático, durante la cual las prendas de vestir se someten a la acción de enzimas amilolíticas con el fin de proporcionar suavidad al tejido y volver el algodón más accesible a las posteriores etapas de acabado enzimático. La invención proporciona métodos para el tratamiento de artículos textiles, p. ej., acabado de prendas de vestir de tela vaquera, desaprestado enzimático y aporte de suavidad a los tejidos mediante el uso de una combinación de enzimas, tales como, las glucanasas (p. ej., endoglucanasas) de la invención. En un aspecto, las enzimas de la invención se pueden utilizar en tratamientos para prevenir el agrisado de un artículo textil.

En un aspecto, se combinan glucanasas (p. ej., endoglucanasas) alcalinas y/o termoestables de la invención en un desaprestado y biodescrudado en un solo baño. Entre las ventajas de combinar el desaprestado y el descrudado se encuentran la reducción de costes y un menor impacto medioambiental debido al ahorro en el uso de energía y agua

y una menor producción de desechos. Las condiciones de aplicación para el desaprestado y el biodescruado pueden estar entre aproximadamente pH 8,5 a pH 10,0 y temperaturas de aproximadamente 40°C y superiores. Se pueden utilizar dosis enzimáticas menores (p. ej., aproximadamente 5 g por tonelada de algodón) y tiempos de reacción más cortos (p. ej. aproximadamente 15 minutos) para obtener el desaprestado y descruado eficaces sin añadir calcio.

Las enzimas de la invención se pueden utilizar en el tratamiento de tejidos que contienen celulosa para la reducción de la aspereza, para el aclarado del color, o para proporcionar una variación localizada del color de tales tejidos. Véase, p. ej., la Patente de los Estados Unidos Núm. 6.423.534. Por ejemplo, las enzimas de la invención se pueden utilizar para reducir la aspereza de tejidos que contienen algodón, p. ej., en forma de aditivo detergente para reducir la aspereza. Las enzimas de la invención se pueden utilizar en el tratamiento de tejidos para conferir un aspecto "lavado a la piedra" en un tejido coloreado a la vez que se reduce la cantidad de redepósito del colorante sobre el tejido.

Los procedimientos de tratamiento de artículos textiles de la invención (utilizando glucanasas de la invención) se pueden utilizar junto con otros tratamientos de artículos textiles, p. ej., descruado y blanqueo. El descruado es la eliminación de material no celulósico de la fibra de algodón, p. ej., la cutícula (consistente principalmente de ceras) y la pared celular primaria (consistente principalmente de pectina, proteína y xiloglucano). Es necesaria una eliminación apropiada de la cera para obtener una alta humectabilidad. Esto es necesario para la tintura. La eliminación de las paredes celulares primarias mediante el procedimientos de la invención mejora la eliminación de la cera y asegura una tintura más uniforme. El tratamiento de artículos textiles con los procedimientos de la invención puede mejorar la blancura en el procedimiento de blanqueado. El producto químico principal utilizado en el descruado es el hidróxido de sodio a altas concentraciones y a elevadas temperaturas. El blanqueado comprende la oxidación del artículo textil. El blanqueado implica típicamente el uso de peróxido de hidrógeno como agente oxidante con el fin de obtener un tejido completamente blanqueado (blanco) o para asegurar un matiz limpio del colorante.

La invención también proporciona glucanasas alcalinas (p. ej. endoglucanasas activas en condiciones alcalinas). Estas tienen aplicaciones de amplio alcance en el tratamiento de artículos textiles, desengomado de las fibras vegetales (p. ej., fibras liberianas vegetales), tratamiento de aguas residuales, p. ej., aguas residuales pécticas, elaboración de papel, y fermentaciones de café y té. Véase, p. ej., Hoondal (2002) Applied Microbiology and Biotechnology 59:409-418.

Los procesos de tratamiento de artículos textiles de la invención pueden incluir también el uso de cualquier combinación de otras enzimas (incluyendo enzimas que degradan carbohidratos) tales como catalasas, otras glucanasas, celulasas, lipasas, endoglicosidasas, endo-beta-1,4-glucanasas, beta-glucanasas, endo-beta-1,3(4)-glucanasas, cutinasas, lacasas, amilasas, glucoamilasas, pectinasas, reductasas, oxidasas, fenoloxidasas, ligninasas, pululanasas, arabinanasas, hemicelulasas, otras mananasas, xiloglucanasas, otras xilanasas, pectín acetil esterases, ramnogalacturonano acetil esterases, proteasas, poligalacturonasas, ramnogalacturonasas, galactanasas, pectato liasas, pectín metilesterases, celobiohidrolasas y/o transglutaminasas. Las enzimas de la invención se pueden utilizar combinadas con otras enzimas degradantes de carbohidratos, p. ej., celulasa, arabinanasa, xiloglucanasa, pectinasa, xilanasa, y similares, para la preparación de fibras o para la limpieza de fibras. Las proteasas se pueden utilizar en una combinación de enzimas de la invención. Estas se pueden utilizar combinadas con detergentes.

Tratamiento de alimentos y elaboración de alimentos

Las glucanasas de la invención tienen numerosas aplicaciones en la industria de elaboración de alimento. Por ejemplo, en un aspecto, las enzimas de la invención se utilizan para mejorar la extracción de aceite a partir de material vegetal rico en aceite, p. ej., semillas ricas en aceite, por ejemplo, aceite de soja a partir de habas de soja, aceite de oliva a partir de aceitunas, aceite de colza a partir de semillas de colza y/o aceite de girasol a partir de semillas de girasol.

Las enzimas de la invención se pueden utilizar en la preparación de zumos, jarabes, extractos y similares de frutas u hortalizas para incrementar el rendimiento. Las enzimas de la invención se pueden utilizar en el tratamiento enzimático (p. ej., hidrólisis de materiales vegetales que comprenden glucano) de diversos materiales derivados de pared celular vegetal o materiales de desecho, p. ej. de cereales, granos, producción de vino o zumo, o residuos agrícolas tales como cáscaras vegetales, cáscaras de haba, pasta de remolacha azucarera, pasta de aceituna, pasta de patata, y similares. Las enzimas de la invención se pueden utilizar para modificar la consistencia y la apariencia de las frutas u hortalizas elaboradas. Las enzimas de la invención se pueden utilizar para tratar material vegetal para facilitar la elaboración del material vegetal, incluyendo alimentos, facilitar la purificación o extracción de componentes vegetales. Las enzimas de la invención se pueden utilizar para mejorar el valor del pienso, disminuir la capacidad de retención de agua, mejorar la degradabilidad en plantas de tratamiento de aguas residuales y/o

mejorar la conversión de material vegetal para el ensilaje, y similares. Las enzimas de la invención se pueden utilizar también en la industria de la fruta y la elaboración de cerveza para la limpieza y el mantenimiento del equipamiento.

En un aspecto, las enzimas, p. ej., las glucanasas de la invención, se utilizan en aplicaciones para panificación, p. ej., galletas y galletas saladas, para hidrolizar glucanos y reducir la viscosidad. Las glucanasas de la invención también se pueden utilizar para crear masas no pegajosas que no sean difíciles de maquinar y para reducir el tamaño de las galletas. El uso de las enzimas de la invención para hidrolizar los glucanos se utiliza para prevenir la rehidratación rápida del producto horneado que da como resultado la pérdida de carácter crujiente y la reducción de la vida útil. En un aspecto, las enzimas de la invención se utilizan como aditivos en la elaboración de masas. En un aspecto, las enzimas de la invención se utilizan en el acondicionamiento de masas, donde en un aspecto las enzimas poseen alta actividad a lo largo de un intervalo de temperatura de alrededor de 25-35°C y a un pH casi neutro (7,0 – 7,5). En un aspecto, las enzimas para el acondicionamiento de masas se puede inactivar a las temperaturas extremas de panificación (>260°C).

Los procedimientos para el tratamiento de alimentos de la invención también pueden incluir el uso de una combinación de otras enzimas tales como catalasas, glucanasas, celulasas, endoglicosidasas, endo-beta-1,4-glucanasas, amiloglucosidasas, glucosa isomerasas, glicosiltransferasas, lipasas, fosfolipasas, lipooxigenasas, beta-glucanasas, endo-beta-1,3(4)-glucanasas, cutinasas, peroxidadas, lacasas, amilasas, glucoamilasas, pectinasas, reductasas, oxidasas, descarboxilasas, fenoloxidasas, ligninasas, pululanadas, arabinanasas, hemicelulasas, mananasas, xiloglucanasas, pectín acetil esteradas, ramnogalacturonano acetil esteradas, proteasas, peptidasas, proteinasas, poligalacturonasas, ramnogalacturonasas, galactanasas, pectato liasas, transglutaminasas, pectín metilesteradas, celobiohidrolasas y/o transglutaminasas.

Tratamiento de papel o pasta

Las glucanasas de la invención pueden estar en el tratamiento de pasta o destintado de papel. Por ejemplo, en un aspecto, la invención proporciona un procedimiento de tratamiento de papel en el que se utiliza una glucanasa de la invención. En un aspecto, una enzima de la invención es aplicable en la reducción de la necesidad de un agente blanqueante químico, tal como dióxido de cloro, y en entornos muy alcalinos y de alta temperatura. En un aspecto, la enzima de la invención es una glucanasa alcalina termoestable que puede efectuar una reducción de más de 25% del requerimiento de dióxido de cloro de la pasta kraft con una pérdida de rendimiento de pasta menor de 0,5%. En un aspecto, los parámetros límite son pH 10, 65-85°C y tiempo de tratamiento menor de 60 minutos a una carga de enzima de menos de 0,001% en peso. Se puede someter a ensayo una reserva de endoglucanasas para determinar la capacidad para hidrolizar el glucano marcado con colorante, por ejemplo, a pH 10 y 60°C. Las enzimas positivas para la prueba en estas condiciones se pueden evaluar a continuación, por ejemplo a pH 10 y 70°C. Alternativamente, las enzimas se pueden someter a ensayo a pH 8 y pH 10 a 70°C. Al descubrir las endoglucanasas deseables en la industria de la pasta y el papel, se eligieron como diana bibliotecas de entornos de alta temperatura o altamente alcalinos. Específicamente, estas bibliotecas se escrutaron para determinar las enzimas que funcionan a pH alcalino y a una temperatura de aproximadamente 45°C. En otro aspecto, las glucanasas de la invención son útiles en la industria de la pasta y el papel en la degradación de una conexión hemicelulosa de la lignina, con el fin de liberar la lignina.

Las glucanasas de la invención se pueden utilizar en la industria del papel y la pasta como se describe p. ej., en las Patentes de los Estados Unidos Núms. 5.661.021; 6.387.690; 6.083.733; 6.140.095 y 6.346.407. Por ejemplo, como en la Patente de los Estados Unidos Núm. 6.140.095, una enzima de la invención puede ser una glucanasa tolerante a los álcalis. Se puede utilizar una enzima de la invención en la industria del papel y la pasta donde la enzima es activa a un intervalo de temperatura de 65°C a 75°C y a un pH de aproximadamente 10. Adicionalmente, una enzima de la invención puede tener actividad a pH ligeramente ácido (5,5 - 6,0) en el intervalo de temperatura de 40°C a 70°C con inactivación a 95°C. En un aspecto una enzima de la invención tiene una actividad óptima entre 40 y 75°C, y pH de 5,5 a 6,0; es estable a 70°C durante al menos 50 minutos, y se inactiva a 96-100°C.

Adicionalmente, las glucanasas de la invención pueden ser útiles para el bioblanqueamiento y el tratamiento de pastas químicas, como se describe, p. ej., en la Patente de los Estados Unidos Núm. 5.202.249, el bioblanqueamiento y el tratamiento de pastas de madera y papel, como se describe p. ej., en las Patentes de los Estados Unidos Núms. 5.179.021, 5.116.746, 5.407.827, 5.405.769, 5.395.765, 5.369.024, 5.457.045, 5.434.071, 5.498.534, 5.591.304, 5.645.686, 5.725.732, 5.759.840, 5.834.301, 5.871.730 y 6.057.438, para reducir la lignina de la madera y modificar la madera, como se describe, por ejemplo, en las Patentes de los Estados Unidos Núms. 5.486.468 y 5.770.012.

En un aspecto, la enzima de la invención se utiliza en un procedimiento de blanqueo para aumentar el brillo de las pastas blanqueadas, p. ej., completamente o parcialmente a partir de madera blanda. Mediante el uso de una enzima de la invención, se puede reducir la cantidad de cloro utilizado en las fases de blanqueamiento.

Los procedimientos con pasta y papel de la invención pueden incluir también el uso de cualquier combinación de otras enzimas tales como catalasas, glucanasas, celulasas, endoglicosidasas, endo-beta-1,4-glucanasas, amiloglicosidasas, glucosa isomerasas, glicosiltransferasas, lipasas, fosfolipasas, lipooxigenasas, beta-glucanasas, endo-beta-1,3(4)-glucanasas, cutinasas, peroxidadas, lacasas, amilasas, glucoamilasas, pectinasas, reductasas, oxidadas, descarboxilasas, fenoloxidasas, ligninasas, pululanadas, arabinanasas, hemicelulasas, mananasas, xiloglucanasas, xilanasas, pectín acetil esteradas, ramnogalacturonano acetil esteradas, proteasas, peptidasas, proteinasas, poligalacturonasas, ramnogalacturonasas, galactanasas, pectato liasas, transglutaminasas, pectín metilesteradas, celobiohidrolasas y/o transglutaminasas.

10 **Piensos para animales y aditivos para alimentos o piensos**

Los métodos para el tratamiento de piensos y alimentos para animales y aditivos para alimentos o piensos pueden utilizar las glucanasas de la invención, para animales incluyendo mamíferos (p. ej., seres humanos), aves, peces y similares. La invención proporciona piensos, alimentos, y aditivos para animales que comprenden las glucanasas de la invención. En un aspecto, el tratamiento de piensos, alimentos y aditivos para animales para animales las utilizando glucanasas de la invención puede ayudar a la disponibilidad de los nutrientes, p. ej., almidón, proteínas, y similares, en los piensos o aditivos para animales. Rompiendo las proteínas difíciles de digerir o indirectamente o directamente desenmascarando el almidón (u otros nutrientes), la glucanasa hace los nutrientes más accesibles a otras enzimas endógenas o exógenas. La enzima de la invención también puede causar simplemente la liberación de nutrientes y azúcares rápidamente digeribles y fácilmente absorbidos. En otro aspecto, las enzimas de la invención se utilizan en piensos para disminuir la viscosidad de los glucanos en un alimento o un pienso, p. ej., una dieta con alto contenido de cebada o alto contenido de trigo, tal como una dieta para volatería. En un aspecto, esto puede minimizar la bajada de humedad.

25 Cuando se añaden a piensos para animales, las glucanasas de la invención mejoran la rotura *in vivo* del material de la pared celular vegetal debido parcialmente a la reducción de la viscosidad intestinal (véase, p. ej., Bedford et al., Proceedings of the 1st Symposium on Enzymes in Animal Nutrition, 1993, págs. 73-77), con lo que se logra una mejor utilización de los nutrientes vegetales por el animal. De este modo, utilizando las glucanasas de la invención en piensos se mejora la tasa de crecimiento y/o la proporción de conversión de pienso (es decir el peso de pienso ingerido con respecto a la ganancia de peso) del animal.

35 El aditivo para piensos para animales de la invención puede ser un producto enzimático granular que se puede remezclar fácilmente con los componentes del pienso. Alternativamente, aditivos para piensos pueden formar un componente de una pre-mezcla. El producto enzimático granular de la invención puede estar revestido o no revestido. El tamaño de partícula de los productos granulados enzimáticos puede ser compatible con el del pienso y los componentes de premezcla. Esto proporciona un medio seguro y conveniente para incorporar enzimas a piensos. Alternativamente, el aditivo para piensos para animales puede ser una composición líquida estabilizada. Esta puede ser una suspensión acuosa o con una base oleosa. Véase, p. ej., la Patente de los Estados Unidos Núm. 6.245.546.

40 Las glucanasas de la presente invención, en la modificación de piensos o alimentos para animales, pueden elaborar el alimento o pienso *in vitro* (modificando los componentes del pienso o alimento) o *in vivo*. Las glucanasas se pueden añadir a las composiciones de piensos o alimentos para animales que contienen altas cantidades de glucanos, p. ej. pienso o alimento que contiene material vegetal de cereales, granos y similares. Cuando se añade al pienso o alimento la glucanasa mejora significativamente la rotura *in vivo* del material que contiene glucano, p. ej., paredes de la célula vegetal, con lo que se logra una mejor utilización de los nutrientes vegetales por el animal (p. ej., ser humano). En un aspecto, se mejora la tasa de crecimiento y/o la proporción de conversión de pienso (es decir el peso de pienso ingerido con respecto a la ganancia de peso) del animal. Por ejemplo una proteína que comprende glucano parcialmente digerible o indigerible es degradada completamente o parcialmente por una glucanasa de la invención, p. ej. en combinación con otra enzima, p. ej., beta-galactosidasa, a péptidos y galactosa y/o galactooligómeros. Estos productos de digestión enzimática son mas digeribles para el animal. De este modo, las glucanasas de la invención pueden contribuir a la energía disponible del pienso o el alimento. Asimismo, mediante la contribución a la degradación de proteínas que comprenden glucano, una glucanasa de la invención puede mejorar la digeribilidad y la absorción de constituyentes del pienso o alimento carbohidratados y no carbohidratados tales como proteínas, grasa y minerales.

60 En otro aspecto, la glucanasa de la invención se puede suministrar expresando las enzimas directamente en los cultivos de pienso transgénico (en forma, p. ej., de plantas transgénicas, semillas y similares), tales como granos, cereales, maíz, haba de soja, semilla de colza, altramuza y similares. Como se ha comentado anteriormente, la invención proporciona plantas transgénicas, partes de plantas y células vegetales que comprenden una secuencia de ácido nucleico que codifica un polipéptido de la invención. En un aspecto, el ácido nucleico se expresa de manera que la glucanasa de la invención es producida en cantidades recuperables. Las glucanasas de la invención se pueden recuperar a partir de cualquier planta o parte de planta. Alternativamente, la planta o parte de planta que

contienen el polipéptido recombinante se pueden utilizar tal cual para mejorar la calidad de un alimento o pienso, p. ej., mejorando su valor nutritivo, palatabilidad, y propiedades reológicas, o para destruir un factor antinutritivo.

5 En un aspecto, la invención proporciona métodos para eliminar oligosacáridos del pienso antes de su consumo por un sujeto animal pueden utilizar una glucanasa de la invención. En este procedimiento se forma un pienso que tiene un valor energético metabolizable incrementado. Además de las glucanasas de la invención, se pueden utilizar galactosidasas, celulasas y combinaciones de las mismas. En un aspecto, la enzima se añade en una cantidad igual a entre alrededor de 0,1% y 1% en peso del material para pienso. En un aspecto, el pienso es un material de un cereal, un trigo, un grano, un haba de soja (p. ej., un haba de soja molida). Véase, p. ej., la Patente de los Estados Unidos Núm. 6.399.123.

10 En otro aspecto, la invención proporciona métodos para utilizar las glucanasas de la invención como suplemento nutricional en las dietas de animales preparando un suplemento nutricional que contiene una enzima glucanasa recombinante de la invención, y administrar el suplemento nutricional a un animal para incrementar la utilización del glucano contenido en el alimento ingerido por el animal.

15 En otro aspecto más, la invención proporciona una matriz de liberación de enzimas peletizada comestible y un método de uso para la liberación de glucanasas de la invención a un animal, por ejemplo en forma de suplemento nutricional. La matriz de liberación de enzimas libera fácilmente una enzima glucanasa, tal como una que tiene una secuencia de aminoácidos de la invención, o al menos 30 aminoácidos contiguos de la misma, en medios acuosos, tales como, por ejemplo, el fluido digestivo de un animal. La matriz de liberación de enzimas se prepara a partir de un portador comestible granulado seleccionado entre componentes tales como germen de grano que es residuo de la extracción de aceite, forraje, alfalfa, heno, cáscara de soja, harina de semilla de girasol y salvado de trigo, y similares, que dispersan fácilmente la enzima recombinante contenida allí en medios acuosos. Para su uso, la matriz de liberación de enzimas peletizada comestible se administra a un animal para la liberación de glucanasa al animal. Los sustratos basados en granos adecuados pueden comprender o estar derivados de cualquier grano comestible adecuado, tal como trigo, maíz, soja, sorgo, alfalfa, cebada, y similares. Un sustrato a base de granos ilustrativo es un sustrato a base de maíz. El sustrato se puede obtener de cualquier parte adecuada del grano, pero es preferiblemente un germen de grano aprobado para su uso en piensos para animales, tal como germen de maíz que se obtiene en un procedimiento de molienda en mojado o en seco. El germen de grano comprende preferiblemente un germen residuo de extracción, que es un germen de grano del cual se ha expulsado el aceite, por ejemplo mediante presión o extracción con hexano u otro disolvente. Alternativamente, el germen de grano es extraído expeliéndolo, esto es, el aceite ha sido separado por presión.

20 La matriz de liberación de enzimas está en forma de partículas plurales discretas, peletes o gránulos. Mediante "gránulos" se quieren significar partículas que son comprimidas o compactadas, por ejemplo mediante peletización, extrusión, o compactación similar para eliminar el agua de la matriz. Tal compresión o compactación de las partículas también promueve la cohesión intrapartícula de las partículas. Por ejemplo, los gránulos se pueden preparar peletizando en sustrato o base de grano en un molino peletizador. Los peletes preparados de ese modo se muelen o desmenuzan hasta un tamaño de gránulo adecuado para su uso como coadyuvante en piensos para animales. Puesto que la propia matriz está aprobada para su uso en piensos para animales, ésta se puede utilizar como diluyente para la liberación de enzimas en piensos para animales.

25 En un aspecto, la matriz de liberación de enzimas está en forma de gránulos que tienen un tamaño de gránulo que oscila de una malla de alrededor de 4 a alrededor de 400 (USS); en un aspecto más, una malla de alrededor de 8 a alrededor de 80; y en otro aspecto adicional una malla de alrededor de 14 a alrededor de 20. Si el germen de grano es extraído mediante extracción con disolvente, puede ser necesario el uso de un agente lubricante tal como aceite de maíz en el peletizador, pero tal agente lubricante normalmente no es necesario si el germen es extraído expeliéndolo. En otros aspectos de la invención, la matriz se prepara mediante otros procedimientos de compactación o compresión tales como, por ejemplo, extrusión del sustrato a base de grano a través de un troquel y molienda del producto extrudido hasta un tamaño de gránulo adecuado.

30 La matriz de liberación de enzimas puede incluir adicionalmente un componente polisacárido como agente de cohesión para mejorar la cohesión de los gránulos de la matriz. Se cree que el agente de cohesión proporciona grupos hidroxilo adicionales, que potencia la unión entre las proteínas del grano dentro del gránulo de la matriz. Se cree adicionalmente que los grupos hidroxilo adicionales funcionan así potenciando los puentes de hidrógeno de las proteínas al almidón y a otras proteínas. El agente de cohesión puede estar presente en cualquier cantidad adecuada para potenciar la cohesión de los gránulos de la matriz de liberación de enzimas. Los agentes de cohesión adecuados incluyen uno o más de dextrinas, maltodextrinas, almidones, tales como almidón de maíz, harinas, derivados celulósicos, hemicelulósicos, y similares. Por ejemplo, el porcentaje de germen de grano y de agente de cohesión en la matriz (sin incluir la enzima) es 78% de harina de germen de maíz y 20% en peso de almidón de maíz.

Puesto que la matriz liberadora en enzimas de la invención se elabora a partir de materiales biodegradables, la matriz puede ser sujeto de degradación, por ejemplo mediante la proliferación de moho. Para prevenir o inhibir tal proliferación del moho, la matriz puede incluir un inhibidor del moho, tal como una sal propionato, que puede estar presente en cualquier cantidad suficiente para inhibir la proliferación del moho en la matriz liberadora de enzimas, proporcionando de ese modo una matriz de liberación en una formulación estable que no requiere refrigeración.

La enzima glucanasa contenida en la matriz y métodos de liberación de enzimas de la invención es preferiblemente una glucanasa termoestable, como se ha descrito en la presente memoria, de manera que resiste la inactivación de la glucanasa durante fabricación donde se pueden emplear temperatura elevadas y/o vapor de agua para preparar matriz de liberación de enzimas paletizada. Durante la digestión del pienso que contiene la matriz de liberación de enzimas, los fluidos digestivo acuosos causarán la liberación de la enzima activa. También se pueden incorporar otros tipos de enzimas termoestables y suplementos nutricionales que son termoestables a la matriz de liberación para su liberación en cualquier tipo de condiciones acuosas.

Se puede aplicar un revestimiento a las partículas de la matriz enzimática para muchos diferentes propósitos, tales como añadir un sabor o suplemento nutricional a los piensos para animales, retrasar la liberación de los suplementos y enzimas de los piensos para animales en condiciones gástricas, y similares. O, el revestimiento se puede aplicar para lograr un objetivo funcional, por ejemplo, siempre que sea deseable ralentizar la liberación de la enzima desde las partículas de la matriz o para controlar las condiciones bajo las que se liberará la enzima. La composición del material de revestimiento puede ser una que sea disgregada selectivamente por un agente al que sea susceptible (tal como calor, ácido o base, enzimas u otros agentes químicos). Alternativamente, se pueden aplicar consecutivamente a las partículas de la matriz dos o más revestimientos susceptibles a tales agentes disgregantes diferentes.

Un procedimiento para la preparación de una matriz liberadora de enzimas comprende proporcionar partículas plurales discretas de un sustrato a base de grano con un tamaño de partícula adecuado para su uso como una matriz liberadora de enzimas, donde las partículas comprenden una enzima glucanasa codificada por una secuencia de aminoácidos de la invención. En un aspecto, el procedimiento incluye compactar o comprimir las partículas de la matriz liberadora de enzimas para formar gránulos, lo que muy preferiblemente se logra mediante peletización. El inhibidor de moho y el agente de cohesión, cuando se utilizan, se pueden añadir en cualquier momento adecuado, y en un aspecto se mezclan con el sustrato a base de grano en las proporciones deseadas antes de la peletización del sustrato a base de grano. El contenido de humedad del pienso en el molino peletizador se encuentra preferiblemente en los intervalos mostrados anteriormente con respecto al contenido de humedad en el producto finalizado, y en un aspecto es de alrededor de 14-15%. En un aspecto, se añade humedad a la materia prima en forma de una preparación acuosa de la enzima para llevar la materia prima a este contenido de humedad. La temperatura en el molino peletizador en un aspecto se lleva a alrededor de 82°C con vapor de agua. El molino peletizador se puede manejar en cualquier condición que confiera suficiente actividad a la materia prima para proporcionar peletes. El propio procedimiento de peletización es un procedimiento económico para eliminar el agua de la composición que contiene enzima.

En un aspecto, el molino peletizador se maneja con una matriz de 3,2 mm por 51 mm (1/8 pulgadas por 2 pulgadas) a una presión de 45 kg/min. (100 lb./min.) a 82°C para proporcionar peletes, que a continuación se desmenuzan en un molino peletizador desmenuzador para proporcionar partículas plurales discretas que tienen un tamaño de partícula susceptible de pasar a través de un tamiz de malla 8 pero que quedan retenidas en un tamiz de malla 20.

Las glucanasas termoestables de la invención se pueden utilizar en los peletes de la invención. Pueden tener temperaturas óptimas altas y resistencia térmica alta de manera que se puede lograr una reacción enzimática a una temperatura no llevada a cabo hasta ahora. Se puede utilizar el gen codificante de la glucanasa de acuerdo con la presente invención (p. ej. como se muestra en cualquiera de las secuencias de la invención) en la preparación de glucanasas (p. ej. utilizando GSSM™ como se ha descrito en la presente memoria) que tienen características diferentes de las de las glucanasas de la invención (en términos de pH óptimo, temperatura óptima, resistencia térmica, estabilidad frente a disolventes, actividad específica, afinidad con el sustrato, capacidad de secreción, velocidad de traducción, control de la transcripción y similares). Además, se puede emplear un polinucleótido de la invención para el escrutinio de glucanasas variantes preparadas mediante los métodos descritos en la presente memoria para determinar aquellas que tienen una actividad deseada, tal como termoestabilidad o termotolerancia mejoradas o modificadas. Por ejemplo, la Patente de los Estados Unidos Núm. 5.830.732, describe un análisis de escrutinio para determinar la termotolerancia de una glucanasa.

En un aspecto, las glucanasas de la invención en piensos para animales son activas en el estómago de los animales. De este modo, en un aspecto, una enzima de la invención, p. ej., en un pienso, tiene actividad a aproximadamente 37°C y a bajo pH para animales monogástricos (pH 2-4) y a pH casi neutro para rumiantes (pH 6,5-7). La enzima de la invención tiene resistencia a las enzimas intestinales del animal, p. ej., proteasas, y estabilidad a las temperaturas más altas implicadas en la peletización del pienso. En un aspecto, las glucanasas de la invención se utilizan en aditivos para piensos, p. ej., piensos para animales monogástricos, y pueden tener una

alta actividad específica, p. ej., actividad a 35-40°C y pH 2-4, una vida media superior a 30 minutos en SGF y una vida media > 5 minutos a 85°C en estado formulado. Para el pienso de rumiantes, las glucanasas de la invención en los aditivos para piensos tienen una actividad específica. p. ej., actividad a 35-40°C y pH de 6,5-7,5, vida media superior a 30 minutos en SRF y estabilidad en forma de polvo seco concentrado.

El pienso para animales y los procedimientos de producción de pienso para animales de la invención pueden incluir cualquier combinación de otras enzimas tales como catalasas, otras glucanasas, celulasas, endoglicosidasas, endo-beta-1,4-glucanasas, amiloglicosidasas, glucosa isomerasas, glicosiltransferasas, lipasas, fosfolipasas, lipooxigenasas, beta-glucanasas, endo-beta-1,3(4)-glucanasas, cutinasas, peroxidasas, lacasas, amilasas, glucoamilasas, pectinasas, reductasas, oxidasas, descarboxilasas, fenoloxidasas, ligninasas, pululanasas, fitasas, arabinanasas, hemicelulasas, otras mananasas, xiloglucanasas, xilanasas, pectín acetil esterases, ramnogalacturonano acetil esterases, poligalacturonasas, ramnogalacturonasas, galactanasas, pectato liasas, transglutaminasas, pectín metilesterases, celobiohidrolasas y/o transglutaminasas.

Tratamiento de Residuos

Las glucanasas de la invención se pueden utilizar en una variedad de aplicaciones industriales diferentes, p. ej., en el tratamiento de residuos (además de, p. ej., la conversión de biomásas en combustibles). Por ejemplo, en un aspecto, la invención proporciona un procedimiento de digestión de residuos sólidos utilizando glucanasas de la invención. Los métodos pueden comprender la reducción de la masa y el volumen de residuos sólidos sustancialmente no tratados. Los residuos sólidos se pueden tratar mediante un procedimiento digestivo enzimático en presencia de una solución enzimática (incluyendo glucanasas de la invención) a una temperatura controlada. Esto da como resultado una reacción sin fermentación bacteriana apreciable de los microorganismos añadidos. El residuo sólido se convierte en un residuo licuado y ningún residuo sólido restante. El residuo líquido resultante se puede separar de dicho cualquier residuo sólido restante. Véase p. ej., la Patente de los Estados Unidos Núm. 5.709.796.

Los procedimientos de tratamiento de residuos de la invención pueden incluir el uso de cualquier combinación de otras enzimas tales como catalasas, otras glucanasas, celulasas, endoglicosidasas, endo-beta-1,4-glucanasas, amiloglicosidasas, glucosa isomerasas, glicosiltransferasas, lipasas, fosfolipasas, lipooxigenasas, beta-glucanasas, endo-beta-1,3(4)-glucanasas, cutinasas, peroxidasas, lacasas, amilasas, glucoamilasas, pectinasas, reductasas, oxidasas, descarboxilasas, fenoloxidasas, ligninasas, pululanasas, fitasas, arabinanasas, hemicelulasas, otras mananasas, xiloglucanasas, xilanasas, pectín acetil esterases, ramnogalacturonano acetil esterases, proteasas, peptidasas, proteinasas, poligalacturonasas, ramnogalacturonasas, galactanasas, pectato liasas, transglutaminasas, pectín metilesterases, celobiohidrolasas y/o transglutaminasas.

Productos para higiene bucal

Los productos para higiene bucal pueden comprender glucanasas de la invención. Los productos para la higiene bucal ilustrativos incluyen pastas dentales, cremas dentales, geles o polvos dentales, odonticos, colutorios, formulaciones de enjuague post- o pre-cepillado, chicles, grageas, o caramelos. Véase, p. ej., la Patente de los Estados Unidos Núm. 6.264.925.

Los productos orales de la invención pueden incluir cualquier combinación de otras enzimas tales como proteasas, peptidasas, proteinasas, glucosa oxidasas, peroxidasas, glucanasas, celulasas, endoglicosidasas, endo-beta-1,4-glucanasas, amiloglicosidasas, endo-beta-1,3(4)-glucanasas, amiloglicosidasas y glucosidasas.

Elaboración y fermentación de cerveza

Los métodos para elaborar (p. ej., fermentar) cerveza pueden comprender glucanasas de la invención. En un procedimiento ilustrativo, las materias primas que contienen almidón se disgregan y se procesan para formar una malta. Se utiliza una enzima de la invención en cualquier momento del procedimiento fermentación. Las glucanasas de la invención se pueden utilizar en la industria cervecera para la degradación de beta-glucanos. Las glucanasas de la invención se utilizan en la industria cervecera para el aclarado de la bebida.

En un aspecto, las glucanasas de la invención se pueden utilizar en el procesamiento de la malta de la cebada. La materia prima principal para la elaboración de cerveza es la malta de cebada. Esto puede consistir en un procedimiento de tres etapas. Primero, el grano de cebada se puede empapar para incrementar el contenido de agua, p. ej., hasta alrededor de 40%. Segundo, se puede hacer germinar el grano mediante incubación de 1 a 25°C durante 3 a 6 días cuando la síntesis enzimática se estimula el control de giberelinas. En un aspecto, las enzimas de la invención se añaden en esta (o cualquier otra) etapa del procedimiento.

En un aspecto, las enzimas de la invención se utilizan en procedimientos de maceración y conversión. En las industrias de elaboración de cerveza y fermentación, los procedimientos de maceración y conversión se realizan a temperaturas que son demasiado bajas para promover la degradación adecuada de los glucanos y xilanos solubles

en agua. Estos polímeros forman sustratos gomosos que pueden causar un aumento de la viscosidad en el mosto en maceración, dando como resultado flujo de maceración, turbidez residual y precipitados mayores en el producto de cerveza final debido a la filtración ineficiente y al bajo rendimiento de extracción. Por estas razones, las enzimas se añaden durante los procedimientos de fabricación de cerveza para romper los glucanos unidos mediante enlaces β -1,4- u β -1,3.

En un aspecto, las enzimas de la invención se utilizan en operaciones de maltería, p. ej., la glucanasa se añade al agua de procesamiento, para acortar los tiempos de germinación y/o para fomentar la conversión de cebada de baja calidad en maltas aceptables. En un aspecto, las enzimas de la invención se utilizan para la maceración, p. ej., se añaden para aumentar la filtrabilidad del mosto y/o para mejorar la clarificación del mismo. En un aspecto, las enzimas de la invención se utilizan en un fermentador y/o tanque de sedimentación para, p. ej., ayudar a aclarar la turbidez y/o mejorar la filtración. En un aspecto, las enzimas de la invención se utilizan como complemento en la fabricación de cerveza, p. ej., una glucanasa de la invención se añade para romper los glucanos de la cebada, el trigo, y/o otros cereales, incluyendo los glicanos de la malta. En un aspecto, las enzimas de la invención se utilizan para la elaboración de cerveza a partir de malta, p. ej., se añade una glucanasa de la invención para modificar las maltas de baja calidad con alto contenido de glucano.

Las glucanasas de la invención se pueden utilizar en cualquier procedimiento de producción de cerveza o bebidas alcohólicas, como se describe, p. ej., en la Patente de los Estados Unidos Núm. 5.762.991; 5.536.650; 5.405.624; 5.021.246; 4.788.066.

Los procedimientos de elaboración de cerveza de la invención pueden incluir el uso de cualquier combinación de otras enzimas tales como otras xilanasas, esterasas, celulasas, pectinasas, pectato liasas, amilasas, descarboxilasas, lacasas, glucanasas, proteasas, peptidasas, proteinasas, amiloglucosidasas, glucosa isomerasas, glucoamilasas, beta-glucanasas, endo-beta-1,3(4)-glucanasas, hemicelulasas, endoglicosidasas, endo-beta-1,4-glucanasas, glicosiltransferasas, fosfolipasas, lipooxigenasas, reductasas, oxidasas, fenoloxidasas, ligninasas, pululaninas, arabinanasas, otras mananasas, xiloglucanasas, pectín acetil esterasas, ramnogalacturonano acetil esterasas, poligalacturonasas, ramnogalacturonasas, galactanasas, transglutaminasas, pectín metilesterasas, celobiohidrolasas y/o transglutaminasas.

Aplicaciones médicas y de investigación

Las glucanasas de la invención se pueden utilizar como agentes antimicrobianos debido a sus propiedades bacteriolíticas y propiedades anti-fúngicas. Las glucanasas de la invención se pueden utilizar para eliminar o proteger animales de la salmonela, como se describe p. ej., en las Solicitudes PCT Núms. WO 0049890 y WO 9903497. Las glucanasas de la invención se pueden utilizar en un método de uso y una composición de una carboxilasa y/o una glucanasa para la fabricación de un agente para tratamientos y/o profilaxis de las coccidiosis. El agente manufacturado puede estar en forma de un pienso para animales basado en cereales (véase, por ejemplo, la Patente de los Estados Unidos Núm. 5.624.678).

Aplicaciones de perforación

Las glucanasas de la invención se pueden utilizar para modificar la viscosidad de un material derivado de plantas. En un aspecto, las enzimas de la invención se utilizan en la industria del petróleo en la que se utilizan goma guar y guar modificada, p. ej., para fracturar líquidos y lodos de perforación. Las enzimas de la invención se pueden utilizar para limpiar pozos de petróleo, p. ej. para romper la estructura de alta viscosidad o de gel en el fluido fractural después de la fractura. En un aspecto, las enzimas de la invención utilizadas en estas aplicaciones tienen una alta estabilidad térmica. En un aspecto, las enzimas de la invención utilizadas en estas aplicaciones son resistentes a las elevadas temperaturas del terreno o generadas en los procedimientos de perforación. Las glucanasas de la invención se pueden utilizar para tratar el lodo de perforación (p. ej., el lodo utilizado).

Otras aplicaciones industriales

Las glucanasas de la invención se pueden utilizar en una amplia gama de aplicaciones para alimentos, piensos para animales y bebidas. Se descubren nuevas glucanasas mediante el escrutinio de bibliotecas existentes y bibliotecas de ADN construidas a partir de diversas localizaciones mesófilas y moderadamente termófilas así como de fuentes elegidas como diana incluyendo flora digestiva, microorganismos en residuos animales, bacterias de suelo y hábitats altamente alcalinos. También son útiles las biotrasas y las estrategias de enriquecimiento primarias que utilizan sustratos de arabinoglucano y/o fracciones de polisacáridos no solubles de material de pienso para animales.

Las glucanasas de la invención se pueden utilizar en la conversión de biomasa en combustibles, y en la producción de etanol, p. ej., como se describe en las Solicitudes PCT Núms. WO 0043496 y WO 8100857. Las glucanasas de la invención se pueden utilizar para producir azúcares fermentables y biomasa que contiene glucano que se pueden convertir en etanol combustible.

Las glucanasas de la invención se pueden utilizar combinadas con otras enzimas implicadas en la digestión de celulosa como las celobiohidrolasas y las betaglucosidasas.

5 Las glucanasas de la invención se pueden utilizar en otras numerosas aplicaciones. Por ejemplo, las glucanasas de la invención se pueden utilizar para mejorar la calidad y la cantidad de la producción de proteína láctea en vacas en período de lactación (véase, por ejemplo, Kung, L., et al., J. Dairy Science, Enero de 2000 83: 115-122), aumentar la cantidad de sacáridos solubles en el estómago y el intestino delgado de cerdos (véase, por ejemplo, van der Meulen, J. et al., Arch. Tierernahr, 2001 54:101-115), mejorando la eficacia en la producción de huevos tardíos y los rendimientos de huevos en gallinas (véase, por ejemplo, Jaroni, D., et al., Poult Sci., Junio de 1999 78:841-847).
10 Adicionalmente las glucanasas de la invención se pueden utilizar como mejoradores de harina, masa y pan (véanse, por ejemplo, las Patentes de los Estados Unidos Núms. 5.108.765 y 5.306.633) como aditivos y/o suplementos para piensos, como se ha mostrado anteriormente (véanse, por ejemplo, las Patentes de los Estados Unidos Núms. 5.432.074, 5.429.828, 5.612.055, 5.720.971, 5.981.233, 5.948.667, 6.099.844, 6.132.727 y 6.132.716), en la fabricación de soluciones de celulosa (véase, por ejemplo, la Patente de los Estados Unidos Núm. 5.760.211). Las composiciones detergentes que comprenden glucanasas de la invención puede ser útiles para compuestos de frutas, hortalizas y/o barro y arcilla (véase, por ejemplo, la Patente de los Estados Unidos Núm. 5.786.316).

20 Los usos adicionales para las glucanasas de la invención incluyen su uso en la producción de fibra dietética soluble en agua (véase, por ejemplo, la Patente de los Estados Unidos Núm. 5.622.738), para mejorar la filtrabilidad, la separación y la producción de almidón (véanse, por ejemplo, las Patentes de los Estados Unidos Núms. 4.960.705 y 5.023.176), en la industria de elaboración de bebidas para mejorar la filtrabilidad del mosto o la cerveza (véase, por ejemplo, la Patente de los Estados Unidos Núm. 4.746.517), en una composición enzimática para promover la secreción de por el ganado y para mejorar la calidad de la leche (véase, por ejemplo, la Patente de los Estados Unidos Núm. 4.144.354), para reducir la viscosidad de un material vegetal (véase, por ejemplo, la Patente de los Estados Unidos Núm. 5.874.274), para mejorar la viscosidad o firmeza de gel de de productos alimenticios tales como confitura, mermelada, jalea, zumo, pasta, sopa, salsa, etc. (véase, por ejemplo, la Patente de los Estados Unidos Núm. 6.036.891). Las glucanasas también se pueden utilizar en la hidrólisis de hemicelulosa para la cual es selectiva, concretamente en presencia de celulosa. Adicionalmente, el producto retenido (retentato) rico en celulosa es adecuado para la hidrólisis de celulosa (véase, por ejemplo, la Patente de los Estados Unidos Núm. 4.725.544).
30

Diversos usos de las glucanasas de la invención incluyen la transformación de un microbio que produce etanol (véase, por ejemplo, la Solicitud PCT Núm. WO99/46362), en la producción de taninos enológicos y composiciones enzimáticas (véase, por ejemplo, la Solicitud PCT Núm. WO0164830), en la estimulación de las defensas naturales de las plantas (véase, por ejemplo, la Solicitud PCT Núm. WO0130161), en la producción de azúcares a partir de sustratos de hemicelulosa (véase, por ejemplo, la Solicitud PCT Núm. WO9203541), en la limpieza de manchas que contienen fruta, hortalizas, barro o arcilla (véase, por ejemplo, la Solicitud PCT Núm. WO9613568), en la limpieza de membranas de filtración de cerveza (véase, por ejemplo, la Solicitud PCT Núm. WO9623579), en un método para eliminar o inhibir células microbianas (véase, por ejemplo, la Solicitud PCT Núm. WO9732480) y para determinar las características de las aguas de tratamiento del blanqueamiento de la pasta de madera mediante la utilización de las proporciones de dos mediciones de absorción UV y la comparación de los espectros (véase, por ejemplo, la Solicitud PCT Núm. WO9840721).
35 40

Cualquier producto o procedimiento de la invención puede incluir cualquier combinación de otras enzimas tales como catalasas, glucanasas, celulasas, endoglicosidasas, endo-beta-1,4-glucanasas, amiloglicosidasas, glucosa isomerasas, glicosiltransferasas, lipasas, estererasas, fosfolipasas, lipooxigenasas, beta-glucanasas, endo-beta-1,3(4)-glucanasas, cutinasas, peroxidadas, lacasas, amilasas, glucoamilasas, pectinasas, reductasas, oxidasas, descarboxilasas, fenoloxidasas, ligninasas, pululanadas, fitasas, arabinanasas, hemicelulasas, mananasas, xiloglucanasas, xilanasas, pectín acetil estererasas, ramnogalacturonano acetil estererasas, poligalacturonasas, ramnogalacturonasas, galactanasas, pectato liasas, transglutaminasas, pectín metilesterasas, celobiohidrolasas y/o transglutaminasas.
45 50

Se utilizan dos formatos de escrutinio (basados en la actividad y basados en la secuencia) en el descubrimiento de glucanasas novedosas. El enfoque basado en la actividad es un escrutinio directo para la actividad glucanasa en placas de agar utilizando un sustrato tal como beta glucano de cebada AZO (Megazyme). Alternativamente se puede utilizar un enfoque basado en la secuencia, que cuenta con la bioinformática y la biología molecular para diseñar sondas para hibridación y bioinmunopurificación. Véanse, por ejemplo, las Patentes de los Estados Unidos Núms. 6.054.267, 6.030.779, 6.368.798, 6.344.328. Los éxitos del escrutinio se purifican, se secuencian, se caracterizan (por ejemplo, determinando la especificidad, la temperatura y el pH óptimo), se analizan utilizando bioinformática, se subclonan y se expresan para determina su caracterización bioquímica. Estos métodos se pueden utilizar en el escrutinio de glucanasas útiles en innumerables aplicaciones, incluyendo acondicionamiento de masas y como enzimas adicionales de piensos para animales.
55 60

Al caracterizar las enzimas obtenidas de un escrutinio, se puede evaluar la utilidad ilustrativa en el procesamiento de masas y aplicaciones para panificación. La caracterización puede incluir, por ejemplo, la medición de la especificidad del sustrato (glucano, arabinoglucano, CMC, BBG), temperatura y estabilidad de pH y la actividad específica. Se puede utilizar una enzima comercial como referencia. En un aspecto, las enzimas de la invención actividad significativa a $\text{pH} \geq 7$ y $25\text{-}35^\circ\text{C}$, son inactivas en glucano insoluble, son estables y activas en sacarosa al 50-67%.

En otro aspecto, la utilidad como aditivos para pienso se puede evaluar a partir de la caracterización de enzimas candidato. La caracterización puede incluir, por ejemplo, medición de la especificidad de sustrato (glucano, CMC, B β G), temperatura y estabilidad de pH, actividad específica y estabilidad gástrica. En un aspecto el pienso se diseña para un animal monogástrico y en otro aspecto el pienso se diseña para un animal rumiante. En un aspecto, las enzimas de la invención tienen actividad significativa a $\text{pH} 2\text{-}4$ y $35\text{-}40^\circ\text{C}$, una vida media superior a 30 minutos en fluido gástrico, vida media en formulación (en tampón o células) superior a 5 minutos a 85°C y se utilizan en forma de un aditivo de pienso para animales monogástricos. En otro aspecto, las enzimas de la invención tienen una o más de las siguientes características: actividad significativa a $\text{pH} 6,5\text{-}7,0$ y $35\text{-}40^\circ\text{C}$, una vida media superior a 30 minutos en el fluido del rumen, estabilidad de formulación tan estable como el polvo seco y se utilizan en forma de un aditivo de pienso para animales rumiantes.

Las enzimas son reactivas con una amplia gama de sustratos naturales y no naturales, posibilitando de ese modo la modificación virtualmente de cualquier compuesto inicial orgánico. Por otra parte, a diferencia de los catalizadores químicos tradicionales, las enzimas son altamente enantio- y regio-selectivas. El alto grado de especificidad de grupo funcional mostrado por las enzimas posibilita el seguimiento de cada reacción en una secuencia sintética que conduce a un nuevo compuesto activo. Las enzimas también son capaces de catalizar muchas reacciones diversas no relacionadas con su función fisiológica en la naturaleza. Por ejemplo, las peroxidasas catalizan la oxidación de fenoles con peróxido de hidrógeno. Las peroxidasas pueden catalizar también reacciones de hidroxilación que no están relacionadas con la función nativa de la enzima. Otros ejemplos son las glucanasas que catalizan la rotura de polipéptidos. En solución orgánica algunas glucanasas pueden también acilar azúcares, una función no relacionada con la función nativa de estas enzimas.

La presente invención explota las propiedades catalíticas únicas de las enzimas. Si bien el uso de biocatalizadores (es decir, enzimas purificadas o brutas, células vivas o no vivas) en transformaciones químicas requiere normalmente de la identificación de un biocatalizador concreto que reacciona con un compuesto de partida específico, la presente invención utiliza biocatalizadores seleccionados y condiciones de reacción que son específicas de los grupos funcionales que están presentes en muchos compuestos de partida. Cada biocatalizador es específico de un grupo funcional, o varios grupos funcionales relacionado y puede reaccionar con muchos compuestos de partida que contienen este grupo funcional. Las reacciones biocatalíticas producen una población de derivados a partir de un solo compuesto de partida. Estos derivados se pueden someter a otra ronda de reacciones biocatalíticas para producir una segunda población de compuestos derivado. Se pueden producir miles de variaciones del compuesto original con cada iteración de la derivatización biocatalítica.

Las enzimas reaccionan en sitios específicos de un compuesto de partida sin afectar al resto de la molécula, un procedimiento que es muy difícil de alcanzar utilizando métodos químicos tradicionales. Este alto grado de especificidad biocatalítica proporciona el medio para identificar un solo compuesto activo en la biblioteca. La biblioteca está caracterizada por la serie de reacciones biocatalíticas utilizadas para producirlo, una denominada "historial biosintético". El escrutinio de la biblioteca para determinar las actividades biológicas y el seguimiento del historial biosintético identifica la secuencia de reacciones específica que produce el compuesto activo. La secuencia de reacciones se repite y se determina la estructura del compuesto sintetizado. Este modo de identificación, a diferencia de otros enfoques de síntesis y escrutinio, no requiere tecnologías de inmovilización y los compuestos se pueden sintetizar y someter a ensayo libres en solución utilizando virtualmente cualquier tipo de análisis de escrutinio. Resulta importante observar, que el alto grado de especificidad de las reacciones enzimáticas sobre los grupos funcionales permite el "seguimiento" de reacciones enzimáticas específicas que constituyen la biblioteca producida biocatalíticamente.

Muchas de las etapas procedimentales se realizan utilizando automatización robótica lo que permite la ejecución de muchos miles de reacciones biocatalíticas y análisis de escrutinio por día así como asegura un alto nivel de exactitud y reproducibilidad. Como resultado, se puede producir una biblioteca de compuestos derivados en cuestión de semanas que podría tardar años en producirse utilizando los métodos químicos actuales. (Para enseñanzas adicionales sobre la modificación de moléculas, incluyendo moléculas pequeñas, véase el documento PCT/US94/09174).

La invención se describirá adicionalmente con referencia a los siguientes Ejemplos; no obstante, se debe entender que la invención no está limitada a tales Ejemplos.

EJEMPLOS

Ejemplo 1: Descubrimiento de la enzima endoglicosidasa basado en el cultivo en placa: escrutinio de expresión

5 El siguiente ejemplo demuestra en aislamiento y la confirmación de la actividad enzimática de enzimas y ácidos nucleicos ilustrativos de la invención. Estos análisis también se pueden utilizar para determinar si un polipéptido tiene la actividad de la enzima requisito (p. ej., glucanasa) para estar dentro del alcance de la invención.

10 Determinación del título de la biblioteca Lambda: Añadir 1,0 µL de provisión de partida de biblioteca amplificada Lambda Zap Express a 600 µL de células MRF' de *E. coli* ($DO_{600} = 1,0$). Diluir la provisión de MRF' con $MgSO_4$ 10 mM. Incubar la mezcla a 37°C durante 15 minutos, a continuación transferir la suspensión a 5-6 mL de agar NZY Top a 50°C y mezclar suavemente. Verter inmediatamente la solución de agar sobre placas grandes (150 mm) de medio NZY y permitir que el agar Top se solidifique completamente (aproximadamente 30 minutos). Invertir la placa. Incubar la placa a 39°C durante 8-12 horas. (El número de placas es aproximado. Determinar el título de fago para producir 50.000 pfu/placa. Diluir una alícuota de la Biblioteca de fagos con tampón SM si fuera necesario.)

15 Escrutinio de sustrato: Añadir Lambda Zap Express (50.000 pfu) de la biblioteca amplificada a 600 µL de células MRF' de *E. coli* ($DO_{600} = 1,0$) e incubar a 37°C durante 15 minutos. Mientras se incuba la suspensión de fagos/células, añadir 1,0 mL del sustrato de polisacárido marcado con colorante deseado (usualmente 1-2% p/v) a 5,0 mL de agar NZY top a 50°C y mezclar cuidadosamente. (Mantener la solución a 50°C hasta que se necesite). Transferir la suspensión celular a una solución de sustrato/agar top y mezclar suavemente. Verter inmediatamente la solución sobre placas grandes (150 mm) de medio NZY. Permitir que el agar Top se solidifique completamente (aproximadamente 30 minutos), a continuación invertir la placa. Incubar la placa a 39°C durante 8-12 horas. Observar las zonas aclaradas (halos) alrededor de las placas. Quitar el núcleo de las placas con halos fuera del agar y transferirlas a microtubos estériles. (Una punta de pipeta grande con orificio de 200 µL funciona bien para retirar (quitar el núcleo) el tapón de agar que contiene la placa deseada.) Resuspender el fago en 500 µL de tampón SM. Añadir 20 µL de cloroformo para inhibir cualquier crecimiento celular adicional.

20 Aislamiento de clones puros: Añadir 5 µL de la suspensión de fagos resuspendida a 500 µL de células MRF' de *E. coli* ($DO_{600} = 1,0$). Incubar a 37°C durante 15 minutos. Mientras se está incubando la suspensión de fagos/células, añadir 600 µL de sustrato de polisacárido marcado con colorante deseado (usualmente 1-2% p/v) a 3,0 mL de agar NZY top a 50°C y mezclar cuidadosamente. (Mantener la solución a 50°C hasta que se necesite). Transferir la suspensión celular a la solución de sustrato/agar top y mezclar suavemente. Verter inmediatamente la solución sobre placas pequeñas (90 mm) de medio NZY y permitir que el agar top se solidifique completamente (aproximadamente 30 minutos), a continuación invertir la placa. Incubar la placa a 39°C durante 8-12 horas. Observar una zona aclarada (halo) de la placa alrededor de una sola placa (clon puro). (Si no se puede aislar una sola placa, ajustar el título y volver a cultivar en placa la suspensión de fagos). Los fagos se resuspenden en 500 µL de tampón SM y se añaden 20 µL de cloroformo para inhibir cualquier crecimiento celular adicional.

30 Escisión de clones puros: Permitir que la suspensión de fagos puros se incuba a temperatura ambiente durante 2 a 3 horas o durante la noche a 4°C. Añadir 100 µL de la suspensión de fagos puros a 200 µL de células MRF' de *E. coli* ($DO_{600} = 1,0$). Añadir 1,0 µL de fago coadyuvante ExAssist ($>1 \times 10^6$ pfu/mL; Stratagene). Incubar la suspensión a 37°C durante 15 minutos. Añadir 3,0 mL de medio 2 x YT a la suspensión celular. Incubar a 37°C durante 2-2.5 horas con sacudimiento. Transferir a un tubo a 70°C durante 20 minutos. Transferir 50-100 µL de la suspensión de fegémidos a un microtubo que contiene 200 µL de células Exp 505 de *E. coli* ($DO_{600} = 1,0$). Incubar la suspensión a 37°C durante 45 minutos. Cultivar en placa 100 µL de suspensión celular en medio $LB_{kan 50}$ (medio LB con 50 µg/mL de Kanamicina). Incubar la placa a 37°C durante 8-12 horas. Observar la placa para determinar las colonias. Cualquiera de las colonias que crecen contienen el fegémido puro. Recoger una colonia y cultivar en un pequeño (3-10 mL) cultivo líquido durante 8-12 horas. El medio de cultivo es $LB_{kan 50}$ líquido.

35 Verificación de la actividad: Transferir 1,0 mL de cultivo líquido a un microtubo estéril. Centrifugar a 13200 rpm (16000 g) durante 1 minuto. Descartar el sobrenadante y añadir 200 µL de tampón de fosfato pH 6,2. Someter a sonicación durante 5 a 10 segundos sobre hielo utilizando una micropunta ("microtip). Añadir 200 µL del sustrato apropiado, mezclar suavemente e incubar a 37°C durante 1,5-2 horas. Se debe ejecutar también un control negativo que contiene solo tampón y sustrato. Añadir 1,0 mL de etanol absoluto (200 proof) a la suspensión y mezclar. Centrifugar a 13200 rpm durante 10 minutos. Observar el color del sobrenadante. La cantidad de coloración puede variar, pero se considera que cualquier tubo con más coloración que el control tiene actividad positiva. Se puede utilizar un espectrofotómetro para esta etapa si así se desea o necesita. (Para Azo-xilano, Megazyme, leer a 590 nm).

40 RFLP de clones puros de las mismas Bibliotecas: Transferir 1,0 mL de cultivo líquido a un microtubo estéril. Centrifugar a 13200 rpm (16000 g) durante 1 minuto. Seguir el protocolo del kit QIAprep spin mini (Qiagen) para el aislamiento del plásmido y utilizar 40 µL de agua purificada como tampón de elución. Transferir 10 µL del ADN plasmídico a un microtubo estéril. Añadir 1,5 µL de Buffer 3 (New England Biolabs), 1,5 µL de solución 100X BSA

(New England Biolabs) y 2,0 μL de agua purificada. Añadir a esto 1,0 μL de endonucleasas de restricción Not 1 y 1,0 μL de Pst 1 (New England Biolabs). Incubar durante 1,5 horas a 37°C. Añadir 3,0 μL 6X tampón de Carga (Invitrogen). Hacer circular 15 μL de muestra digerida sobre un gel de agarosa al 1,0% durante 1-1.5 horas a 120 voltios. Observar el gel con un sistema para tomar imágenes del gel. Realizar el análisis de secuencia sobre todos los clones con un patrón de digestión diferente.

La Figura 5 es una tabla que contiene la caracterización de las enzimas que incluye el resumen de las actividades relativas de varias enzimas en varias condiciones, p. ej., pH y temperatura variables, como se ha comentado anteriormente.

Ejemplo 2: Análisis de actividad

El siguiente ejemplo demuestra la actividad enzimática de ciertas enzimas. Estos análisis se pueden utilizar también para determinar si un polipéptido tiene la actividad de la enzima requisito (v.g., glucanasa) para estar dentro del alcance de la invención.

Se demostró que los polipéptidos que tienen las secuencias mostradas en los SEQ ID NO: enumerados más abajo tenían actividad glucanasa, como se describe más abajo. La actividad específica se determinó sobre β -glucano de cebada (BBG) o carboximetilcelulosa (CMC) utilizando el análisis de azúcares reductores con BCA. 1 Unidad (U) de actividad glucanasa = 1 $\mu\text{mol}/\text{min}^{-1}$ de equivalentes reductores de glucosa liberados a 37°C, pH 5,3.

Glucanasa	Mw (kDa)	pI	Familia GH	Actividad Específica (U/mg)			T _{opt} (°C)	pH _{opt}
				BBG Nativa	BBG con Etiqueta de 6H	CMC con etiqueta de 6H		
SEQ ID NO:6 (codificado por el SEQ ID NO:5)	37,5	5,9	5	22	ND	ND	≥90	5-7
SEQ ID NO:400 (codificado por el SEQ ID NO:399)	37,9	5,5	5	0,85	ND	ND	≥90	5-7
SEQ ID NO:162 (codificado por el SEQ ID NO:161)	34,0	5,2	5	0,95	ND	ND	≥85	ND
SEQ ID NO:84 (codificado por el SEQ ID NO:83)	36,9	6,3	5	>40	ND	ND	80	4-6
SEQ ID NO:172 (codificado por el SEQ ID NO: 171)	29,8	5,0	16	32	ND	ND	50	5-6
SEQ ID NO:104 (codificado por el SEQ ID NO:103)	39,7	5,9	5	ND	3,2	2,8	85	5-6
SEQ ID NO:10 (codificado por el SEQ ID NO:9)	77,7	4,9	5	ND	ND	0,5	85	5-6
SEQ ID NO:222 (codificado por el SEQ ID NO:221)	53,8	9,1	5	>40	16	24	85	5-6
SEQ ID NO:108 (codificado por el SEQ ID NO:107)	78,9	4,3	5	ND	3,8	4,0	75	ND
SEQ ID NO:176 (codificado por el SEQ ID NO:175)	37,2	6,0	5	ND	3,5	21	75	5-6
SEQ ID NO:110 (codificado por el SEQ ID NO:109)	39,9	6,2	5	ND	13	12	ND	ND
SEQ ID NO:268 (codificado por el SEQ ID NO:267)	51,8	4,6	5	ND	3,6	2,8	50	ND
SEQ ID NO:324 (codificado por el SEQ ID NO:323)	49,3	6,1	5	ND	ND	ND	ND	ND
SEQ ID NO:370 (codificado por el SEQ ID NO:369)	42,1	5,8	5	ND	ND	ND	ND	ND
SEQ ID NO:168 (codificado por el SEQ ID NO:167)	37,3	5,7	5	ND	ND	ND	ND	ND

ES 2 401 795 T3

Glucanasa	Mw (kDa)	pI	Familia GH	Actividad Específica (U/mg)			T _{opt} (°C)	pH _{opt}
				BBG Nativa	BBG con Etiqueta de 6H	CMC con etiqueta de 6H		
SEQ ID NO:154 (codificado por el SEQ ID NO:153)	35,6	5,4	5	ND	ND	ND	ND	ND
SEQ ID NO:118 (codificado por el SEQ ID NO:117)	34,5	6,1	5	ND	ND	ND	ND	ND
SEQ ID NO:148 (codificado por el SEQ ID NO:147)	74,1	5,3	5	ND	ND	ND	ND	ND

ND = No determinado

Se demostró que los polipéptidos que tenían la secuencia mostrada en los SEQ ID NO: de más abajo tenían actividad endoglucanasa alcalina/celulasa, con los óptimos de pH y temperatura mostrados más abajo. Esta actividad se demostró utilizando un análisis de la actividad celulasa (un análisis de extremos reductores con BCA), como se describe con detalle en el Ejemplo 3, más abajo.

5

SEQ ID NO:	Tipo	Óptimo de pH	Óptimo de Temperatura
409,410	Endoglucanasa alcalina/celulasa	5	NA
343,344	Endoglucanasa alcalina/celulasa	6	60
319,320	Endoglucanasa alcalina/celulasa	6	70
383, 384	Endoglucanasa alcalina/celulasa	7	60
301, 302	Endoglucanasa alcalina/celulasa	7	60
257, 258	Endoglucanasa alcalina/celulasa	8	42
419, 420	Endoglucanasa alcalina/celulasa	8	70
421, 422	Endoglucanasa alcalina/celulasa	9	60
405, 406	Endoglucanasa alcalina/celulasa	9	50
329, 330	Endoglucanasa alcalina/celulasa	9	50
325, 326	Endoglucanasa alcalina/celulasa	(5-7)*	70
415, 416	Endoglucanasa alcalina/celulasa	(6-10)*	70
303, 304	Endoglucanasa alcalina/celulasa	(6-10)*	60
271, 272	Endoglucanasa alcalina/celulasa	(6-7)*	60
175, 176	Endoglucanasa alcalina/celulasa	(6-7)*	70
9, 10	Endoglucanasa alcalina/celulasa	(6-7)*	70
297, 298	Endoglucanasa alcalina/celulasa	(6-8)*	50
109, 110	Endoglucanasa alcalina/celulasa	(6-8)*	60
267, 268	Endoglucanasa alcalina/celulasa	(6-8)*	70
107, 108	Endoglucanasa alcalina/celulasa	(6-8)*	70
305, 306	Endoglucanasa alcalina/celulasa	(7-10)*	60
417,418	Endoglucanasa alcalina/celulasa	(7-8,5)*	NA
227,228	Endoglucanasa alcalina/celulasa	(7-9)*	60
375, 376	Endoglucanasa alcalina/celulasa	(7-9)*	70
335,336	Endoglucanasa alcalina/celulasa	(7-9)*	50
155, 156	Endoglucanasa alcalina/celulasa	(7-9)*	60
445, 446	Endoglucanasa alcalina/celulasa	(7-9)*	60

SEQ ID NO:	Tipo	Óptimo de pH	Óptimo de Temperatura
259, 260	Endoglucanasa alcalina/celulasa	(8-10)*	50
423, 424	Endoglucanasa alcalina/celulasa	(8-10)*	50
345, 346	Endoglucanasa alcalina/celulasa	(8-9)*	25
285, 286	Endoglucanasa alcalina/celulasa	(8-9)*	50
351, 352	Endoglucanasa alcalina/celulasa	(9-10)*	80

Ejemplo 3: Análisis de la actividad celulasa: análisis de extremos reductores BCA

5 El siguiente ejemplo describe un análisis, un análisis de la actividad celulasa (un análisis de extremos reductores con BCA) que se puede utilizar para determinar si un polipéptido tiene la actividad de la enzima requisito (p. ej., glucanasa), p. ej., una actividad endoglucanasa alcalina/celulasa (véase el Ejemplo 2, anterior) para estar dentro del alcance de la invención.

10 Este análisis se diseñó para medir la cantidad de extremos reductores producidos durante la degradación enzimática de la carboximetilcelulosa (CMC) en un formato de 96 pocillos de muestras múltiples de alto rendimiento.

Materiales

Soluciones sustrato

15 CMC al 1%

20 Disolver 1 g de CMC en 100 ml de tampón Britton-Robinson 50 mM a pH ~4, calentar la solución de CMC en un baño de agua hirviendo, mientras se mezcla, durante 20-40 minutos hasta que se disuelva (la solución todavía aparecerá ligeramente lechosa, pero translúcida). Ajustar al pH deseado con NaOH o HCl 1 M.

Solución A:

25 64 mg/ml de monohidrato de carbonato de sodio
24 mg/ml de bicarbonato de sodio
1,95 mg/ml de BCA (sal de disodio de 4,4'-dicarboxi-2,2'-biquinolina (Sigma Chemical Núm. de cat. D-8284)
Podría ser necesario disolver la BCA mediante calentamiento, sin calentar a más de ~80°C

Solución B:

30 1,24 mg/ml de pentahidrato de sulfato cúprico
1,26 mg/ml de L-serina
Añadir lo anterior a dH₂O

Reactivo de trabajo

1:1 de soluciones A y B, elaborar mezclas de reactivos de trabajo de nueva aportación cada día (elaborar usualmente solo lo necesario para cada análisis), elaborar Soluciones A y B de nueva aportación cada semana.

40

Solución de partida de glucosa

Glucosa 10 mM en dH₂O. Filtro de 0,2 µm, almacenar a 4C.

Patrones de glucosa:

45 Diluir la provisión de partida de glucosa en CMC al 1% al pH deseado; hasta una concentración final de 0, 100, 200, 300, 400, 500 µM. Puesto que se necesita generar la curva para cada placa de puntos temporales de muestra, puesto que el ciclo de calentamiento puede afectar a la cantidad de señal observada.

Método:

50

Configuración:

5 Tomar alícuotas de 1 ml de la solución sustrato (CMC al 1%) en una placa de pocillos profundos (si se utiliza temperatura ambiente) o tubos Acme en bloque caliente, equilibrar a la temperatura deseada (~ 5 min) en un bloque térmico o un baño de agua calentada.

Mientras la solución se está equilibrando, elaborar 10 ml del reactivo de trabajo y tomar alícuotas de 100 µl en una placa de PCR de 96 pocillos. Colocar la placa sobre hielo.

10 **Reacción/Toma de muestras**

Una vez completado el equilibrado de la temperatura, añadir la solución de enzima a la solución de sustrato. Mezclar inmediatamente pipeteando arriba y abajo. Tomar inmediatamente alícuotas de 10 µl en una placa de PCR (esto es t = 0, punto temporal cero). Tomar alícuotas de 10 µl en una placa de PCR en cada punto temporal deseado (p. ej., 0, 2, 5, 10, 15, 20, 30 minutos).

Reservar la última fila de la placa para la adición de 10 µl de patrones de glucosa (es decir, los pocillos deben tener únicamente 100 µl de reactivo de trabajo)

20 **Análisis de desarrollo de color:**

25 Cuando se recogen todos los puntos temporales y se añaden todos los patrones, cubrir la placa y calentar a 100°C durante 10 min. utilizando una máquina de PCR. Enfriar la placa sobre hielo durante 5 – 10 min (o ajustar la máquina de PCR a 1C durante 10 min).

Añadir 100 µl de H₂O a los pocillos. Mezclar. Tomar alícuotas de 100 µl de la mezcla en una placa de 96 pocillos de fondo plano y leer la absorbancia a 560 nm.

30 **Generación de la curva patrón**

Trazar A560 vs. µmoles de glucosa de los pocillos que contienen los patrones de glucosa. Utilizar regresión lineal para calcular la pendiente (S_{std}).

35 **Generación del gráfico de la pendiente de reacción**

Trazar A560 vs puntos temporales. Tomar como cero los puntos temporales de cada muestra contra su propio T = 0 (es decir restar el valor de absorbancia a T = 0 de las muestras de todos los demás puntos temporales de la misma muestra). Generar la pendiente (S_{rxn}) para cada conjunto de puntos temporales de las muestras (A560/tiempo).

40 **Determinación de la actividad:**

Dividir S_{rxn} por S_{std}, y multiplicar por 100 (puesto que los µmoles de producto detectados son la cantidad de extremos reductores en los 10 µl utilizados en el análisis, no la cantidad total generada en la reacción enzimática de 1 ml).

45 **Determinación de la actividad específica:**

Dividir la Actividad (en unidades de mmoles/min) por los mg totales de proteína añadida en la reacción de 1 ml.

50 Determinar la concentración de proteína mediante un análisis Bradford o similar.

Dividir la concentración de proteína por cualquier dilución utilizada.

Multiplicar por el volumen (en ml) utilizado en la reacción.

55 Todos los puntos se deben realizar por duplicado, siendo mejor por triplicado.

El siguiente diagrama expone un conjunto ilustrativo de datos ("datos de muestra") que es ilustrado en forma de gráfico como una "curva patrón" en la Figura 6.

60 **Datos de muestra**

	fecha	mg/ml	Diln.	ul/rxn	0 min	5 min	8 min	12 min	24 min	36 min	45 min
Enz x	06/09	20	500	20	0,1252	0,1654	0,1889	0,2315	0,3386	0,4036	0,4695

	fecha	mg/ml	Diln.	ul/rxn	0 min	5 min	8 min	12 min	24 min	36 min	45 min
<u>Pendiente de la curva patrón:</u>								88,375 A560/umoles de glucosa			
<u>Pendiente de reacción:</u>								0,0076 A560/min			
<u>Actividad (pendiente de reacción/pendiente patrón):</u>								8,70061E-05 umoles/min			
<u>Actividad real / 1ml rxn (=Actividad x 100):</u>								0,0087 umoles/min			
<u>Actividad específica:</u>								10,87 umoles/min, mg			

Ejemplo 4: Optimización de codones

5 El siguiente ejemplo demuestra una optimización de codón ilustrativa. Se puede utilizar cualquier protocolo de optimización de codones conocido en la técnica para la optimización de codones de cualquier ácido nucleico de la invención.

10 Un ácido nucleico que codifica el polipéptido que tiene una secuencia como se muestra en el SEQ ID NO: 6, es decir, el SEQ ID NO: 5, se sometió a optimización de codones para determinar la expresión óptima en *Pichia pastoris*; el ácido nucleico que codifica la enzima de codón optimizado de *Pichia pastoris* es el SEQ ID NO: 463. Además de optimizar los codones del ácido nucleico que codifica la enzima, se modificó un aminoácido (A91V), y la nueva secuencia de polipéptido se muestra como SEQ ID NO: 464.

15 Los análisis de la actividad glucanasa (cuyos datos se ilustran en las Figuras 7 y 8) demostraron una mejora de la expresión en *Pichia pastoris* del SEQ ID NO: 464 (codificado, p. ej., por el SEQ ID NO: 463), que es la versión de codón optimizado del polipéptido que tiene una secuencia mostrada en el SEQ ID NO: 6 (codificado, p. ej., por el SEQ ID NO: 5). El nivel de expresión se mejoró cambiando el pH.

20 En la Figura 7, la actividad glucanasa durante el curso de la fermentación se muestra en U/ml de cultivo. 1 unidad de actividad glucanasa = $1 \mu\text{mol}/\text{min}^{-1}$ de equivalentes reductores de glucosa liberados a 37°C, pH 5,3. Se utilizó la glucanasa de codón optimizado del SEQ ID NO: 464 (codificado, p. ej., por el SEQ ID NO: 463), expresada en *Pichia pastoris*. La fermentación se puso en marcha a 5,0.

25 En la Figura 8, la actividad glucanasa durante el curso de la fermentación se muestra en U/ml de cultivo. 1 unidad de actividad glucanasa = $1 \mu\text{mol}/\text{min}^{-1}$ de equivalentes reductores de glucosa liberados a 37°C, pH 5,3. Se utilizó la glucanasa de codón optimizado del SEQ ID NO: 464 (codificado, p. ej., por el SEQ ID NO: 463), expresada en *Pichia pastoris*. La fermentación se realizó a pH 6,2.

30 Ejemplo 5: Actividad enzimática

El siguiente ejemplo demuestra la confirmación de la actividad enzimática. Estos análisis también se pueden utilizar para determinar si un polipéptido tiene la actividad de la enzima requisito (p. ej., glucanasa) para estar dentro del alcance de la invención.

35 **Actividad específica de la glucanasa codificada por en SEQ ID NO: 6**
 EEI La actividad específica de la enzima que tiene una secuencia mostrada en el SEQ ID NO: 6 (codificada, p. ej., por el SEQ ID NO: 5) se demostró utilizando en siguiente protocolo:

40 La glucanasa codificada por el SEQ ID NO: 6 se purificó hasta homogeneidad utilizando cromatografía de intercambio iónico. Las actividades específicas se determinaron sobre sustrato al 1% en tampón de acetato de sodio de pH 5,3, a 37°C utilizando el análisis de azúcares reductores con BCA. 1 Unidad (U) de actividad glucanasa = $1 \text{ mol}/\text{min}^{-1}$ de equivalentes reductores de glucosa liberados a 37°C, pH 5,3.

- 45
- o Beta Glucano de Cebada (BBG): 30 U/mg
 - o Beta Glucano de Avena (OBG): 30 U/mg
 - o Carboximetilcelulosa (CMC): 40 U/mg
 - o Galactomanano de Algarroba: 0,3 U/m

50 **Perfil de temperatura de la glucanasa codificada por el SEQ ID NO: 6**

El perfil de temperatura se determinó en tres sustratos separados (BBG, OBG y CMC). La glucanasa codificada por el SEC ID NO:6 tenía su mayor actividad a las temperaturas más altas. La actividad específica de la glucanasa codificada por el SEQ ID NO: 6 sobre BBG y CMC a 80°C es 10 veces mejor que la actividad observada a 37°C. En

presencia de manano, la glucanasa codificada por el SEQ ID NO: 6 mostró la actividad más alta a 100°C, como se ilustra en la Figura 9.

5 El perfil de temperatura se determinó incubando BD10 en presencia de sustrato (BBG, OBG y CMC). Las velocidades iniciales se determinaron utilizando el análisis de azúcares reductores con BCA y tampón de acetato de sodio de pH 5,3. Las velocidades iniciales se normalizaron y se trazaron como % de la actividad, como se ilustra en la Figura 9.

Determinación de la vida media de la glucanasa codificada por el SEQ ID NO: 6

10 La vida media de la glucanasa codificada por el SEQ ID NO: 6 se determinó a 85°C y 90°C. La glucanasa codificada por el SEQ ID NO: 6 se sensibilizó con calor durante diferentes tiempos a 85 y 90 grados y la actividad residual se midió a 37°C. La glucanasa codificada por el SEQ ID NO: 6 conservó más de 60% de su actividad al cabo de 10 minutos de incubación a 85°C. A 90°C, no quedó actividad residual al cabo de 2 minutos, como se ilustra en la
15 Figura 10.

Como se ilustra en la Figura 10, la vida media de BD10 se determinó sensibilizando con calor la enzima durante 30
20 seg, 1 min, 2 min, 3 min, 4 min, 5 min, y 10 min a las temperaturas indicadas (85°C y 90°C) y verificando la actividad en condiciones normalizadas utilizando el azúcar reductor BCA.

Si bien la invención ha sido descrita en detalle con referencia a ciertos aspectos Ilustrativos de la misma, se entenderá que las modificaciones y variaciones están dentro del alcance de lo que se describe y reivindica.

REIVINDICACIONES

1. Un ácido nucleico, sintético o recombinante, aislado que comprende:
- 5 (a) una secuencia de ácido nucleico que tiene al menos 90%, 91 %, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, o más o completa (100%) de identidad de secuencia con el SEQ ID NO: 37, donde el ácido nucleico codifica al menos un polipéptido que tiene: (i) una actividad glucanasa, o (ii) actividad inmunogénica y puede generar un anticuerpo que se une específicamente a un polipéptido que tiene una secuencia como la mostrada en el SEQ ID NO: 38;
- 10 (b) una secuencia de ácido nucleico que codifica un polipéptido que tiene una secuencia como la mostrada en el SEQ ID NO: 38;
- (c) una secuencia de ácido nucleico que hibrida en condiciones restrictivas con el complemento de un ácido nucleico que comprende el SEQ ID NO: 37; en donde el ácido nucleico codifica un polipéptido que tiene una actividad glucanasa, y las condiciones restrictivas incluyen una etapa de lavado que comprende un lavado a 0,2x SSC a una temperatura de 65°C
- 15 durante 15 minutos; o
- (d) un ácido nucleico complementario al ácido nucleico de cualquiera de (a) a (c);
2. Una sonda de ácido nucleico para identificar un ácido nucleico que codifica un polipéptido con actividad glucanasa, donde la sonda comprende al menos 10 bases consecutivas, o al menos de 10 a 50, 20 a 60, 30 a 70, 40
- 20 a 80, 60 a 100, o 50 a 150 bases consecutivas, de una secuencia que comprende (a) el SEQ ID NO: 37, o (b) la secuencia de la reivindicación 1, en donde la sonda identifica el ácido nucleico mediante unión o hibridación.
3. Un casete de expresión, un vector o un vehículo de clonación que comprende un ácido nucleico que comprende la secuencia de la reivindicación 1,
- 25 en donde el vehículo de clonación comprende opcionalmente un vector viral que comprende un vector de adenovirus, un vector retroviral o un vector viral adenoasociado, o en donde el vehículo de clonación comprende opcionalmente un cromosoma artificial bacteriano (SAC), un plásmido, un vector derivado del bacteriófago P1 (PAC), un cromosoma artificial de levadura (YAC), o un cromosoma artificial de mamífero (MAC);
- 30 4. Una célula transformada que comprende (a) un ácido nucleico que comprende la secuencia de la reivindicación 1; o (b) el casete de expresión, vector o vehículo de clonación de la reivindicación 3, en donde opcionalmente, la célula transformada es una célula bacteriana, una célula de mamífero, una célula fúngica, una célula de levadura, una célula de insecto o una célula vegetal.
- 35 5. Un animal no humano, planta, parte de planta o semilla vegetal transgénicos que comprenden (a) la secuencia de la reivindicación 1, o el casete de expresión, vector o vehículo de clonación de la reivindicación 3; o (b) el animal no humano transgénico de (a), en donde el animal es un ratón.
- 40 6. Un polipéptido, sintético o recombinante, aislado que comprende:
- (i) una secuencia de aminoácidos que tiene al menos 90%, 91 %, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, o más o completa (100%) de identidad de secuencia con el SEQ ID NO: 38, en donde el polipéptido tiene: (i) una actividad glucanasa, o (ii) actividad inmunogénica y puede generar un anticuerpo que se une específicamente a un polipéptido que tiene una secuencia como la mostrada en el SEQ
- 45 ID NO: 38;
- (ii) una secuencia de ácido nucleico codificada por el ácido nucleico de la reivindicación 1;
- (iii) la secuencia de aminoácidos de (i) o (ii) y que tiene al menos una sustitución de residuo de aminoácido conservativa y conserva una actividad glucanasa, en donde una sustitución conservativa comprende la sustitución de un residuo de aminoácido por otro aminoácido de características similares;
- 50 (iv) la secuencia de aminoácidos de (iii), en donde la sustitución conservativa comprende el remplazo de un aminoácido alifático por otro aminoácido alifático; o, el remplazo de una Serina por una Treonina o viceversa; o, el remplazo de un residuo ácido por otro residuo ácido; o, el remplazo de un residuo que porta un grupo amida por otro residuo que porta un grupo amida; o, el intercambio de un residuo alcalino por otro residuo alcalino; o, el remplazo de un residuo aromático por otro residuo aromático, o una combinación de los mismos;
- 55 (v) la secuencia de aminoácidos de (iv), en donde el residuo alifático comprende Alanina, Valina, Leucina, Isoleucina o un equivalente sintético de los mismos; o, el residuo ácido comprende Ácido aspártico, Ácido glutámico o un equivalente sintético de los mismos; o, el residuo que comprende un grupo amida comprende Ácido aspártico, Ácido glutámico o un equivalente sintético de los mismos; o, el residuo alcalino comprende Lisina, Arginina, o un equivalente sintético de los mismos; o, el residuo aromático comprende Fenilalanina, Tirosina o un equivalente sintético de los mismos;
- 60 (vi) la secuencia de aminoácidos de cualquiera de (i) a (v), en donde el polipéptido carece de una secuencia señal o una secuencia prepro;
- (vii) la secuencia de aminoácidos de cualquiera de (i) a (vi), y que tiene una secuencia señal heteróloga o una secuencia prepro heteróloga, o que tiene una secuencia señal de levadura;

(viii) la secuencia de aminoácidos de cualquiera de (i) a (vii), y que tiene una secuencia heteróloga; o
 (ix) la secuencia de aminoácidos de cualquiera de (i) a (viii), en donde el polipéptido comprende al menos un sitio de glicosilación.

5 7. Una composición que comprende un polipéptido de la reivindicación 6 o el ácido nucleico de la reivindicación 1.

8. Un método para producir un polipéptido recombinante que comprende las etapas de:

(A) (a) proporcionar un ácido nucleico de la reivindicación 1; y (b) expresar el ácido nucleico de la etapa (a) en condiciones que permitan la expresión del polipéptido, produciéndose de ese modo un polipéptido recombinante; o

(B) el método de (A), que comprende adicionalmente la transformación de una célula anfitriona con el ácido nucleico de la etapa (A) (a) seguida de la expresión del ácido nucleico de la etapa (A) (a), produciéndose de ese modo un polipéptido recombinante en una célula transformada.

15 9. Un método para generar una variante de un ácido nucleico que codifica un polipéptido con una actividad glucanasa que comprende las etapas de:

(A)

(a) proporcionar un ácido nucleico molde que comprende la secuencia de la reivindicación 1; y

(b) modificar, suprimir o añadir uno o más nucleóticos a la secuencia molde, o una de sus combinaciones, para generar una variante del ácido nucleico molde; o

(B) el método de (A), que comprende adicionalmente expresar el ácido nucleico variante para generar un polipéptido de glucanasa variante;

(C) el método de (A) o (B), en donde las modificaciones, adiciones o deleciones se introducen mediante un método que comprende PCR propensa a error, barajado, mutagénesis dirigida a oligonucleótidos, PCR de ensamblaje, mutagénesis por PCR sexual, mutagenésis *in vivo*, mutagenésis por inserción de casetes, mutagenésis de ensamblaje recursivo, mutagenésis de conjunto exponencial, mutagénesis de sitio específico, reensamblaje de genes, Mutagénesis por Saturación de Sitio de Genes (GSSM™), reensamblaje de ligación sintética (SLR), recombinación, recombinación de secuencia recursiva, mutagénesis de ADN modificada por fosfotioato, mutagénesis de molde que contiene uracilo, mutagénesis de dúplex discontinuo, mutagénesis por reparación de apareamiento erróneo, mutagénesis de cepa anfitriona de reparación defectuosa, mutagénesis química, mutagénesis radiogénica, mutagénesis por supresión, mutagénesis por restricción-selección, mutagénesis por restricción-purificación, síntesis artificial de genes, mutagénesis de ensamblaje, creación de multímeros de ácidos nucleicos quiméricos, o cualquiera de sus combinaciones; o

(D) el método de (A) o (B) repetido iterativamente hasta que se produzca una glucanasa tenga una actividad alterada o diferente o una estabilidad alterada o diferente de la de un polipéptido codificado por el ácido nucleico molde;

(E) el método de (A), (B), (C) o (D), en donde el péptido de glucanasa variante es termotolerante, y conserva algo de actividad después de ser expuesto a una temperatura elevada;

(F) el método de (A), (B), (C) o (D), en donde el péptido de glucanasa variante tiene incrementada la glicosilación en comparación con la glucanasa codificada por un ácido nucleico molde;

(G) el método de (A), (B), (C) o (D), en donde el péptido de glucanasa variante tiene una actividad glucanasa a alta temperatura, en donde la glucanasa codificada por el ácido nucleico molde no es activo a la alta temperatura;

(H) el método de (A), (B), (C) o (D), en donde el método es repetido iterativamente hasta que se produzca un gen de glucanasa que tenga un nivel de expresión del mensaje o estabilidad superior o inferior al del ácido nucleico molde.

10. Un polipéptido quimérico que comprende

(A) al menos un primer dominio que comprende un péptido señal (SP), una secuencia prepro, un módulo de unión a carbohidratos, y/o un dominio catalítico (CD) del polipéptido de la reivindicación 6, y al menos un segundo dominio que comprende un polipéptido o péptido heterólogo, en donde el polipéptido o péptido heterólogo no está asociado naturalmente con el péptido señal (SP);

(B) el polipéptido quimérico de (A), en donde el polipéptido o péptido heterólogo no es una glucanasa; o

(C) el polipéptido quimérico de (B), en donde el polipéptido o péptido heterólogo es amino terminal, carboxi terminal o está en ambos extremos del péptido señal (SP) o dominio catalítico (CD).

11. Un método para tratar o modificar una composición que comprende poner en contacto una composición con el polipéptido de la reivindicación 6 o un polipéptido codificado por el ácido nucleico de la reivindicación 1,

en donde opcionalmente la composición es una composición farmacéutica, una composición detergente, una solución para lentes de contacto, una composición para el tratamiento de residuos, o un jabón,

o en donde opcionalmente la composición es un combustible, en donde opcionalmente la composición es un alcohol, en donde opcionalmente el alcohol es etanol,

o en donde opcionalmente la composición es un papel, un residuo de papel, un producto de papel reciclado, pasta de papel, producto de papel, madera, producto de madera, pasta de madera, pasta Kraft, artículo textil, tejido, hilo, fibra, o una tela,

5 o en donde opcionalmente la composición comprende una celulosa, una hemicelulosa, una lignina, un xilano, un glucano, o un manano,

o en donde opcionalmente la composición es una bebida, un alimento, un pienso, un producto lácteo, un suplemento nutricional,

o en donde opcionalmente el método es para la producción de alimentos, piensos, suplementos nutricionales o bebidas,

10 o en donde opcionalmente el método es para reducir la lignina en un papel, un producto de papel, una madera o un producto de madera.

12. Un método para producir una bebida, un pienso, un alimento o un suplemento nutricional que comprende el uso del polipéptido de la reivindicación 6 o un polipéptido codificado por el ácido nucleico de la reivindicación 1.

15 13. Un método para elaborar un combustible que comprende el uso del polipéptido de la reivindicación 6 o un polipéptido codificado por el ácido nucleico de la reivindicación 1, en donde opcionalmente el combustible es un alcohol, en donde opcionalmente el alcohol es un etanol.

20 14. Un método para modificar codones en un ácido nucleico que codifica un polipéptido de glucanasa, comprendiendo el método las siguientes etapas:

(a) proporcionar un ácido nucleico que codifica un polipéptido de la reivindicación 6; y

(b) identificar un codón en el ácido nucleico de la etapa (a) y remplazarlo por un codón diferente que codifica el mismo aminoácido que el condón remplazado, modificando de ese modo los codones en un ácido nucleico que

25 codifica una glucanasa.

15. Un método para la perforación, la fractura o el tratamiento de pozos de petróleo que comprende el uso de un polipéptido de la reivindicación 6 o un o un polipéptido codificado por el ácido nucleico de la reivindicación 1.

30

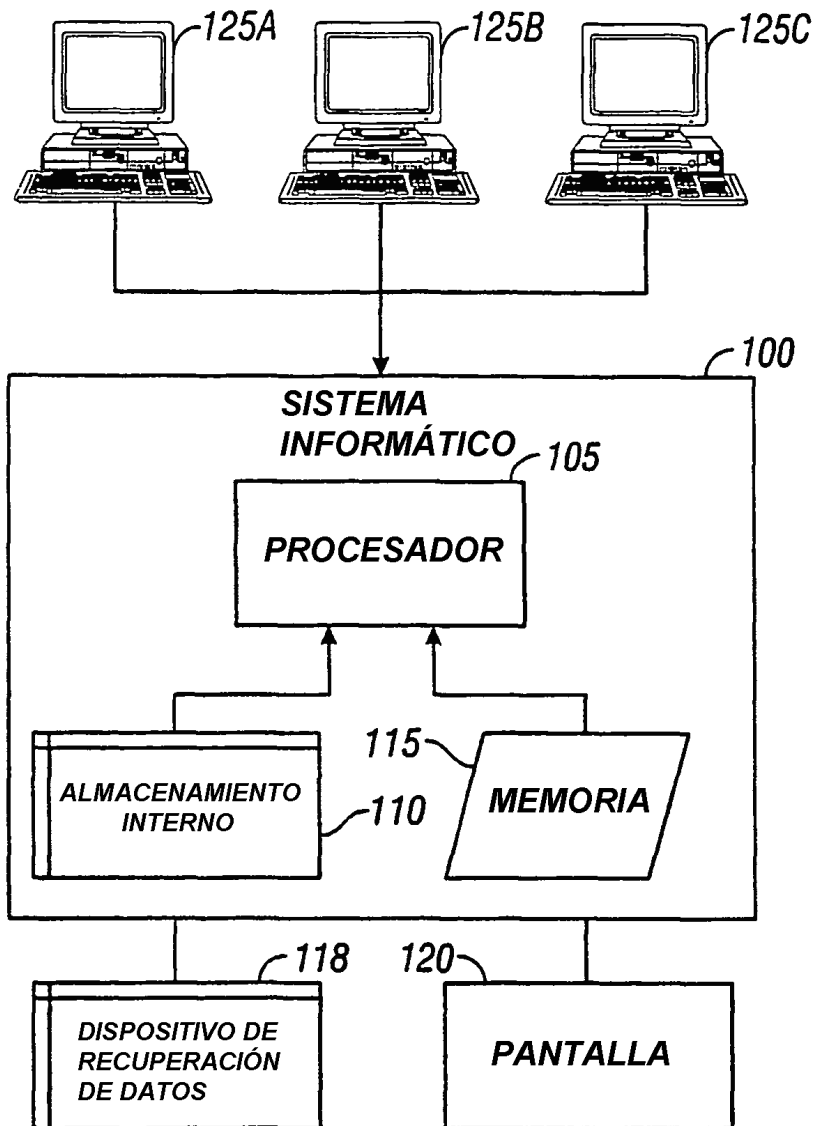


FIG. 1

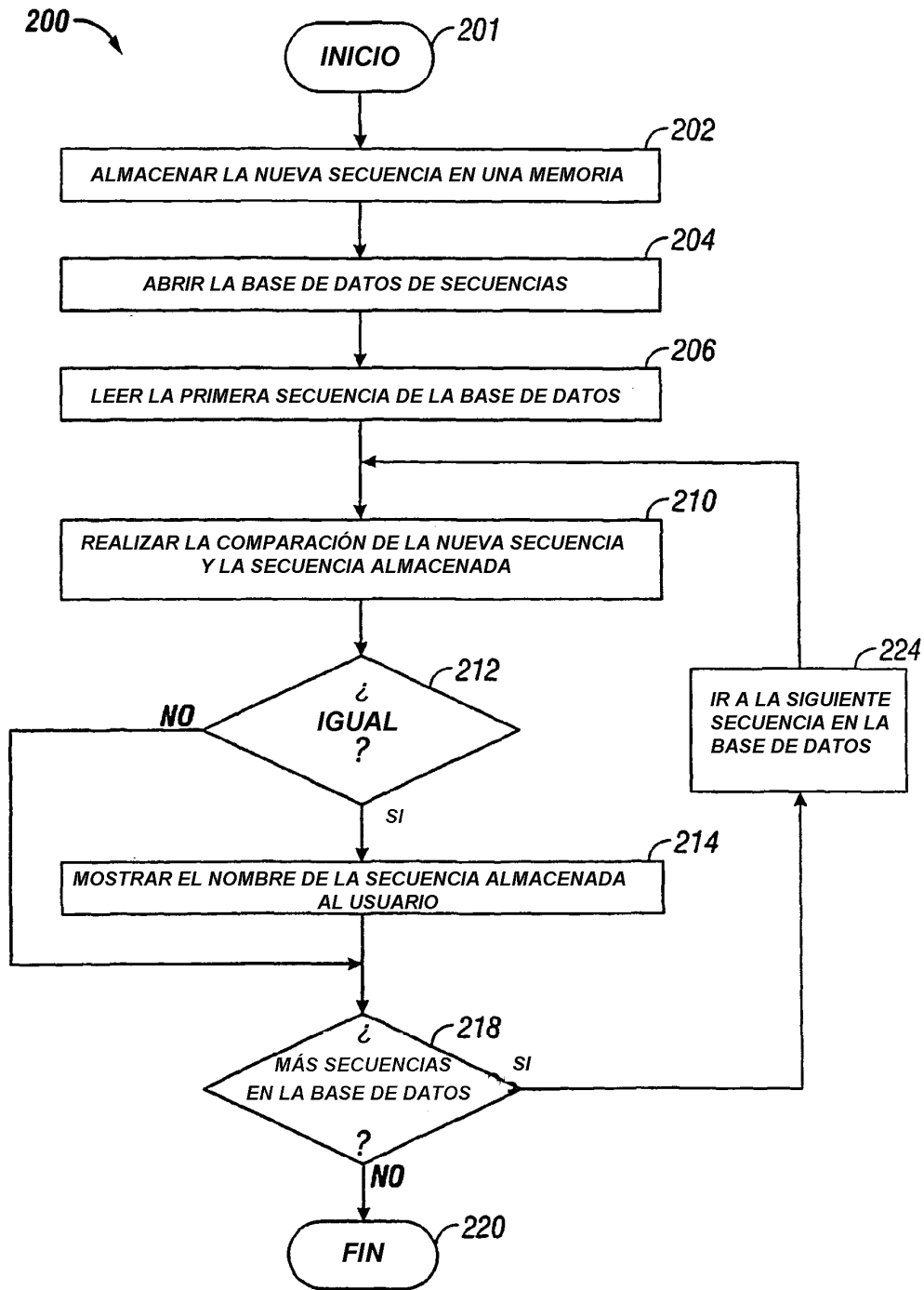


FIG. 2

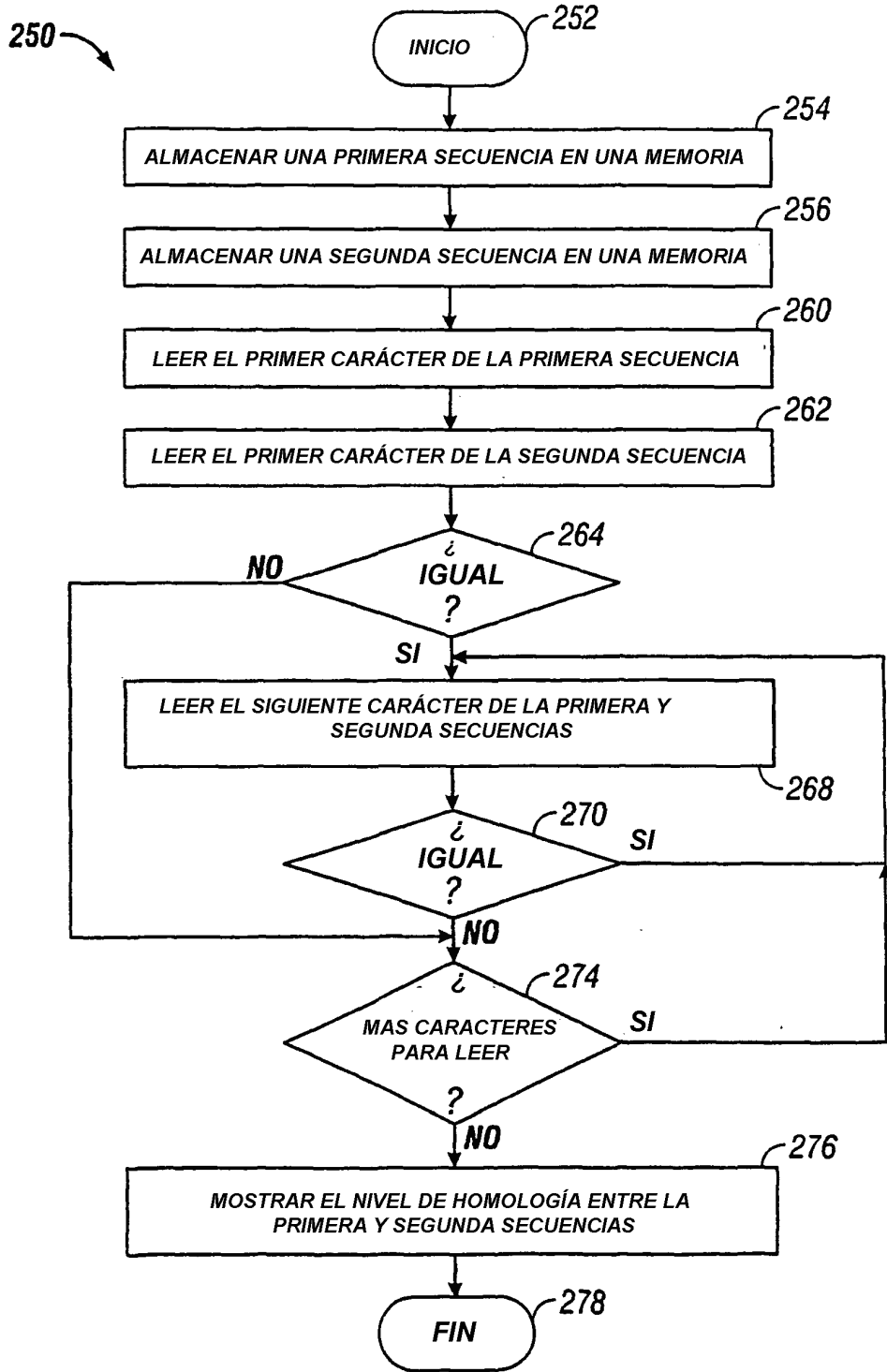


FIG. 3

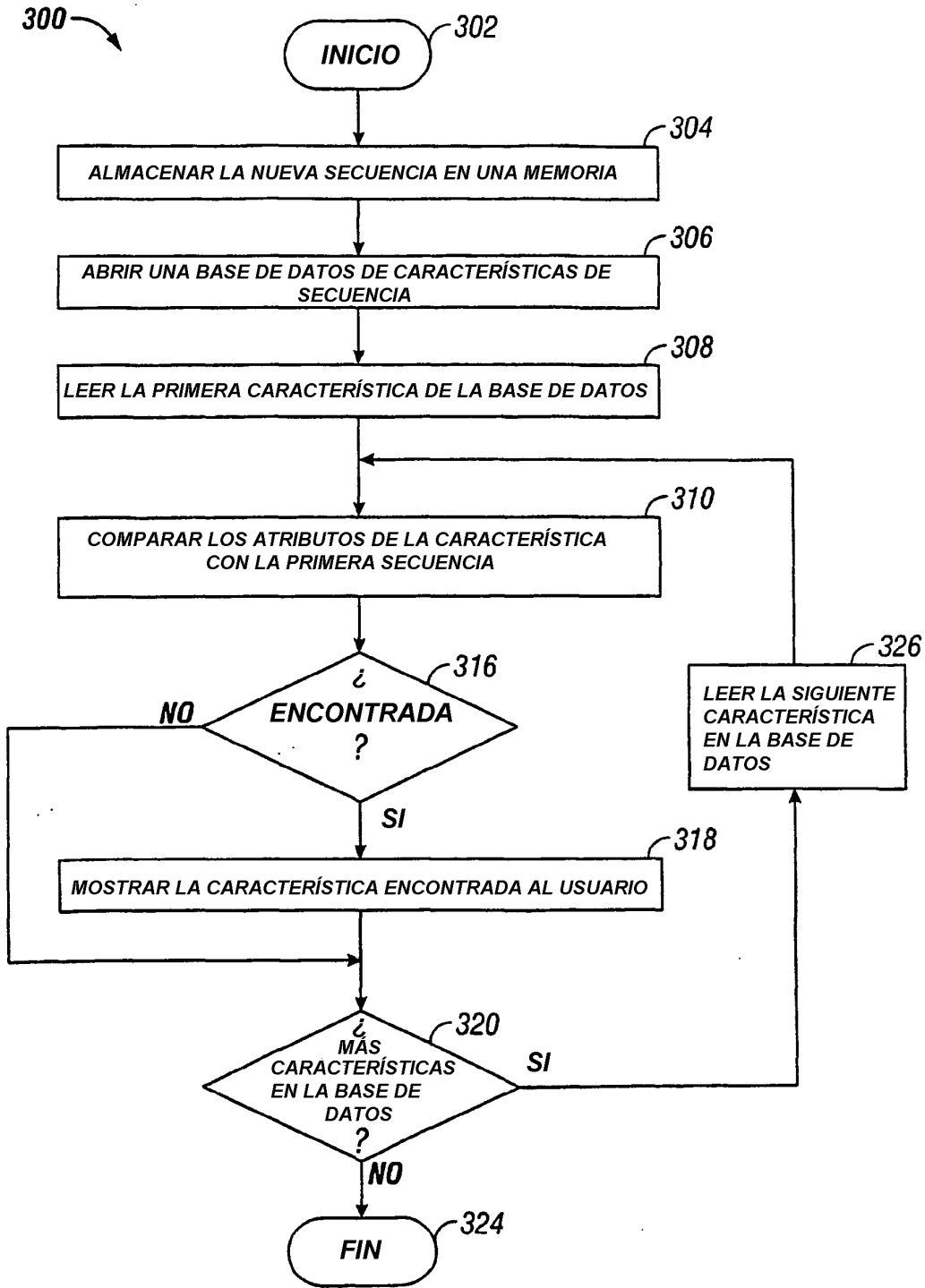


FIG. 4

FIGURA 5

SEQ ID NOS	Familia	Maximo	t1/2	90°C	6 - 8	Actividades Significativas	FA	FA	FA
5, 6	5	>90°C	t1/2 2 min, 90°C	6 - 8	B33 ⁴ Liquenano, Manano, Goma garrofin, Guar	³ 0,0,007, 0,07, 0,85, 1	³ 0,725, 0,91, 1, 0,034		
161, 162	16	³ 85°C	³ 10 min 100°C	5,5*	CMC ⁴	ND			
59, 60	16	³ 85°C	³ 10 min 100°C	5,5*	BBG	ND			-, 0,042, 0,17, 0,46, 0,78, 1
67, 68	9	75°C	< 1 min a 85°C	5,5	BBG	0, 0,05, 1, 0,62, 0,15, 0,09			-, 0,09, 0,27, 0,69, 0,84, 1
105, 106	5	ND	ND	ND	BBG	ND			-, 0, 0,66, 1, 0,94, 0
81, 82	16	70°C	<20 min 70°C	5,0-7,0	BBG	³ 0,006, 0,99, 0,96, 1, 0,83			³ 0,62, 0,84, 3, 0,019
175, 176	5	75°C	< 1 min a 85°C	5,5	BBG	0,1, 0,4, 1, 0,76, 0,34, 0,45			-, 0,03, 0,2, 0,68, 1, 0,88
107, 108	5	ND	ND	ND	BBG	ND			ND
221, 222	5	85°C	< 1 min a 95°C	5,5	BBG	0,08, 0,36, 1, 0,52, 0,12, 0,08			0,003, 0,07, 0,3, 0,74, 0,89, 1
9, 10	5	85°C	> 2 min a 95°C	5,5	BBG	0,006, 0,17, 1, 0,34, 0,01, 0			0, 0,049, 0,21, 0,62, 0,64, 1
219, 220	5	ND	ND	ND	BBG	ND			ND
103, 104	5	ND	ND	ND	BBG	ND			ND
171, 172	16	50°C	< 1 min a 85°C	5,5	BBG	0,21, 0,26, 1, 0,48, 0,24, 0,34			-, 0,77, 1, 0,6, 0,36, 0,49
137, 138	5	37°C	< 1 min a 85°C	7,0	BBG	0, 0, 0,125, 1, 0,25, 0,125			-, 1, 0,53, 0, 0,06, 0,1
143, 144	5	ND	ND	ND	BBG	ND			ND

Figura 6

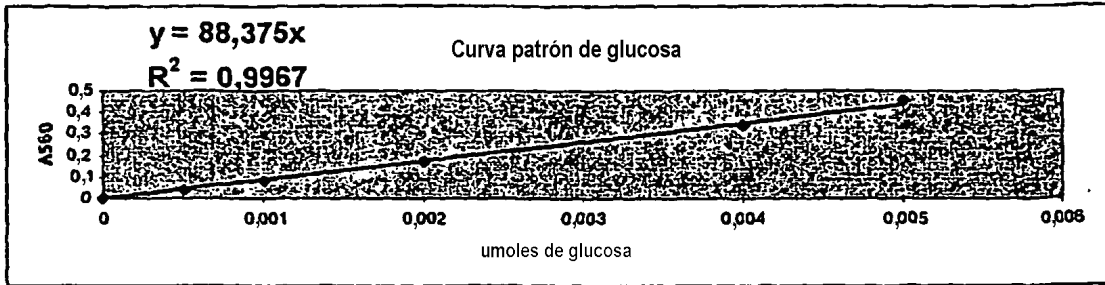


Figura 7

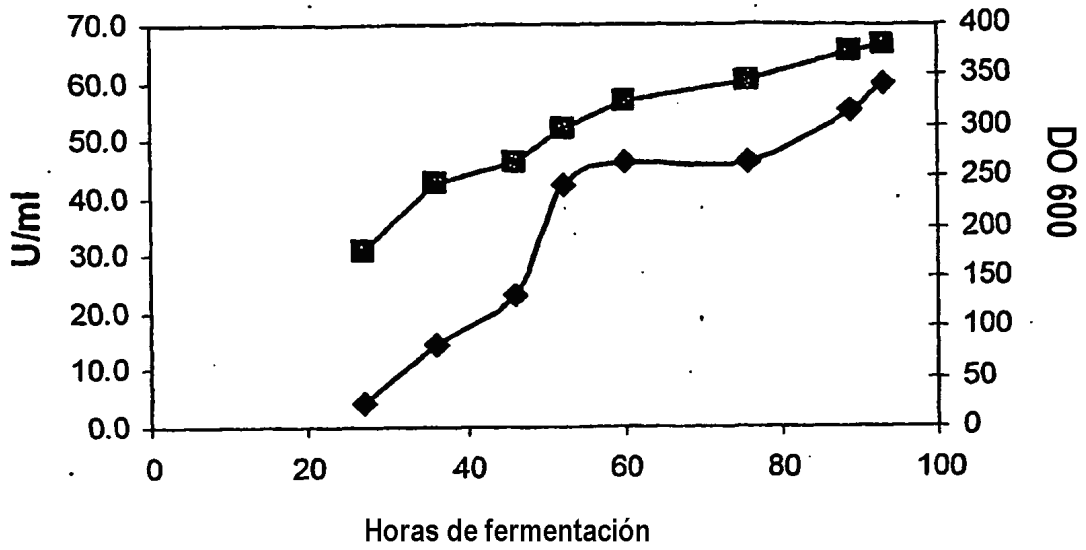


Figura 8

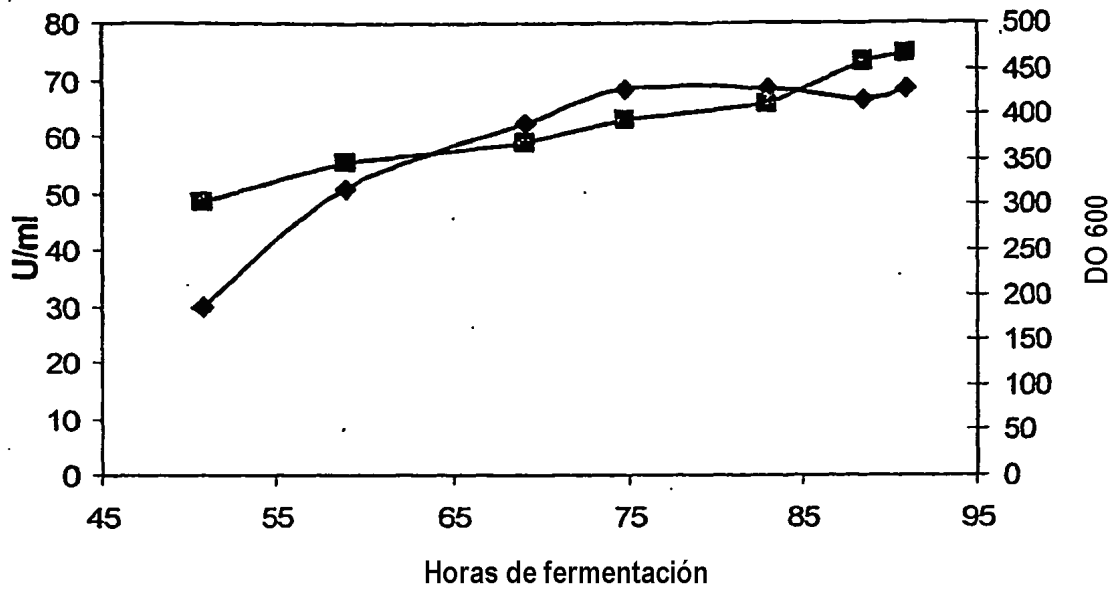


Figura 9

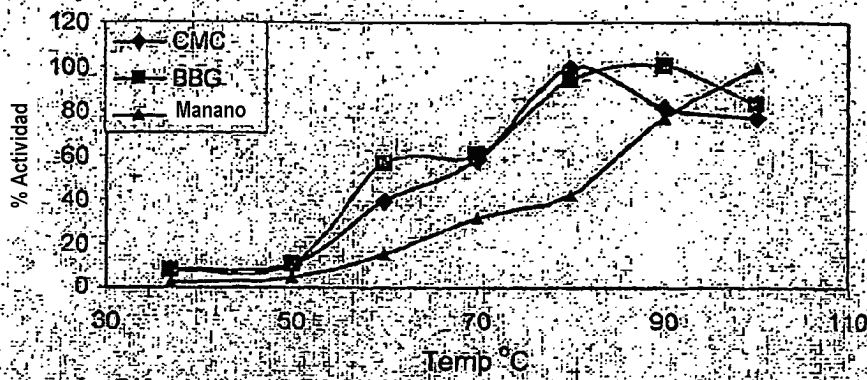


Figura 10

