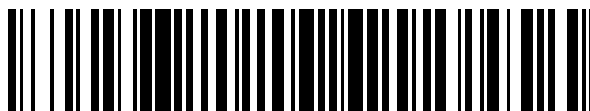


19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 518 916**

51 Int. Cl.:

C12N 9/24 (2006.01)

C12N 9/34 (2006.01)

C07K 14/37 (2006.01)

C12N 15/56 (2006.01)

C12N 15/31 (2006.01)

C12R 1/885 (2006.01)

C12N 15/62 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **08.04.2008 E 12164788 (7)**

97 Fecha y número de publicación de la concesión europea: **11.06.2014 EP 2479265**

54 Título: **Variantes de glucoamilasa con propiedades modificadas**

30 Prioridad:

09.10.2007 WO PCT/US2007/021683

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

05.11.2014

73 Titular/es:

**DANISCO US INC. (100.0%)
925 Page Mill Road
Palo Alto, CA 94304, US**

72 Inventor/es:

**AEHLE, WOLFGANG;
BOTT, RICHARD R.;
NIKOLAEV, IGOR;
SCHEFFERS, MARTIJN;
VAN SOLINGEN, PIET y
VROEMEN, CASPER**

74 Agente/Representante:

RIZZO, Sergio

ES 2 518 916 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

Variantes de glucoamilasa con propiedades modificadas

CAMPO DE LA INVENCION

5 [0001] La presente invención hace referencia a variantes de una glucoamilasa precursora que presentan propiedades modificadas (p. ej., termoestabilidad y/o actividad específica mejoradas). En concreto, la presente invención proporciona composiciones que comprenden las glucoamilasas variantes, entre las que se incluyen composiciones hidrolizantes de almidón, composiciones de pienso animal y composiciones de limpieza. La invención también hace referencia a construcciones de ADN que codifican las variantes y a métodos para producir las variantes de glucoamilasa en células huésped.

10 **ANTECEDENTES DE LA INVENCION**

15 [0002] Las enzimas de glucoamilasa (glucano 1,4- α -glucohidrolasas, EC 3.2.1.3) son carbohidrasas hidrolizantes de almidón que actúan en exo, que catalizan la eliminación de unidades de glucosa sucesivas de los extremos no reductores de moléculas relacionadas con oligosacáridos o polisacáridos o almidón. Las glucoamilasas pueden hidrolizar tanto los enlaces glucosídicos lineales como ramificados de almidón (p.ej., amilosa y amilopectina).

[0003] Las glucoamilasas se producen mediante numerosas cepas de bacterias, hongos, levaduras y plantas. Especialmente interesantes, y comercialmente importantes, las glucoamilasas son enzimas fúngicas que se producen de forma extracelular, por ejemplo, a partir de cepas de *Aspergillus* (Svensson *et al.* (1983) *Carlsberg Res. Commun.* 48:529-544; Boel *et al.*, (1984) *EMBO J.* 3:1097 - 1102; Hayashida *et al.*, (1989) *Agric. Biol. Chem.* 53:923 - 929; patente estadounidense 5.024.941; patente estadounidense 4.794.175 y WO 88/09795); *Talaromyces* (patente estadounidense 4.247.637; patente estadounidense 6.255.084 y patente estadounidense 6.620.924); *Rhizopus* (Ashikari *et al.*, (1986) *Agric. Biol. Chem.* 50:957 - 964; Ashikari *et al.*, (1989) *App. Microbiol. and Biotech.* 32:129 - 133 y patente estadounidense 4.863.864); *Humicola* (WO 05/052148 y patente estadounidense 4.618.579) y *Mucor* (Houghton-Larsen *et al.*, (2003) *Appl. Microbiol. Biotechnol.* 62:210 - 217). Muchos de los genes que codifican estas enzimas han sido clonados y expresados en células bacterianas, fúngicas y/o levadura.

30 [0004] Comercialmente, las glucoamilasas son enzimas muy importantes y se han utilizado en una amplia variedad de aplicaciones que requieren la hidrólisis de almidón (p. ej., para la producción de glucosa y otros monosacáridos a partir de almidón). Las glucoamilasas se utilizan para producir edulcorantes de maíz con alto contenido de fructosa, que comprenden más del 50 % del mercado de edulcorantes en Estados Unidos. En general, las glucoamilasas pueden utilizarse, y normalmente se utilizan, con alfa-amilasas en procesos hidrolizantes de almidón para hidrolizar almidón en dextrinas y posteriormente glucosa. La glucosa puede entonces convertirse en fructosa por parte de otras enzimas (p. ej., isomerasas de glucosa); cristalizada; o utilizada en fermentaciones con el fin de producir numerosos productos finales (p. ej., etanol, ácido cítrico, ácido láctico, succinato, intermediarios de ácido ascórbico, ácido glutámico, glicerol y 1,3-propanodiol). El etanol producido mediante el uso de glucoamilasas en la fermentación de material que contiene celulosa y/o almidón puede utilizarse como una fuente de combustible o para consumo alcohólico.

[0005] WO 2006/060062 expone la glucoamilasa *Trichoderma reesei* TrGA y los usos en conversión de almidón, pienso animal y fermentación de alcohol.

40 [0006] Aunque las glucoamilasas se han utilizado con éxito en aplicaciones comerciales durante muchos años, todavía existe la necesidad de nuevas glucoamilasas con propiedades modificadas.

SUMARIO DE LA INVENCION

45 [0007] En un primer aspecto, la presente invención proporciona una variante de glucoamilasa, que es una variante de una glucoamilasa precursora, presentando dicha glucoamilasa precursora una secuencia de aminoácidos con al menos un 95 % de identidad de secuencia con la secuencia de aminoácidos:

ES 2 518 916 T3

SVDDFISTETPIALNNLLCNVGPDGCRAFGTSAAGAVIASPSTIDPDYYMWTIRDSALVFKNLIDRFTE
TYDAGLQORRIEQYITAQVTLQGLSNPSGSLADGSGLGEPKFELTLKPF TGNWGRPQRDGPALRAIALI
GYSKWLINNNYQSTVSNVIWPIVRNDLNYVAQYWNQTGFDLWEEVNGSSFFFTVANQHRALVEGATLAA
TLGQSGSAYSSVAPQVLCFLQRFVWSSGGYVDSNINTNEGRITGKDVNSVLTSIHTFDPNLGCDAGTFQ
PCSDKALS NLKVVVDSFRSIYGVNKGIPAGAAVAIGRYAEDVYYNGNPWYLATFAAAEQLYDAIYVWK
KTGSITVTATSLAFFQELVPGVTAGTYSSSSSTFTNI INAVSTYADGFLSEAAKYVPADGSLAEQFDR
NSGTPLSALHLTWSYASFLTATARRAGIVPPSWANS SASTIPSTCSGASVVGYSRPTATSFPPSQTP
KPGVPSGTPYTPLPCATPTSVAVTFHELVSTQFGQTVKVAGNAAALGNWSTSAVALDAVNYADNHPL
WIGTVNLEAGDVVEYKYINVGQDGSVTWESDPNHTYTVPAVACVTQVVKEDTWQS (SEQ ID NO:
2),

presentando dicha variante solo una, dos, tres o cuatro sustituciones comparada con dicha precursora, en la que dichas sustituciones comprenden la sustitución T430A, T430E, T430F, T430G, T430H, T430I, T430K, T430M, T430N, T430Q, T430R o T430V en SEQ ID NO:2 o una posición equivalente en dicha glucoamilasa precursora como se determina mediante la alineación de secuencia.

[0008] En algunos casos, la variante de glucoamilasa es una variante de una glucoamilasa precursora que presenta al menos un 97 % o al menos un 98 % de identidad de secuencia con la secuencia de SEQ ID NO:2.

[0009] En algunos casos, dicha variante presenta únicamente dos, tres o cuatro sustituciones comparada con dicha precursora, sustituciones adicionales que se eligen de entre sustituciones en una posición correspondiente a la posición 10, 14, 15, 23, 42, 45, 46, 59, 60, 61, 67, 68, 72, 73, 97, 98, 99, 102, 108, 110, 113, 114, 122, 124, 125, 133, 140, 144, 145, 147, 152, 53, 164, 175, 182, 204, 205, 214, 216, 219, 228, 229, 230, 231, 236, 239, 240, 241, 242, 244, 263, 264, 265, 268, 269, 276, 284, 291, 300, 301, 303, 310, 311, 313, 316, 338, 342, 344, 346, 349, 359, 361, 364, 379, 382, 390, 391, 393, 394, 408, 410, 415, 417, 418, 431, 433, 436, 442, 443, 444, 448 y 451 de SEQ ID NO:2 o una posición equivalente en dicha glucoamilasa precursora como se determina mediante la alineación de secuencia.

[0010] En algunos casos, la variante presenta actividad específica aumentada comparada con la glucoamilasa que comprende la secuencia de SEQ ID NO:2.

[0011] En algunos casos, la variante presenta termoestabilidad aumentada comparada con la glucoamilasa que comprende la secuencia de SEQ ID NO:2.

[0012] En algunos casos, la variante presenta actividad específica aumentada y termoestabilidad aumentada comparada con la glucoamilasa que comprende SEQ ID NO:2.

[0013] En algunos casos, la glucoamilasa precursora se obtiene a partir de una *Trichoderma spp.*

[0014] En un segundo aspecto, la presente invención proporciona un polinucleótido que codifica la variante del primer aspecto de la invención.

[0015] En un tercer aspecto, la presente invención proporciona una célula huésped que comprende un polinucleótido del segundo aspecto de la invención.

[0016] En un cuarto aspecto, la presente invención proporciona una composición de enzima que comprende la variante de glucoamilasa del primer aspecto de la invención.

[0017] En algunos casos, la composición es adecuada para utilizarse en un proceso de conversión de almidón, una composición de pienso animal o un proceso de fermentación de sustrato con almidón.

[0018] En algunos casos, la composición comprende además un alfa amilasa.

[0019] En un quinto aspecto, la presente invención proporciona un método para producir una glucoamilasa variantes en una célula huésped que comprende:

la transformación de una célula huésped con una construcción de ADN que comprende un polinucleótido del segundo aspecto de la invención y el cultivo de la célula huésped en las condiciones adecuadas para la expresión y producción de dicha variante de glucoamilasa y la producción de dicha variante.

[0020] En algunos casos, el método comprende además la recuperación de la variante de glucoamilasa de dicho cultivo.

BREVE DESCRIPCIÓN DE LOS DIBUJOS

[0021]

- 5 La fig. 1A ilustra una glucoamilasa *Trichoderma reesei* (TrGA) que presenta 632 aminoácidos (SEQ ID NO: 1) El péptido señal se subraya, la región catalítica (SEQ ID NO:3) que comienza con los residuos de aminoácidos SVDDFI (SEQ ID NO:12) y que presenta 453 residuos de aminoácidos está en negrita; la región de enlace está en cursiva y el dominio de unión a almidón (SBD) está tanto en cursiva como subrayado. La proteína madura que incluye el dominio catalítico (SEQ ID NO:3), la región de enlace (SEQ ID NO:10) y el dominio de unión a almidón (SEQ ID NO:11) se representan mediante SEQ ID NO:2. La fig. 1B ilustra el ADNc (SEQ ID NO:4) que codifica la TrGA. La fig. 1C ilustra los dominios de TrGA de proteína madura y precursora.
- 10 La fig. 2 ilustra el plásmido de destino pDONR-TrGA que incluye el ADNc (SEQ ID NO:4) de la TrGA. La fig. 3 ilustra el plásmido pTTT-Dest.
- 15 La fig. 4 ilustra el vector de expresión final pTTT-TrGA. Las figs. 5A-5B ilustran una comparación de alineación de los dominios catalíticos de las glucoamilasas precursoras que incluyen glucoamilasa derivada de *Aspergillus awamori* (AaGA) (SEQ ID NO:5); *Aspergillus niger* (AnGA) (SEQ ID NO:6); *Aspergillus oryzae* (AoGA) (SEQ ID NO:7); *Trichoderma reesei* (TrGA) (SEQ ID NO:3); *Humicola grisea* (HgGA) (SEQ ID NO:8) e *Hypocrea vinosa* (HvGA) (SEQ ID NO:9). Los aminoácidos idénticos se indican con un asterisco (*).
- 20 La fig. 5C ilustra la secuencia de proteína madura de glucoamilasa *Talaromyces* (TeGA) (SEQ ID NO:308). La fig. 6 es una comparación de las estructuras tridimensionales de la glucoamilasa *Trichoderma* (negra) (SEQ ID NO:2) y la glucoamilasa *Aspergillus awamori* (gris) vistas desde el lateral. El lateral se mide en referencia al centro activo y la entrada al centro activo se encuentra en "la parte superior" de la molécula.
- 25 La fig. 7 es una comparación de las estructuras tridimensionales de la glucoamilasa *Trichoderma* (negra) y la glucoamilasa *Aspergillus awamori* (gris) vistas desde la parte superior. La entrada al centro activo se encuentra en la "parte superior" de la molécula.

DESCRIPCIÓN DETALLADA DE LA INVENCION

I. Definiciones

- 30 **[0022]** A menos que se defina lo contrario, todos los términos técnicos y científicos utilizados en el presente documento tienen el mismo significado según lo entiende comúnmente alguien especializado en la técnica a la que esta invención pertenece. Singleton, *et al.*, *DICTIONARY OF MICROBIOLOGY AND MOLECULAR BIOLOGY*, 2D ED., John Wiley and Sons, New York (1994), y Hale & Markham, *THE HARPER COLLINS DICTIONARY OF BIOLOGY*, Harper Perennial, N.Y. (1991) proporcionan al experto el significado general de muchos de los términos empleados en el presente documento. Aún así, determinados términos se definen a continuación para una mayor claridad y facilidad de referencia.

[0023] El término "glucoamilasa (EC 3.2.1.3)" se refiere en el presente documento a una enzima que cataliza la liberación de D-glucosa desde los extremos no reductores de almidón y polisacáridos y oligosacáridos relacionados.

- 40 **[0024]** El término "precursor" o "secuencia precursora" hace referencia en el presente documento a una secuencia que presenta una identidad estructural y/o de secuencia con TrGA (p. ej., SEQ ID NO: 1, 2 y/o 3) y se trata de una secuencia natural o que se da de forma natural en una célula huésped.

- [0025]** El término "TrGA" hace referencia en el presente documento a una secuencia de la glucoamilasa *Trichoderma reesei* que presenta la secuencia de proteína madura ilustrada en SEQ ID NO:2, que incluye el dominio catalítico que tiene la secuencia ilustrada en SEQ ID NO:3. El aislamiento, clonación y expresión de la TrGA se describen en WO 2006/060062 y la publicación de patente estadounidense nº 2006/0094080 publicada el 4 de mayo, 2006. La TrGA también se considera una secuencia de glucoamilasa precursora. En algunas formas de realización, la secuencia precursora hace referencia a una TrGA que es el punto inicial para la ingeniería de proteínas. La numeración de los aminoácidos de glucoamilasa en el presente documento se basa en la secuencia de la secuencia de glucoamilasa TrGA (SEQ ID NO:2 y SEQ ID NO:3).

[0026] La frase "forma madura de una proteína o polipéptido" hace referencia a la forma funcional final de la proteína o polipéptido. Con el fin de ejemplificarlo, una forma madura de TrGA incluye el dominio catalítico, la región de enlace y el dominio de unión a almidón que presenta la secuencia de aminoácidos de SEQ ID NO:2.

[0027] El término "homólogos de glucoamilasa *Trichoderma*" hace referencia a glucoamilasas precursoras que

presentan al menos un 80 % de identidad de secuencia de aminoácidos con la secuencia TrGA (SEQ ID NO:1, SEQ ID NO:2 o SEQ ID NO:3) y que retienen las características funcionales de una glucoamilasa.

5 **[0028]** El término una "secuencia homóloga" se refiere en el presente documento a una secuencia de ácido nucleico o de polipéptidos que presenta al menos un 100 %, al menos un 99 %, al menos un 98 %, al menos un 97 %, al menos un 96 %, al menos un 95 %, al menos un 94 %, al menos un 93 %, al menos un 92 %, al menos un 91 %, al menos un 90 %, al menos un 88 %, al menos un 85 %, al menos un 80 %, al menos un 75 %, al menos un 70 %, al menos un 65 %, al menos un 60 %, al menos un 55 %, al menos un 50 % o al menos un 45 % de identidad de secuencia con una secuencia de ácido nucleico o una secuencia de polipéptidos cuando se
10 alinea de forma óptima para compararla, en la que la función de la secuencia de ácido nucleico o secuencia de polipéptidos candidata es básicamente la misma que la secuencia de ácido nucleico o secuencia de polipéptidos con la que se compara dicha secuencia homóloga candidata. En algunas formas de realización, las secuencias homólogas presentan entre un 85 % y un 100 % de identidad de secuencia, mientras que en otras formas de realización existe un 90 % y un 100 % de identidad de secuencia y en otras formas de realización, existe un 95 % y un 100 % de identidad de secuencia. En algunas formas de realización, la secuencia homóloga candidata
15 o precursora se compara con la secuencia de ácido nucleico de TrGA o la secuencia de proteína madura. La identidad de secuencia puede medirse sobre toda la longitud de la secuencia precursora u homóloga.

20 **[0029]** Los términos "variante de glucoamilasa", "variante" y "variante de TrGA" se emplean en el presente documento en referencia a glucoamilasas que son similares a una secuencia de glucoamilasa precursora (p. ej., los homólogos de glucoamilasa *Trichoderma* o TrGA) pero presentan al menos una sustitución, delección o inserción en su secuencia de aminoácidos que hace que sean diferentes de una glucoamilasa precursora por la secuencia. En algunos casos, han sido manipuladas/diseñadas para que incluyan al menos una sustitución, delección o inserción en su secuencia de aminoácidos que las diferencie de una glucoamilasa precursora por la secuencia.

25 **[0030]** El término "dominio catalítico" hace referencia en el presente documento a una región estructural de un polipéptido que contiene el centro activo para la hidrólisis del sustrato.

[0031] El término "enlazador" se refiere a una secuencia de aminoácidos corta que presenta generalmente entre 3 y 40 residuos de aminoácido que unen de forma covalente una secuencia de aminoácidos que comprende un dominio de unión a almidón con una secuencia de aminoácidos que comprende un dominio catalítico.

30 **[0032]** El término "dominio de unión a almidón" se refiere a una secuencia de aminoácidos que se une preferentemente a un sustrato de almidón.

35 **[0033]** Los términos "secuencia mutante" y "gen mutante" se utilizan en el presente documento indistintamente y hacen referencia a una secuencia de polinucleótidos que presenta una alteración en al menos un codón que se da en una secuencia original de una célula huésped. El producto de expresión de la secuencia mutante es una proteína variante con una secuencia de aminoácidos modificada en relación con la precursora. El producto de expresión puede tener una capacidad funcional modificada (p. ej., actividad enzimática mejorada).

40 **[0034]** El término "propiedad" o equivalentes gramaticales del mismo en el contexto de un polipéptido se refiere en el presente documento a cualquier característica o atributo de un polipéptido que pueda elegirse o detectarse. Entre estas propiedades se incluyen, sin carácter limitativo, estabilidad oxidativa, especificidad de sustrato, actividad catalítica, estabilidad térmica, perfil de actividad de pH, resistencia a degradación proteolítica, K_M , K_{CAT} , relación K_{CAT}/K_M , plegamiento de proteínas, capacidad de unirse a un sustrato y capacidad de ser secretado.

45 **[0035]** El término "propiedad" o equivalentes gramaticales del mismo en el contexto de un ácido nucleico se refiere en el presente documento a cualquier característica o atributo de un ácido nucleico que pueda elegirse o detectarse. Entre estas propiedades se incluyen, sin carácter limitativo, una propiedad que afecta a la transcripción de genes (p. ej., fuerza del promotor o reconocimiento del promotor), una propiedad que afecta al procesamiento de ARN (p. ej., corte y empalme de ARN y estabilidad de ARN), una propiedad que afecta a la traducción (p. ej., regulación, unión de ARNm a proteínas ribosómicas).

50 **[0036]** Los términos "estable térmicamente" y "termoestable" hacen referencia a variantes de glucoamilasa de la presente invención que retienen una cantidad específica de actividad enzimática tras la exposición a temperaturas identificadas sobre un periodo de tiempo determinado en condiciones que se mantienen durante la hidrólisis de sustratos de almidón, por ejemplo, mientras se exponen a temperaturas modificadas.

[0037] El término "estabilidad mejorada" en el contexto de una propiedad tal como termoestabilidad se refiere a una mayor actividad hidrolítica de almidón retenido con el tiempo en comparación con otra glucoamilasa de referencia (p. ej., precursora).

[0038] El término "estabilidad disminuida" en el contexto de una propiedad tal como termoestabilidad se refiere

a una menor actividad hidrolítica de almidón retenido con el tiempo en comparación con otra glucoamilasa de referencia.

[0039] El término "actividad específica" La actividad específica se define como la actividad por mg de proteína de glucoamilasa. En algunas formas de realización, la actividad de glucoamilasa se determina mediante el análisis de etanol descrito en el presente documento y expresado como la cantidad de glucosa que se produce del sustrato de almidón. En algunas formas de realización, la concentración de proteína puede determinarse mediante el análisis Caliper descrito en el presente documento.

[0040] Los términos "activo" y "biológicamente activo" se refieren a una actividad biológica asociada con una proteína específica. Por consiguiente, la actividad biológica de una proteína determinada se refiere a cualquier actividad biológica típicamente atribuida a aquella proteína por los expertos en la técnica. Por ejemplo, una actividad enzimática asociada a una glucoamilasa es hidrolítica y, por tanto, una glucoamilasa activa presenta actividad hidrolítica.

[0041] Los términos "polinucleótido" y "ácido nucleico", utilizados indistintamente en el presente documento, se refieren a una forma polimérica de nucleótidos de cualquier longitud, ya sean ribonucleótidos o desoxirribonucleótidos. Entre estos términos se incluyen, aunque sin carácter limitativo, ADN de cadena simple, doble o triple, ADN genómico, ADNc, ARN, híbrido ADN-ARN o un polímero que comprenda bases de pirimidina y purina u otras bases de nucleótidos naturales, modificadas de forma química, bioquímica, no naturales o derivatizadas.

[0042] Los términos "construcción de ADN", "ADN transformante" y "vector de expresión" se usan indistintamente para referirse al ADN utilizado para introducir secuencias en una célula huésped u organismo. El ADN puede generarse *in vitro* mediante PCR o cualquier otra técnica adecuada conocida por los expertos en la técnica. La construcción de ADN, el ADN transformante o el casete de expresión recombinante pueden incorporarse en un plásmido, cromosoma, ADN mitocondrial, ADN plastidial, virus o fragmento de ácido nucleico. Normalmente, la parte del casete de expresión recombinante de un vector de expresión, construcción de ADN o ADN transformante incluye, entre otras secuencias, una secuencia de ácido nucleico que ha de transcribirse y un promotor. En formas de realización, los vectores de expresión tienen la capacidad de incorporar y expresar fragmentos de ADN heterólogos en una célula huésped.

[0043] El término "vector" se refiere en el presente documento a una construcción de polinucleótidos diseñada para introducir ácidos nucleicos en uno o más tipos de células. Entre los vectores se incluyen vectores de clonación, vectores de expresión, vectores transportadores, plásmidos, casetes y similares.

[0044] Tal y como se emplea en el presente documento en el contexto de introducción de una secuencia de ácido nucleico en una célula, el término "introducido" hace referencia a cualquier método adecuado para transferir la secuencia de ácido nucleico a la célula. Tales métodos para la introducción incluyen sin carácter limitativo fusión de protoplasto, transfección, transformación, conjugación y transducción.

[0045] Los términos "transformado" y "transformado de forma estable" hacen referencia en el presente documento a una célula que tiene una secuencia de polinucleótidos no natural (heteróloga) integrada en su genoma o como un plásmido episomal que se mantiene durante al menos dos generaciones.

[0046] Los términos "marcador seleccionable" y "marcador selectivo" hacen referencia en el presente documento a un ácido nucleico (p. ej., un gen) capaz de expresarse en células huésped que permite la facilidad de selección de aquellos huéspedes que contienen el vector. Normalmente, los marcadores seleccionables son genes que conceden resistencia antimicrobiana o una ventaja metabólica en la célula huésped con el fin de permitir que las células que contienen el ADN exógeno se distingan de las células que no han recibido ninguna secuencia exógena durante la transformación.

[0047] El término "promotor" hace referencia en el presente documento a una secuencia de ácido nucleico que funciona para dirigir la transcripción de un gen aguas abajo. El promotor, junto con otras secuencias de ácido nucleico reguladoras de traducción y transcripción (también denominadas "secuencias de control"), es necesario para expresar un gen determinado. En general, entre las secuencias reguladoras de traducción y transcripción se incluyen, sin carácter limitativo, secuencias promotoras, lugares de unión ribosómicos, secuencias de transcripción de inicio y parada, secuencias de traducción de inicio y de parada y secuencias potenciadoras o activadoras.

[0048] Un ácido nucleico está "unido de forma operativa" cuando se coloca en una relación funcional con otra secuencia de ácido nucleico. Por ejemplo, ADN que codifica un líder secretor (es decir, un péptido señal), está unido de forma operativa a un ADN para un polipéptido si se expresa como una preproteína que participa en la secreción del polipéptido. Generalmente, "unido de forma operativa" indica que las secuencias de ADN que están unidas son contiguas y, en el caso de un líder secretor, contiguo y en fase de lectura.

[0049] El término "gen" se refiere en el presente documento a un polinucleótido (p. ej., un segmento de ADN) que codifica un polipéptido e incluye regiones que preceden y siguen a las regiones codificantes así como secuencias intermedias (intrones) entre segmentos codificantes individuales (exones).

5 **[0050]** El término "genes homólogos" se refiere en el presente documento a un par de genes de diferentes especies pero normalmente relacionadas, que se corresponden una a la otra y que son idénticas o muy similares una a la otra. El término abarca genes que se separan mediante especiación (es decir, el desarrollo de nuevas especies) (p. ej., genes ortólogos), así como genes que se han separado mediante duplicación genética (p. ej., genes parálogos).

10 **[0051]** Los términos "ortólogo" y "genes ortólogos" se refieren en el presente documento a genes en diferentes especies que han evolucionado de un gen ancestral común (es decir, un gen homólogo) mediante especiación. Normalmente, los ortólogos conservan la misma función durante el transcurso de la evolución. La identificación de ortólogos es útil a la hora de predecir de forma fiable la función del gen en genomas secuenciados recientemente.

15 **[0052]** Los términos "parálogo" y "genes parálogos" se refieren en el presente documento a genes que están relacionados por duplicación dentro de un genoma. Mientras que los ortólogos conservan la misma función durante el transcurso de la evolución, los parálogos desarrollan nuevas funciones, aunque algunas funciones están relacionadas normalmente con la original. Entre los ejemplos de genes parálogos se incluyen, sin carácter limitativo, genes que codifican tripsina, quimotripsina, elastasa y trombina, que son todas serina proteasas y se dan de forma conjunta en las mismas especies.

20 **[0053]** El término "homología" se refiere en el presente documento a identidad o similitud de secuencia, prefiriéndose la identidad. Esta homología se determina mediante técnicas convencionales conocidas en la técnica (véase p. ej., Smith y Waterman, (1981) *Adv. Appl. Math.*, 2:482; Needleman y Wunsch, (1988) *J. Mol. Biol.*, 48:443; Pearson y Lipman, (1988) *Proc. Natl. Acad. Sci. EEUU* 85:2444; programas como GAP, BESTFIT, FASTA y TFASTA en Wisconsin Genetics Software Package (Genetics Computer Group, Madison, WI); y Devereux *et al.*, (1984) *Nucl. Acid Res.*, 12:387-395).

25 **[0054]** El "porcentaje (%) de identidad de secuencia de ácido nucleico" o "porcentaje (%) de identidad de secuencia de aminoácidos" se define como el porcentaje de residuos de nucleótidos o residuos de aminoácidos en una secuencia candidata que son idénticos a los residuos de nucleótidos o residuos de aminoácidos de la secuencia inicial (es decir, TrGA). La identidad de secuencia puede medirse sobre la longitud total de la secuencia de inicio (p. ej., SEQ ID NO:2 o 3 de TrGA).

30 **[0055]** Las secuencias homólogas se determinan mediante métodos conocidos de alineación de secuencias. Un método de alineación utilizado normalmente es BLAST descrito por Altschul *et al.*, (Altschul *et al.*, (1990) *J. Mol. Biol.*, 215:403-410; y Karlin *et al.*, (1993) *Proc. Natl. Acad. Sci. EEUU* 90:5873-5787). Un programa BLAST especialmente útil es el programa WU-BLAST-2 (Véase, Altschul *et al.*, (1996) *Meth. Enzymol.*, 266:460-480). WU-BLAST-2 utiliza diferentes parámetros de búsqueda, la mayoría de los cuales están establecidos a los valores por defecto. Los parámetros regulables están configurados con los siguientes valores: período de solapamiento = 1, fracción de solapamiento = 0,125, umbral de palabras (T) = 11. Los parámetros HSP S y HSP S2 son valores dinámicos y los establece el propio programa dependiendo de la composición de la secuencia específica y la composición de la base de datos específica frente la que se busca la secuencia de interés. Sin embargo, los valores pueden regularse con el fin de aumentar la sensibilidad. Un valor de % de identidad de secuencia de aminoácidos se determina por el número de residuos idénticos emparejados dividido por el número total de residuos de la secuencia "más larga" en la región alineada. La secuencia "más larga" es la que presenta los residuos más reales en la región alineada (se ignoran los huecos introducidos por WU-Blast-2 con el fin de maximizar la puntuación de alineación).

35 **[0056]** Otros métodos pueden ser útiles para alinear secuencias. Un ejemplo de un algoritmo útil es PILEUP. PILEUP crea una alineación de secuencias múltiple de un grupo de secuencias relacionadas utilizando alineaciones por pares progresivas. También puede trazar un árbol que muestre las relaciones de agrupamiento utilizadas para crear la alineación. PILEUP utiliza una simplificación del método de alineación progresiva de Feng y Doolittle (Feng y Doolittle, (1987) *J. Mol. Evol.*, 35:351-360). El método es similar al descrito por Higgins y Sharp (Higgins y Sharp, (1989) *CABIOS* 5:151-153). Parámetros útiles de PILEUP incluyen un peso del hueco por defecto de 3,00, un peso de la longitud del hueco por defecto de 0,10 y huecos de extremo ponderados.

40 **[0057]** El término "alineación óptima" hace referencia a la alineación que proporciona la mayor puntuación de porcentaje de identidad. Una "posición equivalente" se refiere a una alineación óptima entre dos secuencias. Por ejemplo, utilizando las figs. 5D y 5E, la posición 491 en TrGA (SEQ ID NO:2) es C491; la posición equivalente para *Aspergillus niger* es la posición C509; y la posición equivalente para *Aspergillus awamori* es la posición Q538. Véase la figura 8 para una alineación de ejemplo de la secuencia tridimensional.

[0058] El término "hibridación" se refiere en el presente documento al proceso por el que una hebra de ácido nucleico se une con una hebra complementaria a través del emparejamiento de bases, como se conoce en la técnica.

5 **[0059]** Se considera que una secuencia de ácido nucleico es "hibrizable de forma selectiva" en referencia a una secuencia de ácido nucleico si dos secuencias se hibridan una a la otra de forma específica bajo condiciones de lavado e hibridación con una astringencia de moderada a alta. Las condiciones de hibridación se basan en la temperatura de fusión (T_m) de la sonda o complejo que une ácido nucleico. Por ejemplo, la "astringencia máxima" normalmente se da a aproximadamente T_m-5°C (5° por debajo de la T_m de la sonda); "astringencia alta" a aproximadamente 5-10 °C por debajo de la T_m; "astringencia intermedia" a aproximadamente 10-20 °C por debajo de la T_m de la sonda y "baja astringencia" a aproximadamente 20-25°C por debajo de la T_m. De forma funcional, pueden utilizarse las condiciones de astringencia máxima para identificar secuencias que tengan identidad precisa o casi precisa con la sonda de hibridación; mientras que se puede utilizar una hibridación de astringencia intermedia o baja para identificar o detectar homólogos de secuencia de polinucleótidos.

15 **[0060]** Las condiciones de hibridación de astringencia moderada y alta son ya conocidas en la técnica. Un ejemplo de condiciones de astringencia alta incluye la hibridación a aproximadamente 42 °C en formamida 50 %, 5X SSC, 5X solución Denhardt, SDS 0,5 % y 100 µg/ml de ADN portador desnaturizado seguido de dos lavados en 2X SSC y SDS 0,5 % a temperatura ambiente y dos veces adicionales en 0,1X SSC y SDS 0,5 % a 42 °C. Un ejemplo de condiciones astringentes moderadas incluye una incubación durante la noche a 37 °C en una solución que comprende formamida 20 %, 5 x SSC (150 mM NaCl, 15 mM citrato trisódico), 50 mM fosfato de sodio (pH 7,6), 5 x solución Denhardt, sulfato de dextrano 10 % y 20 mg/ml ADN de esperma de salmón cortado desnaturizado seguido del lavado de los filtros en 1x SSC a aproximadamente 37 - 50 °C. Aquellos expertos en la técnica saben cómo regular la temperatura, la fuerza iónica, etc. según sea necesario para adaptarse a factores tales como la longitud de sonda y similares.

25 **[0061]** El término "recombinante" en el presente documento incluye referencia a una célula o vector que ha sido modificado mediante la introducción de una secuencia de ácido nucleico homóloga o heteróloga o que la célula se ha derivado de una célula modificada de ese modo. Por lo tanto, por ejemplo, las células recombinantes expresan genes que no se hallan en forma idéntica dentro de la forma natural (no recombinante) de la célula o expresan genes naturales que se han expresado de otro modo de forma anormal, subexpresado o no expresado en general como resultado de una intervención humana deliberada.

30 **[0062]** En algunas formas de realización de la presente exposición, las secuencias de ADN mutado se generan con mutagénesis por saturación del lugar en al menos un codón. En otras formas de realización, la mutagénesis por saturación del lugar se lleva a cabo en dos o más codones. En una forma de realización adicional, las secuencias de ADN mutante presentan más del 50 %, más del 55 %, más del 60 %, más del 65 %, más del 70 %, más del 75 %, más del 80 %, más del 85 %, más del 90 %, más del 95 %, más del 98 % o más del 99 % de homología con la secuencia precursora. En formas de realización alternativas, el ADN mutante se genera *in vivo* utilizando cualquier procedimiento mutagénico tal como, por ejemplo, radiación, nitrosoguanidina y similares. A continuación, la secuencia de ADN deseada se aísla y se utiliza en los métodos proporcionados en el presente documento.

40 **[0063]** El término "proteína heteróloga" se refiere en el presente documento a una proteína o polipéptido que no se da de forma natural en la célula huésped.

[0064] El término "proteína homóloga" se refiere en el presente documento a una proteína o polipéptido natural o que se da de forma natural en una célula e incluye proteínas naturales que están sobreexpresadas en la célula ya sea mediante tecnología de ADN recombinante o de forma natural.

45 **[0065]** Una enzima está "sobreexpresada" en una célula huésped si la enzima se expresa en la célula con un nivel mayor que el nivel al que se expresa en una célula de tipo silvestre correspondiente.

50 **[0066]** Los términos "proteína" y "polipéptido" se utilizan indistintamente en el presente documento. En la presente exposición y reivindicaciones, se utilizan los códigos de una letra y tres letras convencionales para residuos de aminoácidos. Los códigos de tres letras para aminoácidos se definen según la Comisión sobre nomenclatura bioquímica IUPAC-IUB (JCBN). También se entiende que un polipéptido puede codificarse mediante más de una secuencia de nucleótidos debido a la degeneración del código genético.

55 **[0067]** Las variantes de la presente exposición se describen mediante la siguiente nomenclatura: [residuo aminoácido original/posición/residuo aminoácido sustituido]. Por ejemplo, la sustitución de leucina por arginina en la posición 76 se representa como R76L. Cuando se sustituye más de un aminoácido en una determinada posición, la sustitución se representa como 1) Q172C, Q172D o Q172R; 2) Q172C, D, o R o c) Q172C/D/R. Cuando se identifica en la misma una posición adecuada para la sustitución sin un aminoácido específico sugerido, se ha de entender que cualquier residuo de aminoácido puede sustituirse por el residuo de aminoácido

presente en la posición. Cuando una glucoamilasa variante contiene una delección en comparación con otras glucoamilasas la delección se indica con "***". Por ejemplo, una delección en la posición R76 se representa como R76*. Una delección de dos o más aminoácidos consecutivos se indica por ejemplo como (76-78)*.

5 **[0068]** Una "prosecuencia" es una secuencia de aminoácidos entre la secuencia señal y la proteína madura que es necesaria para la secreción de la proteína. La escisión de la prosecuencia tendrá como resultado una proteína activa madura.

10 **[0069]** El término "secuencia señal" o "péptido señal" se refiere a cualquier secuencia de nucleótidos y/o aminoácidos que puedan participar en la secreción de las formas precursoras o maduras de la proteína. Esta definición de secuencia señal es una funcional, que pretende incluir todas las secuencias de aminoácidos codificadas por la parte N-terminal del gen de la proteína, que participan en la realización de la secreción de proteína. Normalmente, aunque no de forma universal, están ligados a la parte N-terminal de una proteína o a la parte N-terminal de una proteína precursora. La secuencia señal puede ser endógena o exógena. La secuencia señal puede estar asociada normalmente a la proteína (p. ej., glucoamilasa) o puede ser de un gen que codifica otra proteína secretada.

15 **[0070]** El término forma "precursora" de una proteína o péptido se refiere a la forma madura de la proteína que tiene una prosecuencia unida de forma operativa al carbonilo o amino-terminal de la proteína. El precursor puede también tener una secuencia "señal" unida de forma operativa al amino-terminal de la prosecuencia. El precursor puede también presentar polipéptidos adicionales que estén involucrados en la actividad traduccional posterior (p. ej., polipéptidos escindidos de este para dejar la forma madura de una proteína o péptido).

20 **[0071]** Los términos "cepa huésped" o "célula huésped" se refieren a un huésped adecuado para un vector de expresión que comprende ADN según la presente invención.

25 **[0072]** Los términos "derivado de" y "obtenido de" se refieren no solo a una glucoamilasa producida o que se puede producir mediante una cepa del organismo en cuestión, sino también a una glucoamilasa codificada por una secuencia de ADN aislada de tal cepa y producida en un organismo huésped que contiene dicha secuencia de ADN. Además, el término se refiere a una glucoamilasa que está codificada por una secuencia de ADN de origen sintético y/o ADNc y que tiene las características de identificación de la glucoamilasa en cuestión.

30 **[0073]** Un "derivado" dentro del alcance de la presente definición generalmente conserva la característica actividad hidrolizante observada en la forma de tipo silvestre, natural o precursora hasta el punto de que el derivado es útil para fines similares como la forma precursora, natural o de tipo silvestre. Los derivados funcionales de glucoamilasas abarcan péptidos o fragmentos de péptidos que se dan de forma natural, se producen de forma sintética o recombinante que presentan las características generales de las glucoamilasas de la presente invención.

35 **[0074]** Los términos "aislado" o "purificado" se refieren a un material que se ha extraído de su medio original (p. ej., el medio natural si se da de forma natural). En algunas formas de realización, una proteína aislada presenta más de un 10 % de pureza, preferiblemente más de un 20 % de pureza e incluso más preferiblemente más de un 30 % de pureza, según se determina mediante SDS-PAGE. Aspectos adicionales de la presente exposición incluyen la proteína en forma altamente purificada (es decir, más de un 40 % de pureza, más de un 60 % de pureza, más de un 80 % de pureza, más de un 90 % de pureza, más de un 95 % de pureza, más de un 97 % de pureza y incluso más de un 99 % de pureza), según se determina mediante SDS-PAGE.

40 **[0075]** El término "mutagénesis combinatoria" hace referencia en el presente documento a métodos en los que se generan las bibliotecas de variantes de una secuencia inicial. En estas bibliotecas, las variantes contienen una o varias mutaciones elegidas de un conjunto predefinido de mutaciones. Además, los métodos proporcionan medios para introducir mutaciones aleatorias que no eran elementos del conjunto predefinido de mutaciones. En algunas formas de realización, los métodos incluyen aquellos descritos en la patente estadounidense 6.582.914.
45 En formas de realización alternativas, los métodos de mutagénesis combinatoria incluyen kits disponibles comercialmente (p. ej., QUIKCHANGE® Multisite, Stratagene, San Diego, CA).

[0076] El término "biblioteca de mutantes" hace referencia en el presente documento a una población de células que son idénticas en la mayoría de su genoma pero incluyen diferentes homólogos de uno o más genes. Tales bibliotecas pueden utilizarse, por ejemplo, para identificar genes u operones con rasgos mejorados.

50 **[0077]** El término "contenido de sólidos secos (DS o ds por sus siglas en inglés)" hace referencia en el presente documento a los sólidos totales de una suspensión en % sobre una base de peso en seco.

[0078] El término "resultado inicial" hace referencia en el presente documento a una variante que se identificó mediante la detección de una biblioteca de mutagénesis combinatoria por consenso. En formas de realización, los resultados iniciales presentan características de actuación mejoradas, comparadas con el gen inicial.

[0079] El término "resultado mejorado" hace referencia en el presente documento a una variante que se identificó mediante la detección de una biblioteca de mutagénesis combinatoria por consenso mejorada.

[0080] El término "propiedad diana" hace referencia en el presente documento a la propiedad del gen inicial que ha de modificarse. No se pretende que la presente invención esté limitada por ninguna propiedad diana específica. Sin embargo, en algunas formas de realización, la propiedad diana es la estabilidad de un producto génico (p. ej., resistencia a la desnaturalización, proteólisis u otros factores de degradación), mientras que en otras formas de realización, el nivel de producción en un huésped de producción se modifica. De hecho, se contempla que cualquier propiedad de un gen inicial será útil en la presente invención. Pueden aparecer otras definiciones de los términos en toda la memoria.

[0081] Antes de describir las formas de realización de ejemplo con más detalle, ha de entenderse que la presente invención no está limitada a las formas de realización específicas descritas en el presente documento, ya que tales pueden variar. También ha de entenderse que la terminología empleada en el presente documento tiene el fin de describir formas de realización específicas únicamente y no pretende poseer carácter limitativo.

[0082] En los casos en los que se presente una gama de valores, ha de entenderse que también se indica de forma específica cada valor intermedio, hasta la décima de la unidad del límite inferior a menos que el contexto indique claramente lo contrario, entre los límites superior e inferior de cada gama. Cada gama menor entre cualquier valor indicado o valor intermedio en una gama indicada y cualquier otro valor indicado o intermedio en esa gama indicada se incluye dentro de la invención. Los límites superior e inferior de estas gamas menores pueden incluirse o excluirse de forma independiente en la gama, y cada gama en la que cualquier límite, ambos o ninguno se incluyan en las gamas menores también se incluye dentro de la invención, sujeta a cualquier límite excluido de forma específica en la gama indicada. En los casos en los que la gama indicada incluya uno o ambos límites, las gamas que excluyan alguno o ambos de esos límites incluidos también se incluyen en la invención.

[0083] Tal y como se utilizan en el presente documento y en las reivindicaciones adjuntas, las formas singulares "un/a" y "el/la" incluyen referentes plurales a menos que el contexto indique claramente lo contrario. Por lo tanto, por ejemplo, la referencia a "un gen" incluye una pluralidad de tales agentes candidatos y la referencia a "la célula" incluye referencia a una o más células y equivalentes de estos conocidos por los expertos en la técnica, etc.

[0084] Las publicaciones detalladas en el presente documento se presentan únicamente para su exposición anterior a la fecha de presentación de la presente solicitud. Nada de lo que aparece en el presente documento ha de entenderse como una admisión de que la presente invención no tiene derecho a preceder a tal publicación en virtud de la invención anterior.

Formas de realización detalladas

[0085] Un objetivo de la presente invención era modificar propiedades, tal como la estabilidad térmica y/o la actividad específica de glucoamilasas precursoras y en concreto de la glucoamilasa *Trichoderma reesei* (TrGA) con el fin de obtener variantes de glucoamilasa con propiedades modificadas que serían útiles en diferentes aplicaciones tal como procesos de fermentación de alcohol o conversión de almidón.

Glucoamilasas precursoras:

[0086] En algunas formas de realización, la presente invención proporciona una variante de glucoamilasa de una glucoamilasa precursora como se define en las reivindicaciones. La glucoamilasa precursora comprende una secuencia que presenta una identidad estructural y/o de secuencia con TrGA (SEQ ID NO:2). En algunas formas de realización, la glucoamilasa precursora es un homólogo. En algunas formas de realización la glucoamilasa presenta al menos un 95 % de identidad de secuencia, al menos un 96 % de identidad de secuencia, al menos un 97 % de identidad de secuencia, al menos un 98 % de identidad de secuencia y también al menos un 99 % de identidad de secuencia con la secuencia de aminoácidos de TrGA de SEQ ID NO:2.

[0087] En algunas formas de realización, la glucoamilasa precursora de la presente exposición comprende un dominio catalítico que presenta una secuencia de aminoácidos con al menos un 50 % de identidad de secuencia de aminoácidos, al menos un 60 % de identidad de secuencia de aminoácidos, al menos un 70 % de identidad de secuencia de aminoácidos, al menos un 80 % de identidad de secuencia de aminoácidos, al menos un 85 % de identidad de secuencia de aminoácidos, al menos un 90 % de identidad de secuencia de aminoácidos, al menos un 93 % de identidad de secuencia de aminoácidos, al menos un 95 % de identidad de secuencia de aminoácidos, al menos un 97 % de identidad de secuencia de aminoácidos y al menos un 99 % de identidad de secuencia de aminoácidos con una o más de las secuencias de aminoácidos ilustradas en SEQ ID NO:1, 2, 3, 5, 6, 7 u 8. En otras formas de realización, la glucoamilasa precursora tendrá al menos un 80 % de identidad de secuencia, al menos un 85 % de identidad de secuencia, al menos un 90 % de identidad de secuencia, al menos un 95 % de identidad de secuencia, al menos un 97 % de identidad de secuencia y también al menos un 98 % de

identidad de secuencia con el dominio catalítico de la secuencia de aminoácidos de TrGA de SEQ ID NO:3.

[0088] La glucoamilasa precursora de la presente exposición puede codificarse mediante una secuencia de ADN que hibrida bajo condiciones astringentes, altas o medias con un ADN que codifica una glucoamilasa con una de las secuencias de aminoácidos de SEQ ID NO:1, 2, o 3. En algunas formas de realización, la glucoamilasa precursora con al menos un 50 % de identidad de secuencia, al menos un 60 % de identidad de secuencia de aminoácidos, al menos un 70 % de identidad de secuencia de aminoácidos, al menos un 80 % de identidad de secuencia de aminoácidos, al menos un 90 % de identidad de secuencia de aminoácidos, al menos un 95 % de identidad de secuencia de aminoácidos y al menos un 97 % de identidad de secuencia de aminoácidos presenta también identidad de secuencia con SEQ ID NO:2 y/o 3. Aunque la glucoamilasa precursora es una secuencia natural o que se da de forma natural en una célula huésped, en algunas formas de realización, la glucoamilasa precursora es una variante que se da de forma natural. En algunas formas de realización, la glucoamilasa precursora es una variante que se ha diseñado y/o una glucoamilasa híbrida.

[0089] Las estructuras predichas y las secuencias conocidas de glucoamilasas se conservan entre especies fúngicas (Coutinho *et al.*, (1994) *Protein Eng.*, 7:393 - 400 y Coutinho *et al.*, (1994), *Protein Eng.*, 7: 749-760). En algunas formas de realización, la glucoamilasa precursora es una glucoamilasa fúngica filamentosa. En algunas formas de realización, la glucoamilasa precursora se obtiene de una cepa de *Trichoderma* (p. ej., *T. reesei*, *T. longibrachiatum*, *T. strictipilis*, *T. asperellum*, *T. konilangbra* y *T. hazianum*), una cepa de *Aspergillus* (p. ej., *A. niger*, *A. nidulans*, *A. kawachi*, *A. awamori* y *A. oryzae*), una cepa de *Talaromyces* (p. ej., *T. emersonii*, *T. thermophilus*, y *T. duponti*), una cepa de *Hypocrea* (p. ej., *H. gelatinosa*, *H. orientalis*, *H. vinosa* y *H. citrina*), una cepa de *Fusarium* (p. ej., *F. oxysporum*, *F. roseum*, y *F. venenatum*), una cepa de *Neurospora* (p. ej., *N. crassa*) y una cepa de *Humicola* (p. ej., *H. grisea*, *H. insolens* y *H. lanuginosa*), una cepa de *Penicillium* (p. ej., *P. notatum* o *P. chrysogenum*), o una cepa de *Saccharomycopsis* (p. ej., *S. fibuligera*) de la presente exposición.

[0090] En algunas formas de realización, la glucoamilasa precursora de la presente exposición es una glucoamilasa bacteriana. Por ejemplo, el polipéptido puede obtenerse de una cepa bacteriana gram positiva tal como *Bacillus* (p. ej., *B. alkalophilus*, *B. amyloliquefaciens*, *B. lentus*, *B. licheniformis*, *B. stearothermophilus*, *B. subtilis* y *B. thuringiensis*) o una cepa de *Streptomyces* (p. ej., *S. lividans*).

[0091] En algunas formas de realización, la glucoamilasa precursora presenta al menos un 95 % de identidad de secuencia, al menos un 96 % de identidad de secuencia, al menos un 97 % de identidad de secuencia, al menos un 98 % de identidad de secuencia y también al menos un 99 % de identidad de secuencia con la secuencia de aminoácidos de TrGA de SEQ ID NO:2.

[0092] En formas de realización adicionales, se obtendrá un homólogo de glucoamilasa *Trichoderma* a partir de una cepa de *Trichoderma* o *Hypocrea*. Algunos homólogos de glucoamilasa *Trichoderma* se describen en la patente estadounidense pub. nº 2006/0094080 y se hace referencia de forma específica a secuencias de aminoácidos detallados en SEQ ID NO:17 - 22 y 43 - 47 de dicha referencia.

[0093] En algunas formas de realización, la glucoamilasa precursora es TrGA que comprende la secuencia de aminoácidos de SEQ ID NO:2 o un homólogo de glucoamilasa *Trichoderma* que presenta al menos un 95 %, al menos un 96 %, al menos un 97 %, al menos un 98 %, al menos un 99 % de identidad de secuencia con la secuencia de TrGA.

[0094] Una glucoamilasa precursora de la presente exposición puede aislarse y/o identificarse mediante técnicas de ADN recombinante convencionales. Se puede utilizar cualquier técnica convencional conocida por los expertos. Por ejemplo, se pueden utilizar sondas y/o cebadores específicos para áreas conservadas de la glucoamilasa con el fin de identificar homólogos en células bacterianas o fúngicas (el dominio catalítico, el centro activo, etc.). De forma alternativa, se puede utilizar una PCR con cebadores degenerados para identificar homólogos en células bacterianas o fúngicas. En algunos casos, se pueden analizar secuencias conocidas, tales como en una base de datos, para la homología estructural y/o de secuencia con una de las glucoamilasas conocidas, entre las que se incluyen SEQ ID NO:1, 2 o 3. También se pueden utilizar análisis funcionales para identificar la actividad de glucoamilasa en una célula bacteriana o fúngica. Las proteínas con actividad de glucoamilasa se pueden aislar y cambiar el orden con el fin de aislar la secuencia de ADN correspondiente. El experto en la técnica conoce tales métodos.

50 **Homología estructural de la glucoamilasa:**

[0095] El dogma central de la biología molecular es que la secuencia de ADN que codifica un gen para una enzima específica determina la secuencia de aminoácidos de la proteína, secuencia que a su vez determina el plegamiento tridimensional de la enzima. Este plegamiento reúne los residuos dispares que crean un centro catalítico y una superficie de unión al sustrato y tiene como resultado una alta especificidad y actividad de las enzimas en cuestión.

[0096] Las glucoamilasas consisten en hasta tres dominios estructurales distintos, un dominio catalítico de aproximadamente 450 residuos que está conservado de forma estructural, generalmente seguido de una región de enlace consistente en entre 30 y 80 residuos que están conectados a un dominio de unión a almidón de aproximadamente 100 residuos. La estructura de la glucoamilasa *Trichoderma reesei* con las tres regiones intactas se determinó con la resolución Angstrom 1,8 en el presente documento (véase la tabla 9 y los ejemplos 8 y 9). Al utilizar las coordenadas (véase la tabla 9) la estructura catalítica se alineó con las coordenadas del dominio catalítico de la cepa *Aspergillus awamorii* X100 que se determinó previamente (Aleshin, *et al.*, (1994). *J Mol Biol* 238: 575-591). La estructura de cristal de *Aspergillus awamorii* incluía únicamente el dominio catalítico. Como se ve en las figuras 6 y 7, la estructura de los dominios catalíticos se superpone muy estrechamente y es posible identificar residuos equivalentes basándose en esta superposición estructural. Los inventores creen que todas las glucoamilasas comparten la estructura básica representada en las figuras 6 y 7. La conservación de la estructura en la molécula de glucoamilasa se corresponde con la conservación de la actividad y un mecanismo conservado de acción para todas las glucoamilasas. Teniendo en cuenta esta alta homología, las variantes específicas del lugar de la glucoamilasa *Trichoderma* que tienen como resultado una función modificada tendrían también consecuencias estructurales y, por tanto, funcionales similares en otras glucoamilasas.

[0097] Sin embargo, la presente exposición no está limitada a las variantes de la glucoamilasa precursora *Trichoderma* expuesta en la figura 1, sino que se extiende a variantes de glucoamilasas precursoras que comprenden residuos de aminoácidos en posiciones que son "equivalentes" a los residuos identificados específicos en la glucoamilasa *Trichoderma reesei* (SEQ ID NO:2). En algunas formas de realización, de la presente exposición, la glucoamilasa precursora es una GA *Talaromyces* y las sustituciones se realizan en las posiciones de residuo de aminoácidos equivalentes en la glucoamilasa *Talaromyces* (véase, p. ej., SEQ ID NO:308, figura 5C) a las descritas en el presente documento. En otras formas de realización, la glucoamilasa precursora comprende SEQ ID NO:5; SEQ ID NO:6; SEQ ID NO:7; SEQ ID NO:3; SEQ ID NO:8 y SEQ ID NO:9 (véanse las figuras 5A – 5B). En formas de realización adicionales, la glucoamilasa precursora es una glucoamilasa *Penicillium*, tal como *Penicillium chrysogenum*.

[0098] La identidad estructural determina si los residuos de aminoácidos son equivalentes. La identidad estructural es un equivalente topológico directo cuando las dos estructuras (estructuras de aminoácidos y tridimensionales) se alinean. Una posición de residuos (aminoácidos) de una glucoamilasa es equivalente a un residuo de glucoamilasa *T. reesei* si es tanto homóloga (es decir, se corresponde en posición ya sea en estructura primaria o terciaria) como análoga con un residuo específico o parte de ese residuo en la glucoamilasa *T. reesei* (que tenga la misma capacidad funcional o similar para combinar, reaccionar o interactuar de forma química).

[0099] Con el fin de establecer la identidad con la estructura primaria, la secuencia de aminoácidos de una glucoamilasa puede compararse directamente con la secuencia primaria de glucoamilasa *Trichoderma reesei* y específicamente con un conjunto de residuos que se sabe son invariables en glucoamilasas por cuya secuencia es conocido. Por ejemplo, las figuras 5A y B en el presente documento muestran los residuos conservados entre glucoamilasas. Una vez se han alineado los residuos conservados, y permitiendo las inserciones y deleciones necesarias con el fin de mantener la alineación (es decir, evitando la eliminación de residuos conservados a través de la delección e inserción arbitraria), se definen los residuos equivalentes a aminoácidos específicos en la secuencia primaria de glucoamilasa *Trichoderma reesei*. La alineación de residuos conservados puede conservar el 100 % de tales residuos. Sin embargo, la alineación de más del 75 % o tan solo un 50 % de residuos conservados también es adecuada para definir residuos equivalentes, especialmente cuando se incluya la alineación basada en identidad estructural.

[0100] Por ejemplo, en la figura 5, se alinean glucoamilasas de seis organismos para proporcionar la máxima cantidad de homología entre secuencias de aminoácidos. Una comparación de estas secuencias muestra que existe un número de residuos conservados contenidos en cada secuencia como se indica mediante un asterisco. Por lo tanto, se pueden utilizar estos residuos conservados para definir los correspondientes residuos de aminoácidos equivalentes de glucoamilasa *Trichoderma reesei* en otras glucoamilasas tal como la glucoamilasa de *Aspergillus niger*.

[0101] La identidad estructural supone la identificación de residuos equivalentes entre las dos estructuras. "Los residuos equivalentes" pueden definirse mediante la determinación de homología en el nivel de estructura terciaria (identidad estructural) para una enzima cuya estructura terciaria ha sido determinada mediante cristalografía de rayos X. Los residuos equivalentes se definen como aquellos para cuyas coordenadas atómicas de dos o más de los átomos de la cadena principal de un residuo de aminoácido específico de la glucoamilasa *Trichoderma reesei* (N en N, CA en CA, C en C y O en O) están dentro de 0,13 nm y preferiblemente 0,1 nm tras la alineación. La alineación se consigue después de que el mejor modelo se haya orientado y colocado para proporcionar la máxima superposición de coordenadas atómicas de átomos de proteína no hidrógenos de la glucoamilasa en cuestión con la glucoamilasa *Trichoderma reesei*. El mejor modelo es el modelo cristalográfico que proporciona el factor R más bajo para los datos de difracción experimental con la resolución más alta disponible.

$$\text{Factor } R = \frac{\sum_h |F_o(h)| - |F_c(h)|}{\sum_h |F_o(h)|}$$

[0102] Los residuos equivalentes que son análogos funcionalmente a un residuo específico de la glucoamilasa *Trichoderma reesei* se definen como aquellos aminoácidos de la enzima que pueden adoptar una conformación de tal forma que bien cambian, modifican o contribuyen a la estructura de proteína, unión de sustrato o catálisis de forma definida y atribuida a un residuo específico de la glucoamilasa *Trichoderma reesei*. Además, se trata de aquellos residuos de la enzima (para los que se ha obtenido una estructura terciaria mediante cristalografía de rayos X) que ocupan una posición análoga hasta el punto de que, aunque los átomos de la cadena principal del residuo determinado puedan no cumplir los criterios de equivalencia sobre la base de ocupar una posición homóloga, las coordenadas atómicas de al menos dos de los átomos de la cadena lateral del residuo quedan con 0,13 nm de los correspondientes átomos de la cadena lateral de la glucoamilasa *Trichoderma reesei*. Las coordenadas de la estructura tridimensional de la glucoamilasa *Trichoderma reesei* están establecidas en la tabla 9 y pueden utilizarse como se detalla anteriormente con el fin de determinar residuos equivalentes en el nivel de estructura terciaria.

Variantes:

[0103] Las variantes según la presente exposición incluyen al menos una sustitución, deleción o inserción en la secuencia de aminoácidos de una glucoamilasa precursora que hace que la variante se diferencie de la glucoamilasa precursora por la secuencia. En algunas formas de realización, las variantes de la presente exposición tendrán al menos un 20 %, al menos un 40 %, al menos un 50 %, al menos un 60 %, al menos un 70 %, al menos un 80 %, al menos un 85 %, al menos un 90 %, al menos un 95 %, al menos un 97 % y también al menos un 100 % de la actividad de glucoamilasa de TrGA (SEQ ID NO:2). En algunas formas de realización, las variantes comprenderán una sustitución, deleción o inserción en al menos una posición de aminoácidos de la TrGA precursora (SEQ ID NO:2) o en una posición equivalente en la secuencia de otra glucoamilasa precursora con al menos un 50 %, al menos un 60 %, al menos un 70 %, al menos un 80 %, al menos un 90 % de identidad de secuencia con la secuencia de TrGA, que incluye sin carácter limitativo, al menos un 93 % de identidad de secuencia, al menos un 95 %, al menos un 97 % y al menos un 99 % de identidad de secuencia. En algunas formas de realización, la glucoamilasa precursora tendrá identidad estructural con la secuencia de TrGA.

[0104] En otras formas de realización, la variante según la presente exposición comprenderá una sustitución, deleción o inserción en al menos una posición de aminoácido de un fragmento de la TrGA precursora, en la que el fragmento comprende el dominio catalítico de la secuencia de TrGA (SEQ ID NO:3) o en una posición equivalente en un fragmento que comprende el dominio catalítico de una glucoamilasa precursora con al menos un 50 %, al menos un 60 %, al menos un 70 %, al menos un 80 % de identidad de secuencia con el fragmento de la secuencia de TrGA, al menos un 90 %, al menos un 95 %, al menos un 97 % y al menos un 99% de identidad de secuencia. En algunas formas de realización, el fragmento comprenderá al menos 400, 425, 450 y/o 500 residuos de aminoácidos. En algunas formas de realización, cuando la glucoamilasa precursora incluye un dominio catalítico, región de enlace y dominio de unión a almidón, el fragmento puede incluir parte de la región de enlace. En algunas formas de realización, la variante comprenderá una sustitución, deleción o inserción en la secuencia de aminoácidos de un fragmento de la secuencia de TrGA (SEQ ID NO:2 o SEQ ID NO:3). En algunas formas de realización, la variante tendrá identidad estructural con la secuencia de TrGA (SEQ ID NO:2 o 3).

[0105] La identidad estructural con respecto a una sustitución de aminoácidos indica que la sustitución tiene lugar en la posición de aminoácidos equivalente en la glucoamilasa homóloga o glucoamilasa precursora. El término posición equivalente indica una posición que es común a las dos secuencias precursoras que se basa en una alineación de la secuencia de aminoácidos de la glucoamilasa precursora en cuestión así como la alineación de la estructura tridimensional de la glucoamilasa precursora en cuestión con la secuencia de aminoácidos de la glucoamilasa de referencia TrGA y la secuencia tridimensional. Por ejemplo, en referencia a la figura 5, la posición 24 en TrGA (SEQ ID NO:3) es D24 y la posición equivalente para *Aspergillus niger* (SEQ ID NO:6) es la posición D25 y la posición equivalente para *Aspergillus oryzae* (SEQ ID NO:7) es la posición D26. Véanse las figuras 6 y 7 para una alineación de ejemplo de la secuencia tridimensional.

[0106] En algunas formas de realización, la variante de glucoamilasa tendrá más de una sustitución (p. ej., dos, tres o cuatro sustituciones) en comparación con una glucoamilasa precursora correspondiente.

[0107] En algunas formas de realización, una variante de glucoamilasa comprende una sustitución, deleción o inserción en al menos una posición de aminoácidos en una posición correspondiente a las regiones de aminoácidos no conservados como se ilustra en las figuras 5A y 5B (p. ej., posiciones de aminoácidos correspondientes a aquellas posiciones que no están indicadas con "*" en las figuras 5A y 5B). En algunas formas de realización, la variante comprende una sustitución en al menos una posición de aminoácidos en una posición correspondiente a las regiones de aminoácidos no conservados como se ilustra en las figuras 5A y 5B.

- [0108]** Aunque las variantes pueden estar en cualquier posición en la secuencia de proteína madura (SEQ ID NO:2 o 3), en algunas formas de realización, una variante de glucoamilasa comprende una o más sustituciones en las siguientes posiciones en la secuencia de aminoácidos establecida en SEQ ID NO: 2 o 3: 10, 14, 15, 23, 42, 45, 46, 59, 60, 61, 67, 68, 72, 73, 97, 98, 99, 102, 108, 110, 113, 114, 122, 124, 125, 133, 140, 144, 145, 147, 152, 153, 164, 175, 182, 204, 205, 214, 216, 219, 228, 229, 230, 231, 236, 239, 240, 241, 242, 244, 263, 264, 265, 268, 269, 276, 284, 291, 300, 301, 303, 310, 311, 313, 316, 338, 342, 344, 346, 349, 359, 361, 364, 375, 379, 382, 390, 391, 393, 394, 408, 410, 415, 417, 418, 430, 431, 433, 436, 442, 443, 444, 448 y 451 o en una posición equivalente en una glucoamilasa precursora. En algunas formas de realización, la glucoamilasa precursora presentará al menos un 50 %, al menos un 60 %, al menos un 70 %, al menos un 80 %, al menos un 90 %, al menos un 95 %, al menos un 96 %, al menos un 97 %, al menos un 98 % y al menos un 99 % de identidad de secuencia con SEQ ID NO:2 o SEQ ID NO:3. En otras formas de realización, la glucoamilasa precursora será un homólogo de glucoamilasa *Trichoderma*. En algunas formas de realización, la variante tendrá propiedades modificadas. En algunas formas de realización, la glucoamilasa precursora tendrá identidad estructural con la glucoamilasa de SEQ ID NO:2 o 3.
- [0109]** En algunas formas de realización, la variante de glucoamilasa comprende una o más sustituciones en las siguientes posiciones en la secuencia de aminoácidos establecida en SEQ ID NO:2 o 3: T10, L 14, N 15, P23, T42, P45, D46, F59, K60, N61, T67, E68, A72, G73, S97, L98, A99, S102, K108, E110, L113, K114, R122, Q124, R125, I133, K140, N 144, N 145, Y147, S 152, N 153, N164, F175, N 182, A204, T205, S214, V216, Q219, W228, V229, S230, S231, D236, I239, N240, T241, N242, G244, N263, L264, G265, A268, G269, D276, V284, S291, P300, A301, A303, Y310, A311, D313, Y316, V338, T342, S344, T346, A349, V359, G361, A364, T375, N379, S382, S390, E391, A393, K394, R408, S410, S415, L417, H418, T430, A431, R433, I436, A442, N443, S444, T448 y S451 o una posición equivalente en una glucoamilasa precursora (p. ej., un homólogo de glucoamilasa *Trichoderma*). En algunas formas de realización, la variante presentará propiedades modificadas en comparación con la glucoamilasa precursora.
- [0110]** En otras formas de realización, la variante de una precursora de glucoamilasa comprende una o más sustituciones en las siguientes posiciones en la secuencia de aminoácidos establecida en SEQ ID NO:2 o 3: 10, 14, 15, 23, 59, 60, 61, 65, 67, 68, 72, 73, 97, 98, 99, 102, 110, 113, 133, 140, 144, 145, 147, 152, 153, 164, 182, 204, 205, 214, 216, 219, 228, 229, 230, 231, 236, 239, 241, 242, 263, 264, 265, 268, 269, 276, 284, 291, 300, 301, 303, 311, 338, 342, 344, 346, 349, 359, 361, 364, 375, 379, 382, 390, 391, 393, 394, 410, 417, 418, 430, 431, 433, 442, 444, 448 y 451 o una posición equivalente en una glucoamilasa precursora. En algunas formas de realización, la glucoamilasa precursora tendrá un 50 %, al menos un 60 %, al menos un 70 %, al menos un 80 %, al menos un 90 %, al menos un 95 %, al menos un 96 %, al menos un 97 %, al menos un 98 % y al menos un 99% de identidad de secuencia con SEQ ID NO:2 o SEQ ID NO:3. En otras formas de realización, la glucoamilasa precursora será un homólogo de glucoamilasa *Trichoderma*. En algunas formas de realización, la variante presentará al menos una propiedad modificada en comparación con la glucoamilasa precursora. En algunas formas de realización, la al menos una propiedad modificada es la actividad específica. En algunas formas de realización, la variante presenta una o más sustituciones correspondientes a una de las siguientes posiciones: 61, 67, 72, 97, 102, 133, 205, 219, 228, 230, 231, 239, 263, 268, 291, 342, 394, 430, 431 y 451 de SEQ ID NO:2 y/o 3. En algunas formas de realización, la sustitución en estas posiciones se elige de entre: N61I, T67M, A72Y, S97N, S102M, I133T, T205Q, Q219S, W228H, W228M, S230F, S230G, S230N, S230R, S231L, I239V, I239Y, I263P, A268C, A268G, A268K, S291A, T342V, K394S, T430K, A431 Q y S451K de SEQ ID NO:2 y/o 3. En algunas formas de realización, la variante presenta una o más sustituciones correspondientes a una de las siguientes posiciones: 72, 133, 219, 228, 230, 231, 239, 263, 268 y 451 de SEQ ID NO:2 y/o 3. En algunas formas de realización, la sustitución en estas posiciones se elige de entre: A72Y, I133T, Q219S, W228H, W228M, S230R, S230F, S230G, S231L, I239V, N263P, A268C, A268G y S451 K de SEQ ID NO:2 y/o 3. En algunas formas de realización, la variante presenta al menos una propiedad modificada y la al menos una propiedad modificada es una actividad específica aumentada en comparación con la glucoamilasa precursora.
- [0111]** En formas de realización adicionales, la variante de una precursora de glucoamilasa comprende al menos una de las siguientes sustituciones en las siguientes posiciones en una secuencia de aminoácidos establecida en SEQ ID NO:2 o 3: T10D/F/G/K/L/M/P/R/S; L14E/H; N15D/N; P23A/G; F59A/G; K60F/H; N61D/I/L/QN/W; R65A/C/G/I/K/M/SN/Y; T67C/I/K/M/T; E68I/M/W; A72E/G/L/M/Q/R/W/Y; G73C/L/W; S97F/M/N/P/R/SN/W/Y; L98H/M; A99C/L/M/N/P; S102A/C/I/L/M/N/R/V/W/Y; E110Q/S/W; L113E/N; K114C/D/E/L/M/Q/S/T/V; I133K/R/S/T; K140A/E/F/H/K/L/M/N/Q/R/S/V/W/Y; N144C/D/E/I/K; N145A/C/E/I/K/L/M/Q/R/H/V/W/Y; Y147A/M/R; S152H/M; N153C/D/K/L/W/Y; N164A/G; N182C/E/K/P/R; A204C/D/G/I/M/Q/T; T205A/D/HMC/M/N/P/Q/SN/W/Y; S214P/T; V216C/G/H/K/N/Y; Q219D/G/H/N/P/S; W228A/F/G/H/VL/M/Q/S/TN/Y; V229E/I/M/N/Q; S230C/D/E/F/G/H/I/K/L/N/P/Q/R/TN/Y; S231C/D/F/L/M/N/Q/R/S/V/Y; D236F/G/L/M/N/P/S/T/V; I239M/Q/SN/W/Y; T241C/E/H/L/M/P/S/T/V; N242C/F/H/M/TN/W; N263H/K/P; L264A/C/E/F/L/S; G265E/G/H/I/K/R/T; A268C/D/E/F/G/I/K/L/P/R/TW; G269E; D276S; V284R/TN/Y/A/E/F/H/K/N/P/W; P300K/R; A301 E/K/L/P/S/W; A303C/D/F/H/I/K/L/N/R/TN/W/Y; A311N/P/Q/S/Y; V338P/Q/S/Y; T342N/V; S344A/T; T346G/H/M/N/P/Q/Y; A349L/I/K/M/N/Q/R/W; G361H/L/R; A364M/W; T375C/D/E/H/V/W/Y; N379A/C/D/G/I/M/P/S; S382A/N/P/V/W; S390A/Y; E391A/E/I/K/L/M/Q/R/V/W/Y; A393E/G/H/I/L/M/N/Q/R/S/T/V/W/Y; K394A/H/K/L/M/Q/R/S/T/V/W; S410E/H/N;

L417A/D/E/F/G/I/K/Q/R/S/T/V/W/Y; H418E/M; T430A/E/F/G/H/I/K/M/N/Q/R/V; A431C/E/H/I/L/M/Q/R/S/W/Y; R433A/C/E/F/G/K/L/M/N/S/V/W/Y; I436E/F/G/H/K/P/R/S/TN/Y; S444M/N/P/Q/R/TN/W; T448F/G/I/P/Q/TN y S451 E/H/K/L/Q/T o una sustitución en una posición equivalente en una glucoamilasa precursora.

[0112] En otras formas de realización, la variante de una precursora de glucoamilasa comprende una o más sustituciones en las siguientes posiciones en la secuencia de aminoácidos establecida en SEQ ID NO:2 o 3: 10, 15, 23, 42, 59, 60, 61, 68, 72, 73, 97, 98, 99, 102, 114, 133, 140, 144, 147, 152, 153, 164, 182, 204, 205, 214, 216, 228, 229, 230, 231, 236, 241, 242, 263, 264, 265, 268, 269, 276, 284, 291, 300, 301, 303, 311, 338, 342, 344, 346, 349, 359, 361, 364, 375, 379, 382, 390, 391, 393, 394, 410, 417, 430, 431, 433, 436, 442, 443, 444, 448 y 451 o una posición equivalente en una glucoamilasa precursora (p. ej., homólogo de glucoamilasa *Trichoderma*). En algunas formas de realización, la glucoamilasa precursora tendrá al menos un 50 %, al menos un 60 %, al menos un 70 %, al menos un 80 %, al menos un 90 %, al menos un 95 %, al menos un 96 %, al menos un 97 %, al menos un 98 % y al menos un 99 % de identidad de secuencia con SEQ ID NO:2 o SEQ ID NO:3. En otras formas de realización, la glucoamilasa precursora también muestra identidad estructural con la glucoamilasa de SEQ ID NO:2 y/o 3. En otras formas de realización, la glucoamilasa precursora será un homólogo de glucoamilasa *Trichoderma*. En algunas formas de realización, la variante presenta una o más sustituciones correspondientes a una de las siguientes posiciones: 10, 42, 68, 73, 97, 114, 153, 229, 231, 236, 264, 291, 301, 344, 361, 364, 375, 417 y 433 de SEQ ID NO:2 y/o 3. En algunas formas de realización, la sustitución en estas posiciones se elige de entre: T10S, T42V, E68C, E68M, G73F, G73W, K114M, K114T, N153A, N153S, N153V, W228V, D236R, G361D, G361E, G361P, G361Y, A364D, A364E, A364F, A364G, A364K, A365L, A365R, R433C, R433G, R433L, R433N, R433S, R433V y I436H de SEQ ID NO:2 y/o 3. En algunas formas de realización, la variante presenta una o más sustituciones correspondientes a una de las siguientes posiciones: 42, 68, 73, 114, 153, 236, 361 y 364 de SEQ ID NO:2 y/o 3. En algunas formas de realización, la sustitución en estas posiciones se elige de entre: T42V, E68M, G73F, G73W, K114T, N153S, N153V, D236R, G361D, A364F y A364L de SEQ ID NO:2 y/o 3. En algunas formas de realización, la variante presenta al menos una propiedad modificada y la al menos una propiedad modificada es una termoestabilidad aumentada en comparación con la glucoamilasa precursora.

[0113] En otras formas de realización, la variante de una glucoamilasa precursora comprende una o más sustituciones en las siguientes posiciones en la secuencia de aminoácidos establecida en SEQ ID NO:2 o 3: 10, 15, 59, 61, 68, 72, 73, 97, 99, 102, 133, 140, 153, 182, 204, 205, 214, 228, 229, 230, 231, 236, 241, 242, 264, 265, 268, 275, 284, 291, 300, 301, 303, 311, 338, 344, 346, 359, 361, 364, 375, 370, 382, 391, 393, 394, 410, 417, 430, 431, 433, 444, 448 y 451 y/o una posición equivalente en una glucoamilasa precursora (p. ej., homólogo de glucoamilasa *Trichoderma*). En algunas formas de realización, la glucoamilasa precursora tendrá al menos un 50 %, al menos un 60 %, al menos un 70 %, al menos un 80 %, al menos un 90 %, al menos un 95 %, al menos un 96 %, al menos un 97 %, al menos un 98 % y al menos un 99 % de identidad de secuencia con SEQ ID NO:2 o SEQ ID NO:3. En otras formas de realización, la glucoamilasa precursora también tendrá identidad estructural con la glucoamilasa de SEQ ID NO:2 y/o 3. En otras formas de realización, la glucoamilasa precursora será un homólogo de glucoamilasa *Trichoderma*. En algunas formas de realización, la variante presentará al menos una propiedad modificada en comparación con la glucoamilasa precursora. En algunas formas de realización, la variante presenta una o más sustituciones correspondientes a una de las siguientes posiciones: 228, 230, 231, 268, 291, 417, 433 y 451 de SEQ ID NO:2 o 3. En algunas formas de realización, la sustitución en estas posiciones se elige de entre: W228H, W228M, S230F, S230G, S230R, S231L, A268C, A268G, S291 A, L417R, R433Y y S451K de SEQ ID NO:2 y/o 3. En algunas formas de realización, la variante presenta al menos una propiedad modificada y la al menos una propiedad modificada es una termoestabilidad o actividad específica aumentada en comparación con la glucoamilasa precursora. En algunas formas de realización, la variante presenta tanto una termoestabilidad aumentada como una actividad específica aumentada en comparación con la glucoamilasa precursora.

[0114] Las variantes de glucoamilasa de la presente exposición pueden también incluir glucoamilasas quiméricas o híbridas con, por ejemplo, un dominio de unión a almidón (SBD) de una glucoamilasa y un dominio catalítico y enlazador de otra. Por ejemplo, una glucoamilasa híbrida puede formarse mediante el intercambio del SBD de AnGA con el SBD de TrGA, realizando un híbrido con el SBD de AnGA y el dominio catalítico y enlazador de TrGA. De forma alternativa, el SBD y el enlazador de AnGA pueden intercambiarse por el SBD y el enlazador de TrGA.

[0115] En algunos aspectos, la glucoamilasa variante muestra una termoestabilidad aumentada en comparación con la glucoamilasa precursora. En algunas formas de realización, la propiedad modificada es la actividad específica aumentada en comparación con la glucoamilasa precursora. En algunas formas de realización, la propiedad modificada es la termoestabilidad aumentada con temperaturas inferiores en comparación con la glucoamilasa precursora. En algunas formas de realización, la propiedad modificada es tanto la actividad específica aumentada como la termoestabilidad aumentada en comparación con la glucoamilasa precursora.

[0116] Se ha alineado un número de glucoamilasas con la secuencia de aminoácidos de TrGA. La figura 5 incluye el dominio catalítico de las siguientes glucoamilasas precursoras: *Aspergillus awamori* (AaGA) (SEQ ID

NO:5); *Aspergillus niger* (AnGA) (SEQ ID NO:6); *Aspergillus oryzae* (AoGA) (SEQ ID NO:7), *Humicola grisea* (HgGA) (SEQ ID NO:8) e *Hypocrea vinosa* (HvGA) (SEQ ID NO:9). El % de identidad de los dominios catalíticos se representa en la tabla 1 a continuación.

Tabla 1

	AaGA	AnGA	AoGA	HgGA	HvGA	TrGA
AaGA	100	95	58	53	57	56
AnGA		100	59	53	57	56
AoGA			100	55	56	56
HgGA				100	61	63
HvGA					100	91
TrGA						100

5 **[0117]** En algunas formas de realización, por ejemplo, la glucoamilasa variante se derivará de una glucoamilasa precursora que es una glucoamilasa *Aspergillus* y la variante incluirá al menos una sustitución en una posición equivalente a una posición establecida en SEQ ID NO:2 o SEQ ID NO:3 y especialmente en una posición correspondiente a: T10, L14, N15, P23, T42, P45, D46, F59, K60, N61, R65, T67, E68, A72, G73, S97, L98, A99, S102, K108, E110, L113, K114, R122, Q124, R125, I133, K140, N144, N145, Y147, S152, N153, N164, F175, N182, A204, T205, S214, V216, Q219, W228, V229, S230, S231, D236, 1239, N240, T241, N242, G244, N263, L264, G265, A268, G269, D276, V284, S291, P300, A301, A303, Y310, A311, D313, Y316, V338, T342, S344, T346, A349, V359, G361, A364, T375, N379, S382, S390, E391, A393, K394, R408, S410, S415, L417, H418, T430, A431, R433, I436, A442, N443, S444, T448 y S451.

Propiedades modificadas:

15 **[0118]** La presente exposición también proporciona variantes de glucoamilasa con al menos una propiedad modificada (p. ej., propiedad mejorada) en comparación con una glucoamilasa precursora y en concreto con la TrGA. En algunas formas de realización, se elige al menos una propiedad modificada (p. ej., propiedad mejorada) de entre estabilidad del ácido, estabilidad térmica y actividad específica. En algunas formas de realización, la propiedad modificada es estabilidad ácida aumentada, estabilidad térmica aumentada y/o actividad específica aumentada. En algunas formas de realización, la estabilidad térmica aumentada es a altas temperaturas. En algunas formas de realización, la estabilidad de PH aumentada es con un pH alto. En formas de realización adicionales, la estabilidad de pH aumentada es con un pH bajo.

25 **[0119]** Las variantes de glucoamilasa de la presente exposición pueden también proporcionar índices más altos de hidrólisis de almidón con concentraciones de sustrato bajas en comparación con la glucoamilasa precursora. La variante puede tener una Vmax superior o Km inferior a una glucoamilasa precursora cuando se realiza el análisis con las mismas condiciones. Por ejemplo, la glucoamilasa variante puede tener una Vmax superior con una gama de temperaturas de 25 °C a 70 °C (p. ej., a 25 °C a 35 °C; 30 °C - 35 °C; 40 °C a 50 °C; de 50 °C a 55 °C y de 55 °C a 62 °C). La constante Michaelis-Menten, los valores Km y Vmax pueden determinarse fácilmente utilizando procedimientos conocidos convencionales.

30 **Estabilidad térmica (Variantes termoestables):**

[0120] En algunos aspectos, la invención hace referencia a una glucoamilasa variante con estabilidad térmica modificada en comparación con una precursora (tipo silvestre) tal como se define en las reivindicaciones. La termoestabilidad modificada puede ser con temperaturas aumentadas o temperaturas disminuidas. La termoestabilidad se mide como el % de actividad residual tras la incubación durante 1 hora a 64 grados centígrados en un tampón NaAc pH 4,5. Con estas condiciones, TrGA presenta una actividad residual de entre aproximadamente 15 % y 44 % debido a la variación diaria en comparación con la actividad inicial antes de la incubación. Por lo tanto, en algunas formas de realización, las variantes con termoestabilidad aumentada de la presente exposición presentan una actividad residual que está entre al menos un 1 % y al menos un 50 % superior a la de la precursora (tras incubación durante 1 hora a 64 grados centígrados en tampón NaAc pH 4,5) incluidos 2 %, 3 %, 4 %, 5 %, 6 %, 7 %, 8 %, 9 %, 10 %, 11 %, 12 %, 13 %, 14 %, 15 %, 16 %, 17 %, 18 %, 19 %, 20 %, 21 %, 22 %, 23 %, 24 %, 25 %, 26 %, 27 %, 28 %, 29 %, 30 %, 31 %, 32 %, 33 %, 34 %, 35 %, 36 %, 37 %, 38 %, 39 %, 40 %, 41 %, 42 %, 43 %, 44 %, 45 %, 46 %, 47 %, 48 %, 49 % y 50 % en comparación con la actividad inicial antes de la incubación. Por ejemplo, cuando la actividad residual precursora es un 15 %, una variante con estabilidad térmica aumentada puede tener una actividad residual de entre aproximadamente un 16% y aproximadamente un 75 %. En algunas formas de realización, la variante de glucoamilasa tendrá termoestabilidad mejorada tal como retener al menos 50 %, 60 %, 70 %, 75 %, 80 %, 85 %, 90 %, 92 %, 95 %, 96 %, 97 %, 98 % o 99 % de actividad enzimática tras la exposición a temperaturas modificadas durante un

periodo de tiempo determinado, por ejemplo, al menos 60 minutos, 120 minutos, 180 minutos, 240 minutos, 300 minutos, etc. En algunas formas de realización, la variante presenta estabilidad térmica aumentada en comparación con la glucoamilasa precursora con temperaturas elegidas en el intervalo que va desde aproximadamente 40 a 80 °C, también en el intervalo de 50 a 75 °C y en el intervalo de 60 a 70 °C y con un intervalo de pH de 4,0 a 6,0. En algunas formas de realización, la termoestabilidad se determina tal y como se describe en los ejemplos. En algunas formas de realización, la variante presenta estabilidad térmica aumentada con una temperatura inferior en comparación con la glucoamilasa precursora con una temperatura seleccionada dentro del intervalo de 20 a 50°C, incluyendo 35 a 45 y 30 °C a 40 °C.

[0121] En algunas formas de realización, las variantes con una mejora en la termoestabilidad incluyen una o más delecciones, sustituciones o inserciones y especialmente sustituciones en las siguientes posiciones en la secuencia de aminoácidos establecida en SEQ ID NO:2 o SEQ ID NO:3: T10, N15, P23, T42, F59, K60, N61, E68, A72, G73, S97, L98, A99, S102, K114, I133, K140, N144, Y147, S152, N 153, N164, N 182, A204, T205, S214, V216, W228, V229, S230, S231, D236, T241, N242, N263, L264, G265, A268, G269, D276, V284, S291, P300, A301, A303, A311, V338, T342, S344, T346, A349, V359, G361, A364, T375, N379, S382, S390, E391, A393, K394, S410, L417, T430, A431, R433, I436, A442, N443, S444, T448 y S451 y/o una posición equivalente en una glucoamilasa precursora. En algunas formas de realización, la glucoamilasa precursora será un homólogo de glucoamilasa *Trichoderma* y en formas de realización adicionales, la glucoamilasa precursora tendrá al menos un 50 %, al menos un 60 %, al menos un 70 %, al menos un 80 %, al menos un 90 %, al menos un 95 % y al menos un 98 % de identidad de secuencia con SEQ ID NO:2 o 3. En algunas formas de realización, la glucoamilasa precursora también presentará identidad estructural con SEQ ID NO:2 y/o 3. En algunas formas de realización, la variante con termoestabilidad aumentada cuenta con una sustitución en al menos una de las posiciones: 10, 42, 68, 73, 97, 153, 229, 231, 236, 264, 291, 301, 344, 361, 364, 375 y/o 417 de SEQ ID NO:2 y/o 3. En algunas formas de realización, la variante con termoestabilidad cuenta con una sustitución en al menos una de las posiciones: 42, 68, 73, 153, 236, 344, 361, 364 y 365 de SEQ ID NO:2 o SEQ ID NO:3.

Actividad específica

[0122] Actividad específica en el presente documento es la actividad de la glucoamilasa por mg de proteína. La actividad se determinó mediante el análisis de etanol. La detección identificó variantes con un índice de rendimiento (PI) > 1,0 en comparación con el PI de la TrGA precursora. El PI se calcula a partir de las actividades específicas (actividad/mg enzima) de las enzimas variantes y de tipo silvestre (WT). Es el cociente "actividad específica variante/actividad específica WT" y puede ser una medida del aumento en la actividad específica de la variante. Un PI de 2 debería ser aproximadamente 2 veces mejor que WT. En algunos aspectos, la invención hace referencia a una glucoamilasa variante con actividad específica aumentada en comparación con una glucoamilasa de tipo silvestre o precursora, tal y como se define en las reivindicaciones. La actividad específica aumentada puede definirse como un índice de rendimiento aumentado mayor o igual que aproximadamente 1, incluyendo mayor o igual que aproximadamente 1,1; 1,2; 1,3; 1,4; 1,5; 1,6; 1,7; 1,8; 1,9 y 2. En algunas formas de realización de la presente exposición, la actividad específica aumentada va desde aproximadamente 1,0 a aproximadamente 5,0, incluyendo 1,1; 1,2; 1,3; 1,4; 1,5; 1,6; 1,7; 1,8; 1,9; 2,0; 2,1; 2,2; 2,3; 2,4; 2,5; 2,6; 2,7; 2,8; 2,9; 3,0; 3,1; 3,2; 3,3; 3,4; 3,5; 3,6; 3,7; 3,8; 3,9; 4,0; 4,1; 4,2; 4,3; 4,4; 4,5; 4,6; 4,7; 4,8 y 4,9. En algunas formas de realización, la variante de la presente exposición cuenta con una actividad específica al menos 1 vez más alta que la glucoamilasa precursora, incluyendo al menos 1,1 veces, 1,2 veces, 1,3 veces, 1,4 veces, 1,5 veces, 1,6 veces, 1,7 veces, 1,8 veces, 1,9 veces, 2 veces, 2,2 veces, 2,5 veces, 2,7 veces, 2,9 veces, 3 veces, 4 veces y 5 veces.

[0123] En algunas formas de realización, las variantes de la presente exposición con una mejora en la actividad específica incluyen una o más delecciones, sustituciones o inserciones en las siguientes posiciones en la secuencia de aminoácidos establecida en SEQ ID NO:2 y/o SEQ ID NO:3: T10, L14, N15, P23, F59, K60, N61, T67, E68, A72, G73, S97, L98, A99, S102, E110, L113, I133, K140, N144, N145, Y147, S152, N153, N164, N182, A204, T205, S214, V216, Q219, W228, V229, S230, S231, D236, I239, T241, N242, N263, L264, G265, A268, G269, D276, V284, S291, P300, A301, A311, V338, T342, S344, T346, A349, V359, G361, A364, T375, N379, S382, S390, E391, A393, K394, S410, S415, L417, H418, T430, A431, R433, A442, S444, T448 y/o S451 y/o una posición equivalente en una glucoamilasa precursora. En algunas formas de realización, las variantes de la presente exposición con actividad específica mejorada incluyen una sustitución en las siguientes posiciones en la secuencia de aminoácidos establecida en SEQ ID NO:2 o 3: 61, 67, 72, 97, 102, 133, 205, 219, 228, 230, 231, 239, 263, 268, 291, 342, 394, 430, 431 y 451 y/o una posición equivalente en una glucoamilasa precursora. En algunas formas de realización, la glucoamilasa precursora comprenderá una secuencia con al menos un 50 %, al menos un 60 %, al menos un 70 %, al menos un 80 %, al menos un 90 % o un 95 % de identidad de secuencia con la secuencia de SEQ ID NO:2 o 3. En algunas formas de realización, la glucoamilasa precursora también presentará identidad estructural con SEQ ID NO:2 y/o 3. En algunas formas de realización, la variante con actividad específica aumentada cuenta con una sustitución en al menos una de las posiciones: 72, 133, 219, 228, 230, 231, 239, 263, 268 y 451 de SEQ ID NO:2 y/o 3.

Termoestabilidad y actividad específica combinadas

[0124] En algunos aspectos, la invención hace referencia a una glucoamilasa variante como se define en las reivindicaciones tanto con termoestabilidad modificada como con actividad específica modificada en comparación con una precursora (p. ej., de tipo silvestre). En algunas formas de realización, la actividad específica modificada es una actividad específica aumentada. En algunas formas de realización, la termoestabilidad modificada es una termoestabilidad aumentada con altas temperaturas (p. ej., con temperaturas por encima de los 80 °C) en comparación con la glucoamilasa precursora.

[0125] En algunas formas de realización, las variantes de la presente exposición con una termoestabilidad aumentada y actividad específica aumentada incluyen una o más delecciones, sustituciones o inserciones y sustituciones en las siguientes posiciones en la secuencia de aminoácidos establecida en SEQ ID NO:2 o 3: T10, N15, F59, N61, E68, A72, G73, S97, A99, S102, I133, K140, N153, N182, A204, T205, S214, W228, V229, S230, S231, D236, T241, N242, L264, G265, A268, D276, V284, S291, P300, A301, A303, A311, V338, S344, T346, V359, G361, A364, T375, N379, S382, E391, A393, K394, S410, L417, T430, A431, R433, S444, T448 y/o S451 y/o una posición equivalente en una glucoamilasa precursora. En algunas formas de realización, la glucoamilasa precursora será un homólogo de glucoamilasa *Trichoderma* y en formas de realización adicionales, la glucoamilasa precursora tendrá al menos un 50 %, al menos un 60 %, al menos un 70 %, al menos un 80 %, al menos un 90 %, al menos un 95 % y al menos un 98 % de identidad de secuencia con SEQ ID NO:2 o 3. En algunas formas de realización, la glucoamilasa precursora también presentará identidad estructural con SEQ ID NO:2 y/o 3. En algunas formas de realización, la variante con termoestabilidad y actividad específica aumentadas cuenta con una sustitución en al menos una de las posiciones: 228, 230, 231, 268, 291, 417, 433 y 451 de SEQ ID NO:2 y/o 3.

Polinucleótidos:

[0126] La presente invención también hace referencia a polinucleótidos aislados que codifican una glucoamilasa variante de la presente invención, tal y como se define en las reivindicaciones. Los polinucleótidos pueden prepararse mediante técnicas consolidadas conocidas en la técnica. Los polinucleótidos pueden prepararse de forma sintética, tal como mediante un sintetizador de ADN automático. La secuencia de ADN puede ser de origen sintético y genómico combinados (o ADNc) preparada mediante la ligación conjunta de fragmentos. Los polinucleótidos también pueden prepararse mediante la reacción en cadena de la polimerasa (PCR por sus siglas en inglés) utilizando cebadores específicos. En general, se hace referencia a Minshull J., *et al.*, (2004), "Engineered protein function by selective amino acid diversification", *Methods* 32(4):416 - 427. El ADN también puede ser sintetizado por parte de diferentes empresas comerciales tal como Genearth AG, Regensburg, Alemania.

[0127] La presente exposición también proporciona polinucleótidos aislados que comprenden una secuencia de nucleótidos (i) con al menos un 50 % de identidad con SEQ ID NO:4, incluyendo al menos un 60 %, al menos un 70 %, al menos un 80 %, al menos un 90 %, al menos un 95 % y al menos un 99 % o (ii) capaz de hibridar con una sonda derivada de la secuencia de nucleótidos establecida en SEQ ID NO:4, con condiciones de astringencia intermedia a alta, o (iii) complementario a una secuencia de nucleótidos con al menos un 90 % de identidad de secuencia con la secuencia establecida en SEQ ID NO:4. Sondas útiles según la presente exposición pueden incluir al menos 50, 100, 150, 200, 250, 300 o más nucleótidos contiguos de SEQ ID NO:4. En algunas formas de realización, el polipéptido codificado también presenta identidad estructural con SEQ ID NO:2 y/o 3.

[0128] La presente exposición proporciona además polinucleótidos aislados que codifican glucoamilasas variantes que comprenden una secuencia de aminoácidos que comprende al menos un 50 %, al menos un 60 %, al menos un 70 %, al menos un 80 %, al menos un 90 %, al menos un 93 %, al menos un 95 %, al menos un 97 %, al menos un 98 % o al menos un 99 % de identidad secuencia de aminoácidos con SEQ ID NO:2. De forma adicional, la presente invención proporciona vectores de expresión que comprenden cualquiera de los polinucleótidos, tal y como se define en las reivindicaciones.

[0129] La presente exposición también proporciona fragmentos (es decir, partes) del ADN que codifican las glucoamilasas variantes proporcionadas en el presente documento. Estos fragmentos son útiles a la hora de obtener fragmentos de ADN de longitud parcial capaces de ser usados para aislar o identificar polinucleótidos que codifican enzimas de glucoamilasa maduras descritas en el presente documento a partir de células fúngicas filamentosas (p. ej., *Trichoderma*, *Aspergillus*, *Fusarium*, *Penicillium*, y *Humicola*) o un segmento de estos con actividad de glucoamilasa. En algunas formas de realización, los fragmentos de ADN pueden comprender al menos 50, 100, 150, 200, 250 300 o más nucleótidos contiguos. En algunas formas de realización, partes del ADN proporcionado en SEQ ID NO:4 son útiles a la hora de obtener la glucoamilasa precursora y especialmente homólogos de la glucoamilasa *Trichoderma* de otras especies, tal como hongos filamentosos que codifican una glucoamilasa.

Construcciones de ADN y vectores

[0130] Según una forma de realización de la invención, una construcción de ADN que comprende un polinucleótido según se describe anteriormente y que codifica una glucoamilasa variante incluida en la invención y unida de forma operativa a una secuencia promotora se junta para transferirse a una célula huésped.

5 [0131] La construcción de ADN puede introducirse en una célula huésped utilizando un vector. El vector puede ser cualquier vector que, cuando se introduce en una célula huésped, se introduzca de forma estable. En algunas formas de realización, el vector de la presente exposición se integra en un genoma de célula huésped y se duplica. Entre los vectores se incluyen vectores de clonación, vectores de expresión, vectores transportadores, partículas fágicas, plásmidos, casetes y similares. En algunas formas de realización, el vector es un vector de
10 expresión que comprende secuencias reguladoras unidas de forma operativa a la secuencia que codifica la glucoamilasa.

[0132] Ejemplos de vectores de expresión y/o integración se presentan en Sambrook *et al.*, (1989) *supra*, y Ausubel (1987) *supra*, y van den Hondel *et al.* (1991) en Bennett and Lasure (Eds.) *MORE GENE MANIPULATIONS IN FUNGI*, Academic Press pp. 396-428 y la patente estadounidense N° 5.874.276. También
15 se hace referencia al catálogo de cepas del Centro de Almacenamiento genético fúngico (FGSC, < www.fgsc.net>) para un listado de vectores. Entre los vectores especialmente útiles se incluyen vectores obtenidos de por ejemplo Invitrogen y Promega.

[0133] Plásmidos adecuados para utilizarse en células bacterianas incluyen pBR322 y pUC19 que permiten la replicación en *E.coli* y pE194 por ejemplo que permite la replicación en *Bacillus*. Otros vectores específicos
20 adecuados para utilizarse en células huésped *E. coli* incluyen vectores tales como pFB6, pBR322, pUC18, pUC100, pDONR™201, pDONR™221, pENTR™, pGEM@3Z y pGEM@4Z.

[0134] Entre los vectores específicos adecuados para utilizarse en células fúngicas se incluye pRAX, un vector de expresión con fin general útil en *Aspergillus*, pRAX con un promotor *glaA*, y en *Hypocreal Trichoderma* incluye pTrex3g con un promotor *cbh1*.

25 [0135] En algunas formas de realización, el promotor muestra actividad transcripcional en una célula huésped fúngica o bacteriana y puede derivarse de genes que codifican proteínas ya sean homólogas o heterólogas a la célula huésped. El promotor puede ser un promotor híbrido, truncado y/o mutante. Los promotores mencionados anteriormente son conocidos en la técnica. Ejemplos de promotores adecuados útiles en células fúngicas y especialmente en células fúngicas filamentosas tal como *Trichoderma* o *Aspergillus* incluyen tales promotores de
30 ejemplo como los promotores *cbh1*, *cbh2*, *egl1*, *egl2*, *eg5*, *xln1* y *xln2* de *T. reesei*. Otros ejemplos de promotores útiles incluyen promotores de genes de glucoamilasa *A. awamori* y *A. niger* (*glaA*) (Véase, Nunberg *et al.*, (1984) *Mol. Cell Biol.* 4:2306-2315 y Boel *et al.*, (1984) *EMBO J.* 3:1581-1585), promotor de TAKA amilasa *A. oryzae*, el promotor TPI (triosa fosfato isomerasa) de *S. cerevisiae*, el promotor de genes acetamidasa *Aspergillus nidulans* y genes lipasa *Rhiaomucor miehei*. Ejemplos de promotores adecuados útiles en células bacterianas incluyen aquellos obtenidos del operón lac *E. coli*; gen alfa-amilasa *Bacillus licheniformis* (*amyL*), gen amilasa *B. stearothermophilus* (*amyS*); genes *xylA* y *xylB* *Bacillus subtilis*, el gen beta-lactamasa y el promotor *tac*. En algunas formas de realización, el promotor es uno que es natural de la célula huésped. Por ejemplo, cuando el huésped es *T. reesei*, el promotor es un promotor natural de *T. reesei*. En otras formas de realización, el promotor es uno que es heterólogo a la célula huésped fúngica. En algunas formas de realización, el promotor
40 será el promotor de una glucoamilasa precursora (p. ej., el promotor de TrGA).

[0136] En algunas formas de realización, la construcción de ADN incluye ácidos nucleicos que codifican una secuencia señal, es decir, una secuencia de aminoácidos enlazada al amino terminal del polipéptido que dirige al polipéptido codificado a la trayectoria secretora de la célula. El extremo 5' de la secuencia codificante de la
45 secuencia de ácido nucleico puede incluir de forma natural una región codificante de péptido señal que está enlazada de forma natural en el marco de lectura de traducción con el segmento de la secuencia codificante de glucoamilasa que codifica la glucoamilasa secretada o el extremo 5' de la secuencia codificante de la secuencia de ácido nucleico puede incluir un péptido señal que es ajeno a la secuencia codificante. En algunas formas de realización, la construcción de ADN incluye una secuencia señal que está asociada de forma natural a un gen de glucoamilasa precursora a partir del cual se obtiene la glucoamilasa variante. En algunas formas de realización, la secuencia señal será la secuencia representada en SEQ ID NO:1 o una secuencia con al menos un 90 %, al
50 menos un 94 % y al menos un 98 % de identidad de secuencia con la misma. Secuencias señales eficaces pueden incluir las secuencias señales obtenidas de otras enzimas fúngicas filamentosas, tal como de *Trichoderma* (glucoamilasa *T. reesei*), *Humicola* (celulasa *H. insolens* o glucoamilasa *H. grisea*), *Aspergillus* (glucoamilasa *A. niger* y TAKA amilasa *A. oryzae*) y *Rhizopus*.

55 [0137] En formas de realización adicionales, una construcción de ADN o vector que comprende una secuencia señal y una secuencia promotora que han de introducirse en una célula huésped se derivan de la misma fuente. En algunas formas de realización, se puede utilizar la secuencia señal de glucoamilasa natural de un homólogo

de glucoamilasa *Trichoderma*, tal como una secuencia señal de una cepa de *Hypocrea*.

[0138] En algunas formas de realización, el vector de expresión también incluye una secuencia de terminación. En la presente exposición, puede utilizarse cualquier secuencia de terminación funcional de la exposición en la célula huésped. En algunas formas de realización, la secuencia de terminación y la secuencia promotora se derivan de la misma fuente. En otra forma de realización, la secuencia de terminación es homóloga a la célula huésped. Entre las secuencias de terminación útiles se incluyen secuencias de terminación obtenidas a partir de los genes de *cbh1* *Trichoderma reesei*; glucoamilasa *A. niger* o *A. awamori* (Nunberg *et al.* (1984) *supra*, y Boel *et al.*, (1984) *supra*), antranilato sintasa *Aspergillus nidulans*, Taka amilasa *Aspergillus oryzae* o *A. nidulans trpC* (Punt *et al.*, (1987) *Gene* 56:117 - 124).

[0139] En algunas formas de realización, el vector de expresión incluye un marcador seleccionable. Entre los ejemplos de marcadores seleccionables se incluyen unos que confieren resistencia antimicrobiana (p. ej., higromicina y fleomicina). Los marcadores selectivos nutricionales también son útiles en la presente exposición incluyendo aquellos marcadores conocidos en la técnica como *amdS* (acetamidasa), *argB* (ornitina carbamoiltransferasa) y *pyrG* (orotidina-5' fosfato descarboxilasa). Los marcadores útiles en sistemas de vectores para la transformación de *Trichoderma* son conocidos en la técnica (Véase, p. ej., Finkelstein, capítulo 6 en *BIOTECHNOLOGY OF FILAMENTOUS FUNGI*, Finkelstein *et al.* (1992) Eds. Butterworth-Heinemann, Boston, MA; Kinghorn *et al.* (1992) *APPLIED MOLECULAR GENETICS OF FILAMENTOUS FUNGI*, Blackie Academic and Professional, Chapman and Hall, London; Berges y Barreau (1991) *Curr. Genet.* 19:359-365; y van Hartingsveldt *et al.*, (1987) *Mol. Gen. Genet.* 206:71 - 75). En algunas formas de realización, el marcador selectivo es el gen *amdS*, que codifica la enzima acetamidasa, lo que permite que las células transformadas crezcan en acetamida como una fuente de nitrógeno. El uso del gen *amdS* *A. nidulans* como marcador selectivo se describe en Kelley *et al.*, (1985) *EMBO J.* 4:475 - 479 y Penttilä *et al.*, (1987) *Gene* 61:155-164.

[0140] Los métodos utilizados para ligar la construcción de ADN que comprende una secuencia de ácido nucleico que codifica una glucoamilasa variante, un promotor, una terminación y otras secuencias y para insertarlos en un vector adecuado son conocidos en la técnica. El enlace se consigue generalmente mediante la ligación en los lugares de restricción convenientes. Si no existen estos lugares, los enlazadores de oligonucleótidos sintéticos se utilizan según la práctica convencional. (Véase, Sambrook (1989) *supra*, y Bennett y Lasure, *MORE GENE MANIPULATIONS IN FUNGI*, Academic Press, San Diego (1991) pp 70 - 76.). De forma adicional, los vectores pueden construirse utilizando técnicas de recombinación conocidas (p. ej., Invitrogen Life Technologies, Gateway Technology).

Células huésped:

[0141] La presente invención también hace referencia a células huésped que comprenden un polinucleótido que codifica una glucoamilasa variante de la invención tal y como se define en las reivindicaciones. En algunas formas de realización, las células huésped de la presente exposición se eligen de células bacterianas, fúngicas, vegetales y de levadura. El término célula huésped incluye tanto las células, la progenie de las células y los protoplastos creados a partir de las células que se utilizan para producir una glucoamilasa variante según la presente exposición.

[0142] En algunas formas de realización, las células huésped de la presente exposición son células fúngicas y preferiblemente células huésped fúngicas filamentosas. El término "hongos filamentosos" hace referencia a todas las formas filamentosas de la subdivisión Eumycotina (Véase Alexopoulos, C. J. (1962), *INTRODUCTORY MYCOLOGY*, Wiley, Nueva York). Estos hongos se caracterizan por un micelio vegetativo con una pared celular compuesta de quitina, celulosa y otros polisacáridos complejos de la exposición. Los hongos filamentosos de la presente exposición son distintos morfológica, fisiológica y genéticamente de las levaduras. El crecimiento vegetativo por hongos filamentosos se realiza mediante elongación de las hifas y el catabolismo de carbono es aerobio obligado. En la presente exposición, la célula precursora fúngica filamentosa puede ser una célula de una especie de entre, sin carácter limitativo, *Trichoderma*, (p. ej., *Trichoderma reesei*, la mutación asexual de *Hypocrea jecorina*, clasificada anteriormente como *T. longibrachiatum*, *Trichoderma viride*, *Trichoderma koningii*, *Trichoderma harzianum*) (Sheir-Neirs *et al.*, (1984) *Appl. Microbiol. Biotechnol* 20:46 - 53; ATCC N° 56765 y ATCC N° 26921); *Penicillium sp.*, *Humicola sp.* (p. ej., *H. insolens*, *H. lanuginosa* y *H. grisea*); *Chrysosporium sp.* (p. ej., *C. lucknowense*), *Gliocladium sp.*, *Aspergillus sp.* (p. ej., *A. oryzae*, *A. niger*, *A. sojae*, *A. japonicus*, *A. nidulans*, y *A. awamori*) (Ward *et al.*, (1993) *Appl. Microbiol. Biotechnol.* 39:738 - 743 y Goedegebuur *et al.*, (2002) *Genet* 41:89 -98), *Fusarium sp.*, (p. ej., *F. roseum*, *F. graminum* *F. cerealis*, *F. oxysporum* y *F. venenatum*), *Neurospora sp.*, (*N. crassa*), *Hypocrea sp.*, *Mucor sp.*, (*M. miehei*), *Rhizopus sp.* y *Emericella sp.* (Véase también, Innis *et al.*, (1985) *Sci.* 228:21 -26). Los términos "*Trichoderma*" o "*Trichoderma sp.*" o "*Trichoderma spp.*" hacen referencia a cualquier género fúngico clasificado previamente o actualmente como *Trichoderma*.

[0143] En algunas formas de realización, las células huésped serán células bacterianas gram positivas. Entre

los ejemplos sin carácter limitativo se incluyen cepas de *Streptomyces*, (p. ej., *S. lividans*, *S. coelicolor* y *S. griseus*) y *Bacillus*. Tal y como se utiliza en el presente documento, "el género *Bacillus*" incluye todas las especies dentro del género "*Bacillus*", según lo conocen los expertos en la técnica, entre las que se incluyen sin carácter limitativo *B. subtilis*, *B. licheniformis*, *B. lentus*, *B. brevis*, *B. stearothermophilus*, *B. alkalophilus*, *B. amyloliquefaciens*, *B. clausii*, *B. halodurans*, *B. megaterium*, *B. coagulans*, *B. circulans*, *B. lautus* y *B. thuringiensis*. Se reconoce que el género *Bacillus* continúa experimentando reorganización taxonómica. Por lo tanto, se pretende que el género incluya especies que han sido reclasificadas, incluyendo sin carácter limitativo tales organismos como *B. stearothermophilus*, que ahora se denomina "*Geobacillus stearothermophilus*".

[0144] En algunas formas de realización, la célula huésped es una cepa bacteriana gram negativa, tal como *E. coli* o *Pseudomonas sp.* En otras formas de realización, las células huésped pueden ser células de levadura tal como *Saccharomyces sp.*, *Schizosaccharomyces sp.*, *Pichia sp.* o *Candida sp.*

[0145] En otras formas de realización, la célula huésped será una célula huésped diseñada genéticamente en la que los genes naturales se han desactivado, por ejemplo, mediante delección en células bacterianas o fúngicas. Cuando se desee obtener una célula huésped fúngica con uno o más genes desactivados, se pueden utilizar métodos conocidos (p. ej., los métodos expuestos en la patente estadounidense N° 5.246.853, patente estadounidense N° 5.475.101 y WO 92/06209). La desactivación génica puede conseguirse mediante una delección completa o parcial, mediante desactivación por inserción o mediante cualquier otro medio que proporcione un gen no funcional para el fin destinado (de forma que se impida la expresión del gen a partir de una proteína funcional). En algunas formas de realización, cuando la célula huésped es una célula *Trichoderma* y especialmente una célula huésped *T. reesei*, los genes *cbh1*, *cbh2*, *egl1* y *egl2* se desactivarán y/o eliminarán. Células huésped *Trichoderma reesei* de ejemplo con proteínas con delección cuádruple se establecen y describen en la patente estadounidense 5.847.276 y WO 05/001036. En otras formas de realización, la célula huésped es una cepa carente de proteasa o proteasa negativa.

Transformación de células huésped:

[0146] La introducción de una construcción de ADN o vector en una célula huésped incluye técnicas tales como transformación, electroporación, microinyección nuclear, transducción, transfección (p. ej., transfección mediada por lipofección y mediada por DEAE-Dextrina); incubación con precipitado de ADN fosfato calcio; bombardeo de alta velocidad con microproyectiles revestidos de ADN y fusión de protoplastos. Las técnicas de transformación generales se conocen en la técnica (Véase, p. ej., Ausubel *et al.*, (1987), *supra*, capítulo 9; y Sambrook (1989) *supra*, y Campbell *et al.*, (1989) *Curr. Genet.* 16:53-56).

[0147] Los métodos de transformación para *Bacillus* se exponen en numerosas referencias entre las que se incluyen Anagnostopoulos C y J. Spizizen (1961) *J. Bacteriol.* 81:741 - 746 y WO 02/14490.

[0148] Los métodos de transformación para *Aspergillus* se describen en Yelton *et al.* (1984) *Proc. Natl. Acad. Sci. EEUU* 81:1470 - 1474; Berka *et al.*, (1991) en *Applications of Enzyme Biotechnology*, Eds. Kelly and Baldwin, Plenum Press (NY); Cao *et al.*, (2000) *Sci.* 9:991 - 1001; Campbell *et al.*, (1989) *Curr. Genet.* 16:53-56 y EP 238 023. La expresión de proteína heteróloga en *Trichoderma* se describe en la patente estadounidense 6.022.725; patente estadounidense 6.268.328; Harkki *et al.* (1991); *Enzyme Microb. Technol.* 13:227-233; Harkki *et al.*, (1989) *Bio Technol.* 7:596-603; EP 244.234; EP 215.594; y Nevalainen *et al.*, "The Molecular Biology of *Trichoderma* and its Application to the Expression of Both Homologous and Heterologous Genes", en *MOLECULAR INDUSTRIAL MYCOLOGY*, Eds. Leong y Berka, Marcel Dekker Inc., NY (1992) pp. 129 - 148). También se hace referencia a WO96/00787 y Bajar *et al.* (1991) *Proc. Natl. Acad. Sci. EEUU* 88:8202 - 28212 para la transformación de cepas *Fusarium*.

[0149] En una forma de realización específica, la preparación de *Trichoderma sp.* para la transformación implica la preparación de protoplastos a partir de micelios fúngicos (Véase, Campbell *et al.* (1989) *Curr. Genet.* 16:53-56; Pentilla *et al.*, (1987) *Gene* 61:155 - 164). La transformación mediada por *Agrobacterium tumefaciens* de hongos filamentosos es conocida. (Véase, de Groot *et al.*, (1998) *Nat. Biotechnol.* 16:839 - 842). También se hace referencia a la patente estadounidense 6.022.725 y a la patente estadounidense 6.268.328 para la transformación de procedimientos utilizados con huéspedes fúngicos filamentosos.

[0150] En algunas formas de realización, los transformantes estables genéticamente se construyen con sistemas de vectores por los que el ácido nucleico que codifica la glucoamilasa variante se integra de forma estable en un cromosoma de la cepa huésped. A continuación, se purifican los transformantes con técnicas conocidas.

[0151] En algunas formas de realización adicionales, las células huésped son células vegetales, tal como células de una planta monocotiledónea (p. ej., maíz, trigo y sorgo) o células de una planta dicotiledónea (p. ej., soja). Se conocen los métodos para realizar construcciones de ADN útiles para la transformación de plantas y métodos para la transformación vegetal. Algunos de estos métodos incluyen transferencia génica mediada por

Agrobacterium tumefaciens, bombardeo de microproyectiles, transformación de protoplastos mediada por PEG, electroporación y similares. Se hace referencia a la patente estadounidense 6.803.499, la patente estadounidense 6.777.589; Fromm *et al.* (1990) *Biotechnol.* 8:833-839; Potrykus *et al.* (1985) *Mol.Gen. Genet.* 199:169-177).

5 **Producción de proteínas:**

[0152] La presente invención hace referencia además a métodos para producir las glucoamilasas variantes tal y como se define en las reivindicaciones, que comprende la transformación de una célula huésped con un vector de expresión que comprende un polinucleótido que codifica una glucoamilasa variante de acuerdo con la invención, el cultivo de la célula huésped en condiciones adecuadas para la expresión y producción de la glucoamilasa variante y la recuperación de forma opcional de la glucoamilasa variante.

[0153] En los métodos de producción y expresión de la presente invención, las células huésped se cultivan bajo condiciones adecuadas en cultivo con agitación del matraz, fermentaciones a pequeña o gran escala (incluyendo fermentaciones continuas, por lotes y discontinuas) en fermentadores industriales o de laboratorio, con el medio adecuado que contenga sales fisiológicas y nutrientes (Véase, p. ej., Pourquie, J. *et al.*, *BIOCHEMISTRY AND GENETICS OF CELLULOSE DEGRADATION*, eds. Aubert, J. P. *et al.*, Academic Press, pp. 71-86, 1988 e Ilmen, M. *et al.*, (1997) *Appl. Environ. Microbiol.* 63:1298-1306). Los medios comunes preparados comercialmente (p. ej., caldo de extracto de malta levadura, caldo Luria Bertani y el caldo Sabouraud Dextrosa) son útiles en la presente invención. Las condiciones de cultivo para células fúngicas filamentosas y bacterianas son conocidas en la técnica y pueden hallarse en la literatura científica y/o desde la fuente de los hongos tal como American Type Culture Collection y Fungal Genetics Stock Center. En los casos en los que una secuencia codificante de glucoamilasa esté bajo el control de un promotor inducible, el agente inductor (p. ej., un azúcar, una sal metálica o antimicrobiano) se añade al medio con una concentración eficaz para inducir la expresión de glucoamilasa.

[0154] En algunas formas de realización, la presente exposición hace referencia a métodos para producir la glucoamilasa variante en un huésped vegetal que comprende la transformación de una célula vegetal con un vector que comprende un polinucleótido que codifica una variante de glucoamilasa de acuerdo con la invención y el crecimiento de la célula vegetal bajo condiciones adecuadas para la expresión y producción de la variante.

[0155] En algunas formas de realización, los ensayos se llevan a cabo para evaluar la expresión de una glucoamilasa variante mediante una línea celular que ha sido transformada con un polinucleótido que codifica una glucoamilasa variante incluida en la invención. Los ensayos pueden llevarse a cabo en el nivel de proteína, el nivel de ARN y/o mediante el uso de bioanálisis funcionales específicos de la actividad y/o producción de glucoamilasa. Algunos de estos ensayos incluyen la prueba de Northern blot, dot blot (análisis de ADN o ARN), RT-PCR (reacción en cadena de la polimerasa con transcriptasa inversa), hibridación *in situ* utilizando una sonda marcada de forma adecuada (según la secuencia codificante de ácido nucleico) y la prueba de Southern blot convencional y autorradiografía.

[0156] Además, la producción y/o expresión de una glucoamilasa variante puede medirse en una muestra directamente, por ejemplo, mediante ensayos que miden directamente los azúcares reductores tal como glucosa en el medio de cultivo y mediante ensayos que miden la actividad, expresión y/o producción de glucoamilasa. En concreto, la actividad de glucoamilasa puede analizarse mediante el método del ácido 3,5-dinitrosalicílico (DNS) (Véase, Goto *et al.*,

[0157] (1994) *Biosci. Biotechnol. Biochem.* 58:49 - 54). En formas de realización adicionales, la expresión de proteína se evalúa mediante métodos inmunológicos, tal como la coloración inmunohistoquímica de células, secciones de tejido o inmunoanálisis de un medio de cultivo de tejido (p. ej., mediante Western blot o ELISA). Tales inmunoanálisis pueden utilizarse para evaluar de forma cualitativa y cuantitativa la expresión de una glucoamilasa. Los expertos en la técnica conocen los detalles de tales métodos y muchos reactivos para practicar tales métodos están disponibles comercialmente.

[0158] Las glucoamilasas de la presente invención pueden recuperarse o purificarse del medio de cultivo mediante una variedad de procedimientos conocidos en la técnica entre los que se incluyen la centrifugación, la filtración, la extracción, la precipitación y similares.

50 **Composiciones:**

[0159] Las glucoamilasas variantes de la invención pueden utilizarse en composiciones de enzima entre las que se incluyen sin carácter limitativo composiciones para la hidrólisis y sacarificación de almidón, composiciones de limpieza y detergente (p. ej., detergentes para la colada, detergentes para lavar los platos y composiciones de limpieza de superficies duras), composiciones para la fermentación del alcohol y en composiciones para pienso animal. Además, las glucoamilasas variantes pueden utilizarse en aplicaciones de cocción, tal como la

producción de pan y pasteles, destilación, atención sanitaria, textil, procesos de conversión de residuos medioambientales, procesamiento de biopulpeo y aplicaciones de conversión de biomasa.

[0160] En algunas formas de realización, se utilizará de forma opcional una composición de enzima de la presente exposición que comprende una glucoamilasa variante incluida en la invención junto con cualquiera de las siguientes enzimas o una combinación de estas - alfa-amilasas, proteasas, pululanasa, isoamilasas, celulasas, hemicelulasas, xilanasas, ciclodextrina glicotransferasas, lipasas, fitasas, laccasas, oxidasas, esterasas, cutinasas, xilanasas, enzimas hidrolizantes de almidón granular y otras glucoamilasas.

[0161] En algunas formas de realización, la composición de enzima incluirá un alfa-amilasa, tal como alfa-amilasas fúngicas (p. ej., *Aspergillus sp.*) o alfa-amilasas bacterianas (p. ej., *Bacillus sp.* tal como *B. stearothermophilus*, *B. amyloliquefaciens* y *B. licheniformis*) y variantes e híbridos de estas. En algunas formas de realización, el alfa-amilasa es un alfa-amilasa estable ácida. En algunas formas de realización, la alfa-amilasa es alfa-amilasa *Aspergillus kawachi* (Akaa), véase US 7.037.704. Las alfa-amilasas disponibles comercialmente contempladas para utilizarse en las composiciones de la invención son conocidas e incluyen GZYME G997, SPEZYME FRED, SPEZYME XTRA, (Danisco US, Inc, Genencor Division), TERMAMYL 120-L y SUPRA (Novozymes, Biotech).

[0162] En algunas formas de realización, la composición de enzima incluirá una proteasa fúngica ácida. En una forma de realización adicional, la proteasa fúngica ácida se deriva de una *Trichoderma sp.* y puede ser cualquiera de las proteasas expuestas en US 2006/015342, publicada el 7/13/2006 (p. ej., SEQ ID NO:10). En una forma de realización adicional, la composición de enzima incluirá una fitasa (p. ej., una *Buttiauxiella spp.* o una variante de la misma (p. ej., véase la publicación de patente PCT WO 2006/043178).

[0163] En otras formas de realización, las glucoamilasas variantes de la invención pueden combinarse con otras glucoamilasas. En algunas formas de realización, las glucoamilasas de la invención se combinarán con una o más glucoamilasas derivadas de cepas de *Aspergillus* o variantes de esta, tal como *A. oryzae*, *A. niger*, *A. kawachi*, y *A. awamori*; glucoamilasas derivadas de cepas de *Humicola* o variantes de esta, especialmente *H. grisea*, tal como la glucoamilasa con al menos un 90 %, 93 %, 95 %, 96 %, 97 %, 98 % y 99% de identidad de secuencia con SEQ ID NO:3 expuesta en WO 05/052148; glucoamilasas derivadas de cepas de *Talaromyces* o variantes de esta, especialmente *T. emersonii*; glucoamilasas derivadas de cepas de *Athelia* y especialmente *A. rolfii*; glucoamilasas derivadas de cepas de *Penicillium*, especialmente *P. chrysogenum*.

Usos:

[0164] En concreto, las glucoamilasas variantes expuestas en el presente documento pueden utilizarse para los procesos de conversión de almidón y especialmente en la producción de dextrosa para jarabes de fructosa, azúcares especializados y en alcohol y la producción de otros productos finales (p. ej., ácido orgánico, ácido ascórbico y aminoácidos) a partir de la fermentación de sustratos que contienen almidón (G.M.A van Beynum *et al.*, Eds. (1985) *STARCH CONVERSION TECHNOLOGY*, Marcel Dekker Inc. NY). Las dextrinas producidas utilizando composiciones de glucoamilasa variante de la presente exposición pueden tener como resultado una producción de glucosa de al menos un 80 %, al menos un 85 %, al menos un 90 % y al menos un 95 %. La producción de alcohol a partir de la fermentación de sustratos de almidón con glucoamilasas incluidas en la invención puede incluir la producción de carburante alcohólico o alcohol de boca. En algunas formas de realización, la producción de alcohol será mayor cuando la glucoamilasa variante se utilice con las mismas condiciones que la glucoamilasa precursora. En algunas formas de realización, la producción de alcohol será entre aproximadamente un 0,5 % y 2,5 % mejor, incluyendo sin carácter limitativo 0,6 %; 0,7 %; 0,8 %; 0,9 %; 1,0%; 1,1 %; 1,2 %; 1,3 %; 1,4 %; 1,5 %; 1,6 %; 1,7 %; 1,8 %; 1,9 %; 2,0 %; 2,1 %; 2,2 %; 2,3 % y 2,4 % más de alcohol que la glucoamilasa precursora.

[0165] En algunas formas de realización, las glucoamilasas variantes de la invención serán útiles en la hidrólisis de almidón de diferentes sustratos con base vegetal, que se utilizan para la producción de alcohol. En algunas formas de realización, los sustratos con base vegetal incluirán maíz, trigo, cebada, centeno, sorgo, arroz, caña de azúcar, patatas y combinaciones de estos. En algunas formas de realización, el sustrato con base vegetal se fraccionará en material vegetal, por ejemplo, un grano de cereal, tal como maíz, que se fracciona en componentes tal como fibra, germen, proteína y almidón (endospermo) (patente estadounidense 6.254.914 y patente estadounidense 6.899.910). Los métodos para las fermentaciones de alcohol se describen en *THE ALCOHOL TEXTBOOK, A REFERENCE FOR THE BEVERAGE, FUEL AND INDUSTRIAL ALCOHOL INDUSTRIES*, 3ª Ed., Eds K.A. Jacques *et al.*, 1999, Nottingham University Press, Reino Unido. En determinadas formas de realización, el alcohol será etanol. En concreto, los procesos de producción de fermentación de alcohol se caracterizan como procesos de molienda húmeda o seca. En algunas formas de realización, la glucoamilasa variante se utilizará en un proceso de fermentación de molienda húmeda y en otras formas de realización la glucoamilasa variante será útil en un proceso de molienda seca.

[0166] La molienda de grano húmeda implica un número de etapas básicas, que incluyen generalmente:

triturado, cocción, licuefacción, sacarificación, fermentación y separación de líquidos y sólidos con el fin de producir alcohol y otros subproductos. Se tritura material vegetal y especialmente granos de cereales integrales, tal como maíz, trigo o centeno. En algunos casos, el grano se puede fraccionar primero en partes componentes. El material vegetal triturado puede molerse para obtener una partícula fina o gruesa. El material vegetal triturado se mezcla con líquido (p. ej., agua y/o finos residuos de la destilería) en un tanque de suspensión. La suspensión está sujeta a altas temperaturas (p. ej., de 90 °C a 105 °C o superiores) en un cocedor a chorro junto con enzimas de licuefacción (p. ej., alfa-amilasas) para solubilizar e hidrolizar el almidón en el grano a dextrinas. La mezcla se enfría y se trata además con enzimas de sacarificación, tal como glucoamilasas incluidas en la presente invención, para producir glucosa. A continuación, la mezcla que contiene glucosa puede fermentarse durante aproximadamente 24 a 120 horas en presencia de microorganismos de fermentación, tal como microorganismos que producen etanol y especialmente levadura (*Saccharomyces spp*). Los sólidos en la mezcla se separan de la fase líquida y se obtiene alcohol tal como etanol y subproductos útiles tal como granos de destilación.

[0167] En algunas formas de realización, la etapa de sacarificación y la etapa de fermentación se combinan y se hace referencia al proceso como sacarificación y fermentación simultáneas o sacarificación, propagación de levadura y fermentación simultáneas.

[0168] En otras formas de realización, la glucoamilasa variante se utiliza en un proceso para la hidrólisis de almidón en el que la temperatura del proceso está entre 30 °C y 75 °C, en algunas formas de realización, entre 40 °C y 65 °C. En algunas formas de realización, la glucoamilasa variante se utiliza en un proceso para la hidrólisis de almidón con un pH de entre pH 3,0 y pH 6,5. Los procesos de fermentación en algunas formas de realización incluyen la molienda de un grano de cereal o grano fraccionado y la combinación del grano de cereal molido con líquido con el fin de formar una suspensión que a continuación se mezcla en un único recipiente con una glucoamilasa variante de acuerdo con la invención y de forma opcional otras enzimas tal como, sin carácter limitativo, alfa-amilasas, otras glucoamilasas, fitasas, proteasas, pululaninas, isoamilasas u otras enzimas con actividad hidrolizante de almidón granular y levadura con el fin de producir etanol y otros subproductos (véase, p. ej., la patente estadounidense 4.514.496, WO 04/081193 y WO 04/080923).

[0169] En algunas formas de realización, la presente exposición pertenece a un método para sacarificar una solución de almidón líquida, que comprende una etapa de sacarificación enzimática con una glucoamilasa variante de la invención.

[0170] La presente exposición también presenta una composición o formulación de pienso animal que comprende al menos una glucoamilasa variante incluida en la invención. Los métodos para usar una enzima de glucoamilasa en la producción de piensos que comprenden almidón se presentan en WO 03/049550. En resumen, la variante de glucoamilasa se mezcla con un pienso que comprende almidón. La glucoamilasa es capaz de degradar almidón resistente para ser usado por el animal.

[0171] Otros objetos y ventajas de la presente invención resultan evidentes a partir de la presente memoria.

EXPERIMENTAL

[0172] En la exposición y la sección experimental que se presenta a continuación, se aplican las siguientes abreviaturas: GA (glucoamilasa); GAU (unidad glucoamilasa); % p (porcentaje en peso); °C (grados centígrados); rpm (revoluciones por minuto); H₂O (agua); dH₂O (agua desmineralizada), dH₂O (agua desmineralizada, filtración Milli-Q) aa o AA (aminoácido); pb (pares de bases); kb (pares de kilobases); kD (kilodaltons); g o gm (gramos); µg (microgramos); mg (miligramos); µL (microlitros); ml y mL (mililitros); mm (milímetros), µm (micrómetro); M (molar); mM (milimolar); µM (micromolar); U (unidades); V (voltios); PM (peso molecular); s (segundo/segundos); min(s) o m(s) (minuto/minutos); hr(s) o h(s) (hora/horas); OD (oxígeno disuelto); ABS (absorbancia); EtOH (etanol); SF (suero fisiológico); m/v (masa/volumen) y PMT (placa microtitulación); N (Normal); DP1 (monosacáridos); DP2 (disacáridos); DP>3 (oligosacáridos, azúcares con un grado de polimerización superior a 3); ppm (partes por millón).

[0173] Los análisis y métodos utilizados en los ejemplos se presentan a continuación. Sin embargo, cabe destacar que pueden utilizarse diferentes métodos para proporcionar variantes de una molécula precursora y la invención no está limitada a los métodos utilizados en los ejemplos. Se pretende que se pueda utilizar cualquier medio adecuado para realizar variantes y para la selección de variantes.

Análisis de la actividad glucoamilasa pNPG para placas de microtitulación de 96 pocillos:

[0174] Las soluciones de reactivos fueron: tampón NaAc (200 mM de tampón de acetato de sodio pH 4,5); sustrato (50 mM p-nitro-fenil-α-D-glucopiranosido (Sigma N-1377) en tampón NaAc (0,3 g/20 ml) y solución de cese de reacción (800 mM tampón glicina-NaOH pH 10). Se colocaron 30 µl de sobrenadante filtrado en una PMT nueva con base plana de 96 pocillos. Se añadieron a cada pocillo 50 µl de tampón NaAc y 120 µl de

sustrato y se incubaron durante 30 minutos a 50 °C (incubadora/agitador HT iEMS de Thermolab systems). Se finalizó la reacción añadiendo 100 µl de solución de cese de reacción. La absorbancia se midió a 405 nm en un lector de PMT (Spectramax 384 plus de Molecular Devices) y la actividad se calculó utilizando un coeficiente de extinción molar de 0,011 µM/cm.

5 **Análisis de estabilidad térmica:**

[0175] Se añadió sobrenadante crudo (8 µl) a 280 µl 50 mM tampón NaAc pH 4,5. La muestra diluida se dividió equitativamente sobre 2 PMT. Una PMT (placa inicial) se incubó durante 1 hr a 4 °C y la otra PMT (placa residual) se incubó a 64 °C (incubadora/agitador HT iEMS de Thermolab systems) durante 1 hr. La placa residual se enfrió durante 10 min en hielo. Se midió la actividad de ambas placas utilizando el análisis de detección de etanol descrito a continuación. Se añadieron 60 µl de la placa inicial o placa residual a 120 µl 4 % almidón de maíz soluble y se incubó durante 2 horas a 32 °C 900 rpm (incubadora/agitador HT iEMS de Thermolab systems).

[0176] Se calculó la estabilidad térmica como el % de actividad residual tal y como se muestra a continuación:

$$\frac{\text{ABS (340) residual} - \text{vacío}}{\text{ABS (340) inicial} - \text{vacío}} \times 100\%$$

15 [0177] Se analizó el material sobrenadante crudo para detectar la glucosa restante en el medio de cultivo tras el periodo de crecimiento. Si se hallaba glucosa restante, no se calculaba la termoestabilidad.

[0178] Según la actividad residual del WT y mutante, se calculó el índice de rendimiento (PI) para detectar la estabilidad térmica. El PI de una variante es el cociente "actividad residual variante/actividad residual WT". El PI de WT es 1,0 y una variante con un PI > 1,0 tiene una actividad específica que es mayor que WT.

20 **Determinación de proteína Caliper**

[0179] Análisis de datos y cálculo del índice de rendimiento del ensayo de aplicación de etanol

[0180] Se midieron los niveles de proteína utilizando un instrumento de electroforesis de microfluidos (Caliper Life Sciences, Hopkinton, MA, EEUU). Se prepararon las muestras de proteína y el chip de microfluidos según las instrucciones del fabricante (LABCHIP® HT Protein Express, P/N 760301). Se prepararon los sobrenadantes de cultivo y se almacenaron en placas de microtitulación de 96 pocillos a -20 °C hasta que se utilizaron, momento en el que se descongelaron mediante el calentamiento en una incubadora a 37 °C durante 30 minutos. Después de una breve agitación, se transfirieron 2 µl de cada muestra de cultivo a una placa de PCR de 96 pocillos (Bio-Rad, Hercules, CA, EEUU) con muestras de 7 µl de solución amortiguadora (Caliper) seguido del calentamiento de la placa a 90 °C durante 5 minutos sobre un calentador de placa controlado de forma termostática. Se dejó que la placa se enfriara antes de añadir 40 µl de agua a cada muestra. A continuación, la placa se colocó en el instrumento junto con un ejemplo de proteína suministrado y calibrado por el fabricante. Cuando las proteínas pasan por un punto focal en el chip, se registró la señal fluorescente y se determinó la concentración de proteína mediante la cuantificación de la señal en relación a la señal generada por el conjunto calibrado de ejemplos de proteína.

35 [0181] Después de la determinación de proteína Caliper, se procesaron los datos de la siguiente forma. Se comprueban las escaleras de calibración para la exactitud del patrón pico. Si la escalera de calibración que estaba asociada a la secuencia no era suficiente, se reemplazaba con una escalera de calibración de una secuencia adyacente. Para la detección del pico, se empleó la configuración por defecto de la opción para hallar el pico global del software caliper. El pico de interés se eligió a 75 kDa +/- 10 %.

40 [0182] El resultado se exportó a un programa de hojas de cálculo y el área pico se relacionó con la actividad correspondiente (ABS340-medición en blanco) en el ensayo de aplicación de etanol. Con los números de actividad y el área de 12 muestras de tipo silvestre, se realizó una línea de calibración utilizando la ecuación "Cinética de enzimas" del programa Grafit Version 5 (Erithacus Software, Horley, Reino Unido) junto con una función de ajuste no lineal. Se empleó la configuración por defecto para calcular los parámetros de Vmax y Km. Según estos dos parámetros, se realizó una línea de referencia Michaelis-Menten y se calculó la actividad específica de cada variante según esta línea de referencia.

45 [0183] Según la actividad específica del WT y variantes, se calculó el índice de rendimiento (PI) para detectar la actividad específica. El PI de una variante es el cociente "actividad específica variante/actividad específica WT". El PI de WT es 1,0 y una variante con un PI > 1,0 tenía una actividad específica que es mayor que WT.

50

Análisis de actividad de hexoquinasa:

[0184] Cóctel de hexoquinasa: 10-15 minutos antes de usarse, se añadieron 90 ml de agua a un recipiente BoatlL de glucosa HK R1 (kit de análisis de glucosa IL (HK), Instrumento de Laboratorio nº 182507-40) y se mezcló gradualmente. Se añadieron 100 µl de cóctel de hexoquinasa a 85 µl de dH₂O. Se añadieron 15 µl de muestra a las mezclas y se incubaron durante 10 minutos en la oscuridad a temperatura ambiente. Se leyó la absorbancia a 340 nm en un lector de PMT. Las concentraciones de glucosa se calcularon según una curva estándar de glucosa (0 - 2 mg/ml).

Condiciones de análisis del análisis de detección de etanol:

[0185] 8 % de solución madre: Se suspendieron 8 g de almidón de maíz soluble (Sigma nº S4180) en 40 ml dH₂O a temperatura ambiente. Se añadieron 50 ml de dH₂O hirviendo a la suspensión en un matraz de 250 ml y se coció durante 5 minutos. La solución de almidón se enfrió a 25 °C y se reguló el volumen a 100 ml con dH₂O.

[0186] Solución de cese de reacción: 800 mM tampón glicina-NaOH, pH 10

[0187] 4 % (m/v) solución de trabajo de almidón soluble: la solución madre se diluyó (1:1) con 100 mM tampón acetato de sodio, pH 3,7.

[0188] Se diluyeron 5 µl de sobrenadante crudo con 175 µl 50 mM de tampón NaAc pH 4,5 en una PMT de 96 pocillos de base plana. Se añadieron 60 µl de esta solución a 120 µl 4 % almidón de maíz soluble y se incubaron durante 2 horas a 32 °C 900 rpm (incubadora/agitador HT iEMS de Thermolabsystems). Se finalizó la reacción añadiendo 90 µl de solución de cese de reacción enfriada a 4 °C. La muestra se colocó en hielo. El almidón se centrifugó a 1118xg a 15 °C durante 5 minutos (SIGMA 6K15) y se utilizaron 15 µl de sobrenadante en el análisis de actividad de Hexoquinasa descrito anteriormente con el fin de determinar el contenido de glucosa.

[0189] Se analizó el material sobrenadante crudo para detectar la glucosa restante en el medio de cultivo tras el periodo de crecimiento. Si se hallaba glucosa restante, la cantidad de glucosa producida por la glucoamilasa no se calculaba.

EJEMPLOS

[0190] Los siguientes ejemplos se presentan con el fin de demostrar e ilustrar además determinadas formas de realización y aspectos de la presente invención y no han de considerarse como limitadores del alcance de la misma.

Ejemplo 1: Construcción de bibliotecas de evaluación del lugar TrGA (SEL por sus siglas en inglés) en el vector pTTT para la expresión en *Trichoderma reesei*

[0191] Se clonó una secuencia de ADNc *Trichoderma reesei* (SEQ ID NO:4) en pDONR™201 mediante la reacción de recombinación Gateway® BP (Invitrogen, Carlsbad, CA, EEUU) teniendo como resultado el vector de entrada pDONR-TrGA (Fig. 2). La secuencia de ADNc (SEQ ID NO:4) codifica el péptido señal TrGA, la prosequencia y la proteína madura, incluyendo el dominio catalítico, la región de enlace y el dominio de unión a almidón (SEQ ID NO:1). SEQ ID NO:4 y SEQ ID NO:1 se muestran en la figura 1B y 1A. La figura 1C ilustra los dominios de TrGA de proteína madura y precursora.

[0192] Con el fin de expresar la proteína TrGA en *Trichoderma reesei*, la secuencia codificante TrGA (SEQ ID NO:4) se clonó en el vector de destino compatible Gateway pTTT-Dest (Fig. 3) mediante la reacción de recombinación GATEWAY® LR. El vector de expresión contenía las regiones terminadoras y promotoras derivadas de cbhl *T. reesei* que permitían una fuerte expresión inducible de un gen de interés. El vector también contenía el marcador selectivo amdS *Aspergillus nidulans* que permitía el crecimiento de los transformantes en acetamida como una única fuente de nitrógeno. El vector de expresión también contenía regiones de telómero *T. reesei* que permitían el mantenimiento plásmido no cromosómico en una célula fúngica. En el plásmido pTTT-Dest de destino, las regiones de promotor y terminador de cbhl se separaron mediante el gen de resistencia cloranfenicol, Cm^R y el gen letal *E. coli*, ccdB, flanqueado por lugares de recombinación específicos basados en bacteriófago lambda attR1, attR2. Esta configuración permitía la selección directa de recombinantes que contenían el gen TrGA bajo el control de los elementos reguladores cbhl en la orientación adecuada mediante la reacción de recombinación GATEWAY® LR. El vector de expresión final pTTT-TrGA se muestra en la figura 4.

[0193] Los SEL se crearon utilizando el vector de entrada pDONR-TrGA (fig. 2) como plantilla y los cebadores enumerados en la tabla 2. Todos los cebadores utilizados en los experimentos de la mutagénesis contenían el triplete NNS (N = A,C,T,G y S=C o G) en la posición que se alinea con el codón de la secuencia TrGA diseñada para mutarse (SEQ ID NO:1), lo que permite una incorporación aleatoria de nucleótidos en la posición preseleccionada. La construcción de cada biblioteca SEL empezó con dos amplificaciones por PCR

independientes en el vector de entrada pDONR-TrGA: una utilizando el Gateway F (pDONR201 - FW) y un cebador de mutagénesis específico R (tabla 2) y la otra, el cebador Gateway R (pDONR201 - RV) y un cebador de mutagénesis específico F (tabla 2). La polimerasa de ADN PHUSION de alta fidelidad (Finnzymes OY, Espoo, Finlandia) se utilizó en una reacción de amplificación por PCR que incluía 0,2 μM cebadores. Las reacciones se llevaron a cabo durante 25 ciclos según el protocolo proporcionado por Finnzymes. 1 μl alícuotas de los fragmentos de PCR obtenidos se utilizaron como plantillas para una reacción por PCR de fusión posterior junto con los cebadores Gateway FW y Gateway RV (Invitrogen). La amplificación por PCR, después de 22 ciclos, produjo una población de fragmentos de ADN de TrGA lineal de longitud completa mutados de forma aleatoria en la posición de codón específica. Los fragmentos se flanquearon por los lugares de recombinación attL1, attL2 específicos de Gateway en ambos extremos. Los fragmentos de ADN se purificaron con un kit de limpieza CHARGESWITCH® PCR (Invitrogen, Carlsbad EEUU) y a continuación se recombinaron con 100 ng del vector de destino pTTT (fig. 3) utilizando la mezcla de enzimas LR CLONASE™ II según el protocolo proporcionado por Invitrogen. Los productos de recombinación que se generaron se transformaron en *E.coli* Max Efficiency DH5α, según describe el proveedor (Invitrogen). Las construcciones de expresión finales pTTT-TrGA con mutaciones en la posición deseada se eligieron mediante el recubrimiento de bacterias en placas de agar 2xYT (16 g/L Bacto Tryptone (Difco), 10 g/L Bacto Yeast Extract (Difco), 5 g/L NaCl, 16 g/L Bacto Agar (Difco) con 100 μg/ml ampicilina.

[0194] Se cultivaron 96 colonias únicas de cada biblioteca durante 24 hrs a 37 °C en una PMT con 200 μl de medio 2xYT con 100 μg/ml ampicilina. Los cultivos se utilizaron directamente para amplificar fragmentos por PCR incluidos en la región en la que se introducía una mutación específica. Los productos por PCR específica obtenidos se secuenciaron utilizando un analizador de secuencias AB13100 (Applied Biosystems). Cada biblioteca contenía de 15 a 19 variantes de TrGA diferentes en el vector de expresión final. Estas variantes se transformaron de forma individual en *T. reesei*, como se describe a continuación. Las bibliotecas se enumeran desde 1 a 182 en referencia al residuo de aminoácido específico en la secuencia de TrGA que se mutó de forma aleatoria.

Tabla 2 - Cebadores utilizados para generar los SEL de TrGA

D = Directo; I = Inverso		
Posición AA	D/I	Secuencia ADN 5' a 3'
pDONR2 01-	D	TCGCGTTAACGCTAGCATGGATCTC (SEQ ID NO:13)
pDONR2 01-	I	TCGCGTTAACGCTAGCATGGATCTC (SEQ ID NO:14)
4	D	CGTCACCAAGAGGTCTGTTGACNNS TTCATCAGCACCGAGACGCC (SEQ ID NO:15)
4	I	GTCAACAGACCTCTTGGTGACGTCG (SEQ ID NO:16)
5	D	ACCAAGAGGTCTGTTGACGACNNSATCAGCACCGAGACGCCTATTGC SEQ ID NO:17)
5	I	GTCGTCAACAGACCTCTTGGTGAC (SEQ ID NO:18)
10	D	TGACGACTTCATCAGCACCGAGNNSCCTATTGCACTG (SEQ ID NO:19)
10	I	CTCGGTGCTGATGAAGTCGTC (SEQ ID NO:20)
12	D	TCATCAGCACCGAGACGCCTNNSGCACTGAACAATCTTCTTTGCA (SEQ ID NO:21)
12	I	AGGCGTCTCGGTGCTGATGAAGTGG (SEQ ID NO:22)
14	D	CAGCACCGAGACGCCTATTGCANNSAACAATCTTCTT (SEQ ID NO:23)
14	I	TGCAATAGGCGTCTCGGTGCT (SEQ ID NO:24)
15	D	CACCGAGACGCCTATTGCACTGNNSAATCTTCTTTGC (SEQ ID NO:25)
1S	I	CAGTGCAATAGGCGTCTCGGT (SEQ ID NO:26)
23	D	CAATCTTCTTTGCAATGTTGGTNNSGATGGATGCCGT (SEQ ID NO:27)
23	I	ACCAACATTGCAAAGAAGATTG (SEQ ID NO:28)
24	D	TTCTTTGCAATGTTGGTCCTNNSGGATGCCGTGCATTCGGCACAT (SEQ ID NO:29)
24	I	AGGACCAACATTGCAAAGAAGATTG (SEQ ID NO:30)
29	D	GTCCTGATGGATGCCGTGCANNSGGCACATCAGCTGGTGCGGTGA (SEQ ID NO:31)

ES 2 518 916 T3

29	I	TGCACGGCATCCATCAGGACCAACA (SEQ ID NO:32)
42	D	TGCGGTGATTGCATCTCCAGCNNSATTGACCCGGAC (SEQ ID NO:33)
42	I	GCTGGGAGATGCAATCACCGCA (SEQ ID NO:34)
43	D	TGATTGCATCTCCCAGCACANNSGACCCGGACTACTATTACATGT (SEQ ID NO:35)
43	I	TGTGCTGGGAGATGCAATCACCGCA (SEQ ID NO:36)
44	D	TTGCATCTCCCAGCACAAATTNNSCCGGACTACTATTACATGTGGA (SEQ ID NO:37)
44	I	AATTGTGCTGGGAGATGCAATCACC (SEQ ID NO:38)
45	D	CATCTCCCAGCACAAATTGACNNSGACTACTATTACATGTGGACGC (SEQ ID NO:39)
45	I	GTCAATTGTGCTGGGAGATGCAATC (SEQ ID NO:40)
46	D	CTCCCAGCACAAATTGACCCGNNSTACTATTACATGTGGACGCGAGA (SEQ ID NO:41)
46	I	CGGGTCAATTGTGCTGGGAGATGCA (SEQ ID NO:42)
47	D	CCAGCACAAATTGACCCGGACNNSTATTACATGTGGACGCGAGATA (SEQ ID NO:43)
47	I	GTCCGGGTCAATTGTGCTGGGAGAT (SEQ ID NO:44)
49	D	CAATTGACCCGGACTACTATNNSATGTGGACGCGAGATAGCGCTC (SEQ ID NO:45)
49	I	ATAGTAGTCCGGGTCAATTGTGCTG (SEQ ID NO:46)
51	D	ACCCGGACTACTATTACATGNNSACGCGAGATAGCGCTCTTGTCT (SEQ ID NO:47)
51	I	CATGTAATAGTAGTCCGGGTCAATT (SEQ ID NO:48)
59	D	GACGCGAGATAGCGCTCTTGTCTNNSAAGAACCTCATC (SEQ ID NO:49)
59	I	GACAAGAGCGCTATCTCGCGT (SEQ ID NO:50)
60	D	GCGAGATAGCGCTCTTGTCTTCNNSAACCTCATCGAC (SEQ ID NO:51)
60	I	GAAGACAAGAGCGCTATCTCG (SEQ ID NO:52)
61	D	AGATAGCGCTCTTGTCTTCAAGNNSCTCATCGACCGC (SEQ ID NO:53)
61	I	CTTGAAGACAAGAGCGCTATC (SEQ ID NO:54)
65	D	TGTCTTCAAGAACCTCATCGACNNSTCACCGAAACG (SEQ ID NO:55)
65	I	GTCGATGAGGTTCTTGAAGAC (SEQ ID NO:56)
67	D	CAAGAACCTCATCGACCGCTTCNNSGAAACGTACGAT (SEQ ID NO:57)
67	I	GAAGCGGTCGATGAGGTTCTT (SEQ ID NO:58)
68	D	GAACCTCATCGACCGCTTACCNNNSACGTACGATGCG (SEQ ID NO:59)
68	I	GGTGAAGCGGTCGATGAGGTT (SEQ ID NO:60)
70	D	TCGACCGCTTACCGAAACGNNSGATGCGGGCCTGCAGCGCCGCA (SEQ ID NO:61)
70	I	CGTTTCGGTGAAGCGGTCGATGAGG (SEQ ID NO:62)
72	D	CGCTTACCGAAACGTACGATNNSGGCCTGCAGCGC (SEQ ID NO:63)
72	I	ATCGTACGTTTCGGTGAAGCGG (SEQ ID NO:64)
73	D	CTTACCGAAACGTACGATGCGNNSCTGCAGCGCCGC (SEQ ID NO:65)
73	I	GGCATCGTACGTTTCGGTGAA (SEQ ID NO:66)
75	D	AAACGTACGATGCGGGCCTGNNSCGCCGCATCGAGCAGTACATTA (SEQ ID NO:67)
75	I	AGGCCCGCATCGTACGTTTCGGTG (SEQ ID NO:68)
76	D	CGTACGATGCGGGCCTGCAGNNSCGCATCGAGCAGTACACTACTG (SEQ ID NO:69)

ES 2 518 916 T3

76	I	CTGCAGGCCCGCATCGTACGTTTCG (SEQ ID NO:70)
94	D	CTCTCCAGGGCCTCTTAACNNSTCGGGCTCCCTCGCGGACGGCT (SEQ ID NO:71)
94	I	GTTAGAGAGGCCCTGGAGAGTGACC (SEQ ID NO:72)
97	D	GGGCCTCTCTAACCCCTCGGGCNSCTCGCGGACGGC (SEQ ID NO:73)
97	I	GCCCCAGGGGTTAGAGAGGCC (SEQ ID NO:74)
98	D	CTCTCTAACCCCTCGGGCTCCNNSGCGGACGGCTCT (SEQ ID NO:75)
98	I	GGAGCCCGAGGGGTTAGAGAG (SEQ ID NO:76)
99	D	CTCTAACCCCTCGGGCTCCCTCNNSGACGGCTCTGGT (SEQ ID NO:77)
99	I	GAGGGAGCCCGAGGGGTTAGA (SEQ ID NO:78)
100	D	ACCCCTCGGGCTCCCTCGCGNNSGGCTCTGGTCTCGGCGAGCCCA (SEQ ID NO:79)
100	I	CGCGAGGGAGCCCGAGGGGTTAGAG (SEQ ID NO:80)
102	D	CTCGGGCTCCCTCGCGGACGGCNSGGTCTCGGCGAG (SEQ ID NO:81)
102	I	GCCGTCCGCGAGGGAGCCCA (SEQ ID NO:82)
110	D	TGGTCTCGGCGAGCCCAAGTTTNNSTTGACCCTGAAG (SEQ ID NO:83)
110	I	AAACTTGGGCTCGCCGAGACCA (SEQ ID NO:84)
111	D	TCTCGGCGAGCCCAAGTTTGAGNNSACCCTGAAGCCT (SEQ ID NO:85)
111	I	CTCAAATTGGGCTCGCCGAG (SEQ ID NO:86)
113	D	CGAGCCCAAGTTTGAGTTGACCNNSAAGCCTTTCACC (SEQ ID NO:87)
113	I	GGTCAACTCAAATTGGGCTC (SEQ ID NO:88)
114	D	GCAAGTTTGAGTTGACCCTGNNSCCTTTCACCGGCAACTGGGGTC (SEQ ID NO:89)
114	I	CAGGGTCAACTCAAATTGGGCTCG (SEQ ID NO:90)
116	D	TTGAGTTGACCCTGAAGCCTNNSACCGGCAACTGGGGTTCGACCGCA (SEQ ID NO:91)
116	I	AGGCTTCAGGGTCAACTCAAATTG (SEQ ID NO:92)
119	D	CCCTGAAGCCTTTCACCGGCNNSTGGGGTTCGACCGCAGCGGGATG (SEQ ID NO:93)
119	I	GCCGGTGAAAGGCTTCAGGGTCAAC (SEQ ID NO:94)
122	D	CTTTCACCGGCAACTGGGGTNNSCCGCAGCGGGATGGCCCAGCTC (SEQ ID NO:95)
122	I	ACCCAGTTGCCGGTGAAAGGCTTC (SEQ ID NO:96)
125	D	GCAACTGGGGTCGACCGCAGNNSGATGGCCCAGCTCTGCGAGCCA (SEQ ID NO:97)
125	I	CTGCGGTTCGACCCAGTTGCCGGTG (SEQ ID NO:98)
133	D	GGATGGCCAGCTCTGCGAGCCNNSGCTTGATTGGA (SEQ ID NO:99)
133	I	GGCTCGCAGAGCTGGGCCATCC (SEQ ID NO:100)
137	D	TGCGAGCCATTGCCCTGATTNNSTACTCAAAGTGGCTCATCAACA (SEQ ID NO:101)
137	I	AATCAAGGCAATGGCTCGCAGAGCT (SEQ ID NO: 102)
140	D	CATTGCCTTGATTGGATACTCANNSTGGCTCATCAAC (SEQ ID NO:103)
140	I	TGAGTATCCAATCAAGGCAATG (SEQ ID NO:104)
144	D	TGGATACTCAAAGTGGCTCATCNNSAACAATATCAG (SEQ ID NO:105)
144	I	GATGAGCCACTTTGAGTATCC (SEQ ID NO:106)
145	D	ATACTCAAAGTGGCTCATCAACNNSAATATCAGTCG (SEQ ID NO:107)
145	I	GTTGATGAGCCACTTTGAGTA (SEQ ID NO:108)

ES 2 518 916 T3

146	D	CAAAGTGGCTCATCAACAACNNSTATCAGTCGACTGTGTCCAACG (SEQ ID NO:109)
146	I	GTTGTTGATGAGCCACTTTGAGTAT (SEQ ID NO:110)
147	D	AAAGTGGCTCATCAACAACAACNNSCAGTCGACTGTG (SEQ ID NO:111)
147	I	GTTGTTGTTGATGAGCCACTT (SEQ ID NO:112)
148	D	GCTCATCAACAACAACATATNNSTCGACTGTGTCCAACGTCATCT (SEQ ID NO:113)
148	I	ATAGTTGTTGTTGATGAGCCACTTT (SEQ ID NO:114)
152	D	CAACAACATCAGTCGACTGTGNNSAACGTCATCTGG (SEQ ID NO:115)
152	I	CACAGTCGACTGATAGTTGTT (SEQ ID NO:116)
153	D	CAACTATCAGTCGACTGTGTCCNNSGTCATCTGGCCT (SEQ ID NO:117)
153	I	GGACACAGTCGACTGATAGTT (SEQ ID NO:118)
164	D	GCCTATTGTGCGCAACGACCTCNNSTATGTTGCCAGT (SEQ ID NO:119)
164	I	GAGGTGTTGCGCACAATAGG (SEQ ID NO:120)
169	D	ACCTCAACTATGTTGCCAGNNSTGGAACCAAACCGGCTTTGACC (SEQ ID NO:121)
169	I	CTGGGCAACATAGTTGAGGTCGTTG (SEQ ID NO:122)
172	D	ATGTTGCCAGTACTGGAACNNSACCGGCTTTGACCTCTGGGAAG (SEQ ID NO:123)
172	I	GTTCCAGTACTGGGCAACATAGTTG (SEQ ID NO:124)
175	D	AGTACTGGAACCAAACCGGCNNSGACCTCTGGGAAGAAGTCAATG (SEQ ID NO:125)
175	I	GCCGGTTTGGTTCCAGTACTGGGCA (SEQ ID NO:126)
178	D	ACCAAACCGGCTTTGACCTCNNSGAAGAAGTCAATGGGAGCTCAT (SEQ ID NO:127)
178	I	GAGGTCAAAGCCGGTTTGGTTCCAG (SEQ ID NO:128)
180	D	CCGGCTTTGACCTCTGGGAANNSGTCAATGGGAGCTCATTCTTTA (SEQ ID NO:129)
180	I	TTCCCAGAGGTCAAAGCCGGTTTGG (SEQ ID NO:130)
181	D	GCTTTGACCTCTGGGAAGAANNSAATGGGAGCTCATTCTTTACTG (SEQ ID NO:131)
181	I	TTCTTCCCAGAGGTCAAAGCCGGTT (SEQ ID NO:132)
182	D	CTTTGACCTCTGGGAAGAAGTCNNSGGGAGCTCATTC (SEQ ID NO:133)
182	I	GACTTCTTCCCAGAGGTCAAAG (SEQ ID NO:134)
204	D	TGTCGAGGGCGCCACTCTTGCTNNSACTCTTGCCAG (SEQ ID NO:135)
204	I	AGCAAGAGTGGCGCCCTCGAC (SEQ ID NO:136)
205	D	CGAGGGCGCCACTCTTGCTGCCNNSCTTGCCAGTCG (SEQ ID NO:137)
205	I	GCAGCAAGAGTGGCGCCCTC (SEQ ID NO:138)
208	D	CTCTTGCTGCCACTCTTGGCNNSTCGGAAGCGCTTATTCATCTG (SEQ ID NO:139)
208	I	GCCAAGAGTGGCAGCAAGAGTGGCG (SEQ ID NO:140)
211	D	CCACTCTTGCCAGTCGGGANNSGCTTATTCATCTGTTGCTCCCC (SEQ ID NO:141)
211	I	TCCCAGTGGCCAAGAGTGGCAGCA (SEQ ID NO:142)
214	D	TGGCCAGTCGGGAAGCGCTTATNNSTCTGTTGCTCCC (SEQ ID NO:143)
214	I	ATAAGCGCTTCCCGACTGGCC (SEQ ID NO:144)
216	D	GTCGGGAAGCGCTTATTCATCTNNSGCTCCCCAGGTT (SEQ ID NO:145)
216	I	AGATGAATAAGCGCTTCCCGA (SEQ ID NO:146)
219	D	CGCTTATTCATCTGTTGCTCCNNSGTTTTGTGCTTT (SEQ ID NO:147)

ES 2 518 916 T3

219	I	GGGAGCAACAGATGAATAAGC (SEQ ID NO:148)
228	D	TGTGCTTTCTCCAACGATTCCNNSGTGTCGTCTGGTGGATACGTCG (SEQ ID NO:149)
228	I	GAATCGTTGGAGAAAGCACAAAACCT (SEQ ID NO: 150)
229	D	GTGCTTTCTCCAACGATTCTGGNNSTCGTCTGGTGGGA (SEQ ID NO:151)
229	I	CCAGAATCGTTGGAGAAAGCA (SEQ ID NO:152)
230	D	CTTTCTCCAACGATTCTGGGTGNNSTCTGGTGGATACG (SEQ ID NO:153)
230	I	CACCCAGAATCGTTGGAGAAA (SEQ ID NO:154)
231	D	TCTCCAACGATTCTGGGTGTCGNNSGGTGGATACGTC (SEQ ID NO:155)
231	I	CGACACCCAGAATCGTTGGAGA (SEQ NO:156)
236	D	GGTGTCTGTGGTGGATACGTCNNSTCCAACATCAACAC (SEQ ID NO:157)
236	I	GACGTATCCACCAGACGACAC (SEQ ID NO:158)
239	D	TGGTGGATACGTCGACTCCAACNNSAACACCAACGAG (SEQ ID NO: 159)
239	I	GTTGGAGTCGACGTATCCACC (SEQ ID NO:160)
240	D	TGGATACGTCGACTCCAACATCNNSACCAACGAGGGCA (SEQ ID NO: 161)
240	I	GATGTTGGAGTCGACGTATCCA (SEQ ID NO:162)
241	D	ATACGTCGACTCCAACATCAACNNSAACGAGGGCAGGAC (SEQ ID NO:163)
241	I	GTTGATGTTGGAGTCGACGTA (SEQ ID NO: 64)
242	D	TCGACTCCAACATCAACACCNNSGAGGGCAGGACTGGCAAGGATG (SEQ ID NO:165)
242	I	GGTGTTGATGTTGGAGTCGACGTAT (SEQ ID NO:166)
243	D	ACTCCAACATCAACACCAACNNSGGCAGGACTGGCAAGGATGTCA (SEQ ID NO:167)
243	I	GTTGGTGTGATGTTGGAGTCGACG (SEQ ID NO:168)
244	D	CTCCAACATCAACACCAACGAGNNSAGGACTGGCAAG (SEQ ID NO:169)
244	I	CTCGTTGGTGTGATGTTGGAGT (SEQ ID NO:170)
245	D	ACATCAACACCAACGAGGGCNNSACTGGCAAGGATGTCAACTCCG (SEQ ID NO:171)
245	I	GCCCTCGTTGGTGTGATGTTGGAGT (SEQ ID NO:172)
263	D	TTCCATCCACACCTTCGATCCNNSCTTGGCTGTGAC (SEQ ID NO:173)
263	I	GGGATCGAAGGTGTGGATGGA (SEQ ID NO:174)
264	D	CATCCACACCTTCGATCCCAACNNSGGCTGTGACGCA (SEQ ID NO:175)
264	I	GTTGGGATCGAAGGTGTGGAT (SEQ ID NO:176)
265	D	CCACACCTTCGATCCCAACCTTNNSTGTGACGCAGGC (SEQ ID NO:177)
265	I	AAGGTTGGGATCGAAGGTGTG (SEQ ID NO:178)
268	D	CGATCCCAACCTTGGCTGTGACNNSGGCACCTTCCAGC (SEQ ID NO:179)
268	I	GTCACAGCCAAGGTTGGGATC (SEQ ID NO:180)
269	D	TCCCAACCTTGGCTGTGACGCANNSACCTTCCAGCCA (SEQ ID NO:181)
269	I	TGCGTCACAGCCAAGGTTGGG (SEQ ID NO:182)
276	D	AGGCACCTTCCAGCCATGCAGTNNSAAAGCGCTCTCC (SEQ ID NO:183)
276	I	ACTGCATGGCTGGAAGGTGCC (SEQ ID NO:184)
284	D	CAAAGCGCTCTCCAACCTCAAGNNSGTTGTGACTCCT (SEQ ID NO:185)
284	I	CTTGAGGTTGGAGAGCGCTTT (SEQ ID NO:186)
291	D	GGTTGTTGTGACTCCTTCCGCNNSATCTACGGCGTG (SEQ ID NO:187)
291	I	GCGGAAGGAGTCGACAACAAC (SEQ ID NO:188)
292	D	TTGTGACTCCTTCCGCTCCNNSACGGCGTGAACAAGGGCATT (SEQ ID NO:189)

ES 2 518 916 T3

		NO:189)
292	I	GGAGCGGAAGGAGTCGACAACAACC (SEQ ID NO:190)
294	D	ACTCCTTCCGCTCCATCTACNNSGTGAACAAGGGCATTCTGCGG (SEQ ID NO:191)
294	I	GTAGATGGAGCGGAAGGAGTCGACA (SEQ ID NO:192)
297	D	GCTCCATCTACGGCGTGAACNNSGGCATTCTGCCGGTCTGCGG (SEQ ID NO:193)
297	I	GTTACACCCGTAGATGGAGCGGAAG (SEQ ID NO:194)
300	D	CTACGGCGTGAACAAGGGCATTNNSGCCGGTCTGCGG (SEQ ID NO:195)
300	I	AATGCCCTTGTTACGCGTA (SEQ ID NO:196)
301	D	CGGCGTGAACAAGGGCATTCTNNSGGTCTGCGGTC (SEQ ID NO:197)
301	I	AGGAATGCCCTTGTTACGCC (SEQ ID NO:198)
303	D	GAACAAGGGCATTCTGCCGGTNSGCCGTCGCCATT (SEQ ID NO:199)
303	I	ACCGGCAGGAATGCCCTTGTT (SEQ ID NO:200)
309	D	GTGCTGCGGTCGCCATTGGCANNSTATGCAGAGGATGTGTACTACA (SEQ ID NO:201)
309	I	GCCAATGGCGACGGCAGCACCGCA (SEQ ID NO:202)
310	D	CTGCCGTCGCCATTGGCCGGNNSGCAGAGGATGTGTACTACAACG (SEQ ID NO:203)
310	I	CCGGCCAATGGCGACGGCAGCACCG (SEQ ID NO:204)
311	D	TGCCGTCGCCATTGGCCGGTATNNSGAGGATGTGTAC (SEQ ID NO:205)
311	I	ATACCGGCCAATGGCGACGGC (SEQ ID NO:206)
313	D	CCATTGGCCGGTATGCAGAGNNSGTGTACTACAACGGCAACCCTT (SEQ ID NO:207)
313	D	CCATTGGCCGGTATGCAGAGNNSGTGTACTACAACGGCAACCCTT (SEQ ID NO:208)
313	I	CTCTGCATACCGGCCAATGGCGACG (SEQ ID NO:209)
313	I	CTCTGCATACCGGCCAATGGCGACG (SEQ ID NO:210)
314	D	TTGGCCGGTATGCAGAGGATNNSACTACAACGGCAACCCTTGGT (SEQ ID NO:211)
314	I	ATCCTCTGCATACCGGCCAATGGCG (SEQ ID NO:212)
315	D	GCCGGTATGCAGAGGATGTGNNSACAACGGCAACCCTTGGTATC (SEQ ID NO:213)
315	I	CACATCCTCTGCATACCGGCCAATG (SEQ ID NO:214)
316	D	GGTATGCAGAGGATGTGTACNNSAACGGCAACCCTTGGTATCTTG (SEQ ID NO:215)
316	I	GTACACATCCTCTGCATACCGGCCAAT (SEQ ID NO:216)
317	D	ATGCAGAGGATGTGTACTACNNSGGCAACCCTTGGTATCTTGCTA (SEQ ID NO:217)
317	D	ATGCAGAGGATGTGTACTACNNSGGCAACCCTTGGTATCTTGCTA (SEQ ID NO:218)
317	I	GTAGTACACATCCTCTGCATACCGGC (SEQ ID NO:219)
317	I	GTAGTACACATCCTCTGCATACCGGC (SEQ ID NO:220)
321	D	TGTAACAACGGCAACCCTNNSACTCTTGCTACATTTGCTGCTG (SEQ ID NO:221)
321	D	TGTAACAACGGCAACCCTNNSACTCTTGCTACATTTGCTGCTG (SEQ ID NO:222)
321	I	AGGGTTGCCGTTGTAGTACACATCC (SEQ ID NO:223)
321	I	AGGGTTGCCGTTGTAGTACACATCC (SEQ ID NO:224)

ES 2 518 916 T3

338	D	GCAGCTGTACGATGCCATCTACNNSTGGAAGAAGACG (SEQ ID NO:225)
338	I	GTAGATGGCATCGTACAGCTG (SEQ ID NO:226)
340	D	ACGATGCCATCTACGTCTGNNNSAAGACGGGCTCCATCACGGTGA (SEQ ID NO:227)
340	I	CCAGACGTAGATGGCATCGTACAGC (SEQ ID NO:228)
341	D	ATGCCATCTACGTCTGGAAGNNSACGGGCTCCATCACGGTGACCG (SEQ ID NO:229)
341	I	CTTCCAGACGTAGATGGCATCGTACAGC (SEQ ID NO:230)
342	D	ATGCCATCTACGTCTGGAAGAAGNNSGGCTCCATCACG (SEQ ID NO:231)
342	I	CTTCTTCCAGACGTAGATGGC (SEQ ID NO:232)
344	D	CTACGTCTGGAAGAAGACGGGCNNSATCACGGTGACC (SEQ ID NO:233)
344	I	GCCCGTCTTCTTCCAGACGTAG (SEQ ID NO:234)
346	D	CTGGAAGAAGACGGGCTCCATCNNSGTGACCGCCACCTC (SEQ ID NO:235)
346	I	GATGGAGCCCGTCTTCTTCCA (SEQ ID NO:236)
349	D	GACGGGCTCCATCACGGTGACNNSACCTCCCTGGCC (SEQ ID NO:237)
349	I	GGTCACCGTGATGGAGCCCGT (SEQ ID NO:238)
350	D	GCTCCATCACGGTGACCGCCNNSCTCCCTGGCCTTCTTCCAGGAGC (SEQ ID NO:239)
350	I	GGCGGTCACCGTGATGGAGCCCGTC (SEQ ID NO:240)
356	D	CCACCTCCCTGGCCTTCTTNNNSGAGCTTGTTCTGGCGTGACGG (SEQ ID NO:241)
356	I	GAAGAAGGCCAGGGAGGTGGCGGTC (SEQ ID NO:242)
359	D	CCTGGCCTTCTTCCAGGAGCTTNNNSCTGGCGTGACG (SEQ ID NO:243)
359	I	AAGCTCCTGGAAGAAGGCCAG (SEQ ID NO:244)
361	D	CTTCTTCCAGGAGCTTGTTCTTNNNSGTGACGGCCGGG (SEQ ID NO:245)
361	I	AGGAACAAGCTCCTGGAAGAA (SEQ ID NO:246)
363	D	AGGAGCTTGTTCTTGGCGTGNNSGCCGGGACCTACTCCAGCAGCT (SEQ ID NO:247)
363	I	CACGCCAGGAACAAGCTCCTGGAAG (SEQ ID NO:248)
364	D	GGAGCTTGTTCTTGGCGTGACGNNSGGGACCTACTCC (SEQ ID NO:249)
364	I	CGTCACGCCAGGAACAAGCTC (SEQ ID NO:250)
368	D	GCGTGACGGCCGGGACCTACNNSAGCAGCTCTTCGACCTTTACCA (SEQ ID NO:251)
368	I	GTAGGTCCCGGCCGTACGCCAGGA (SEQ ID NO:252)
369	D	TGACGGCCGGGACCTACTCCNNSAGCTCTTCGACCTTTACCAACA (SEQ ID NO:253)
369	I	GGAGTAGGTCCCGGCCGTACGCCA (SEQ ID NO:254)
375	D	CTCCAGCAGCTCTTCGACCTTNNNSAACATCATCAACG (SEQ ID NO:255)
375	I	AAAGGTCGAAGAGCTGCTGGA (SEQ ID NO:256)
376	D	GCAGCTCTTCGACCTTTACNNSATCATCAACGCCGTCTCGACAT (SEQ ID NO:257)
376	I	GGTAAAGGTCGAAGAGCTGCTGGAG (SEQ ID NO:258)
379	D	TTCGACCTTTACCAACATCATCNNSGCCGTCTCGACA (SEQ ID NO:259)
379	I	GATGATGTTGGTAAAGGTCGA (SEQ ID NO:260)
382	D	TACCAACATCATCAACGCCGTNNNSACATACGCCGAT (SEQ ID NO:261)
382	I	GACGGCGTTGATGATGTTGGT (SEQ ID NO:262)
390	D	GACATACGCCGATGGCTTCTTNNNSGAGGCTGCCAAG (SEQ ID NO:263)

ES 2 518 916 T3

390	I	GAGGAAGCCATCGGCGTATGT (SEQ ID NO:264)
391	D	ATACGCCGATGGCTTCCTCAGCNNSGCTGCCAAGTAC (SEQ ID NO:265)
391	I	GCTGAGGAAGCCATCGGCGTA (SEQ ID NO:266)
393	D	CGATGGCTTCCTCAGCGAGGCTNNSAAGTACGTCCCC (SEQ ID NO:267)
393	I	AGCCTCGCTGAGGAAGCCATC (SEQ ID NO:268)
394	D	TGGCTTCCTCAGCGAGGCTGCCNNSTACGTCCCCGCC (SEQ ID NO:269)
394	I	GGCAGCCTCGCTGAGGAAGCC (SEQ ID NO:270)
395	D	TCCTCAGCGAGGCTGCCAAGNNSGTCCCCGCCGACGGTTCGCTGG (SEQ ID NO:271)
395	I	CTTGGCAGCCTCGCTGAGGAAGCCA (SEQ ID NO:272)
398	D	AGGCTGCCAAGTACGTCCCCNNSGACGGTTCGCTGGCCGAGCAGTT(SEQ ID NO:273)
398	I	GGGGACGTACTTGGCAGCCTCGCTG (SEQ ID NO:274)
401	D	AGTACGTCCCCGCCGACGGTNNNSCTGGCCGAGCAGTTTGACCGCA (SEQ ID NO:275)
401	I	ACCGTCGGCGGGGACGTACTTGGCAG (SEQ ID NO:276)
408	D	CGCTGGCCGAGCAGTTTGACNNSAACAGCGGCACTCCGCTGTCTG (SEQ ID NO:277)
408	I	GTCAACTGCTCGGCCAGCGAACCG (SEQ ID NO:278)
409	D	TGGCCGAGCAGTTTGACCGNNSAGCGGCACTCCGCTGTCTGCGC (SEQ ID NO:279)
409	I	GCGGTCAAACCTGCTCGGCCAGCGAA (SEQ ID NO:280)
410	D	GGCCGAGCAGTTTGACCGCAACNNSGGCACTCCGCTG (SEQ ID NO:281)
410	I	GTTGCGGTCAAACCTGCTCGGC (SEQ ID NO:282)
412	D	AGTTTGACCGCAACAGCGGCNNSCCGCTGTCTGCGCTTACCTGA (SEQ ID NO:283)
412	I	GCCGCTGTTGCGGTCAAACCTGCTCG (SEQ NO:284)
415	D	GCAACAGCGGCACTCCGCTGNNSGCGTTACCTGACGTGGTCGT (SEQ ID NO:285)
415	I	CAGCGGAGTGCCGCTGTTGCGGTCA (SEQ ID NO:286)
417	D	CAGCGGCACTCCGCTGTCTGCGNNSCACCTGACGTGGT (SEQ ID NO:287)
417	I	CGCAGACAGCGGAGTGCCGCT (SEQ ID NO:288)
418	D	GCACTCCGCTGTCTGCGCTNNSCTGACGTGGTTCGTACGCCTCGT (SEQ ID NO:289)
418	I	AAGCGCAGACAGCGGAGTGCCGCTG (SEQ ID NO:290)
421	D	TGTCTGCGCTTACCTGACGNNSCTGACGCCTCGTTCTTGACAG (SEQ ID NO:291)
421	I	CGTCAGGTGAAGCGCAGACAGCGGA (SEQ ID NO:292)
430	D	GTACGCCTCGTTCTTGACAGCCNNSGCCGTCGGGCT (SEQ ID NO:293)
430	I	GGCTGTCAAGAACGAGGCGTA (SEQ ID NO:294)
431	D	CGCCTCGTTCTTGACAGCCACGNNSCGTCGGGCTGGC (SEQ ID NO:295)
431	I	CGTGGCTGTCAAGAACGAGGC (SEQ ID NO:296)
433	D	TCTTGACAGCCACGGCCCGTNNNSGCTGGCATCGTGCCCCCTCGT (SEQ ID NO:297)
433	I	ACGGGCCGTGGCTGTCAAGAACGAG (SEQ ID NO:298)
436	D	CCACGGCCGTCGGGCTGGCNNSGTGCCCCCTCGTGGGCCAACA (SEQ ID NO:299)
436	I	GCCAGCCCGACGGCCGTGGCTGTC (SEQ ID NO:300)

442	D	TGGCATCGTGCCCCCTCGTGGNNSAACAGCAGCGCT (SEQ ID NO:301)
442	I	CCACGAGGGGGGCACGATGCC (SEQ ID NO:302)
443	D	CATCGTGCCCCCTCGTGGGCCNNSAGCAGCGCTAGC (SEQ ID NO:303)
443	I	GCCCACGAGGGGGGCACGAT (SEQ ID NO:304)
444	D	CGTGCCCCCTCGTGGGCCAACNNSAGCGCTAGCACG (SEQ ID NO:305)
444	I	GTTGGCCCACGAGGGGGGCAC (SEQ ID NO:306)
448	D	GTGGCCAACAGCAGCGCTAGCNNSATCCCCTCGACG (SEQ ID NO:307)

Ejemplo 2: Transformación de SEL de TrGA en *Trichoderma reesei*

[0195] Los SEL se transformaron en *T. reesei* utilizando el método protoplasto PEG. Los clones *E. coli* de los SEL confirmados por análisis de secuencia se cultivaron durante la noche a 37 °C en placas de microtitulación de pocillos hondos (Greiner Art. N° 780271) con 1,200 µl del medio 2xYT con ampicilina (100 µg/ml) y kanamicina (50 µg/ml). Los ADN plasmídicos se aislaron de los cultivos utilizando el kit CHEMAGIC® Plasmid Mini (Chemagen - Biopolymer Technologie AG, Baesweiler, Alemania) y se transformaron de forma individual en una cepa huésped *T. reesei* derivada de RL-P37 con cuatro deleciones de gen (Δ cbh1, Δ cbh2, Δ egl1, Δ egl2, es decir, "con deleción cuádruple"; véase la patente estadounidense 5.847.276, WO 92/06184 y WO 05/001036) utilizando el método PEG-Protoplasto (Penttilä *et al.* (1987) *Gene* 61:155-164) con las siguientes modificaciones.

1) Para la preparación de protoplasto, se cultivaron esporas durante 16-24 horas a 24 °C en un medio mínimo (MM) de *Trichoderma*. (20 g/L glucosa, 15 g/L KH_2PO_4 , pH 4,5, 5g/L $(\text{NH}_4)_2\text{SO}_4$, 0,6 g/L $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$, 0,6 g/L $\text{CaCl}_2 \cdot 2\text{H}_2\text{O}$, 1 ml de 1000X solución de oligoelementos de *T. reesei* {5 g/L $\text{FeSO}_4 \cdot 7\text{H}_2\text{O}$, 1,4 g/L $\text{ZnSO}_4 \cdot 7\text{H}_2\text{O}$, 1,6 g/L $\text{MnSO}_4 \cdot \text{xH}_2\text{O}$, 3,7 g/L $\text{CoCl}_2 \cdot 6\text{H}_2\text{O}$ }) con agitación a 150 rpm. Las esporas germinantes se recolectaron mediante centrifugación y se trataron con 15 mg/ml de solución β -D-glucanasa-G (Interspex-Art.N° 0439-1) para lisar las paredes de las células fúngicas. Se llevó a cabo una preparación adicional de protoplastos mediante un método convencional, como lo describe Penttilä *et al.* (1987 *supra*). 2) El método de transformación se redujo 10 veces. En general, las mezclas de transformación que contenían hasta 600 ng de ADN y $1\text{-}5 \times 10^5$ protoplastos en un volumen total de 25 µl se trataron con 200 ml de 25 % solución PEG, diluido con 2 volúmenes de 1,2 M solución sorbitol, mezclado con 3 % MM agarosa superior selectiva con acetamida (el mismo medio mínimo que se ha mencionado anteriormente pero $(\text{NH}_4)_2\text{SO}_4$ se sustituyó con 20 mM acetamida) y se vertió en 2 % agarosa selectiva con acetamida bien en placas de microtitulación de 24 pocillos o en una bandeja Q de 20x20 cm dividida en 48 pocillos. Las placas se incubaron a 28 °C durante un periodo de 5 a 8 días. Las esporas de la población total de transformantes regeneradas en cada pocillo individual se recolectaron de las placas utilizando una solución de 0,85 % NaCl, 0,015 % Tween 80. Las suspensiones de esporas se utilizaron para inocular fermentaciones en PMT de 96 pocillos. En el caso de las PMT de 24 pocillos, se introdujo una etapa adicional de recubrimiento sobre una PMT de 24 pocillos nueva con MM acetamida selectiva con el fin de enriquecer el número de esporas.

Ejemplo 3: Fermentación de transformantes de *T. reesei* que expresan variantes de TrGA en un formato PMT

[0196] Los transformantes se fermentaron y los sobrenadantes con las proteínas de TrGA variante expresadas se analizaron para la detección de diferentes propiedades. En resumen, las placas de filtrado de 96 pocillos (Coming Art. N° 3505) con 200 µl de medio LD-GSM (5,0 g/L $(\text{NH}_4)_2\text{SO}_4$, 33 g/L 1,4-Piperazinabis(ácido propanosulfónico), pH 5,5, 9,0 g/L Casamino ácidos, 1,0 g/L KH_2PO_4 , 1,0 g/L $\text{CaCl}_2 \cdot 2\text{H}_2\text{O}$, 1,0 g/L $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$, 2,5 ml/L de 1000X oligoelementos *T. reesei*, 20 g/L glucosa, 10 g/L soforosa) se inocularon por cuadruplicado con suspensiones de esporas de transformantes *T. reesei* que expresan variantes de TrGA (más de 10^4 esporas por pocillo). Las placas se incubaron a 28 °C con agitación a 230 rpm y 80 % de humedad durante 6 días. Los sobrenadantes de cultivo se recolectaron mediante filtración al vacío. Los sobrenadantes se utilizaron en diferentes análisis para la detección de variantes con propiedades mejoradas.

Ejemplo 4: Preparación de las muestras de caldo integrales a partir de transformantes que producen GA

[0197] Los transformantes que producen TrGA se precultivaron inicialmente en matraces de agitación de 250 ml con 30 ml de medio Proflo. El medio Proflo contenía: 30 g/L α -lactosa, 6,5 g/L $(\text{NH}_4)_2\text{SO}_4$, 2 g/L KH_2PO_4 , 0,3 g/L $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$, 0,2 g/L $\text{CaCl}_2 \cdot 2\text{H}_2\text{O}$, 1 ml/L 1000X solución salina de oligoelementos como se ha mencionado anteriormente, 2 ml/L 10% Tween 80, 22,5 g/L harina de semilla de algodón ProFlo (Traders protein, Memphis, TN), 0,72 g/L CaCO_3 . Tras dos días de cultivo a 28 °C y 140 rpm, 10 % del cultivo Proflo se transfirió a un matraz de agitación de 250 ml con 30 ml de Medio definido de lactosa. La composición del medio definido de lactosa era como se muestra a continuación: 5 g/L $(\text{NH}_4)_2\text{SO}_4$, 33 g/L tampón 1,4-Piperazinabis (ácido propanosulfónico), pH 5,5, 9 g/L casamino ácidos, 4,5 g/L KH_2PO_4 , 1,0 g/L $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$, 5 ml/L Mazu DF60-P antiespumante (Mazur Chemicals, IL), 1 ml/L of 1000X solución de oligoelementos. 40 ml/L de 40% (p/v) solución lactosa se añadió al medio tras la esterilización. Los matraces de agitación con el medio definido de lactosa se incubaron a 28 °C,

140 rpm durante 4 - 5 días.

[0198] Se extrajo el micelio de las muestras de cultivo mediante centrifugación y se analizó el sobrenadante para detectar el contenido de proteína total (BCA Protein Assay Kit, Pierce Cat. Nº 23225) y la actividad GA, tal y como se describe anteriormente en la sección experimental.

5 [0199] El perfil proteínico de las muestras de caldo integrales se determinó mediante electroforesis SDS-PAGE. Las muestras del sobrenadante de cultivo se mezclaron con el mismo volumen de 2X tampón de carga de muestra con agente reductor y se separaron en Gel NUPAGE® Novex 10% Bis-Tris con tampón de migración MES SDS (Invitrogen, Carlsbad, CA, EEUU). Las bandas de polipéptidos se visualizaron en el gel SDS con SIMPLYBLUE SafeStain (Invitrogen, Carlsbad, CA, EEUU).

10 **Ejemplo 5: Variantes con estabilidad térmica mejorada**

[0200] La molécula TrGA precursora presentaba una actividad residual entre 15 y 44 % (variación diaria) según las condiciones descritas. El índice de rendimiento se calculó según la termoestabilidad de TrGA del mismo lote. Los índices de rendimiento son los cocientes $PI = (\text{actividad residual variante})/(\text{actividad residual TrGA})$. Un índice de rendimiento > 1 indicaba una estabilidad mejorada. Las variantes que presentaban un índice de rendimiento (PI) de estabilidad térmica de más de 1,0 se muestran en la siguiente tabla 3.

15

Tabla 3 - Detección estabilidad térmica

Variante	Índice Rendimiento de Estabilidad Térmica		
T010F	1,11	K114C	1,19
T010G	1,13	K114D	1,17
T010M	1,12	K114E	1,16
T010Q	1,06	K114L	1,10
T010R	1,06	K114M	1,21
T010S	1,24	K114Q	1,25
T042V	1,31	I133V	1,21
F059A	1,03	K140D	1,09
F059G	1,07	K140Q	1,06
F059L	1,06	K140S	1,05
F059M	1,12	K140W	1,04
F059Q	1,10	N144A	1,11
F059V	1,03	N144F	1,06
N061V	1,06	S152C	1,04
E068C	1,23	S152G	1,09
E068F	1,14	S152I	1,09
E068G	1,15	S152N	1,12
E068I	1,19	N153A	1,28
E068K	1,01	N153D	1,06
E068M	1,29	N153E	1,29
E068N	1,18	N153F	1,16
E068Q	1,15	N153H	1,01
E068W	1,09	N153L	1,06
A072E	1,07	N153M	1,27
A072Q	1,02	N153S	1,31
G073F	1,44	N153V	1,34
G073M	1,01	N153W	1,19
G073N	1,10	N182R	1,02
G073W	1,36	A204D	1,02
S097E	1,08	A204M	1,08
S097G	1,03	T205C	1,02
S097T	1,17	T205D	1,06
S097V	1,03	T205N	1,09
L098C	1,07	T205P	1,17
A099C	1,07	T205S	1,04
A099F	1,02	T205V	1,06
A099I	1,14	T205Y	1,07
A099L	1,09	S214C	1,02
A099Q	1,07	S214E	1,03
A099R	1,02	S214N	1,07
A099S	1,11	S214P	1,04
A099T	1,02	S214Q	1,13
S102A	1,02	S214T	1,06
		S214V	1,13
		S214W	1,03

ES 2 518 916 T3

S214Y	1,04
V216I	1,13
W228A	1,01
W228F	1,12
W228G	1,06
W228H	1,05
W228I	1,06
W228L	1,14
W228M	1,04
W228Q	1,15
W228S	1,15
W228T	1,06
W228V	1,21
W228Y	1,10
V229A	1,16
V229D	1,18
V229E	1,16
V229F	1,17
V229G	1,18
V229H	1,13
V229I	1,21
V229L	1,20
V229M	1,11
V229N	1,07
V229P	1,13
V229Q	1,10
V229R	1,07
V229S	1,12
V229T	1,11
V229W	1,07
V229Y	1,08
S230C	1,09
S230D	1,08
S230E	1,11
S230F	1,08
S230G	1,09
S230H	1,03
S230K	1,02
S230L	1,02
S230M	1,08
S230N	1,16
S230P	1,12
S230Q	1,20
S230R	1,11
S230T	1,07
S230V	1,11
S230Y	1,05
S231A	1,06
S231C	1,07
S231D	1,18
S231E	1,10
S231F	1,14
S231K	1,09
S231L	1,15

S231M	1,09
S231N	1,13
S231Q	1,18
S231R	1,13
S231T	1,16
S231V	1,21
S231W	1,06
S231Y	1,10
D236A	1,10
D236C	1,16
D236E	1,06
D236F	1,11
D236G	1,07
D236H	1,16
D236I	1,14
D236K	1,13
D236L	1,08
D236M	1,15
D236N	1,15
D236P	1,06
D236R	1,29
D236T	1,16
D236V	1,18
D236Y	1,07
T241V	1,05
N242F	1,03
N242H	1,01
N263A	1,18
N263C	1,14
N263L	1,02
N263M	1,10
N263R	1,04
N263T	1,07
L264A	1,03
L264D	1,20
L264I	1,14
L264K	1,24
L264M	1,15
L264P	1,02
L264R	1,02
L264V	1,04
L264Y	1,15
G265D	1,08
G265E	1,03
G265F	1,08
G265H	1,09
G265I	1,09
G265K	1,05
G265L	1,07
G265N	1,15
G265P	1,07
G265Q	1,15
G265R	1,12
G265S	1,13

ES 2 518 916 T3

G265T	1,17
G265V	1,15
G265W	1,11
G265Y	1,10
A268C	1,16
A268D	1,20
A268E	1,10
A268F	1,07
A268G	1,11
A268I	1,12
A268K	1,11
A268L	1,11
A268M	1,18
A268N	1,10
A268P	1,13
A268R	1,15
A268S	1,09
A268T	1,14
A268W	1,05
G269K	1,01
G269N	1,08
G269P	1,08
G269Q	1,02
G269R	1,06
D276E	1,03
D276Q	1,03
D276S	1,02
D276V	1,17
V284C	1,09
V284E	1,14
V284G	1,13
V284I	1,10
V284Q	1,08
V284R	1,09
V284S	1,06
V284T	1,08
V284W	1,08
V284Y	1,12
S291A	1,21
S291D	1,02
S291E	1,02
S291F	1,24
S291G	1,05
S291H	1,25
S291I	1,10
S291K	1,13
S291M	1,21
S291N	1,08
S291P	1,11
S291Q	1,08
S291R	1,07
S291T	1,21
S291V	1,06
S291W	1,15

P300A	1,10
P300N	1,06
P300Q	1,05
P300R	1,02
P300T	1,07
P300V	1,02
P300W	1,11
P300Y	1,04
A301A	1,13
A301K	1,11
A301L	1,07
A301P	1,22
A301R	1,21
A301S	1,12
A301T	1,14
A301V	1,07
A301W	1,14
A301Y	1,14
A303E	1,02
A303I	1,02
A303L	1,03
A303Q	1,01
A311C	1,05
A311E	1,06
A311G	1,11
A311H	1,09
A311I	1,04
A311K	1,13
A311L	1,07
A311R	1,07
A311S	1,01
A311Y	1,02
V338H	1,12
V338I	1,21
V338L	1,17
V338M	1,13
V338N	1,23
V338P	1,18
V338Q	1,20
V338S	1,19
V338Y	1,10
T342C	1,06
T342I	1,04
T342L	1,17
T342P	1,08
S344A	1,05
S344C	1,16
S344D	1,02
S344F	1,10
S344K	1,12
S344M	1,26
S344N	1,16
S344P	1,20
S344Q	1,22

ES 2 518 916 T3

S344R	1,22
S344T	1,19
S344V	1,20
S344W	1,03
T346D	1,03
T346L	1,02
T346M	1,07
T346N	1,15
T346P	1,13
T346Q	1,12
T346S	1,13
T346V	1,09
T346W	1,06
T346Y	1,07
A349D	1,15
A349E	1,13
A349F	1,09
A349G	1,12
A349H	1,11
A349K	1,14
A349L	1,16
A349M	1,07
A349N	1,09
A349P	1,03
A349Q	1,09
A349R	1,04
A349T	1,09
A349V	1,08
A349W	1,04
A349Y	1,04
V359Q	1,02
V359R	1,13
V359Y	1,01
G361A	1,13
G361C	1,16
G361D	1,35
G361E	1,24
G361F	1,20
G361H	1,11
G361I	1,20
G361K	1,16
G361L	1,22
G361M	1,28
G361P	1,27
G361R	1,19
G361S	1,22
G361T	1,16
G361V	1,15
G361W	1,22
G361Y	1,26
A364C	1,18
A364D	1,25
A364E	1,27
A364F	1,33

A364G	1,25
A364K	1,27
A364L	1,31
A364M	1,21
A364Q	1,19
A364R	1,28
A364S	1,23
A364T	1,23
A364V	1,23
A364W	1,23
T375A	1,17
T375E	1,03
T375F	1,06
T375G	1,05
T375H	1,13
T375K	1,10
T375L	1,03
T375M	1,17
T375N	1,20
T375P	1,15
T375V	1,16
T375W	1,11
T375Y	1,17
N379A	1,11
N379C	1,03
N379D	1,06
N379F	1,07
N379G	1,09
N379H	1,07
N379I	1,04
N379K	1,04
N379L	1,04
N379M	1,08
N379P	1,17
N379Q	1,11
N379R	1,07
N379T	1,09
N379V	1,13
N379W	1,09
N379Y	1,06
S382A	1,06
S382D	1,04
S382E	1,05
S382G	1,06
S382I	1,07
S382K	1,03
S382N	1,07
S382P	1,10
S382R	1,05
S382T	1,02
S382Y	1,04
S390M	1,06
S390Q	1,02
S390R	1,02

ES 2 518 916 T3

E391L	1,04
E391R	1,02
E391S	1,02
E391W	1,04
E391Y	1,02
A393D	1,11
A393E	1,03
A393F	1,09
A393G	1,12
A393H	1,05
A393I	1,06
A393K	1,05
A393L	1,17
A393M	1,07
A393N	1,11
A393Q	1,02
A393R	1,03
A393S	1,13
A393T	1,08
A393V	1,09
A393W	1,10
A393Y	1,12
K394A	1,08
K394C	1,06
K394E	1,07
K394F	1,09
K394G	1,05
K394H	1,08
K394L	1,08
K394M	1,06
K394Q	1,07
K394R	1,06
K394T	1,03
K394V	1,02
S410E	1,05
S410H	1,11
S410I	1,06
S410K	1,11
S410L	1,05
S410M	1,02
S410N	1,10
S410Q	1,15
S410R	1,18
S410T	1,04
S410V	1,13
L417I	1,04
L417K	1,20
L417M	1,05
L417Q	1,04
L417R	1,20
L417V	1,07
L417Y	1,01
T430A	1,05
T430E	1,02

T430F	1,06
T430H	1,10
T430I	1,04
T430K	1,08
T430M	1,17
T430N	1,13
T430Q	1,05
T430R	1,13
T430S	1,17
T430V	1,05
A431I	1,03
A431N	1,03
A431P	1,08
A431R	1,08
A431V	1,03
R433A	1,13
R433C	1,24
R433E	1,22
R433F	1,17
R433G	1,23
R433K	1,12
R433L	1,23
R433M	1,10
R433N	1,28
R433S	1,23
R433V	1,28
R433W	1,16
R433Y	1,18
I436E	1,07
I436F	1,02
I436G	1,09
I436H	1,20
I436K	1,14
I436P	1,15
I436R	1,16
I436S	1,17
I436T	1,18
I436V	1,12
I436Y	1,05
A442N	1,06
A442R	1,04
A442T	1,09
S444E	1,04
S444K	1,07
S444M	1,05
S444Q	1,04
T448A	1,02
T448E	1,09
T448F	1,12
T448I	1,12
T448L	1,09
T448M	1,11
T448Q	1,17
T448R	1,09

T448S	1,07
T448V	1,13
T448W	1,01
T448Y	1,08
S451E	1,04

S451H	1,18
S451K	1,08
S451L	1,01
S451Q	1,06
S451T	1,02

La tabla 3 incluye aquellas variantes que, cuando se analizaron, mostraron un índice de rendimiento aumentado sobre la glucoamilasa precursora. Estas incluían los siguientes lugares: 10, 42, 59, 60, 61, 68, 72, 73, 97, 98, 99, 102, 114, 133, 140, 144, 147, 152, 153, 164, 182, 204, 205, 214, 216, 228, 229, 230, 231, 236, 241, 242, 263, 264, 265, 268, 269, 276, 284, 291, 300, 301, 303, 311, 338, 342, 344, 346, 349, 359, 361, 364, 375, 379, 382, 390, 391, 393, 394, 410, 417, 430, 431, 433, 436, 442, 443, 444, 448 y 451. Los lugares con sustituciones que mostraban el aumento más alto incluían: : T10S, T42V, E68C, E68M, G73F, G73W, K114M, K114T, I133V, N153A, N153S, N153V, W228V, V229I, V229L, S231V, D236R, L264D, S291M, S291T, A301P, A301R, S344M, S344P, S344Q, S344V, G361 D, G361 E, G361 F, G361I, G361 L, G361 M, G361 P, G361 S, G361 W, G361 Y, A364D, A364E, A364F, A364G, A364K, A364L, A354R, A364S, A364V, A364W, T375N, R433C, R433G, R433N, R433S, R433V y I436H.

Ejemplo 6: Variantes con actividad específica mejorada (AE) en un análisis de detección de etanol

[0201] Se analizaron las variantes en un análisis de detección de etanol utilizando los análisis descritos anteriormente. La tabla 4 muestra los resultados del análisis de detección para las variantes con un índice de rendimiento (PI) > 1,0 en comparación con el PI de la TrGA precursora. El PI se calcula a partir de las actividades específicas (actividad/mg enzima) de las enzimas variantes y WT. Es el cociente "actividad específica-variante/actividad específica-WT". El PI de la actividad específica para TrGA de tipo silvestre fue 1,0 y una variante con un PI > 1,0 tenía una actividad específica que era mayor que la TrGA precursora. La actividad específica era la actividad medida mediante análisis de detección de etanol dividida por los resultados obtenidos en el análisis Caliper descrito anteriormente.

Tabla 4 - Detección Etanol

Variante	PI para Actividad Específica
T010D	1,09
T010F	1,06
T010G	1,12
T010K	1,05
T010L	1,07
T010M	1,05
T010P	1,05
T010R	1,08
T010S	1,09
L014E	1,03
L014H	1,05
N015D	1,02
P023A	1,16

P023F	1,13
P023N	1,05
F059A	1,17
F059F	1,05
F059G	1,05
K060F	1,06
K060H	1,03
N061D	1,05
N061I	1,21
N061L	1,18
N061Q	1,08
N061V	1,11
N061W	1,02
R065A	1,17
R065C	1,08
R065G	1,08
R065I	1,11
R065K	1,09
R065M	1,07

ES 2 518 916 T3

R065S	1,12
R065V	1,14
R065Y	1,01
T067C	1,14
T067I	1,13
T067K	1,05
T067M	1,22
E068I	1,06
E068M	1,10
E068W	1,03
A072E	1,11
A072G	1,02
A072L	1,03
A072M	1,11
A072Q	1,10
A072R	1,10
A072W	1,06
A072Y	1,30
G073C	1,02
G073L	1,07
G073W	1,03
S097F	1,11
S097M	1,11
S097N	1,23
S097P	1,18
S097R	1,07
S097V	1,12
S097W	1,09
S097Y	1,18
L098H	1,04
L098M	1,09
A099C	1,07
A099L	1,01
A099M	1,03
A099N	1,11
A099P	1,08
S102A	1,20
S102C	1,04
S102I	1,04
S102L	1,05
S102M	1,25
S102N	1,19
S102R	1,21
S102V	1,07
S102W	1,06
S102Y	1,10
E110Q	1,02
E110S	1,07
E110W	1,11
L113E	1,15
L113N	1,08
I133K	1,04
I133R	1,16
I133S	1,08

I133T	1,29
K140A	1,04
K140E	1,04
K140F	1,03
K140H	1,14
K140L	1,10
K140M	1,11
K140N	1,15
K140Q	1,08
K140R	1,12
K140S	1,13
K140V	1,15
K140W	1,07
K140Y	1,06
N144C	1,05
N144D	1,15
N144E	1,16
N144I	1,13
N144K	1,05
N145A	1,07
N145C	1,09
N145E	1,03
N145I	1,20
N145K	1,05
N145L	1,03
N145M	1,07
N145Q	1,14
N145R	1,11
N145V	1,12
N145W	1,14
N145Y	1,05
Y147A	1,02
Y147M	1,02
Y147R	1,12
S152H	1,08
S152M	1,10
N153C	1,09
N153D	1,20
N153K	1,13
N153L	1,12
N153W	1,07
N153Y	1,13
N164A	1,02
N164G	1,03
N182C	1,12
N182E	1,13
N182K	1,07
N182P	1,01
N182R	1,03
A204C	1,04
A204D	1,09
A204G	1,02
A204I	1,06
A204M	1,09

ES 2 518 916 T3

A204Q	1,09
A204T	1,05
T205A	1,03
T205D	1,03
T205H	1,03
T205I	1,05
T205K	1,09
T205M	1,05
T205N	1,09
T205P	1,17
T205Q	1,25
T205S	1,10
T205V	1,06
T205W	1,05
T205Y	1,18
S214P	1,08
S214T	1,07
V216C	1,08
V216G	1,05
V216H	1,03
V216K	1,02
V216N	1,13
V216Y	1,09
Q219D	1,05
Q219G	1,06
Q219H	1,03
Q219N	1,08
Q219P	1,16
Q219S	1,29
W228A	1,20
W228F	1,22
W228G	1,17
W228H	1,33
W228I	1,18
W228L	1,12
W228M	1,35
W228T	1,19
V229E	1,01
V229I	1,02
V229M	1,03
V229N	1,01
V229Q	1,02
S230C	1,23
S230D	1,13
S230E	1,10
S230F	1,63
S230G	1,77
S230H	1,05
S230I	1,18
S230K	1,04
S230L	1,20
S230N	1,23
S230P	1,13
S230Q	1,20

S230R	1,84
S230T	1,11
S230V	1,12
S230Y	1,08
S231C	1,13
S231D	1,08
S231F	1,17
S231L	1,29
S231M	1,08
S231N	1,04
S231Q	1,05
S231R	1,02
S231V	1,07
S231Y	1,07
D236F	1,14
D236G	1,05
D236L	1,11
D236M	1,07
D236N	1,03
D236P	1,06
D236S	1,03
D236T	1,14
D236V	1,04
I239M	1,04
I239Q	1,08
I239S	1,11
I239V	1,52
I239W	1,02
I239Y	1,25
T241C	1,07
T241E	1,03
T241H	1,10
T241L	1,04
T241M	1,05
T241P	1,02
T241S	1,08
T241V	1,05
N242C	1,08
N242F	1,06
N242M	1,04
N242T	1,08
N242V	1,03
N242W	1,05
N263H	1,05
N263K	1,02
N263P	1,40
L264A	1,04
L264C	1,08
L264E	1,16
L264F	1,03
L264S	1,05
G265E	1,10
G265H	1,12
G265I	1,06

ES 2 518 916 T3

G265K	1,03
G265R	1,06
G265T	1,10
A268C	1,50
A268D	1,14
A268E	1,18
A268F	1,15
A268G	1,35
A268I	1,15
A268K	1,23
A268L	1,06
A268P	1,08
A268R	1,14
A268T	1,18
A268W	1,05
G269E	1,01
D276S	1,01
V284R	1,06
V284T	1,05
V284Y	1,07
S291A	1,26
S291E	1,09
S291F	1,13
S291H	1,13
S291K	1,07
S291N	1,04
S291P	1,12
S291W	1,04
P300K	1,10
P300P	1,12
P300R	1,11
A301A	1,01
A301E	1,09
A301K	1,09
A301L	1,05
A301P	1,02
A301S	1,03
A301W	1,04
A303C	1,06
A303D	1,04
A303F	1,09
A303H	1,05
A303I	1,09
A303K	1,02
A303L	1,05
A303N	1,04
A303R	1,10
A303T	1,11
A303V	1,04
A303W	1,15
A303Y	1,07
A311N	1,04
A311P	1,09
A311Q	1,19

A311S	1,01
A311Y	1,06
V338P	1,04
V338Q	1,12
V338S	1,14
V338Y	1,05
T342N	1,06
T342V	1,23
S344A	1,05
S344T	1,01
T346G	1,14
T346H	1,07
T346M	1,06
T346N	1,09
T346P	1,07
T346Q	1,04
T346Y	1,06
A349L	1,02
V359I	1,14
V359K	1,08
V359M	1,09
V359N	1,02
V359Q	1,14
V359R	1,15
V359W	1,04
G361H	1,11
G361L	1,04
G361R	1,04
A364M	1,05
A364W	1,07
T375C	1,12
T375D	1,01
T375E	1,02
T375H	1,05
T375V	1,04
T375W	1,02
T375Y	1,05
N379A	1,05
N379C	1,10
N379D	1,05
N379G	1,07
N379I	1,01
N379M	1,04
N379P	1,06
N379S	1,01
S382A	1,02
S382N	1,02
S382P	1,10
S382S	1,09
S382V	1,10
S382W	1,10
S390A	1,05
S390Y	1,03
E391A	1,17

E391E	1,10
E391I	1,13
E391K	1,18
E391L	1,18
E391M	1,05
E391Q	1,04
E391R	1,08
E391V	1,05
E391W	1,12
E391Y	1,08
A393E	1,05
A393G	1,14
A393H	1,10
A393I	1,07
A393K	1,09
A393L	1,12
A393M	1,07
A393N	1,18
A393Q	1,02
A393R	1,09
A393S	1,10
A393T	1,13
A393V	1,16
A393W	1,04
A393Y	1,03
K394A	1,11
K394H	1,11
K394K	1,08
K394L	1,01
K394M	1,14
K394Q	1,09
K394R	1,19
K394S	1,22
K394T	1,08
K394V	1,05
K394W	1,13
S410E	1,01
S410H	1,06
S410N	1,04
L417A	1,12
L417D	1,19
L417E	1,10
L417F	1,08
L417G	1,19
L417I	1,10
L417K	1,02
L417Q	1,04
L417R	1,30
L417S	1,05
L417T	1,10

L417V	1,21
L417W	1,05
L417Y	1,10
H418E	1,01
H418M	1,12
T430A	1,19
T430E	1,15
T430F	1,09
T430G	1,16
T430H	1,15
T430I	1,06
T430K	1,24
T430M	1,16
T430N	1,07
T430Q	1,15
T430R	1,04
T430V	1,09
A431C	1,04
A431E	1,08
A431H	1,11
A431I	1,20
A431L	1,21
A431M	1,12
A431Q	1,22
A431R	1,11
A431S	1,09
A431W	1,04
A431Y	1,13
R433A	1,09
R433M	1,17
R433W	1,06
R433Y	1,22
A442A	1,13
S444K	1,03
S444M	1,13
S444N	1,04
S444P	1,08
S444Q	1,08
S444R	1,03
S444T	1,15
S444V	1,14
S444W	1,16
T448F	1,02
T448G	1,08
T448I	1,10
T448P	1,08
T448Q	1,04
T448V	1,04
T451K	1,29

[0202] La tabla 4 proporciona las variantes con un índice de rendimiento de al menos 1,0. Estas incluían los siguientes lugares: 10, 14, 15, 23, 59, 60, 61, 65, 67, 68, 72, 73, 97, 98, 99, 102, 110, 113, 133, 140, 144, 145, 147, 152, 153, 164, 182, 204, 205, 214, 216, 219, 228, 229, 230, 231, 236, 239, 241, 242, 263, 264, 265, 268, 269, 276, 284, 291, 300, 301, 303, 311, 338, 342, 344, 346, 349, 359, 361, 364, 375, 379, 382, 390, 391, 393, 394, 410, 417, 418, 430, 431, 433, 442, 444, 448 y 451. Los lugares que muestran la actividad específica más

alta (con un índice de rendimiento por encima de 1,2) incluían: 61, 67, 72, 97, 102, 133, 145, 153, 205, 219, 228, 230, 231, 239, 263, 268, 291, 311, 342, 394, 430, 431 y 451.

Ejemplo 7: Variantes de actividad específica y de termoestabilidad combinadas

5 [0203] La tabla 5 muestra las variantes que presentaban un índice de rendimiento de 1,0 o mejor en comparación con la precursora para ambas propiedades: actividad específica y termoestabilidad. Estas incluían los siguientes lugares. 10, 15, 59, 61, 68, 72, 73, 97, 99, 102, 133, 140, 153, 182, 204, 205, 214, 228, 229, 230, 231, 236, 241, 242, 264, 265, 268, 275, 284, 291, 300, 301, 303, 311, 338, 344, 346, 359, 361, 364, 375, 370, 382, 391, 393, 394, 410, 417, 430, 431, 433, 444, 448 y 451. Los lugares que muestran la actividad específica y termoestabilidad combinadas más altas incluían: 228, 230, 231, 268, 291, 417, 433 y 451.

Tabla 5: Variantes combinadas

Variante	PI de Actividad Específica	PI de Estabilidad Térmica
T010F	1,06	1,11
T010G	1,12	1,13
T010M	1,05	1,12
T010R	1,08	1,06
T010S	1,09	1,24
N015N	1,06	1,06
F059A	1,17	1,03
F059G	1,05	1,07
N061V	1,11	1,06
E068I	1,06	1,19
E068M	1,10	1,29
E068W	1,03	1,09
A072E	1,11	1,07
A072Q	1,10	1,02
G073W	1,03	1,36
S097V	1,12	1,03
A099C	1,07	1,07
A099L	1,01	1,09
S102A	1,20	1,02
K140Q	1,08	1,06
K140S	1,13	1,05
K140W	1,07	1,04
N153D	1,20	1,06
N153L	1,12	1,06
N153W	1,07	1,19
N182R	1,03	1,02

A204D	1,09	1,02
A204M	1,09	1,08
T205D	1,03	1,06
T205N	1,09	1,09
T205P	1,17	1,17
T205S	1,10	1,04
T205V	1,06	1,06
T205Y	1,18	1,07
S214P	1,08	1,04
S214T	1,07	1,06
W228A	1,20	1,01
W228F	1,22	1,12
W228G	1,17	1,06
W228H	1,33	1,05
W228I	1,18	1,06
W228L	1,12	1,14
W228M	1,35	1,04
W228T	1,19	1,06
V229E	1,01	1,16
V229I	1,02	1,21
V229M	1,03	1,11
V229N	1,01	1,07
V229Q	1,02	1,10
S230C	1,23	1,09
S230D	1,13	1,08
S230E	1,10	1,11
S230F	1,63	1,08
S230G	1,77	1,09
S230H	1,05	1,03

10

ES 2 518 916 T3

S230K	1,04	1,02
S230L	1,20	1,02
S230N	1,23	1,16
S230P	1,13	1,12
S230Q	1,20	1,20
S230R	1,84	1,11
S230T	1,11	1,07
S230V	1,12	1,11
S230Y	1,08	1,05
S231C	1,13	1,07
S231D	1,08	1,18
S231F	1,17	1,14
S231L	1,29	1,15
S231M	1,08	1,09
S231N	1,04	1,13
S231Q	1,05	1,18
S231R	1,02	1,13
S231V	1,07	1,21
S231Y	1,07	1,10
D236F	1,14	1,11
D236G	1,05	1,07
D236L	1,11	1,08
D236M	1,07	1,15
D236N	1,03	1,15
D236P	1,06	1,06
D236T	1,14	1,16
D236V	1,04	1,18
T241V	1,05	1,05
N242F	1,06	1,03
L264A	1,04	1,03
G265E	1,10	1,03
G265H	1,12	1,09
G265I	1,06	1,09
G265K	1,03	1,05
G265R	1,06	1,12
G265T	1,10	1,17
A268C	1,50	1,16
A268D	1,14	1,20
A268E	1,18	1,10
A268F	1,15	1,07
A268G	1,35	1,11
A268I	1,15	1,12
A268K	1,23	1,11
A268L	1,06	1,11
A268P	1,08	1,13
A268R	1,14	1,15
A268T	1,18	1,14
A268W	1,05	1,05
D276S	1,01	1,02
V284R	1,06	1,09
V284T	1,05	1,08
V284Y	1,07	1,12
S291A	1,26	1,21
S291E	1,09	1,02

S291F	1,13	1,24
S291H	1,13	1,25
S291K	1,07	1,13
S291N	1,04	1,08
S291P	1,12	1,11
S291W	1,04	1,15
P300R	1,11	1,02
A301K	1,09	1,11
A301L	1,05	1,07
A301P	1,02	1,22
A301S	1,03	1,12
A301W	1,04	1,14
A303I	1,09	1,02
A303L	1,05	1,03
A311S	1,01	1,01
A311Y	1,06	1,02
V338P	1,04	1,18
V338Q	1,12	1,20
V338S	1,14	1,19
V338Y	1,05	1,10
S344A	1,05	1,05
S344T	1,01	1,19
T346M	1,06	1,07
T346N	1,09	1,15
T346P	1,07	1,13
T346Q	1,04	1,12
T346Y	1,06	1,07
A349L	1,02	1,16
V359Q	1,14	1,02
V359R	1,15	1,13
G361H	1,11	1,11
G361L	1,04	1,22
G361R	1,04	1,19
A364M	1,05	1,21
A364W	1,07	1,23
T375E	1,02	1,03
T375H	1,05	1,13
T375V	1,04	1,16
T375W	1,02	1,11
T375Y	1,05	1,17
N379A	1,05	1,11
N379C	1,10	1,03
N379D	1,05	1,06
N379G	1,07	1,09
N379I	1,01	1,04
N379M	1,04	1,08
N379P	1,06	1,17
S382A	1,02	1,06
S382N	1,02	1,07
S382P	1,10	1,10
E391L	1,18	1,04
E391R	1,08	1,02
E391W	1,12	1,04
E391Y	1,08	1,02

A393E	1,05	1,03
A393G	1,14	1,12
A393H	1,10	1,05
A393I	1,07	1,06
A393K	1,09	1,05
A393L	1,12	1,17
A393M	1,07	1,07
A393N	1,18	1,11
A393Q	1,02	1,02
A393R	1,09	1,03
A393S	1,10	1,13
A393T	1,13	1,08
A393V	1,16	1,09
A393W	1,04	1,10
A393Y	1,03	1,12
K394A	1,11	1,08
K394H	1,11	1,08
K394L	1,01	1,08
K394M	1,14	1,06
K394Q	1,09	1,07
K394R	1,19	1,06
K394T	1,08	1,03
K394V	1,05	1,02
S410E	1,01	1,05
S410H	1,06	1,11
S410N	1,04	1,10
L417I	1,10	1,04
L417K	1,02	1,20
L417Q	1,04	1,04

L417R	1,30	1,20
L417V	1,21	1,07
L417Y	1,10	1,01
T430A	1,19	1,05
T430E	1,15	1,02
T430F	1,09	1,06
T430H	1,15	1,10
T430I	1,06	1,04
T430K	1,24	1,08
T430M	1,16	1,17
T430N	1,07	1,13
T430Q	1,15	1,05
T430R	1,04	1,13
T430V	1,09	1,05
A431I	1,20	1,03
A431R	1,11	1,08
R433A	1,09	1,13
R433M	1,17	1,10
R433W	1,06	1,16
R433Y	1,22	1,18
S444K	1,03	1,07
S444M	1,13	1,05
S444Q	1,08	1,04
T448F	1,02	1,12
T448I	1,10	1,12
T448Q	1,04	1,17
T448V	1,04	1,13
S451K	1,29	1,08

Referencia

Ejemplo 8: Estructura de cristal de TrGA

- 5 **[0204]** La estructura tridimensional completa de la glucoamilasa *Trichoderma reesei* (*Hypocrea jecorina*) (TrGA) se determinó con la resolución 1,9 Å. La tabla 6 muestra las coordenadas para la estructura de cristal de la glucoamilasa *Trichoderma*. TrGA se cristalizó de forma intacta con 599 residuos y todas las modificaciones traduccionales posteriores que normalmente tendrían lugar en el huésped natural. Se produjo la estructura de cristal y se analizó de la siguiente manera:
- 10 **[0205]** Expresión proteínica y purificación - El gen codificante de GA *H. jecorina* se clonó y se expresó según los protocolos descritos en la publicación de solicitud de patente nº US 2006/0094080 A1, por Dunn-Coleman *et al.* (Mayo 4/2006).
- 15 **[0206]** El material de proteína TrGA usado para los experimentos de cristalización se purificó inicialmente en una etapa mediante cromatografía de intercambio de aniones como se muestra a continuación: se prepararon sobrenadantes de cultivo concentrados de TrGA expresada, consistentes en 180 mg/ml de proteína total mediante la dilución de la muestra 1:10 en un tampón de 25 mM Tris-HCl, pH 8,0. Se empleó una columna FF Sefarosa Q HIPREP 16/10 (GE Healthcare) para la purificación de intercambio aniónico. La columna HIPREP se equilibró con 4 volúmenes de columna (VC) de tampón de inicio (25 mM Tris-HCl, pH 8,0) seguido de una aplicación de 10 ml de la muestra proteínica diluida. Se aplicó un gradiente lineal 8 VC de 0 a 140 mM NaCl en el tampón de migración (25 mM Tris-HCl, pH 8,0) con el fin de eluir la proteína unida. Se eluyó la TrGA unida de la columna sefarosa Q HIPREP con una concentración de sal de aproximadamente 80 mM NaCl. Las fracciones con proteína TrGA pura se acumularon y concentraron a 50 mg/ml utilizando un tubo de concentración centrífuga VIVASPIN (Viva Science) con un límite de peso molecular (LPM) de 10 kD. Se cambió el tampón del material de TrGA purificado y concentrado utilizando una columna desaladora DG-10 (Bio-Rad) equilibrada con 50 mM tampón acetato de sodio, pH 4,3. Las concentraciones de proteína se determinaron mediante la medición de la absorbancia a 280 nm. El contenido proteico TrGA concentrado y purificado inicialmente se almacenó después a -20°C.
- 25 **[0207]** Se introdujeron dos etapas de purificación adicionales, una purificación mediante intercambio de aniones adicional y una purificación por exclusión de tamaño, con el fin de mejorar la tendencia del material proteínico TrGA a formar cristales. Estas dos etapas de purificación adicionales se llevaron a cabo de la siguiente forma: En la primera etapa de purificación por intercambio de aniones se utilizó una columna MONOQ 10 ml (GE
- 30

Healthcare). Una muestra de 1 ml del material TrGA congelado y purificado inicialmente (50 mg proteína) se descongeló y el tampón se cambió a 20 mM Tris-HCL, pH 8,0, mediante dilución repetida de la muestra a 6 ml en el nuevo tampón, seguido de una concentración de la muestra de nuevo a 0,5 ml utilizando un tubo de concentración de 6 ml 5 kD LPM. La muestra TrGA se diluyó tras la última etapa de concentración en agua destilada hasta que se alcanzó una conductividad de la muestra proteínica que correspondía a la conductividad del tampón inicial de la purificación de aniones, es decir, 25 mM Tris-HCl, pH 8,0. La columna MONOQ se equilibró primero con tampón de inicio de 4 volúmenes de columna (VC), seguido de la aplicación de la muestra proteínica diluida a la columna. Se eluyó la proteína unida de la columna MONOQ mediante dos gradientes diferentes. En el primero, se aplicó un gradiente pH lineal 4 VC donde el pH del tampón de inicio se disminuyó de 8,0 a 6,0. En el segundo gradiente, se aplicó un gradiente de sal largo 8 VC en el que la concentración de sal se aumentó de 0 a 350 mM NaCl en el tampón de migración (25 mM Tris-HCl, pH 6,0). Se halló que la TrGA unida se eluía de la columna durante el segundo gradiente de sal con una concentración de NaCl aproximada de 150 mM. Las fracciones con TrGA se acumularon y concentraron a 2 ml utilizando un tubo de concentración VIVASPIN 6 ml 5 kD LPM. A continuación, la muestra TrGA concentrada se aplicó a una columna de exclusión de tamaño 16/60 SUPERDEX 200 (GE Healthcare) equilibrada con 4 VC de 20 mM Tris-Cl, pH 8,0 y 50 mM NaCl, que también se utilizó como tampón de migración. Las fracciones del principal pico de elución tras la purificación por exclusión de tamaño se acumularon y concentraron a una concentración de proteína aproximada de 7,5 mg/ml utilizando un tubo de concentración VIVASPIN de 6 ml 5 kD LPM.

[0208] Cristalización de proteína - La muestra proteínica que se utilizó para hallar las condiciones de cristalización de TrGA iniciales fue una muestra del material TrGA que se purificó una vez mediante purificación por intercambio de aniones y a partir de ahí se almacenó a -20 °C. La muestra proteínica TrGA se descongeló y se diluyó con 50 mM tampón acetato de sodio, pH 4,3 a aproximadamente 12 mg/ml antes de los experimentos de cristalización iniciales. Se halló que los cristales TrGA se cultivaban en solución consistente en 25 % PEG 3350, 0,20M acetato de amonio, 0,10M Bis-Tris pH 5,5 (solución de reserva), utilizando el método de difusión por vapor con gotas colgantes (McPherson 1982), a 20° C. Las gotas de cristalización se prepararon mediante la mezcla de iguales cantidades de solución proteínica (12 mg/ml) y solución de reserva para un volumen final de 10 µl. Se halló que los cristales TrGA pertenecen al grupo espacial ortorrómbico P212121 con las dimensiones celulares aproximadas: $a = 52,2 \text{ \AA}$, $b = 99,2 \text{ \AA}$, $c = 121,2 \text{ \AA}$ y presentan una V_m calculada de 2,3 (Matthews 1968) con una molécula en la unidad asimétrica.

[0209] Recogida de datos de rayos X - Los dos conjuntos de datos TrGA ortorrómbicos se recogieron de cristales únicos fijados en tubos capilares sellados, a temperatura ambiente. El conjunto de datos de rayos X TrGA ortorrómbicos de baja resolución inicial, utilizado para solucionar la estructura mediante métodos de reemplazo molecular (RM), se recogió en una fuente de rayos X doméstica, un detector de placa de imagen Raxis IV++ MSC/Rigaku (Molecular Structures Corp., The Woodlands, Texas) con espejos de enfoque con una radiación Cu K α de un generador de ánodos giratorio Rigaku RU200. Este conjunto de datos se procesó, se extrajo y se realizó la media utilizando el software d*trek suministrado por MSC/Rigaku. El conjunto de datos de rayos X ortorrómbicos se utilizó para solucionar la estructura TrGA mediante reemplazo molecular (RM) y el conjunto de datos ortorrómbicos de alta resolución se utilizó para el modelo de estructura TrGA del grupo espacial ortorrómbico final. El conjunto de datos monoclinicos centrados C se recogió de un único cristal TrGA congelado a 100K, equilibrado en un agente crioprotector compuesto de PEG 3350 25 %, 15% Glicerol 50 mM CaCl₂ y 0,1 M Bis-Tris pH 5,5 como crioprotector, fijado en bucles de fibra rayón y sumergido congelado en nitrógeno líquido antes de su transporte al sincrotrón. El conjunto de datos ortorrómbicos de alta resolución (1,9 Å) y el conjunto de datos monoclinicos centrados C (1,8 Å) se recogieron ambos en una fuente de sincrotrón, línea de haz 911:5 en MAX LAB en Lund, Suecia. Ambos conjuntos de datos recogidos en una fuente de sincrotrón se procesaron con MOSFLM y se organizaron con el programa SCALA incluido en el paquete de programas CCP4 (Collaborative Computational Project Number 4 1994). El procesamiento de datos posterior se llevó a cabo utilizando el paquete de programas CCP4 (Collaborative Computational Project Number 4 1994), a menos que se indicara lo contrario. Un conjunto del 5 % de las reflexiones de cada conjunto de datos se deja apartado y se utiliza para controlar el libre R (Brunger, A. (1992) *Nature* 355:472-475).

[0210] Solución de estructura - La estructura TrGA se solucionó inicialmente mediante RM con el programa de reemplazo automático MOLREP (Collaborative Computational Project Number 4 1994), incluido en el paquete de programas CCP4, utilizando el conjunto de datos ortorrómbicos de resolución baja inicial y utilizando las coordenadas de la variante *Aspergillus awamori* GA (AaGA) X100 (pdb entrada 1GLM (Aleshin *et al.* (1994) *J Mol Biol.* 238:575-591) como modelo de investigación. Se editó el modelo de investigación *A. awamori* GA para eliminar todas las fracciones de glicosilación fijadas a la molécula de proteína como glicosilaciones N- y O- y todas las moléculas de solventes antes de llevar a cabo los experimentos RM. Todas las reflexiones entre 36,8 y 2,8 Å de resolución, del conjunto de datos de baja resolución inicial, se utilizaron para la solución RM. El programa RM halló una única solución de función de rotación, con un máximo de 11,1 σ por encima de lo ya registrado, el siguiente valor más alto fue 3,8 σ por encima de lo ya registrado. La solución de función de traducción proporcionó un factor R de 48,7 % y tuvo un factor de contraste de 17,4. La solución RM se refinó durante 10 ciclos de refinamiento de cuadrados mínimos limitado utilizando el programa Refmac 5,0 (Murshudov, *et al.* (1997) *Acta Crystallogr.*, D53:240-255). Esto redujo el factor R cristalográfico a 31,1 % mientras que el valor

libre R cayó desde 42,2 % a 41,1 %.

5 **[0211]** Refinamiento y encaje modelo - El modelo de solución RM refinado se utilizó para calcular un mapa de densidad inicial a partir del conjunto de datos TrGA ortorrómbicos de baja resolución. La densidad de electrones para un puente disulfuro entre los residuos 19 y 26 de TrGA, un puente disulfuro no presente en el modelo de estructura de la variante *A. awamori* X100, podría identificarse fácilmente en este mapa de densidad de electrones. Esto se tomó como una indicación de que el mapa de densidad de electrones tenía la suficiente calidad para utilizarse para construir un modelo de estructura de TrGA a partir de su secuencia de aminoácidos. El modelo de estructura TrGA inicial, según el conjunto de datos de baja resolución, se refinó con ciclos alternos de construcción de modelo mediante Coot (Emsley *et al.* (2004) *Acta Crystallogr D Biol Crystallogr* 60:2126-32), y 10 la máxima probabilidad de refinamiento utilizando Refmac 5.0.

15 **[0212]** La resolución del modelo de estructura TrGA inicial se extendió a la resolución del conjunto de datos ortorrómbicos de alta resolución (1,9 Å) mediante el refinamiento del modelo de estructura TrGA inicial contra el conjunto de datos de alta resolución durante 10 ciclos de refinamiento limitado utilizando el programa Refmac 5.0. La mayoría de las moléculas de agua en los modelos de estructura se ubicaron de forma automática mediante el uso de protocolos de recogida de agua en los programas de refinamiento y, a continuación, de forma manual se eligieron o descartaron mediante inspección visual. Todas las comparaciones estructurales se realizaron con Coot (Emsley *et al.* (2004) *supra*) u O (Jones *et al.* (1991) *Acta Crystallogr.* A47: 110-119), y las figuras se prepararon con PyMOL (DeLano W.L. (2002) *The PyMOL Molecular Graphics System*, Palo Alto, CA EEUU; DeLano Scientific).

20 **[0213]** A partir de estos resultados, puede verse que el segmento principal catalítico TrGA seguía la misma topología de barriles (α/α)₆ descrita por Aleshin *et al.* (1992, *supra*) para AaGA, consistente en un doble barril de hélices alfa con el C-terminal de la hélice externa hacia el N-terminal de una hélice interna. Fue posible identificar diferencias clave en la densidad de electrones tal como el puente disulfuro entre los residuos 19 y 26 y en una inserción (residuos 257-260) en relación con AaGA. El segmento que comprende 80-100 también experimentó una reconstrucción de modelo extensiva. Se identificó un lugar de glicosilación principal en Asn 171 que tenía hasta cuatro fracciones de glucósidos fijadas. Se identificó un lugar de glicosilación similar en AaGa. De forma adicional, el núcleo catalítico con tres cis péptidos entre los residuos 22-23, 44-45 y 122-123 se conservó entre TrGA y AaGA. En general, existía una variación rms de 0,535 Å entre 409 de 453 átomos Ca cuando se compararon las coordenadas de los núcleos catalíticos de TrGA y AaGA.

30 Referencia **Ejemplo 9: Homología entre TrGA y *Aspergillus awamori* GA**

35 **[0214]** La estructura de cristal del TrGA identificada en el ejemplo 8 se superpuso sobre la estructura de cristal identificada previamente de la *Aspergillus awamori* GA (AaGA). La estructura de cristal AaGA se obtuvo a partir de la base de datos proteínica (BDP) y la forma de AaGA que se cristalizó fue la forma con solo un dominio catalítico. La estructura de la glucoamilasa *Trichoderma reesei* con las tres regiones intactas se determinó con la resolución 1,8 Angstrom en el presente documento (véase la tabla 6 y el ejemplo 8). Al utilizar las coordenadas (véase la tabla 6) la estructura se alineó con las coordenadas del dominio catalítico de la cepa *Aspergillus awamori* X100 que se determinó previamente (Aleshin, A.E., Hoffman, C., Firsov, L.M., y Honzatko, R.B. (1994) *J Mol Biol* 238:575-591 y la BDP). Como puede observarse en las figuras 6 y 7, la estructura del dominio catalítico se superpuso muy estrechamente y permitió la identificación de los residuos equivalentes según esta 40 superposición estructural.

45 **[0215]** Según el presente análisis, se identificaron los lugares que podían mutarse en TrGA y tener como resultado una estabilidad y/o actividad específica aumentada. Estos lugares incluyen 108, 124, 175 y 316 en el lugar activo. También se identificaron variantes por pares específicos Y47W/Y31 5F e Y47F/Y315W. Se identificaron otros lugares 143, D44, P45, D46, R122, R125, V181, E242, Y310, D313, V314, N317, R408 y N409. Debido a la alta homología estructural se espera que las variantes beneficiosas halladas en los lugares en *Trichoderma reesei* GA tengan consecuencias similares en *Aspergillus awamori* y otras glucoamilasas homólogas.

Reivindicaciones

1. Una variante de glucoamilasa, que es una variante de una glucoamilasa precursora, presentando dicha glucoamilasa precursora una secuencia de aminoácidos con al menos un 95 % de identidad de secuencia con la secuencia de aminoácidos:

```

SVDDFI STETPIALNLLCNVGPDCRAFSGT SAGAVIASPSTIDPDYYMWTRDSALVFKNL
IDRFTETYDAGLQRRIEQYI TAQVTLQGLSNP SGLADGSGLGEPKFELTLKPFTGNWGRPQ
RDGPALRAIALIGYSKWL INNNYQSTVSNVIWPIVRNDLNYVAQYWNQTGF DLWEEVNGSSF
FTVANQHRALVEGATLAATLGQSGSAYSSVAPQVLCFLQRFWVSSGGYVDSNINTNEGRIGK
DVNSVLTSIHTFDPNLGCDAGTFQPCSDKALS NLKVVVDSFRS IYGVNKGIPAGAAVAIGRY
AEDVYYNGNPWYLATFAAAEQLYDAIYVWKKTGSITVTATSLAFFQELVPGV TAGTYSSSSS
TFTNI INAVSTYADGFLSEAAKYVPADGSLAEQFDRNSGTPLSALHLTWSYASFLTATARRA
GIVPPSWANSSASTIPSTCSGASVVGSYSRPTATSFPPSQTPKPGVPSGTPYTPLPCATPTS
YAVTFHELVTQFGQTVKVAGNAAALGNWSTSA AVALDAVNYADNHPLWIGTVNLEAGDVVE
YKYINVGQDGSVTWESDPNHTYTVPAVACVTQVVKEDTWQS (SEQ ID NO: 2),
    
```

5

presentando dicha variante únicamente una, dos, tres o cuatro sustituciones en comparación con dicha precursora, en la que dichas sustituciones comprenden la sustitución T430A, T430E, T430F, T430G, T430H, T430I, T430K, T430M, T430N, T430Q, T430R o T430V en SEQ ID NO:2 o una posición equivalente en dicha glucoamilasa precursora como se determina mediante la alineación de secuencia.

10

2. Variante de glucoamilasa de acuerdo con la reivindicación 1, en la que dicha glucoamilasa precursora presenta al menos un 97 % o al menos un 98 % de identidad de secuencia con la secuencia de SEQ ID NO:2.

15

3. Variante de glucoamilasa de acuerdo con la reivindicación 1, en la que dicha variante presenta únicamente dos, tres o cuatro sustituciones en comparación con dicha precursora, sustituciones adicionales que se eligen de sustituciones en una posición correspondiente a la posición 10, 14, 15, 23, 42, 45, 46, 59, 60, 61, 67, 68, 72, 73, 97, 98, 99, 102, 108, 110, 113, 114, 122, 124, 125, 133, 140, 144, 145, 147, 152, 153, 164, 175, 182, 204, 205, 214, 216, 219, 228, 229, 230, 231, 236, 239, 240, 241, 242, 244, 263, 264, 265, 268, 269, 276, 284, 291, 300, 301, 303, 310, 311, 313, 316, 338, 342, 344, 346, 349, 359, 361, 364, 379, 382, 390, 391, 393, 394, 408, 410, 415, 417, 418, 431, 433, 436, 442, 443, 444, 448 y 451 de SEQ ID NO:2 o una posición equivalente en dicha glucoamilasa precursora como se determina mediante la alineación de secuencia.

20

4. Variante de glucoamilasa de acuerdo con la reivindicación 1, en la que la variante presenta actividad específica aumentada en comparación con la glucoamilasa que comprende la secuencia de SEQ ID NO:2.

25

5. Variante de glucoamilasa de acuerdo con la reivindicación 1, en la que la variante presenta termoestabilidad aumentada en comparación con la glucoamilasa que comprende la secuencia de SEQ ID NO:2.

6. Variante de glucoamilasa de acuerdo con la reivindicación 1, en la que dicha variante presenta actividad específica aumentada y termoestabilidad aumentada en comparación con la glucoamilasa que comprende SEQ ID NO:2.

30

7. Variante de glucoamilasa de acuerdo con la reivindicación 1, en la que la glucoamilasa precursora se obtiene a partir de una *Trichoderma spp.*

8. Un polinucleótido que codifica la variante de acuerdo con la reivindicación 1.

9. Una célula huésped que comprende un polinucleótido de acuerdo con la reivindicación 8.

10. Una composición de enzima que comprende la variante de glucoamilasa de acuerdo con la reivindicación 1.

35

11. Composición de enzima de acuerdo con la reivindicación 10, en la que la composición es adecuada para utilizarse en un proceso de conversión de almidón, una composición de pienso animal o un proceso de fermentación de sustrato que contiene almidón.

12. Composición de enzima de acuerdo con la reivindicación 10, que comprende además un alfa-amilasa.

- 13.** Un método para producir una glucoamilasa variante en una célula huésped que comprende:

la transformación de una célula huésped con una construcción de ADN que comprende un polinucleótido de acuerdo con la reivindicación 8 y el cultivo de la célula huésped con las condiciones adecuadas para la expresión y producción de dicha variante de glucoamilasa y la producción de dicha variante.

5

- 14.** Método de acuerdo con la reivindicación 13, que comprende además la recuperación de la variante de glucoamilasa de dicho cultivo.

FIG. 1A

Proteína precursora TrGA (632 aminoácidos) (SEQ ID NO:1)

```

1  MHVLSTAVLL  GSVAVQKVLG  RPGSSGLSDV  TKRSVDDFIS  TETPIALNNL
51  LCNVGPDGCR  AFGTSAGAVI  ASPSTIDPDY  YMWTRDSAL  VFKNLIDRFT
101  ETYDAGLQRR  IEQYITAQVT  LQGLSNPSGS  LADGSGLGEP  KFELTLKPFT
151  GNWGRPQRDG  PALRAIALIG  YSKWLINNNY  QSTVSNVIWP  IVRNDLNYVA
201  QYWNQTGFDL  WEEVNGSSFF  TVANQHRALV  EGATLAATLG  QSGSAYSSVA
251  PQVLCFLQRF  WVSSGGYVDS  NINTNEGRTG  KDVNSVLTSI  HTFDPNLGCD
301  AGTFQPCSDK  ALSNLKVVVD  SFRSIYGVNK  GIPAGAAVAI  GRYAEDVYYN
351  GNPWYLATFA  AAEQLYDAIY  VWKKTGSITV  TATSLAFFQE  LVPGVTAGTY
401  SSSSSTFTNI  INAVSTYADG  FLSEAAKYVP  ADGSLAEQFD  RNSGTPLSAL
451  HLTWSYASFL  TATARRAGIV  PPSWANSSAS  TIPSTCSGAS  VVGSYSRPTA
501  TSFPPSQTPK  PGVPSGTPYT  PLPCATPTSV  AVTFHELVST  QFGQTVKVAG
551  NAAALGNWST  SAAVALDAVN  YADNHPLWIG  TVNLEAGDVV  EYKYINVGQD
601  GSVTWESDPN  HTYTVPVAVAC  VTQVVKEDTW  QS

```

FIG. 1B

Secuencia codificante ADN de TrGA (1899 pb) (SEQ ID NO:4)

```

1 ATGCACGTCC TGTCGACTGC GGTGCTGCTC GGCTCCGTTG CCGTTCAAAA
51 GGTCCCTGGGA AGACCAGGAT CAAGCGGTCT GTCCGACGTC ACCAAGAGGT
101 CTGTTGACGA CTTCATCAGC ACCGAGACGC CTATTGCACT GAACAATCTT
151 CTTTGCAATG TTGGTCCTGA TGGATGCCGT GCATTCGGCA CATCAGCTGG
201 TGCGGTGATT GCATCTCCCA GCACAATTGA CCCGGACTAC TATTACATGT
251 GGACGCGAGA TAGCGCTCTT GTCTTCAAGA ACCTCATCGA CCGCTTACC
301 GAAACGTACG ATGCGGGCCT GCAGCGCCGC ATCGAGCAGT ACATTACTGC
351 CCAGGTCACT CTCCAGGGCC TCTCTAACC CTGCGGCTCC CTCGCGGACG
401 GCTCTGGTCT CGGCGAGCCC AAGTTTGAGT TGACCCTGAA GCCTTTCACC
451 GGCAACTGGG GTCGACCPCA GCGGGATGGC CCAGCTCTGC GAGCCATTGC
501 CTTGATTGGA TACTCAAAGT GGCTCATCAA CAACAATAT CAGTCGACTG
551 TGTCCAACGT CATCTGGCCT ATTGTGCGCA ACGACCTCAA CTATGTTGCC
601 CAGTACTGGA ACCAAACCGG CTTTGACCTC TGGGAAGAAG TCAATGGGAG
651 CTCATTCTTT ACTGTTGCCA ACCAGCACCG AGCACTTGTC GAGGGCGCCA
701 CTCTTGCTGC CACTCTTGGC CAGTCGGGAA GCGCTTATTC ATCTGTTGCT
751 CCCCAGGTTT TGTGCTTCT CCAACGATC TGGGTGTCGT CTGGTGGATA
801 CGTCGACTCC AACATCAACA CCAACGAGGG CAGGACTGGC AAGGATGTCA
851 ACTCCGTCCT GACTTCCATC CACACCTTCG ATCCCAACCT TGGCTGTGAC
901 GCAGGCACCT TCCAGCCATG CAGTGACAAA GCGCTCTCCA ACCTCAAGGT
951 TGTGTGTCGAC TCCTTCCGCT CCATCTACGG CGTGAACAAG GGCATTCCCTG
1001 CCGGTGCTGC CGTCGCCATT GGCCGGTATG CAGAGGATGT GTACTTCAAC
1051 GGCAACCCTT GGTATCTTGC TACATTTGCT GCTGCCGAGC AGCTGTACGA
1101 TGCCATCTAC GTCTGGAAGA AGACGGGCTC CATCACGGTG ACCGCCACCT
1151 CCCTGGCCTT CTTCCAGGAG CTTGTTCCCTG GCGTGACGGC CGGGACCTAC
1201 TCCAGCAGCT CTTGACCTT TACCAACATC ATCAACGCCG TCTCGACATA
1251 CGCCGATGGC TTCCCTCAGCG AGGCTGCCAA GTACGTCCCC GCCGACGGTT
1301 CGCTGGCCGA GCAGTTTGAC CGCAACAGCG GCACTCCGCT GTCTGCGCTT
1351 CACCTGACGT GGTGCTACGC CTCGTTCTTG ACAGCCACGG CCCGTGGGC
1401 TGGCATCGTG CCCCCTCGT GGGCCAACAG CAGCGCTAGC ACGATCCCCT
1451 CGACGTGCTC CGGCGGTCC GTGGTCGGAT CCTACTCGCG TCCCACCGCC
1501 ACGTCATTCC CTCCGTCGCA GACGCCAAG CCTGGCGTGC CTTCCGGTAC
1551 TCCCTACACG CCCCTGCCCT GCGCGACCCC AACCTCCGTG GCCGTACCT
1601 TCCACGAGCT CGTGTGACA CAGTTTGGCC AGACGGTCAA GGTGGCGGGC
1651 AACGCCCGCG CCCTGGGCAA CTGGAGCACG AGCGCCGCCG TGGCTCTGGA
1701 CGCCGTCAAC TATGCCGATA ACCACCCCT GTGGATTGGG ACGGTCAACC
1751 TCGAGGCTGG AGACGTCGTG GAGTACAAGT ACATCAATGT GGGCCAAGAT
1801 GGCTCCGTGA CCTGGGAGAG TGATCCCAAC CACACTTACA CGGTTCTCTG
1851 GGTGGCTTGT GTGACGCAGG TTGTCAAGGA GGACACCTGG CAGTCGTAA

```

FIG. 1C

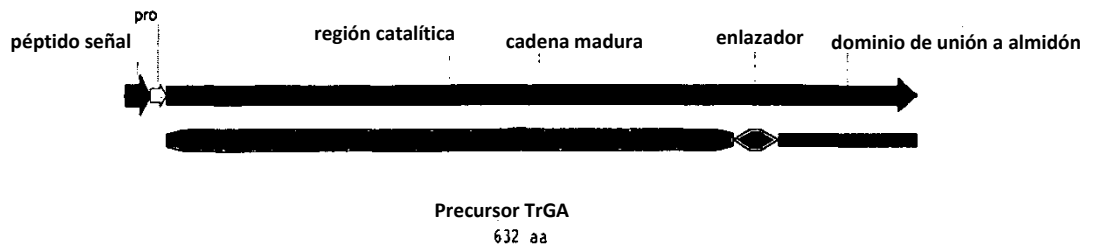


FIG. 2

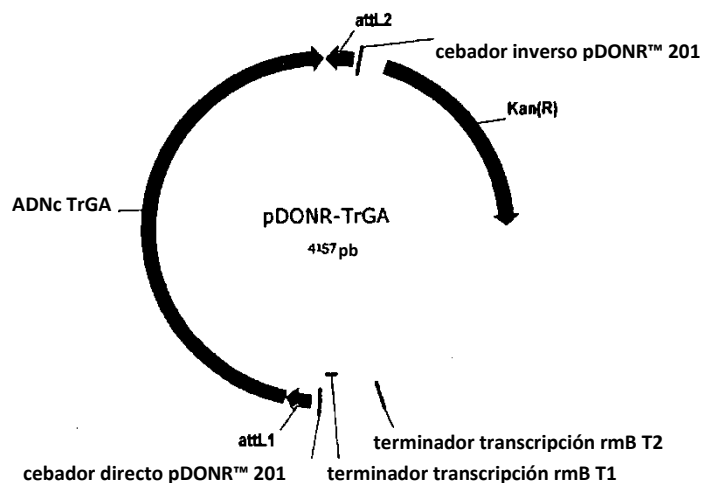


FIG. 3

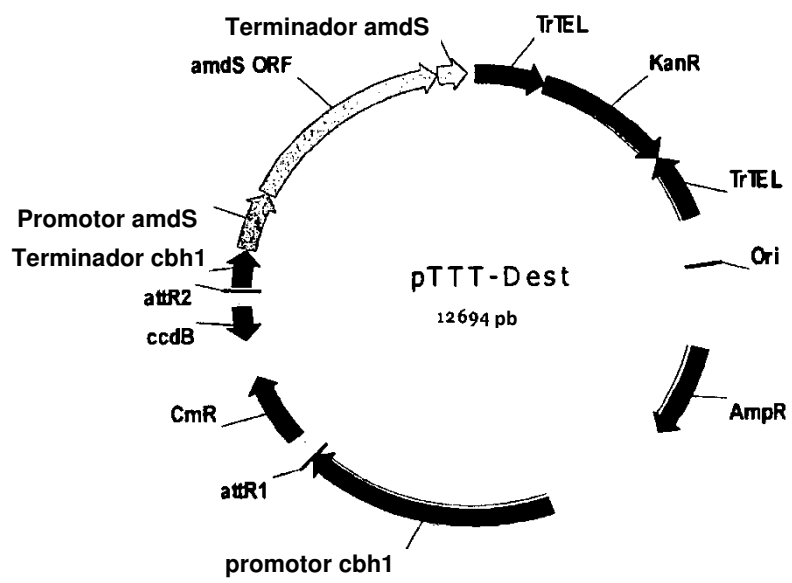


FIG. 4

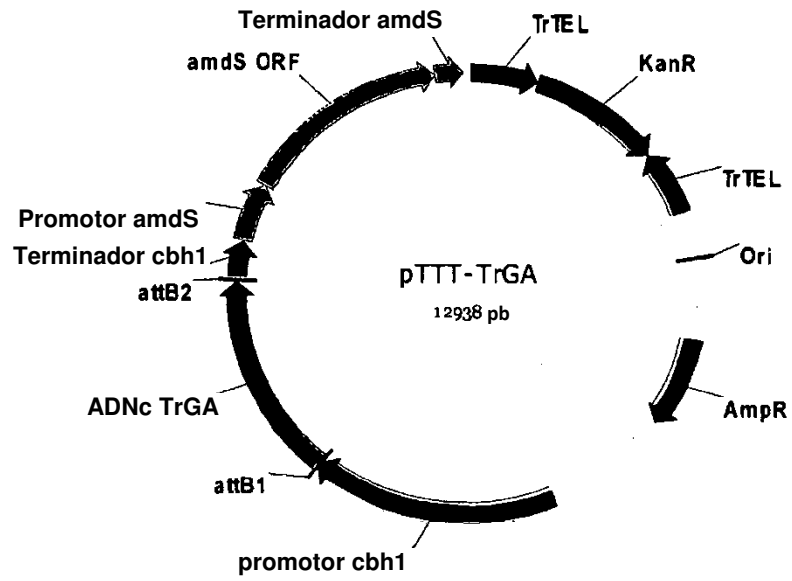


FIG. 5A

AaGA (1) -ATLDSWLSNEATVARTAILNNIGADGAWVSGADSGIIVVASPSTDNPDYF
AnGA (1) -ATLDSWLSNEATVARTAILNNIGADGAWVSGADSGIIVVASPSTDNPDYF
AoGA (1) QSDLNAFIEAQTPIAKQGYLNNIGADGKLVGAAAGIVYASPSKSNPDYF
HgGA (1) -AAVDTFINTEKPIAWNKL LANIGPNGKAAPGAAAGVVIASPSRTDPPYF
HvGA (1) --SVDDFINTQTPIALNNLLCNVGPDCRAFGTSAAGAVIASPSTTDPDYF
TrGA (1) --SVDDFISTETPIALNNLLCNVGPDCRAFGTSAAGAVIASPSTIDPDYF
* * * * *

AaGA (50) YTWTRDSGLVIKTLVDFRNGD TD-LLSTIENYISSQAI VQGISNPSGDL
AnGA (50) YTWTRDSGLVLKTLVDFRNGD TS-LLSTIENYISAQAI VQGISNPSGDL
AoGA (51) YTWTRDAGLTMEEYIEQFIGGDAT-LESTIQNYVDSQANEQAVSNPSGGL
HgGA (50) FTWTPDAALVLTGIIESLGHNYNT-----TLQQVSNPSGTF
HvGA (49) YMWTRDSALVFNKIVDRFTQQYDAGLQRRIEQYISAQVTLQGISNPSGSL
TrGA (49) YMWTRDSALVFNKILDRFTETYDAGLQRRIEQYITAOVTLQGLSNPSGSL
* * * *

AaGA (99) SSGG-LGEPKFNVD EYATYTG SWGRPQRDGPALRATAMIGFRQWLLDNGYT
AnGA (99) SSGAGLGEPKFNVD EYATYTG SWGRPQRDGPALRATAMIGFGQWLLDNGYT
AoGA (100) SDGSGLAEPKFY YNISQFTD SWGRPQRDGPALRASALIAYGNLSISSDKQ
HgGA (86) ADGSGLGEAKFNVDLTAFTGEWGRPQRDGPPLRAIALIQYAKWLIANGYK
HvGA (99) SDGSGLGEPKFELT L SQFTGNWGRPQRDGPALRAIALIGYSKWLINNNYQ
TrGA (99) ADGSGLGEPKFELT LKPF TGNWGRPQRDGPALRAIALIGYSKWLINNNYQ
* * * * *

AaGA (148) SAATEI VWPLVRNDLSYVAQYWNQ TGYDLWEEVNGSSFFTTI AVQHRLVE
AnGA (149) STATDI VWPLVRNDLSYVAQYWNQ TGYDLWEEVNGSSFFTTI AVQHRLVE
AoGA (150) SVVKANI WPIYQNDLSYVQYWNQ TGF DLWEEVQGSFFTTI AVQHKALVE
HgGA (136) STAKSVVWVVKNDL AYTAQYWNQ TGF DLWEEVPGSSFFTTI ASSHRALTE
HvGA (149) STVSNI IWPIVRNDL NYVAQYWNQ TGF DLWEEVNGSSFFTTI VANQHRALVE
TrGA (149) STVSNI WPIVRNDL NYVAQYWNQ TGF DLWEEVNGSSFFTTI VANQHRALVE
* * * * *

AaGA (198) GSAFATAV GSSCSWCDSQAPQILCYLQSF WTG--EYILANFDSS--RSGK
AnGA (199) GSAFATAV GSSCSWCDSQAPEILCYLQSF WTG--SFILANFDSS--RSGK
AoGA (200) GDAFAKALGEECQACS-VAPQILCHLQDFWNG--SAVLSNLPTNG-RSGL
HgGA (186) GAYLAAQLDTECP PCTTVAPQVLCFQAFWNSKGNV VVSTSTAGEYRSGK
HvGA (199) GATLAATLGQSGSTYSVAPQILCF LQRFWVS-GGYIDS NINTNEGR TGK
TrGA (199) GATLAATLGQSGSAYSSVAPQVLCF LQRFWVSSGGYVDSNINTNEGR TGK
* * * * *

AaGA (244) DTNTLLGSIHTFDPEAGCDDSTFQPCSPRALANHK EVVDSFRSIYTLNDG
AnGA (245) DANTLLGSIHTFDPEAACDDSTFQPCSPRALANHK EVVDSFRSIYTLNDG
AoGA (246) DTNSLLGSIHTFDPAACDDTTFQPCSSRALSNHKL VVDSFRSVYGINNG
HgGA (236) DANSILASIHNFDP EAGCDNLTFQPCSERALANHKAYVDSFRNLYA INKG
HvGA (248) DANSLLASIHTFDPSLGCDASTFQPCSDKALS NLKVVVDSFRSIYGVNKG
TrGA (249) DVNSVLTSIHTFDPNLGC DAGTFQPCSDKALS NLKVVVDSFRSIYGVNKG
* * * * *

AaGA (294) LSDSEAVAVGRYPKDSY YNGNPWF LCTLAAAEQLYDALYQWDKQGSLEIT
AnGA (295) LSDSEAVAVGRYPEDTY YNGNPWF LCTLAAAEQLYDALYQWDKQGSLEVT
AoGA (296) RGAGKAAAVGPYAEDTY QGGNPWYL TTVLVA AE LLYDALYQWDKQGVNVT
HgGA (286) IAQGGKAVAVGRYSE DVY YNGNPWY LANFAAAEQLYDAIYVWVKQGSITVT
HvGA (298) IPAGSAVAIGRYPEDVY FNGNPWY LAF AA AEQLYDSVYVWVKKTSITVT
TrGA (299) IPAGAAVAIGRYAEDVY YNGNPWY LAF AA AEQLYDAIYVWVKKTSITVT
* * * * *

FIG. 5B

```

AaGA (344) DVSLDFFQALYSDAATGTYSSSSSTYSSIVDAVKTFADGFVSIIVETHAAS
AnGA (345) DVSLDFFKALYSDAATGTYSSSSSTYSSIVDAVKTFADGFVSIIVETHAAS
AoGA (346) ETSLPFFKDLSSNVTTGSYAKSSSAYESLTSVAVKTYADGFISVVQEYTPD
HgGA (336) SVSLPFFRDLVSSVSTGTYSKSSSTFTNIVNAVKAYADGFIEVAAKYTPS
HvGA (348) STSSAFFQELVPGVAAGTYSSSQSTFTSIINAISTYADGFLSEAAKYVPA
TrGA (349) ATSLAFFQELVPGVTAGTYSSSSSTFTNIINAVSTYADGFLSEAAKYVPA
          *  **  *      *  *  *  *      *      ****

AaGA (394) NGSLSSEQYDKSDGDELSARDLTWSYAALLTANNRRNSVMPPSWGETSAS-
AnGA (395) NGSMSSEQYDKSDGEQLSARDLTWSYAALLTANNRRNSVVPASWGETSAS-
AoGA (396) GGALAEQYSRDQGTPVSASDLTWSYAFLSAVGRNRGTVPASWGSSTAN-
HgGA (386) NGALAEQYDRNTGKPDSAADLTWSYSAFLSAIDRRAGLVPPSWRASVAKS
HvGA (398) DGSLAEQFDRNTGTPLSAVHLTWSYASFLTAAARRAGVVPVPSWASSGAN-
TrGA (399) DGSLAEQFDRNSGTPLSALHLTWSYASFLTATARRAGIVPPSWANSSAS-
          *  **      *  **  *****  *  *  **      *  **  *

AaGA (443) SVPGTC
AnGA (444) SVPGTC
AoGA (445) AVPSQC
HgGA (436) QLPSTC
HvGA (447) TVPSSC
TrGA (448) TIPSTC
          •  *
    
```

FIG. 5C

gsldsflatetpialqgvlnnigpngadvagasagivvaspsrsdpdyfyswtrdaaltakylvdafia
gnkdleqtigeyisagaqvqtisnpsgdltggglgepkfnvetaftgpwgrpqrddgpalarataliaya
nylidngqastadeiipivqndlsyvtqywnsstfdlweevegssffttavqhralvegnalatrlnh
tcpncvsqapqvlcflqsywtgsyvlanfggsgrsgkdvnsilgsihtfdpaggcdstfqpcsarala
nhkvvt dsfrsvyavnsgiaegsavavgrypedvyqggnpwylataaaaelydaiyqwnkigsisitd
vslaffqdiypsaavgtynsgsstfndiisavqtyadgylsiiekytspdgslteqfsrsdgtplsasg
ltwsyaslltaaarqsivpaswgessassvpavcsatsatgpystatntawpssgsgpstttsvpctt
ptsvavtfdeivsttygetiylagsipelgnwspssaiplradaytssnplwyvtlnlpagtsfeykff
kktedgtivweddpnrstypaycgqttailddswq

FIG. 6

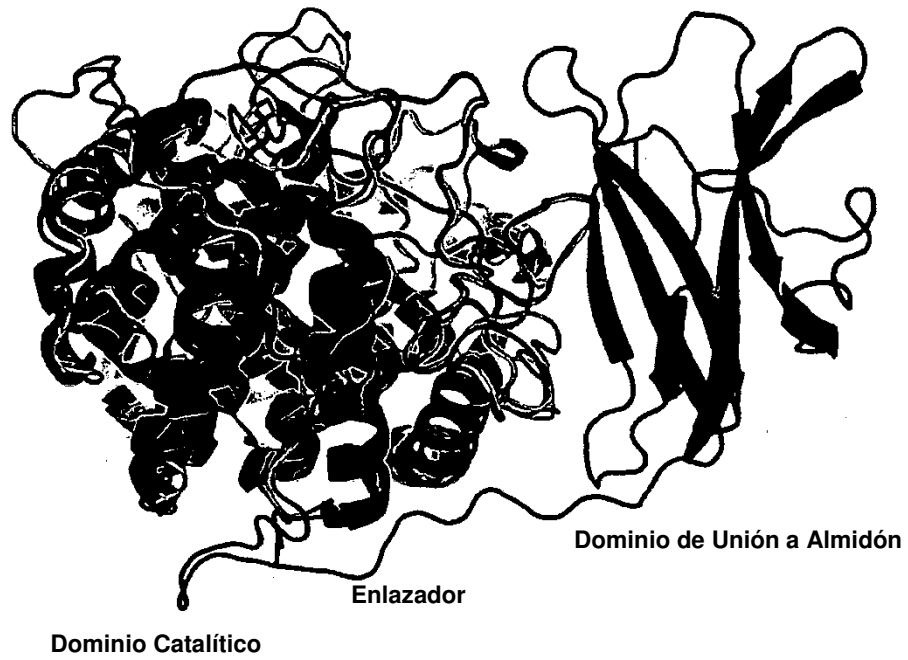


FIG. 7

