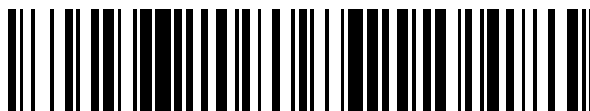


19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 538 396**

51 Int. Cl.:

C12P 21/06 (2006.01)

C12N 9/42 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **30.09.2010** **E 10821289 (5)**

97 Fecha y número de publicación de la concesión europea: **21.01.2015** **EP 2483415**

54 Título: **C1 β -glucosidasa recombinante para la producción de azúcares a partir de biomasa celulósica**

30 Prioridad:

30.09.2009 US 247379 P

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

19.06.2015

73 Titular/es:

**CODEXIS, INC. (100.0%)
200 Penobscot Drive
Redwood City, CA 94063, US**

72 Inventor/es:

**CLARK, LOUIS;
BAIDYARROY, DIPNATH y
SZABO, LORAND**

74 Agente/Representante:

IZQUIERDO BLANCO, María Alicia

ES 2 538 396 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).

C1 β -glucosidasa recombinante para la producción de azúcares a partir de biomasa celulósica**Descripción**

5 La invención se refiere a la expresión de una β -glucosidasa recombinante y su uso en la producción de azúcares fermentables a partir de biomasa celulósica.

ANTECEDENTES

10 La biomasa celulósica es un recurso renovable significativo para la generación de azúcares. La fermentación de estos azúcares puede producir productos finales comercialmente valiosos, incluidos biocombustibles y productos químicos actualmente derivados del petróleo. Aunque la fermentación de azúcares simples a etanol es relativamente sencilla, la conversión eficaz de biomasa celulósica en azúcares fermentables tales como la glucosa, es un reto. Véase, por ejemplo, Ladisch *et al.*, 1983, Enzyme Microb. Technol. 5:82. El material celulósico puede pretratarse químicamente, mecánicamente o de otras formas para aumentar la susceptibilidad de la celulosa a la hidrólisis. Tal pretratamiento puede ir seguido de la conversión enzimática de la celulosa en glucosa, celobiosa, celo-oligosacáridos y similares, utilizando enzimas que están especializadas en romper los enlaces β -1-4 glicosídicos de la celulosa. Estas enzimas son conocidas colectivamente como "celulasas".

20 Las celulasas se dividen en tres subcategorías de enzimas: 1,4- β -D-glucano glucanohidrolasa ("endoglucanasa" o "EG"); 1,4- β -D-glucano celobiohidrolasa ("exoglucanasa", "celobiohidrolasa" o "CBH") y β -D-glucósido glucosidolasa (" β -glucosidasa", "celobiasa" o "BG"). Las endoglucanasas atacan al azar las partes interiores y fundamentalmente las regiones amorfas de la celulosa, produciendo principalmente glucosa, celobiosa y celotriosa. Las exoglucanasas acortan progresivamente las moléculas de glucano uniéndose a los extremos del glucano y liberando principalmente unidades de celobiosa de los extremos del polímero de celulosa. Las β -glucosidasas dividen la celobiosa, un dímero de glucosa unido en β -1,4 soluble en agua, en dos unidades de glucosa. La producción eficaz de celulasas para su uso en el procesamiento de biomasa celulósica reduciría los costos y aumentaría la eficacia de producción de biocombustibles y otros compuestos con valor comercial. En el documento WO2008/008070 y en Gusakov *et al.*, 2007 (Carbohydrate research, 343:48-55) se describe una β -glucosidasa de *Chrysosporium lucknowense*.

RESUMEN DE LA INVENCION

35 La invención proporciona un polinucleótido que codifica un polipéptido β -glucosidasa recombinante, en el que dicho polipéptido:

a) es al menos un 90% idéntico a la SEQ ID NO: 3 y comprende una secuencia amino terminal que se expone como IESRK (SEQ ID NO: 5), ESRK (SEQ ID NO: 29) o SRK; y/o es una proteína secretada derivada de una preproteína, en el que la preproteína es al menos un 90% idéntica a la SEQ ID NO: 4 y comprende una secuencia amino terminal que comprende la SEQ ID NO: 6 o NO: 7; y/o
40 b) es al menos un 90% idéntico a la SEQ ID NO: 3 y comprende al menos uno, al menos dos, o tres residuos de aminoácidos seleccionados de entre ácido aspártico en la posición 358 (358ASP), glutamina en la posición 381 (381 GLN) y ácido glutámico en la posición 385 (385GLU), en los que los números de posición de los aminoácidos corresponden a la SEQ ID NO: 4.

45 La invención también proporciona vectores de expresión que comprenden el polinucleótido de la invención, unido operativamente a un promotor heterólogo.

50 La invención también proporciona un método de producción de un polipéptido β -glucosidasa secretado, que comprende cultivar una célula que comprende un polinucleótido de la invención, unido operativamente a un promotor heterólogo, o un vector de expresión de la invención, en condiciones en las que se produce β -glucosidasa.

55 La invención también proporciona una célula hospedadora recombinante que comprende un polinucleótido de la invención que codifica un polipéptido β -glucosidasa recombinante, en el que la secuencia codificante no comprende intrones.

60 La invención también proporciona una composición que comprende una célula que comprende una secuencia polinucleotídica que codifica un polipéptido β -glucosidasa recombinante unida operativamente a un promotor heterólogo, en el que dicho polipéptido:

a) es al menos un 90% idéntico a la SEQ ID NO: 3 y comprende una secuencia amino terminal que se expone como IESRK (SEQ ID NO: 5), ESRK (SEQ ID NO: 29) o SRK; y/o es una proteína secretada derivada de una preproteína, en el que la preproteína es al menos un 90% idéntica a la SEQ ID NO: 4 y comprende una secuencia amino terminal que comprende la SEQ ID NO: 6 o NO: 7; y/o
65 b) es al menos un 90% idéntico a la SEQ ID NO: 3 y comprende al menos uno, al menos dos, o tres residuos de aminoácidos seleccionados de entre ácido aspártico en la posición 358 (358ASP), glutamina en la posición 381

(381 GLN) y ácido glutámico en la posición 385 (385GLU), en los que los números de posición de los aminoácidos corresponden a la SEQ ID NO: 4, y una proteína β -glucosidasa secretada por la célula.

La invención también proporciona un método de producción de glucosa a partir de celobiosa, que comprende poner en contacto celobiosa con una célula recombinante que comprende una secuencia polinucleotídica que codifica un polipéptido β -glucosidasa recombinante, unida operativamente a un promotor heterólogo, en el que dicho polipéptido:

a) es al menos un 90% idéntico a la SEQ ID NO: 3 y comprende una secuencia amino terminal que se expone como IESRK (SEQ ID NO: 5), ESRK (SEQ ID NO: 29) o SRK; y/o es una proteína secretada derivada de una preproteína, en el que la preproteína es al menos un 90% idéntica a la SEQ ID NO: 4 y comprende una secuencia amino terminal que comprende la SEQ ID NO: 6 o NO: 7; y/o

b) es al menos un 90% idéntico a la SEQ ID NO: 3 y comprende al menos uno, al menos dos, o tres residuos de aminoácidos seleccionados de entre ácido aspártico en la posición 358 (358ASP), glutamina en la posición 381 (381 GLN) y ácido glutámico en la posición 385 (385GLU), en los que los números de posición de los aminoácidos corresponden a la SEQ ID NO: 4, en condiciones en las que la β -glucosidasa es expresada y secretada por la célula y dicha celobiosa es convertida enzimáticamente en glucosa por dicha β -glucosidasa.

En un aspecto, la invención proporciona un método de producción de un polipéptido β -glucosidasa secretado, mediante el cultivo de una célula que comprende una secuencia polinucleotídica que codifica C1 β -glucosidasa (SEQ ID NO: 3) o una variante enzimáticamente activa de la misma según la invención, unida operativamente a un promotor heterólogo, en condiciones en las cuales se produce β -glucosidasa. La célula puede ser una célula de la cepa C1. Como alternativa, la célula puede ser distinta de una célula de la cepa C1. En algunos casos, el promotor heterólogo es el promotor de C1 celobiohidrolasa 1a (CBH1a). En algunos casos, la secuencia polinucleotídica no codifica una o más de entre la SEQ ID NO: 11, la SEQ ID NO: 30 o la SEQ ID NO: 31. Opcionalmente, la secuencia polinucleotídica codifica la SEQ ID NO: 4.

En un aspecto, la invención proporciona un método de producción de glucosa a partir de celobiosa, poniendo en contacto celobiosa con una célula recombinante que comprende una secuencia polinucleotídica que codifica C1 β -glucosidasa (SEQ ID NO: 3) o una variante enzimáticamente activa de la misma según la invención, unida operativamente a un promotor heterólogo, en condiciones en las que la β -glucosidasa es expresada y secretada por la célula y dicha celobiosa es convertida enzimáticamente en glucosa por dicha β -glucosidasa. La célula puede ser una célula de la cepa C1. Como alternativa, la célula puede ser distinta de una célula de la cepa C1. En algunos casos, el promotor heterólogo es el promotor de C1 celobiohidrolasa 1a (CBH1a). En algunos casos, la secuencia polinucleotídica no codifica una o más de entre la SEQ ID NO: 11, la SEQ ID NO: 30 o la SEQ ID NO: 31. Opcionalmente, el polinucleótido codifica la SEQ ID NO: 4.

También se contempla un método de conversión de un sustrato de biomasa en un azúcar fermentable, combinando la β -glucosidasa recombinante generada según la invención, con el sustrato de biomasa (por ejemplo, celobiosa) en condiciones adecuadas para la producción del azúcar fermentable. En algunos casos, el método incluye la etapa de recuperar la β -glucosidasa a partir del medio en el que se cultiva la célula. En un aspecto, se proporciona una composición que comprende la β -glucosidasa recombinante.

En un aspecto, la invención proporciona una célula hospedadora recombinante que comprende una secuencia polinucleotídica que codifica C1 β -glucosidasa (SEQ ID NO: 3), o una variante de la misma según la invención, unida operativamente a un promotor heterólogo. Opcionalmente, el polinucleótido codifica la SEQ ID NO: 4. En una forma de realización, la célula hospedadora recombinante expresa al menos otra enzima celulasa recombinante. También se contempla un método de conversión de un sustrato de biomasa en un azúcar fermentable, combinando la célula recombinante con el sustrato de biomasa en condiciones adecuadas para la producción del azúcar fermentable.

También se proporciona una célula que comprende una secuencia polinucleotídica que codifica C1 β -glucosidasa (SEQ ID NO: 3), o una variante enzimáticamente activa de la misma como se describe en el presente documento, en el que la secuencia codificante no comprende un intrón, y, opcionalmente, está operativamente unida a un promotor heterólogo, que puede ser un promotor de C1 celobiohidrolasa 1a (CBH1a). En una forma de realización, la célula expresa al menos una enzima celulasa recombinante adicional.

También se proporciona una composición que contiene una célula que comprende una secuencia polinucleotídica que codifica C1 β -glucosidasa (SEQ ID NO: 3) o una variante enzimáticamente activa de la misma, tal como se describe en el presente documento, y una proteína β -glucosidasa secretada por la célula. En un aspecto, una célula recombinante descrita en el presente documento se pone en contacto con un sustrato de biomasa (por ejemplo, celobiosa) y el sustrato se convierte en un azúcar fermentable (por ejemplo, glucosa).

BREVE DESCRIPCIÓN DE LAS FIGURAS

La Figura 1 muestra las secuencias genómica (SEQ ID NO: 1) y de ADNc (SEQ ID NO: 2) que codifican la proteína C1 β -glucosidasa 1 (BGL1). La secuencia genómica está por encima de y alineada con la secuencia de ADNc inferior. Los huecos en la secuencia inferior muestran las posiciones de los intrones.

La Figura 2 muestra la secuencia de ADNc de BGL1 (SEQ ID NO: 2) y la secuencia proteica correspondiente. El péptido señal predicho y la secuencia nucleotídica correspondiente están en negrita. La secuencia de aminoácidos que incluye la secuencia del péptido señal es la SEQ ID NO: 4. La secuencia de aminoácidos que no incluye la secuencia del péptido señal es la SEQ ID NO: 3.

La Figura 3 ilustra la estructura del exón del gen *bgl1* (flecha inferior) en comparación con una estructura de exón descrita anteriormente (flecha superior).

La Figura 4 muestra la secuencia (promotora) 5' del gen C1 *cbh1a* (SEQ ID NO: 8).

La Figura 5 muestra una secuencia polinucleotídica artificial que codifica una proteína secretada C1 β -glucosidasa 1 (BGL1) (SEQ ID NO: 9). La secuencia tiene una preferencia de uso de codones para optimizar la expresión en *Saccharomyces cerevisiae*.

DESCRIPCIÓN DETALLADA

I. DEFINICIONES

Las siguientes definiciones se proporcionan para ayudar al lector. A menos que se defina lo contrario, todos los términos de la técnica pretenden tener los significados que entienden comúnmente los expertos en materia de microbiología y biología molecular. En algunos casos, los términos con significados entendidos comúnmente se definen en el presente documento para mayor claridad y/o para una fácil referencia, y la inclusión de tales definiciones en el presente documento no debe interpretarse necesariamente que represente una diferencia sustancial con respecto a la definición del término en su acepción general en la técnica.

El término "celulasa" se refiere a una categoría de enzimas capaces de hidrolizar la celulosa (enlaces β -1,4-glucano o β -D-glucosídico) a oligosacáridos más cortos, celobiosa y/o glucosa.

El término " β -glucosidasa" o "celobiasa" que se utiliza indistintamente en el presente documento se refiere a una β -D-glucósido glucohidrolasa que cataliza la hidrólisis de un dímero de azúcar, incluido pero no limitado a celobiosa, con la liberación de un monómero de azúcar correspondiente. En una forma de realización, una β -glucosidasa es una β -glucosidasa glucohidrolasa de la clasificación E.C. 3.2.1.21 que cataliza la hidrólisis de la celobiosa a glucosa. Algunas de las β -glucosidasas tienen la capacidad de hidrolizar también P-D-galactósidos, β -L-arabinósidos y/o β -D-fucósidos y, adicionalmente, algunas β -glucosidasas pueden actuar sobre sustratos con enlaces α -1,4 tales como el almidón. Puede medirse la actividad β -glucosidasa mediante métodos bien conocidos en la técnica, incluidos los ensayos que se describen más adelante en el presente documento.

La expresión "polipéptido β -glucosidasa" se refiere en el presente documento a un polipéptido que tiene actividad β -glucosidasa.

La expresión "polinucleótido β -glucosidasa" se refiere a un polinucleótido que codifica un polipéptido que tiene actividad β -glucosidasa.

"Actividad celulolítica" abarca la actividad exoglucanasa (CBH), la actividad endoglucanasa (EG) y/o la actividad β -glucosidasa.

El término "exoglucanasa", "exocelobiohidrolasa" o "CBH" se refiere a un grupo de enzimas celulasas clasificadas como E.C. 3.2.1.91. Estas enzimas hidrolizan la celobiosa desde el extremo reductor o no reductor de la celulosa.

El término "endoglucanasa" o "EG" se refiere a un grupo de enzimas celulasas clasificadas como E.C. 3.2.1.4. Estas enzimas hidrolizan los enlaces β -1,4 glucosídicos internos de la celulosa.

Tal como se utiliza en el presente documento, el término "aislado" se refiere a un ácido nucleico, polinucleótido, polipéptido, proteína, u otro componente que está parcial o completamente separado de los componentes con los que está normalmente asociado (otras proteínas, ácidos nucleicos, células, reactivos sintéticos, etc.).

La expresión "de tipo silvestre" tal como se aplica a un polipéptido (proteína) se refiere a un polipéptido (proteína) expresado por un microorganismo natural tal como bacterias u hongos filamentosos. Tal como se aplica a un microorganismo, la expresión "de tipo silvestre" se refiere al microorganismo no recombinante original.

Un ácido nucleico (tal como un polinucleótido), un polipéptido o una célula es "recombinante" cuando es artificial o manipulado, o se deriva de o contiene una proteína o ácido nucleico artificial o manipulado. Un ácido nucleico recombinante, o de forma equivalente, polinucleótido, es aquel que se inserta en una ubicación heteróloga

de manera que no está asociado con secuencias nucleotídicas que normalmente flanquean el ácido nucleico tal como se encuentra en la naturaleza. Por ejemplo, un polinucleótido que se inserta en un vector o en cualquier otra ubicación heteróloga, por ejemplo, en un genoma de un organismo recombinante, de manera que no esté asociado con secuencias nucleotídicas que normalmente flanquean el polinucleótido tal como se encuentra en la naturaleza, es un polinucleótido recombinante. Los ejemplos de ácidos nucleicos recombinantes incluyen una secuencia de ADN que codifica una proteína que (i) está unida operativamente a un promotor heterólogo y/o (ii) codifica un polipéptido de fusión con una secuencia proteica y una secuencia de péptido señal heterólogo. Una proteína expresada *in vitro* o *in vivo* a partir de un polinucleótido recombinante es un ejemplo de polipéptido recombinante. Asimismo, una secuencia polinucleotídica que no aparece en la naturaleza, por ejemplo una variante de un gen natural, es recombinante.

Los ácidos nucleicos "hibridan" cuando se asocian, por lo general en solución. Los ácidos nucleicos hibridan debido a diversas fuerzas fisicoquímicas bien caracterizadas, tales como enlaces de hidrógeno, exclusión de disolvente, apilamiento de bases y similares. Tal como se utiliza en el presente documento, la expresión "condiciones de lavado de hibridación rigurosas" en el contexto de los experimentos de hibridación de ácidos nucleicos, tal como hibridaciones Southern y Northern, son dependientes de la secuencia, y son diferentes bajo diferentes parámetros ambientales. Se encuentra una extensa guía para la hibridación de ácidos nucleicos en Tijssen (1993) "Laboratory Techniques in biochemistry and Molecular Biology-Hybridization with Nucleic Acid Probes", parte I, capítulo 2 (Elsevier, Nueva York). Para polinucleótidos de al menos 100 nucleótidos de longitud, las condiciones de rigurosidad baja a muy alta se definen de la siguiente manera: prehibridación e hibridación a 42°C en 5xSSPE, SDS al 0,3%, 200 µg/ml de ADN de esperma de salmón cortado y desnaturalizado, y, formamida al 25% para rigurosidades bajas, formamida al 35% para rigurosidades media y media-alta, o formamida al 50% para rigurosidades alta y muy alta, siguiendo los procedimientos de transferencia de Southern convencionales. Para polinucleótidos de al menos 100 nucleótidos de longitud, el material transportador se lava finalmente tres veces, cada una durante 15 minutos, utilizando 2xSSC, SDS al 0,2% al menos a 50°C (baja rigurosidad), al menos a 55°C (rigurosidad media), al menos a 60°C (rigurosidad media-alta), al menos a 65°C (alta rigurosidad) y al menos a 70°C (rigurosidad muy alta).

El término "cultivar" o "cultivo" se refiere a hacer crecer una población de células microbianas en condiciones adecuadas en un medio líquido o sólido. En algunas formas de realización, el cultivo se refiere a la bioconversión fermentativa de un sustrato celulósico hasta un producto final.

La expresión "poner en contacto" se refiere a colocar una enzima respectiva lo suficientemente cerca de un sustrato respectivo para permitir que la enzima convierta el sustrato en un producto. Los expertos en la materia reconocerán que mezclar una solución de la enzima con el sustrato respectivo logrará el contacto.

Tal como se utiliza en el presente documento, el término "transformado" o "transformación" utilizado en referencia a una célula, se refiere a que una célula tiene una secuencia de ácido nucleico no original integrada en su genoma o tiene un plásmido episómico que se mantiene a través de múltiples generaciones.

El término "introducido" en el contexto de la inserción de una secuencia de ácido nucleico en una célula se refiere a una célula transfectada, transducida o transformada (conjuntamente "transformada") y procariota en la que el ácido nucleico se incorpora en el genoma de la célula.

Tal como se utiliza en el presente documento, "C1" se refiere a una cepa fúngica descrita por Garg, A., 1966, "An addition to the genus *Chrysosporium corda*", Mycopathologia 30:3-4. "*Chrysosporium lucknowense*" incluye las cepas descritas en las patentes de EE.UU. N° 6.015.707; 5.811.381 y 6.573.086; las publicaciones de patente de EE.UU. N° 2007/0238155, US 2008/0194005, US 2009/0099079; las publicaciones de patente internacional N° WO 2008/073914 y WO 98/15633, y cualquiera de sus derivados, e incluye, sin limitación, *Chrysosporium lucknowense* Garg 27K, VKM-F 3500 D (número de registro VKM F-3500-D), la cepa C1 UV13-6 (número de registro VKM F-3632 D), la cepa C1 NG7C-19 (número de registro VKM F-3633 D) y la cepa C1 UV18-25 (VKM F-3631 D), todas las cuales han sido depositadas en la All-Russian Collection of Microorganisms de la Academia Rusa de las Ciencias (VKM), Bakhurhina St. 8, Moscú, Rusia, 113184, y cualquiera de sus derivados. Aunque se ha descrito inicialmente como *Chrysosporium lucknowense*, C1 puede considerarse actualmente una cepa de *Myceliophthora thermophila*. Otras cepas C1 incluyen las células depositadas con los números de registro ATCC 44006, CBS (Centraalbureau voor Schimmelcultures) 122188, CBS 251.72, CBS 143.77, CBS 272.77 y VKM F-3500D. Las cepas C1 ejemplares incluyen organismos modificados en los que se han delecionado una o más secuencias o genes endógenos y/o se han introducido una o más secuencias o genes heterólogos. Los derivados incluyen UV18#100fΔalpl, UV18#100f, Δpyr5 Δalpl, UV18#100.f Δalpl Δpep4 Δalp2, UV18#100.f Δpyr5 Δalpl Δpep4 Δalp2 y UV18#100.f Δpyr4 Δjyr5 Δalp1 Δpep4 Δalp2, como se describe en el documento WO2008073914.

Cuando se dice que dos elementos, por ejemplo, un promotor y una secuencia codificante, están "unidos operativamente", se refiere a que la yuxtaposición de los dos les permite ser funcionalmente activos. Por lo tanto, un promotor u otra secuencia de control está "unido/a operativamente" a una secuencia codificante cuando el promotor o secuencia de control está colocado en una posición con respecto a la secuencia codificante de la secuencia de ADN de manera que controle la transcripción de la secuencia codificante.

Tal como se utiliza en el presente documento, los términos "idéntico" o porcentaje de "identidad", en el contexto de la descripción de dos o más secuencias polinucleotídicas o polipeptídicas, se refieren a dos o más secuencias que son iguales o tienen un porcentaje especificado de nucleótidos o residuos que son iguales cuando se comparan y alinean para una correspondencia máxima sobre una ventana de comparación o región indicada (por ejemplo, toda la longitud de la secuencia de referencia), tal como se mide mediante alineamiento manual e inspección visual o utilizando uno de los siguientes algoritmos de comparación de secuencias. Los métodos de alineamiento de secuencias para la comparación son bien conocidos en la técnica. En una forma de realización, se alinean dos secuencias (por ejemplo, manualmente) para la identidad óptima sin introducir huecos en ninguna de las dos secuencias, o con no más de 1, 2 ó 3 huecos de por lo general menos de 10 residuos cada uno.

Una "ventana de comparación", tal como se utiliza en el presente documento, incluye la referencia a un segmento de cualquiera del número de posiciones contiguas seleccionado del grupo que consiste en 20 a 500, normalmente de aproximadamente 50 a aproximadamente 300, también de aproximadamente 50 a 250, y también de aproximadamente 100 a aproximadamente 200 en la que puede compararse una secuencia con una secuencia de referencia del mismo número de posiciones contiguas después de haber alineado de manera óptima las dos secuencias. Como se ha señalado, en algunas formas de realización la comparación es entre toda la longitud de las dos secuencias, o, si una secuencia es un fragmento de la otra, toda la longitud de la más corta de las dos secuencias.

Para la comparación de secuencias, por lo general una secuencia hace de secuencia de referencia, con la que se comparan las secuencias de ensayo. Cuando se utiliza un algoritmo de comparación de secuencias, se introducen en un ordenador las secuencias de ensayo y de referencia, se asignan las coordenadas de subsecuencia, en caso necesario, y se asignan los parámetros de programa de algoritmo de secuencia. Pueden utilizarse parámetros de programa por defecto, o pueden asignarse parámetros alternativos. Los ejemplos de algoritmos que son adecuados para determinar el porcentaje de identidad de secuencia y de similitud de secuencia son los algoritmos BLAST y BLAST 2.0, que se describen en Altschul *et al.*, (1990) J. Mol. Biol. 215: 403-410 y Altschul *et al.* (1977) Nucleic Acids Res. 25: 3389-3402, respectivamente. El software para realizar los análisis BLAST está a disposición del público en la página web del National Center for Biotechnology Information, ncbi.nlm.nih.gov. El algoritmo BLAST también realiza un análisis estadístico de la similitud entre dos secuencias (véase, por ejemplo, Karlin y Altschul, Proc. Natl. Acad. Sci. EE.UU. 90: 5873-5787 (1993)). Una medida de similitud proporcionada por el algoritmo BLAST es la probabilidad de suma más pequeña ($P(N)$), que proporciona una indicación de la probabilidad por la cual se produciría por casualidad una coincidencia entre dos secuencias nucleotídicas o de aminoácidos. Por ejemplo, un ácido nucleico se considera similar a una secuencia de referencia si la probabilidad de suma más pequeña en una comparación del ácido nucleico de ensayo con el ácido nucleico de referencia es inferior a aproximadamente 0,2, más preferentemente inferior a aproximadamente 0,01, y lo más preferentemente inferior a aproximadamente 0,001.

Cuando se utiliza en el presente documento, la expresión "secuencia codificante" pretende incluir una secuencia nucleotídica que especifica directamente la secuencia de aminoácidos de su producto proteico. Los límites de la secuencia codificante vienen determinados generalmente por un marco de lectura abierto, que normalmente empieza con el codón de inicio ATG. La secuencia codificante incluye por lo general un ADN, ADNc, y/o secuencia nucleotídica recombinante.

Un promotor u otra secuencia de control es "heterólogo", cuando está unido operativamente a una secuencia que codifica una secuencia proteica con la que el promotor no está asociado en la naturaleza. Por ejemplo, en un constructo recombinante en el que el promotor de C1 Cbh1a está unido operativamente a una secuencia codificante de proteína distinta de la secuencia codificante de la proteína C1 Cbh1a, el promotor es heterólogo. Por ejemplo, en un constructo que comprende un promotor de C1 Cbh1a unido operativamente a una secuencia codificante de C1 β -glucosidasa 1, el promotor es heterólogo.

Tal como se utiliza en el presente documento, el término "expresión" incluye cualquier etapa implicada en la producción del polipéptido, incluida pero no limitada a, transcripción, modificación postranscripcional, traducción, modificación postraducciona y secreción.

La expresión "vector de expresión" se refiere en el presente documento a una molécula de ADN, lineal o circular, que comprende un segmento que codifica un polipéptido de la invención, y que está unida operativamente a segmentos adicionales que proporcionan su transcripción.

Un polipéptido es "enzimáticamente activo" cuando tiene actividad β -glucosidasa.

El término "preproteína" se refiere a una proteína secretada con una región de péptido señal amino terminal fijada. El péptido señal es escindido de la preproteína por una peptidasa señal antes de la secreción para dar como resultado la proteína "madura" o "secretada".

Tal como se utiliza en el presente documento, un "codón de inicio" es el codón ATG que codifica el primer residuo de aminoácido (metionina) de una proteína.

II. INTRODUCCIÓN

La C1 fúngica produce diversas enzimas que actúan concertadamente para catalizar la descomposición y la hidrólisis de la celulosa para producir azúcares solubles. El genoma C1 se ha secuenciado al menos parcialmente, como se indica en las publicaciones de patente de Estados Unidos US 2007/0238155, US 2008/0194005 y US 2009/0099079. La figura 13 de la publicación de patente de EE.UU. 2007/0238155 proporciona información de la secuencia para el gen de C1 β -glucosidasa 1 (*bgl1*) y la proteína codificada (BGL1). Como se analiza en los Ejemplos 1 y 2, más adelante, se ha descubierto recientemente que esta secuencia del gen *bgl1* publicada anteriormente incluía errores de secuenciación. Por otra parte, inesperadamente, la estructura del exón de la secuencia *bgl1* publicada es incorrecta. A consecuencia de estos errores, la secuencia de la proteína C1 BGL1 disponible incluye errores, particularmente en el péptido señal y en el extremo amino terminal de la proteína BGL1 madura (es decir, secretada).

Tal como se analiza con mayor detalle más adelante, la secuencia correcta de la preproteína BGL1 se expone en el presente documento como SEQ ID NO: 3 y la secuencia correcta de la proteína secretada BGL1 se expone como SEQ ID NO: 4. La Tabla 1 presenta un resumen de las secuencias a las que frecuentemente se hace referencia en la presente descripción.

TABLA 1

SEQ ID NO:	Descripción	
1	Secuencia del gen C1 <i>bgl</i>	nucleótido
2	Secuencia de ADNc de C1	nucleótido
3	Proteína secretada BGL	aminoácido
4	Preproteína BGL	aminoácido
5	Extremo amino terminal de la proteína secretada BGL (IESRK)	aminoácido
6	Péptido señal de BGL (MKAAALSCFLFGSTLAVAGA)	aminoácido
7	Extremo amino terminal de la preproteína BGL (MKAAALSCFLFGSTLAVAGAIESRK)	aminoácido
8	Secuencia promotora de C1 Cbh1a	nucleótido
9	Polinucleótido artificial que codifica la proteína C1 BGL1 (codón optimizado para la expresión en <i>S. cerevisiae</i>)	nucleótido
13-17	Secuencia flanqueante en sentido 5' respecto al codón de inicio	nucleótido
18	Secuencia intrónica de <i>bgl1</i>	nucleótido
19	Fragmento de la secuencia intrónica de <i>bgl1</i>	nucleótido
20	Secuencia del exón 1 de <i>bgl1</i>	nucleótido
21	Exón 1 e intrón 1 de <i>bgl1</i>	nucleótido

La preproteína y la proteína secretada C1 β -glucosidasa, y los polinucleótidos que las codifican, pueden utilizarse en diversas aplicaciones en las que se desea la actividad β -glucosidasa, tales como las que se describen más adelante en el presente documento. En aras de la simplicidad, y como resultará evidente a partir del contexto, las referencias a la "proteína C1 BGL1" y similares pueden utilizarse para referirse tanto a la forma madura (secretada) de la proteína como a la preproteína.

En una forma de realización, una secuencia que codifica una proteína C1 BGL1 descrita en el presente documento está unida operativamente a un promotor no asociado con BGL1 en la naturaleza (es decir, un promotor heterólogo), para, por ejemplo, mejorar la eficacia de la expresión de la proteína BGL1 cuando se expresa en una célula hospedadora. En una forma de realización, la célula hospedadora es un hongo, tal como un hongo filamentoso. En una forma de realización, la célula hospedadora es una célula de la cepa C1. En una forma de realización, la célula hospedadora es C1 y el promotor es el promotor de C1 Cbh1a. Ventajosamente, las células C1 cultivadas transfectadas con un vector que comprende una secuencia que codifica C1 β -glucosidasa unida operativamente a un promotor de Cbh1a producía aproximadamente 35 veces mayor actividad β -glucosidasa que la de las células de control que expresaban *bgl1* endógeno.

El sistema de expresión de C1 β -glucosidasa que se describe en el presente documento es particularmente útil para la producción de azúcares fermentables a partir de biomasa celulósica. En un aspecto, la invención se refiere a un método de producción de glucosa poniendo en contacto una composición que comprende celobiosa con una C1 β -glucosidasa expresada de manera recombinante en condiciones en las que la celobiosa se convierte enzimáticamente en glucosa. Puede ponerse en contacto proteína β -glucosidasa recombinante purificada o

parcialmente purificada con la celobiosa. Como alternativa, pueden ponerse en contacto células hospedadoras recombinantes que expresan β -glucosidasa con celobiosa. En un aspecto de la presente invención, dicho "poner en contacto" comprende cultivar una célula hospedadora recombinante en un medio que contenga celobiosa producida a partir de materia prima celulósica, en el que la célula recombinante comprende una secuencia codificante de C1 β -glucosidasa unida operativamente a un promotor heterólogo.

En otro aspecto de la invención, el péptido señal de C1 β -glucosidasa (SEQ ID NO: 6) puede fusionarse con el extremo amino terminal de un polipéptido que no sea una C1 β -glucosidasa (es decir, un polipéptido "heterólogo") para mejorar la secreción, la estabilidad u otras propiedades del polipéptido cuando se expresa en una célula hospedadora, por ejemplo, una célula fúngica tal como una célula C1.

En las siguientes secciones se describen diversos aspectos de la invención.

III. PROPIEDADES DE LAS PROTEÍNAS β -GLUCOSIDASA PARA SU USO EN LOS MÉTODOS DE LA INVENCION

En un aspecto, la descripción proporciona un método para expresar una proteína β -glucosidasa mediante el cultivo de una célula hospedadora que comprende un vector que comprende una secuencia de ácido nucleico que codifica C1 BGL1 de tipo silvestre (SEQ ID NO: 3) unida operativamente a un promotor heterólogo, en condiciones en las que se expresa la proteína β -glucosidasa, una variante de la misma, o un fragmento enzimáticamente activo de la misma. Generalmente, la proteína expresada comprende un péptido señal, que puede ser la SEQ ID NO: 6 (el péptido señal de C1 BGL1) o puede ser un péptido señal diferente. En una forma de realización, la proteína no comprende una o más de las secuencias MQLPAAQWLLTPAKASL (SEQ ID NO: 11), ADNHR (SEQ ID NO: 30) o MQLPAAQWLLTPAKASLADNHR (SEQ ID NO: 31).

En algunas formas de realización, el polipéptido BGL1 incluye secuencias adicionales que no modifican la actividad codificada de una β -glucosidasa. Por ejemplo, la β -glucosidasa puede estar unida a un epítopo de identificación o a otra secuencia útil para la purificación de la β -glucosidasa.

En algunas formas de realización, la proteína BGL1 es una variante que se diferencia de la SEQ ID NO: 3 en una o más posiciones. Por ejemplo, en algunas formas de realización, la secuencia de ácido nucleico codifica una variante de BGL1 enzimáticamente activa que se diferencia de la SEQ ID NO: 3 ó 4 en una o más posiciones, y que:

a) es sustancialmente idéntica a la SEQ ID NO: 3 (proteína secretada) y/o es sustancialmente idéntica a la SEQ ID NO: 4 (preproteína);
y

b) comprende una secuencia amino terminal que se expone como IESRK (SEQ ID NO: 5), ESRK (SEQ ID NO: 29) o SRK; y/o codifica una preproteína con una secuencia amino terminal que comprende la SEQ ID NO: 6 o NO: 7; y/o comprende al menos uno, al menos dos, o tres de los siguientes residuos de aminoácidos: ácido aspártico en la posición 358 (358ASP), glutamina en la posición 381 (381 GLN) y ácido glutámico en la posición 385 (385GLU) en los que los números de posición corresponden a la SEQ ID NO: 4.

En una forma de realización, la proteína variante de BGL1 tiene una identidad de secuencia sustancial con la SEQ ID NO: 3 o la SEQ ID NO: 4. Por ejemplo, la variante de BGL1 puede tener una secuencia de aminoácidos que es al menos aproximadamente un 80% idéntica, al menos aproximadamente un 81% idéntica, al menos aproximadamente un 82% idéntica, al menos aproximadamente un 83% idéntica, al menos aproximadamente un 84% idéntica, al menos aproximadamente un 85% idéntica, al menos aproximadamente un 86% idéntica, secuencia que es al menos aproximadamente un 87% idéntica, al menos aproximadamente un 88% idéntica, al menos aproximadamente un 89% idéntica, al menos aproximadamente un 90% idéntica, al menos aproximadamente un 91% idéntica, al menos aproximadamente un 92% idéntica, al menos aproximadamente un 93% idéntica, secuencia que es al menos aproximadamente un 94% idéntica, al menos aproximadamente un 95% idéntica, al menos aproximadamente un 96% idéntica, al menos aproximadamente un 97% idéntica, al menos aproximadamente un 98% idéntica, o al menos aproximadamente un 99% idéntica a la SEQ ID NO: 3 o a la SEQ ID NO: 4.

Las variantes de BGL1 ejemplares pueden tener otras inserciones, deleciones y/o sustituciones (incluidas una o más sustituciones conservadoras), tal como se describe más adelante en el presente documento, sin dejar de ser sustancialmente idénticas a la proteína de tipo silvestre. Pueden aplicarse métodos de evolución dirigida y mutagénesis a los polinucleótidos que codifican una proteína C1 BGL1 para obtener variantes enzimáticamente activas con propiedades deseables.

Los métodos adecuados de evolución dirigida y mutagénesis son bien conocidos en la técnica. Por ejemplo, se han utilizado técnicas de recombinación *in vitro* tales como el barajado de ADN, el proceso de extensión escalonada (STEP), la generación aleatoria de quimeras en moldes transitorios (RACHITT), el truncamiento iterativo para la creación de enzimas híbridas (ITCHY), la extensión recombinada en moldes truncados (RETT), y otros para producir proteínas con las propiedades deseadas. En general, los polipéptidos se expresan a partir de secuencias

polinucleotídicas mutagenizadas y se ensayan para la actividad beta-galactosidasa (véase más adelante) y/u otras propiedades deseables.

Pueden generarse y cribarse genotecas de estas variantes de polipéptidos β -glucosidasa utilizando cribado de alto rendimiento para la presencia de actividad β -glucosidasa. Por ejemplo, un polinucleótido que codifica una β -glucosidasa de referencia (por ejemplo, la SEQ ID NO: 3 o la SEQ ID NO: 4) se somete a procesos de mutagénesis (por ejemplo mutagénesis aleatoria y recombinación) para introducir mutaciones en el polinucleótido. El polinucleótido mutado se expresa y traduce, generando así enzimas β -glucosidasa manipuladas con modificaciones en el polipéptido. Tal como se utiliza en el presente documento, "modificaciones" incluye sustituciones, deleciones e inserciones de aminoácidos incluídas, por ejemplo, las que se describen más adelante. Puede introducirse en el polipéptido enzimáticamente activo natural cualquier modificación o una combinación de modificaciones para generar enzimas manipuladas, que a continuación se criban mediante diversos métodos para identificar los polipéptidos con una mejora deseada en una propiedad enzimática específica (por ejemplo, actividad enzimática potenciada). Pueden someterse uno o más polinucleótidos que codifican una β -glucosidasa manipulada con una propiedad mejorada a rondas adicionales de tratamientos de mutagénesis para generar polipéptidos con mejoras adicionales en la propiedad enzimática deseada y/o en otras propiedades. Puede determinarse la actividad β -glucosidasa utilizando cualquier ensayo de β -glucosidasa conocido en la técnica, tal como los ensayos de para-nitrofenil- β -D-glucopiranosido (pNPG) y celobiosa que se describen más adelante.

En algunas formas de realización, el polipéptido β -glucosidasa puede tener una sustitución, deleción y/o inserción, con respecto a la SEQ ID NO: 3. Pueden realizarse diversas modificaciones conservando al mismo tiempo (o aumentando) la actividad enzimática. Por lo general, el polipéptido β -glucosidasa es sustancialmente idéntico a la SEQ ID NO: 3 o la SEQ ID NO: 4.

En algunas formas de realización, la variante de proteína BGL1 se diferencia de la SEQ ID NO: 3 o la SEQ ID NO: 4 por tener una inserción o uno o más aminoácidos. Tales inserciones pueden ser en el extremo amino terminal, en el extremo carboxilo terminal o en una porción no terminal de la proteína. En algunas formas de realización, puede insertarse una deleción de 1 a 2, o de 1 ó 2 a 3, 4, 5, 6, 7, 8, 9, 10, 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 ó 50 residuos de aminoácidos (en el extremo N-terminal, C-terminal y/o una región no terminal de la proteína).

En una forma de realización, la proteína BGL1 comprende sustituciones conservadoras de aminoácidos (con respecto a la SEQ ID NO: 3 o la SEQ ID NO: 4) en una o más posiciones. El concepto de sustitución conservadora es bien conocido y se describe en el presente documento. La presente invención incluye variantes modificadas de manera conservadora de las β -glucosidasas descritas en el presente documento. Estas variantes tienen sustituciones conservadoras realizadas en sus secuencias de aminoácidos. Los ejemplos de sustituciones conservadoras están dentro del grupo de aminoácidos básicos (arginina, lisina e histidina), aminoácidos ácidos (ácido glutámico y ácido aspártico), aminoácidos polares (glutamina y asparagina), aminoácidos hidrófobos (leucina, isoleucina y valina), aminoácidos aromáticos (fenilalanina, triptófano y tirosina), y aminoácidos pequeños (glicina, alanina, serina, treonina, prolina, cisteína y metionina). Las sustituciones de aminoácidos que normalmente no modifican la actividad específica son conocidas en la técnica y son descritas, por ejemplo, por H. Neurath y R.L. Hill, 1979, en "The Proteins", Academic Press, Nueva York. Los intercambios que se producen con más frecuencia son Ala/Ser, Val/Ile, Asp/Glu, Thr/Ser, Ala/Gly, Ala/Thr, Ser/Asn, Ala/Val, Ser/Gly, Tyr/Phe, Ala/Pro, Lys/Arg, Asp/Asn, Leu/Ile, Leu/Val, Ala/Glu y Asp/Gly así como éstos a la inversa.

En algunas formas de realización, las variantes de polipéptidos β -glucosidasa de la presente invención incluyen sustituciones de un pequeño porcentaje, por lo general menos de un 5%, más generalmente menos de un 2%, y con frecuencia menos de un 1% de los aminoácidos de la secuencia polipeptídica, por ejemplo, con un aminoácido seleccionado de manera conservadora del mismo grupo de sustitución conservadora.

En algunas formas de realización, la variante de BGL1 incluye secuencias adicionales. La adición de secuencias que no modifican la actividad codificada de una β -glucosidasa, tal como la adición de una secuencia no funcional o no codificante, se considera una variación conservadora del polipéptido/polinucleótido β -glucosidasa. Por ejemplo, la β -glucosidasa puede estar unida a un epítipo de identificación o a otra secuencia útil para la purificación de la β -glucosidasa.

Péptido señal

En general, los polipéptidos β -glucosidasa son secretados por la célula hospedadora en la que se expresan (por ejemplo, C1) y se expresan como una preproteína que incluye un péptido señal, es decir, una secuencia de aminoácidos unida al extremo amino terminal de un polipéptido y que dirige al polipéptido codificado a la vía secretora de la célula. En una forma de realización, el péptido señal es el péptido señal de C1 β -glucosidasa endógeno que tiene la secuencia que se expone como SEQ ID NO: 6. En otras formas de realización, se utilizan péptidos señal de otras proteínas secretadas por C1.

Pueden utilizarse otros péptidos señal, dependiendo de la célula hospedadora y de otros factores. Las regiones codificantes del péptido señal eficaces para las células hospedadora de hongos filamentosos incluyen pero no se limitan a las regiones codificantes del péptido señal obtenidas de la TAKA amilasa de *Aspergillus oryzae*, amilasa neutra de *Aspergillus niger*, glucoamilasa de *Aspergillus niger*, proteinasa aspártica de *Rhizomucor miehei*, celulasa de *Humicola insolens*, lipasa de *Humicola lanuginosa* y celobiohidrolasa II de *T. reesei*. Por ejemplo, puede utilizarse una secuencia de C1 BGL1 con diversos péptidos señal de hongos filamentosos conocidos en la técnica. Los péptidos señal útiles para las células hospedadoras de levadura también incluyen aquellos de los genes para el factor alfa de *Saccharomyces cerevisiae* y la invertasa de *Saccharomyces cerevisiae*. Se describen todavía otras regiones codificantes del péptido señal útiles en Romanos *et al.*, 1992, Yeast 8: 423-488. Las regiones codificantes del péptido señal eficaces para las células hospedadoras bacterianas son las regiones codificantes del péptido señal obtenidas de los genes para la amilasa maltogénica de *Bacillus* NCIB 11837, alfa-amilasa de *Bacillus stearothermophilus*, subtilisina de *Bacillus licheniformis*, β -lactamasa de *Bacillus licheniformis*, proteasas neutras de *Bacillus stearothermophilus* (nprT, nprS, nprM) y prsA de *Bacillus subtilis*. Se describen otros péptidos señal en Simonen y Palva, 1993, Microbiol Rev 57: 109-137. Las variantes de estos péptidos señal y otros péptidos señal resultan adecuadas.

Actividad β -glucosidasa

Las proteínas β -glucosidasa utilizadas en el método de la invención son enzimáticamente activas o son precursoras de la proteína enzimáticamente activa. Puede determinarse la actividad β -glucosidasa mediante métodos conocidos en la técnica. En una forma de realización, la actividad β -glucosidasa se determina utilizando un ensayo de para-nitrofenil- β -D-glucopiranosido (pNPG). En una forma de realización, la actividad β -glucosidasa se determina utilizando un ensayo de celobiosa. Por lo general un polipéptido β -glucosidasa enzimáticamente activo tiene al menos un 50% de la actividad enzimática de la SEQ ID NO: 3.

Por ejemplo, puede utilizarse un ensayo colorimétrico basado en pNPG (p-nitrofenil- β -D-glucopiranosido) para medir la actividad β -glucosidasa. Uno de tales ensayos se describe en el Ejemplo 5, más adelante. En otro ensayo de pNPG ejemplar, en un volumen total de 100 μ l, se añaden 20 μ l de sobrenadante del medio transparente que contiene la enzima β -glucosidasa a una solución de pNPG 4 mM (Sigma-Aldrich, Inc. St. Louis, MO) en tampón de fosfato de sodio 50 mM a pH 6,5. Las reacciones se incuban a pH 6,5, 45°C durante 1 hora. La mezcla de reacción se inactiva con 100 μ l de solución pH 11 de carbonato de sodio 1 M. Se mide la absorbancia de la solución a 405 nm para determinar la conversión del pNPG en p-nitrofenol. Se mide la liberación de p-nitrofenol ($\epsilon = 17.700 \text{ M}^{-1} \text{ cm}^{-1}$) a 405 nm para calcular la actividad β -glucosidasa. Se observa actividad β -glucosidasa detectable en condiciones de cribado de alto rendimiento (pH 7, 50°C). Véase Breves *et al.* (1997) Appl. Environmental Microbiol. 63: 3902.

Como alternativa, puede determinarse la actividad β -glucosidasa utilizando un ensayo de celobiosa, que utiliza celobiosa como sustrato. En un volumen total de 100 μ l, se añaden 25 μ l de sobrenadante del medio transparente que contiene la enzima β -glucosidasa a 10 g/l de celobiosa (Fluka Cat. N° 22150, Sigma-Aldrich, Inc., St. Louis, MO) en tampón de fosfato de sodio 100 mM (pH 6-7) o tampón de acetato de sodio (pH 5-5,5). Se incubaba la reacción a 45°C-70°C durante un tiempo apropiado (de 25 minutos a toda la noche en función de la concentración de enzima) mientras se agita. Se determina la producción de glucosa utilizando un ensayo de glucosa enzimático (K-GLUC, Megazyme, Irlanda). Se añaden 10 μ l de cada reacción a 190 μ l de reactivo GOPOD (suministrado como parte del kit de ensayo K-GLUC). Se incubaba la reacción a 45°C durante 20 minutos y se mide la absorbancia de la solución a 510 nm. El reactivo GOPOD contiene tampón de fosfato de potasio 50 mM pH 7,4, ácido p-hidroxibenzoico 0,011 M, azida de sodio al 0,008% p/v, glucosa oxidasa (>12.000 U/l), peroxidasa (>650 U/l) y 80 mg/l de 4-aminoantipirina. La enzima glucosa oxidasa en el reactivo reacciona con cualquier glucosa presente en la muestra y produce peróxido de hidrógeno que, a continuación, reacciona con la 4-aminoantipirina para producir un colorante quinoneimina en cantidades proporcionales a la cantidad de glucosa presente y que puede medirse mediante espectrofotometría a 510 nm.

IV. POLINUCLEÓTIDOS β -GLUCOSIDASA Y SISTEMAS DE EXPRESIÓN

La presente invención proporciona secuencias polinucleotídicas que codifican C1 β -glucosidasa (es decir, las SEQ ID NO: 3 y 4), así como el péptido señal de C1 β -glucosidasa (SEQ ID NO: 6). Las secuencias genómica y de ADNc de C1 que codifican β -glucosidasa se exponen como SEQ ID NO: 1 y SEQ ID NO: 2, respectivamente. La secuencia de ADNc que codifica el péptido señal de C1 β -glucosidasa es la secuencia del nucleótido 1 al 57 de la SEQ ID NO: 2. Los expertos en la materia entenderán fácilmente que debido a la degeneración del código genético, existe una multitud de secuencias nucleotídicas que codifican polipéptidos β -glucosidasa de la presente invención. La Tabla 2 proporciona el código genético en tripletes convencional para cada aminoácido. Por ejemplo, los codones AGA, AGG, CGA, CGC, CGG y CGU codifican todos ellos el aminoácido arginina. Por lo tanto, en cada posición en los ácidos nucleicos de la invención en la que una arginina esté especificada por un codón, el codón puede cambiarse a cualquiera de los codones correspondientes descritos anteriormente sin modificar el polipéptido codificado. Se entiende que U en una secuencia de ARN corresponde a T en una secuencia de ADN. La invención contempla y proporciona todas y cada una de las posibles variaciones de secuencia de ácido nucleico que codifica

un polipéptido de la invención que podrían realizarse seleccionando combinaciones basadas en posibles elecciones de codones.

También puede diseñarse una secuencia de ADN para los codones con alta preferencia de uso de codones (codones que se utilizan con mayor frecuencia en las regiones codificantes de la proteína que otros codones que codifican el mismo aminoácido). Pueden determinarse los codones preferentes con relación al uso de codones en un único gen, un conjunto de genes de origen o función común, genes altamente expresados, la frecuencia de codones en el conjunto de las regiones codificantes de la proteína de todo el organismo, la frecuencia de codones en el conjunto de las regiones codificantes de la proteína de organismos relacionados, o combinaciones de las mismas. Los codones cuya frecuencia aumenta en función del nivel de expresión génica son por lo general los codones óptimos para la expresión. En particular, puede optimizarse una secuencia de ADN para la expresión en un organismo hospedador particular. Las referencias que proporcionan información de preferencia para una gran variedad de organismos están fácilmente disponibles. Véase, por ejemplo, Henaut y Danchin en "Escherichia Salmonella", Neidhardt, *et al.* Eds., ASM Pres, Washington D.C. (1996), págs. 2047-2066. Con fines de ilustración, y no de limitación, la Figura 5 muestra una secuencia polinucleotídica que codifica C1 BGL1 diseñada con preferencia codónica para la expresión en *Saccharomyces cerevisiae*.

TABLA 2: CÓDIGO GENÉTICO

Aminoácido			Codón
Alanina	Ala	A	GCA GCC GCG GCU
Cisteína	Cys	C	UGC UGU
Ácido aspártico	Asp	D	GAC GAU
Ácido glutámico	Glu	E	GAA GAG
Fenilalanina	Phe	F	UUC UUU
Glicina	Gly	G	GGA GGC GGG GGU
Histidina	His	H	CAC CAU
Isoleucina	Ile	I	AUA AUC AUU
Lisina	Lys	K	AAA AAG
Leucina	Leu	L	UUA UUG CUA CUC CUG CUU
Metionina	Met	M	AUG
Asparagina	Asn	N	AAC AAU
Prolina	Pro	P	CCA CCC CCG CCU
Glutamina	Gln	Q	CAA CAG
Arginina	Arg	R	AGA AGG CGA CGC CGG CGU
Serina	Ser	S	AGC AGU UCA UCC UCG UCU
Treonina	Thr	T	ACA ACC ACG ACU
Valina	Val	V	GUA GUC GUG GUU
Triptófano	Trp	W	UGG
Tirosina	Tyr	Y	UAC UAU

Se conocen diversos métodos para determinar la frecuencia de codones (por ejemplo, uso de codones, uso relativo de codones sinónimos) y la preferencia codónica en organismos específicos, incluido el análisis multivariante, por ejemplo, que utiliza el análisis por conglomerados o el análisis de correspondencias, y el número efectivo de codones utilizados en un gen (véase GCG CodonPreference, Genetics Computer Group Wisconsin Package; Codon W, John Peden, Universidad de Nottingham; McInerney, J. O, 1998, Bioinformatics. 14: 372-73; Stenico *et al.*, 1994, Nucleic Acids Res. 22:437-46; Wright, F., 1990, Gene 87: 23-29; Wada *et al.*, 1992, Nucleic Acids Res. 20: 2111-2118; Nakamura *et al.*, 2000, Nucl. Acids Res. 28: 292; Henaut y Danchin, "Escherichia coli and Salmonella", 1996, Neidhardt, *et al.* Eds, ASM Press, Washington D.C., p. 2047-2066). La fuente de datos para obtener el uso de codones puede basarse en cualquier secuencia nucleotídica disponible capaz de codificar una proteína. Estos conjuntos de datos incluyen secuencias de ácidos nucleicos que se sabe codifican realmente proteínas expresadas (por ejemplo, secuencias codificantes de proteínas completas - CDS), marcadores de secuencia expresada (EST), o regiones codificantes predichas de secuencias genómicas (véase, por ejemplo, Mount, D., Bioinformatics: Sequence and Genome Analysis, capítulo 8, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 2001; Uberbacher, E. C., 1996, Methods Enzymol. 266:259-281; Tiwari *et al.*, 1997, Comput. Appl. Biosci. 13:263-270).

Vectores de expresión

La presente invención utiliza constructos recombinantes que comprenden una secuencia que codifica una β -glucosidasa como se ha descrito anteriormente. En un aspecto particular, la presente invención proporciona un vector de expresión que comprende un polinucleótido β -glucosidasa unido operativamente a un promotor heterólogo. Los vectores de expresión de la presente invención pueden utilizarse para transformar una célula hospedadora apropiada para permitir que el hospedador exprese la proteína β -glucosidasa. Los métodos para la expresión recombinante de proteínas en hongos y otros organismos son bien conocidos en la técnica, y se dispone de o pueden construirse una serie de vectores de expresión utilizando métodos de rutina. Véase, por ejemplo, Tkacz y

Lange, 2004, ADVANCES IN FUNGAL BIOTECHNOLOGY FOR INDUSTRY, AGRICULTURE, AND MEDICINE, KLUWER ACADEMIC/PLENUM PUBLISHERS. Nueva York; Zhu *et al.*, 2009, Construction of two Gateway vectors for gene expression in fungi Plasmid 6:128-33; Kavanagh, K. 2005, FUNGI: BIOLOGY AND APPLICATIONS Wiley.

Los constructos de ácido nucleico de la presente invención comprenden un vector, tal como un plásmido, un cósmido, un fago, un virus, un cromosoma artificial bacteriano (BAC), un cromosoma artificial de levadura (YAC), y similares, en los que se ha insertado una secuencia de ácido nucleico de la invención. Los polinucleótidos de la presente invención pueden incorporarse en cualquiera de diversos vectores de expresión adecuados para la expresión de un polipéptido. Los vectores adecuados incluyen secuencias de ADN cromosómico, no cromosómico y sintético, por ejemplo, derivados de SV40; plásmidos bacterianos; ADN de fago; baculovirus; plásmidos de levadura; vectores derivados de combinaciones de plásmidos y ADN de fago, ADN viral tal como vaccinia, adenovirus, virus de la viruela aviar, de la pseudorrabia, adenovirus, virus adenoasociado, retrovirus y muchos otros. Puede utilizarse cualquier vector que transduzca material genético a una célula, y, si se desea la replicación, que sea replicable y viable en el hospedador pertinente.

En un aspecto, preferente de esta forma de realización, el constructo comprende adicionalmente secuencias reguladoras, incluido, por ejemplo, un promotor, unido operativamente a la secuencia codificante de proteína. Los expertos en la materia conocen un gran número de vectores y promotores adecuados.

Constructos de promotor/gen

Tal como se ha analizado anteriormente, para obtener altos niveles de expresión en un hospedador particular, con frecuencia resulta útil expresar C1 β -glucosidasa bajo el control de un promotor heterólogo (tal como el promotor del gen C1 *cbh1a*). Por lo general, puede unirse operativamente una secuencia promotora a la región 5' de la secuencia codificante de C1 β -glucosidasa. Se reconocerá que en la generación de un constructo de este tipo no es necesario definir los límites de un promotor mínimo. En su lugar, la secuencia de ADN en sentido 5' respecto al codón de inicio de la C1 β -glucosidasa puede sustituirse con la secuencia de ADN que está en sentido 5' respecto al codón de inicio de un determinado gen heterólogo (por ejemplo, *cbh1a*). Esta secuencia 5' "heteróloga" incluye por lo tanto, además de los elementos promotores *per se*, una señal de inicio de la transcripción y la secuencia de la porción 5' no traducida del ARNm quimérico transcrito. Por lo tanto, el constructo de promotor-gen y el ARNm resultante comprenderán una secuencia que codifica β -glucosidasa y una secuencia 5' heteróloga cadena arriba del codón de inicio de la secuencia que codifica β -glucosidasa. En algunos, pero no todos, los casos la secuencia 5' heteróloga empalmará inmediatamente el codón de inicio de la secuencia que codifica β -glucosidasa.

Las SEQ ID NO: 13 y 15 representan la secuencia del gen inmediatamente en sentido 5' respecto al codón de inicio de la preproteína C1 BGL1 natural, y las SEQ ID NO: 14, 16 y 17 son fragmentos de la SEQ ID NO: 13 que terminan un poco cadena arriba del codón de inicio. En una forma de realización, la C1 β -glucosidasa recombinante se expresa a partir de un gen recombinante como una preproteína que incluye el péptido señal de BGL1 natural, y la secuencia del gen recombinante en sentido 5' respecto al codón de inicio de β -glucosidasa no es la SEQ ID NO: 13. En términos más generales, la SEQ ID NO: 14 no se encuentra dentro de las 100 bases en sentido 5' respecto al codón de inicio de β -glucosidasa en la secuencia recombinante. Es decir, el gen de la β -glucosidasa recombinante codifica un ARN que comprende una secuencia que codifica la SEQ ID NO: 4 o una variante de la misma, pero no comprende una secuencia no traducida 5' que contiene la SEQ ID NO: 13, o, como alternativa, no comprende la SEQ ID NO: 14 o no comprende la SEQ ID NO: 15. En otras formas de realización la secuencia no traducida 5' no contiene la SEQ ID NO: 15, o, como alternativa, la SEQ ID NO: 16, o, como alternativa, la SEQ ID NO: 17.

En una forma de realización del constructo génico de la presente invención, la C1 β -glucosidasa se expresa a partir del constructo como una preproteína con un péptido señal heterólogo, en el que, en el constructo, la SEQ ID NO: 13 no está en sentido 5' con respecto al codón de inicio del péptido señal heterólogo. En términos más generales, en el constructo, la SEQ ID NO: 14 no se encuentra dentro de las 100 bases en sentido 5' respecto al codón de inicio del péptido señal heterólogo. Es decir, el constructo génico de β -glucosidasa recombinante codifica un ARN que comprende una secuencia que codifica la SEQ ID NO: 3, o una variante de la misma y un péptido señal heterólogo, pero no comprende una secuencia no traducida 5' que contiene la SEQ ID NO: 13. En otras formas de realización la secuencia no traducida 5' no contiene la SEQ ID NO: 14, o, como alternativa, no contiene la SEQ ID NO: 15, o, como alternativa, no contiene la SEQ ID NO: 16, o, como alternativa, no contiene la SEQ ID NO: 17.

En algunas formas de realización, el promotor heterólogo está unido operativamente a una secuencia genómica de C1 β -glucosidasa que comprende la SEQ ID NO: 18, que es la secuencia del intrón 1 de *bgl1*.

En algunas formas de realización, el promotor heterólogo está unido operativamente a una secuencia de ADNc de C1 β -glucosidasa y no comprende la SEQ ID NO: 19 (secuencia intrónica) y/o no comprende una secuencia de aminoácidos codificada en la SEQ ID NO: 19.

En algunas formas de realización, el constructo incluye la SEQ ID NO: 20 (el primer exón de C1 β -glucosidasa) y/o la secuencia de aminoácidos codificada en la SEQ ID NO: 20.

En algunas formas de realización, el constructo incluye la SEQ ID NO: 21.

Los ejemplos de promotores útiles para la expresión de polinucleótidos β -glucosidasa incluyen promotores fúngicos. Por ejemplo, pueden utilizarse secuencias promotoras que dirigen la expresión de genes distintos del gen de β -glucosidasa 1 en C1. Por ejemplo, puede utilizarse un promotor fúngico de un gen que codifica celobiohidrolasa. En una forma de realización, se utiliza el promotor asociado con el gen de C1 celobiohidrolasa 1 (*cbh1a*). En la figura 4 se proporciona una secuencia promotora asociada con *cbh1a*. Véase también la publicación PCT WO 01/79507 y la publicación de patente de EE.UU. US 2003/0187243, que proporcionan la secuencia de ADN del gen de C1 CBH1a completo, incluidas las secuencias promotora y terminadora.

Se comprenderá que puede utilizarse una subsecuencia de la SEQ ID NO: 8 o una variante del promotor de CBH1a para dirigir la expresión de β -glucosidasa. En determinadas formas de realización, el promotor heterólogo unido operativamente a la secuencia que codifica β -glucosidasa comprende la SEQ ID NO: 8, una subsecuencia de la SEQ ID NO: 8 con actividad promotora o una secuencia de ADN capaz de hibridar con el complemento de la SEQ ID NO: 8 y que tiene actividad promotora. Puede identificarse una subsecuencia de la SEQ ID NO: 8 con actividad promotora utilizando métodos bien conocidos. En un enfoque, se une una supuesta secuencia promotora en sentido 5' respecto a una secuencia que codifica una proteína indicadora, se transfecta el constructo a la célula hospedadora (por ejemplo, C1) y se mide el nivel de expresión del indicador. La expresión del indicador puede determinarse midiendo, por ejemplo, los niveles de ARNm de la secuencia indicadora, una actividad enzimática de la proteína indicadora, o la cantidad de proteína indicadora producida. Por ejemplo, la actividad promotora puede determinarse utilizando la proteína fluorescente verde como secuencia codificante (Henriksen *et al.*, 1999, Microbiology 145: 729-34) o un gen indicador lacZ (Punt *et al.*, 1997, Gene, 197: 189-93). Pueden derivarse promotores funcionales a partir de secuencias promotoras naturales mediante métodos de evolución dirigida. Véase, por ejemplo, Wright *et al.*, 2005, Human Gene Therapy, 16: 881-892. En algunas formas de realización, el ADN del promotor comprende los nucleótidos 1 a 1817 de la SEQ ID NO: 8, los nucleótidos 100 a 1800 de la SEQ ID NO: 8, los nucleótidos 200 a 1800 de la SEQ ID NO: 6, los nucleótidos 300 a 1800 de la SEQ ID NO: 8, los nucleótidos 400 a 1800 de la SEQ ID NO: 8, los nucleótidos 500 a 1800 de la SEQ ID NO: 8, los nucleótidos 600 a 1800 de la SEQ ID NO: 8, los nucleótidos 700 a 1800 de la SEQ ID NO: 8, los nucleótidos 800 a 1800 de la SEQ ID NO: 8, los nucleótidos 900 a 1800 de la SEQ ID NO: 8; los nucleótidos 1000 a 1800 de la SEQ ID NO: 8, los nucleótidos 1100 a 1800 de la SEQ ID NO: 8, los nucleótidos 1200 a 1800 de la SEQ ID NO: 8, los nucleótidos 1300 a 1800 de la SEQ ID NO: 8, los nucleótidos 1400 a 1800 de la SEQ ID NO: 8, los nucleótidos 1500 a 1800 de la SEQ ID NO: 8, los nucleótidos 1600 a 1800 de la SEQ ID NO: 8, los nucleótidos 1700 a 1800 de la SEQ ID NO: 8, o es un polinucleótido que hibrida en condiciones de baja rigurosidad con el complemento de la secuencia o subsecuencia identificada anteriormente, hibrida preferentemente en condiciones de rigurosidad media y lo más preferentemente hibrida en condiciones de alta rigurosidad.

Los ejemplos de otros promotores adecuados útiles para dirigir la transcripción de los constructos de nucleótidos de la presente invención en una célula hospedadora de hongo filamentosos son promotores obtenidos a partir de genes para TAKA amilasa de *Aspergillus oryzae*, proteinasa aspártica de *Rhizomucor miehei*, alfa-amilasa neutra de *Aspergillus niger*, alfa-amilasa estable en medio ácido de *Aspergillus niger*, glucoamilasa de *Aspergillus niger* o *Aspergillus awamori* (glaA), lipasa de *Rhizomucor miehei*, proteasa alcalina de *Aspergillus oryzae*, triosa fosfato isomerasa de *Aspergillus oryzae*, acetamidasa de *Aspergillus nidulans*, y proteasa de tipo tripsina de *Fusarium oxysporum* (WO 96/00787), así como el promotor NA2-tpi (un híbrido de los promotores de los genes para alfa-amilasa neutra de *Aspergillus niger* y triosa fosfato isomerasa de *Aspergillus oryzae*), promotores tales como *cbh1*, *cbh2*, *egl1*, *egl2*, *pepA*, *hfb1*, *hfb2*, *xyn1*, *amy* y *glaA* (Nunberg *et al.*, Mol Cell Biol., 4: 2306 -2315 (1984), Boel *et al.*, EMBO J. 3: 1581-1585 ((1984) y EPA 137280), y promotores mutantes, truncados e híbridos de los mismos. En un hospedador de levadura, los promotores útiles pueden ser de los genes para enolasa de *Saccharomyces cerevisiae* (ENO-1), galactokinasa de *Saccharomyces cerevisiae* (GAL1), alcohol deshidrogenasa/glicerolaldehído-3-fosfato deshidrogenasa (ADH₂/GAP) de *Saccharomyces cerevisiae*, y 3-fosfoglicerato quinasa de *Saccharomyces cerevisiae*. Se describen otros promotores útiles para células hospedadoras de levadura en Romanos *et al.*, 1992, Yeast 8: 423-488. Pueden utilizarse promotores asociados con la producción de quitinasa en hongos. Véase, por ejemplo, Blaiseau y Lafay, 1992, Gene 120243-248 (hongo filamentosos *Aphanocladium album*); Limon *et al.*, 1995, Curr. Genet, 28: 478-83 [*Trichoderma harzianum*].

Los promotores que se sabe controlan la expresión de genes en células procariotas o eucariotas o sus virus y que pueden utilizarse en algunas formas de realización de la invención incluyen el promotor SV40, el promotor lac o trp de *E. coli*, el promotor P_L del fago lambda, el promotor tac, promotor de T7, y similares. En las células hospedadoras bacterianas, los promotores adecuados incluyen los promotores obtenidos del operón lac de *E. coli*, gen de agarasa (*dagA*) de *Streptomyces coelicolor*, gen de levansucrasa (*sacB*) de *Bacillus subtilis*, gen de alfa-amilasa (*amyI*) de *Bacillus licheniformis*, gen de amilasa maltogénica de *Bacillus stearothermophilus* (*amyM*), gen de alfa-amilasa (*amyQ*) de *Bacillus amyloliquefaciens*, genes *xylA* y *xylB* de *Bacillus subtilis* y gen de β -lactamasa procariota.

Un vector de expresión contiene opcionalmente un sitio de unión al ribosoma para la iniciación de la traducción y un terminador de la transcripción, tal como PinII. El vector también incluye opcionalmente secuencias apropiadas para amplificar la expresión, por ejemplo, un potenciador.

Además, los vectores de expresión de la presente invención contienen opcionalmente uno o más genes marcadores seleccionables para proporcionar un rasgo fenotípico para la selección de las células hospedadoras transformadas. Los genes marcadores adecuados incluyen aquellos que codifican resistencia a antibióticos, tal como resistencia a ampicilina (amp^R), kanamicina, cloranfenicol o tetraciclina. Otros ejemplos incluyen el antibiótico espectinomina o estreptomina (por ejemplo, el gen aadA), el gen de estreptomina fosfotransferasa (SPT) que codifica resistencia a estreptomina, el gen de neomicina fosfotransferasa (NPTII) que codifica resistencia a kanamicina o a geneticina, el gen de higromina fosfotransferasa (HPT) que codifica resistencia a higromina. Los genes marcadores seleccionables adicionales incluyen resistencia a dihidrofolato reductasa o a neomicina para el cultivo de células eucariotas, y resistencia a tetraciclina o a ampicilina en *E. coli*.

En el Ejemplo 3 se describe un vector de expresión ejemplar para la expresión de polipéptidos β -glucosidasa de la invención, que se presenta más adelante en el presente documento.

Síntesis y manipulación de polinucleótidos β -glucosidasa

Pueden prepararse polinucleótidos que codifican β -glucosidasa utilizando métodos que son bien conocidos en la técnica. Por lo general, los oligonucleótidos de hasta aproximadamente 40 bases se sintetizan individualmente, a continuación se unen (por ejemplo, mediante métodos de ligación enzimática o química, o métodos mediados por polimerasa) para formar esencialmente cualquier secuencia continua deseada. Por ejemplo, pueden prepararse polinucleótidos de la presente invención mediante síntesis química utilizando, por ejemplo, el clásico método de la fosforamidita descrito por Beaucage, *et al.*, 1981, Tetrahedron Letters, 22:1859-1869, o el método descrito por Matthes, *et al.*, 1984, EMBO J. 3:801-05. Estos métodos se ponen en práctica por lo general en métodos de síntesis automatizados. Según el método de la fosforamidita, los oligonucleótidos se sintetizan, por ejemplo, en un sintetizador de ADN automático, se purifican, hibridan, ligan y clonan en vectores apropiados.

Además, puede encargarse esencialmente cualquier ácido nucleico a cualquiera de diversas fuentes comerciales, tales como The Midland Certified Reagent Company (Midland, TX), The Great American Gene Company (Ramona, CA), ExpressGen Inc. (Chicago, IL), Operon Technologies Inc. (Alameda, CA), y muchos otros.

Los polinucleótidos también pueden sintetizarse mediante técnicas bien conocidas como se describe en la literatura técnica. Véase, por ejemplo, Carruthers, *et al.*, 1982, Cold Spring Harbor Symp. Quant. Biol., 47:411-18 y Adams *et al.*, 1983, J. Am. Chem. Soc. 105:661. A continuación, pueden obtenerse fragmentos de ADN bicatenario ya sea sintetizando la hebra complementaria e hibridando las hebras juntas en condiciones apropiadas, o añadiendo la hebra complementaria utilizando ADN polimerasa con una secuencia de cebador apropiada.

Los textos generales que describen técnicas de biología molecular que son útiles en el presente documento, que incluyen el uso de vectores, promotores, protocolos suficientes para dirigir a los expertos a través de métodos de amplificación *in vitro*, incluida la reacción en cadena de la polimerasa (PCR) y la reacción en cadena de la ligasa (LCR), y muchos otros métodos pertinentes, incluyen Berger y Kimmel, Guide to Molecular Cloning Techniques, Methods in Enzymology, volumen 152, Academic Press, Inc., San Diego, CA (Berger); Sambrook *et al.*, Molecular Cloning - A Laboratory Manual (2^a ed.), vol. 1-3, Cold Spring Harbor Laboratory, Cold Spring Harbor, Nueva York, 1989 ("Sambrook") y Current Protocols in Molecular Biology, F.M. Ausubel *et al.*, Eds., Current Protocols, una empresa conjunta entre Greene Publishing Associates, Inc. y John Wiley & Sons, Inc., (con suplementos hasta 1999) ("Ausubel"). Se hace referencia a Berger, Sambrook y Ausubel, así como a Mullis *et al.*, (1987) patente de Estados Unidos N° 4.683.202; PCR Protocols A Guide to Methods and Applications (Innis *et al.* eds.) Academic Press Inc. San Diego, CA (1990) (Innis); Arnheim y Levinson (1 de octubre de 1990) C&EN 36-47; The Journal Of NIH Research (1991) 3, 81-94; Kwok *et al.* (1989) Proc. Natl. Acad. Sci. EE.UU. 86, 1173; Guatelli *et al.* (1990) Proc. Natl. Acad. Sci. EE.UU. 87, 1874; Lomell *et al.* (1989) J. Clin. Chem. 35, 1826; Landegren *et al.*, (1988) Science 241, 1077-1080; Van Brunt (1990) Biotechnology 8, 291-294; Wu y Wallace, (1989) Gen 4, 560; Barringer *et al.* (1990) Gen 89, 117, y Sooknandan y Malek (1995) Biotechnology 13: 563-564. Se describen métodos para la clonación *in vitro* de ácidos nucleicos amplificados en Wallace *et al.*, patente de Estados Unidos N° 5.426.039.

Hospedadores de expresión

La presente invención también proporciona células hospedadoras manipuladas (recombinantes) que se transforman con un vector de expresión o constructo de ADN que codifica β -glucosidasa. Opcionalmente, la expresión de β -glucosidasa en la célula está bajo el control de un promotor heterólogo. Las células hospedadoras de la invención pueden utilizarse para producir polipéptidos β -glucosidasa. Por lo tanto, la presente invención se refiere a una célula hospedadora que comprende cualquier polinucleótido β -glucosidasa de la presente invención que se haya descrito anteriormente en el presente documento. Tal como se utiliza en el presente documento, una célula hospedadora genéticamente modificada o recombinante incluye la progenie de dicha célula hospedadora que comprende un polinucleótido β -glucosidasa que codifica un polipéptido recombinante de la invención. En algunas formas de realización, la célula hospedadora genéticamente modificada o recombinante es una célula eucariota. Las células hospedadoras eucariotas adecuadas incluyen, pero no se limitan a, células fúngicas, células de algas, células de insectos y células vegetales. En algunos casos, las células hospedadoras pueden modificarse para

5 aumentar la expresión, secreción o estabilidad de las proteínas, o para conferir otras características deseadas. Las células (por ejemplo, de hongos) que se han mutado o seleccionado para que tengan baja actividad proteasa son particularmente útiles para la expresión. Por ejemplo, pueden utilizarse cepas C1 en las que se ha delecionado o alterado el locus *alp1* (proteasa alcalina).

10 Las células hospedadoras fúngicas adecuadas incluyen, pero no se limitan a, Ascomycota, Basidiomycota, Deuteromycota, Zygomycota, Fungi imperfecti. Las células hospedadoras fúngicas particularmente preferentes son células de levadura y células de hongos filamentosos. Las células hospedadoras de hongos filamentosos de la presente invención incluyen todas las formas filamentosas de la subdivisión Eumycotina y Oomycota (véase, por ejemplo, Hawksworth *et al.*, En Ainsworth and Bisby's Dictionary of The Fungi, 8ª edición, 1995, CAB International, University Press, Cambridge, Reino Unido). Los hongos filamentosos se caracterizan por un micelio vegetativo con una pared celular compuesta por quitina, celulosa y otros polisacáridos complejos. Las células hospedadoras de hongos filamentosos de la presente invención son morfológicamente distintas de la levadura.

15 En una forma de realización, la célula hospedadora es una célula C1. En una forma de realización, la célula hospedadora es una célula de una especie de *Myceliophthora*, tal como *Myceliophthora thermophila*. En algunas formas de realización, la célula hospedadora de hongo filamentoso puede ser una célula de una especie de, pero no limitada a, *Achlya*, *Acremonium*, *Aspergillus*, *Aureobasidium*, *Bjerkandera*, *Ceriporiopsis*, *Cephalosporium*, *Chrysosporium*, *Cochliobolus*, *Corynascus*, *Cryphonectria*, *Cryptococcus*, *Coprinus*, *Coriolus*, *Diplodia*, *Endothia*, *Fusarium*, *Gibberella*, *Gliocladium*, *Humicola*, *Hypocrea*, *Myceliophthora*, *Mucor*, *Neurospora*, *Penicillium*, *Podospora*, *Phlebia*, *Piromyces*, *Pyricularia*, *Rhizomucor*, *Rhizopus*, *Schizophyllum*, *Scytalidium*, *Sporotrichum*, *Talaromyces*, *Thermoascus*, *Thielavia*, *Trametes*, *Tolypocladium*, *Trichoderma*, *Verticillium*, *Volvariella* o teleomorfos o anamorfos y sinónimos o equivalentes taxonómicos de las mismas. En algunas formas de realización, la célula hospedadora es una célula distinta de C1 o distinta de una especie de *Myceliophthora*.

25 En algunas formas de realización de la invención, la célula hospedadora de hongo filamentoso es de especies de *Aspergillus*, especies de *Ceriporiopsis*, especies de *Chrysosporium*, especies de *Corynascus*, especies de *Fusarium*, especies de *Humicola*, especies de *Neurospora*, especies de *Penicillium*, especies de *Tolypocladium*, especies de *Trametes*, o especies de *Trichoderma*.

30 En algunas formas de realización de la invención, la célula hospedadora de hongo filamentoso es de especies de *Trichoderma*, por ejemplo, *T. longibrachiatum*, *T. viride* (por ejemplo, ATCC 32098 y 32086), *Hypocrea jecorina* o *T. reesei* (NRRL 15709, ATCC 13631, 56764, 56765, 56466, 56767 y RL-P37 y derivados de las mismas - Véase Sheir-Neiss *et al.*, 1984, Appl. Microbiol. Biotechnology, 20:46-53), *T. koningii*, y *T. harzianum*. Además, el término "*Trichoderma*" se refiere a cualquier cepa fúngica que se haya clasificado anteriormente como *Trichoderma* o que se clasifica actualmente como *Trichoderma*.

40 En algunas formas de realización de la invención, la célula hospedadora de hongo filamentoso es de especies de *Aspergillus*, por ejemplo, *A. awamori*, *A. funigatus*, *A. japonicus*, *A. nidulans*, *A. niger*, *A. aculeatus*, *A. foetidus*, *A. oryzae*, *A. sojae*, y *A. kawachi*. (Se hace referencia a Kelly y Hynes, 1985, EMBO J. 4,475479; NRRL 3112, ATCC 11490, 22342, 44733 y 14331; Yelton *et al.*, 1984, Proc. Natl. Acad. Sci. EE.UU., 81, 1470-1474; Tilburn *et al.*, 1982, Gene 26,205-221; y Johnston *et al.*, 1985, EMBO J. 4, 1307 -1311).

45 En algunas formas de realización de la invención, la célula hospedadora de hongo filamentoso es de especies de *Fusarium*, por ejemplo, *F. bactridioides*, *F. cerealis*, *F. crookwellense*, *F. culmorum*, *F. graminearum*, *F. graminum*, *F. oxysporum*, *F. roseum*, y *F. venenatum*. En algunas formas de realización de la invención, la célula hospedadora de hongo filamentoso es de especies de *Neurospora*, por ejemplo, *N. crassa*. Se hace referencia a Case, M.E. *et al.*, (1979) Proc. Natl. Acad. Sci. EE.UU., 76, 5259-5263; USP 4.486.553; y Kinsey, J.A. y J.A. Rambosek (1984) Molecular and Cellular Biology 4, 117-122. En algunas formas de realización de la invención, la célula hospedadora de hongo filamentoso es de especies de *Humicola*, por ejemplo, *H. insolens*, *H. grisea*, y *H. lanuginosa*. En algunas formas de realización de la invención, la célula hospedadora de hongo filamentoso es de especies de *Mucor*, por ejemplo, *M. miehei* y *M. circinelloides*. En algunas formas de realización de la invención, la célula hospedadora de hongo filamentoso es de especies de *Rhizopus*, por ejemplo, *R. oryzae* y *R. niveus*. En algunas formas de realización de la invención, la célula hospedadora de hongo filamentoso es de especies de *Penicillium*, por ejemplo, *P. purpurogenum*, *P. chrysogenum* y *P. verruculosum*. En algunas formas de realización de la invención, la célula hospedadora de hongo filamentoso es de especies de *Thielavia*, por ejemplo, *T. terrestris*. En algunas formas de realización de la invención, la célula hospedadora de hongo filamentoso es de especies de *Tolypocladium*, por ejemplo, *T. inflatum* y *T. geodes*. En algunas formas de realización de la invención, la célula hospedadora de hongo filamentoso es de especies de *Trametes*, por ejemplo, *T. villosa* y *T. versicolor*.

60 En algunas formas de realización de la invención, la célula hospedadora de hongo filamentoso es de especies de *Chrysosporium*, por ejemplo, *C. lucknowense*, *C. keratinophilum*, *C. tropicum*, *C. merdarium*, *C. inops*, *C. pannicola*, y *C. zonatum*. En una forma de realización particular, el hospedador es *C. lucknowense*.

65 En la presente invención, una célula hospedadora de levadura puede ser una célula de una especie de, pero no limitada a, *Candida*, *Hansenula*, *Saccharomyces*, *Schizosaccharomyces*, *Pichia*, *Kluyveromyces*, y *Yarrowia*.

En algunas formas de realización de la invención, la célula de levadura es *Hansenula polymorpha*, *Saccharomyces cerevisiae*, *Saccharomyces carlsbergensis*, *Saccharomyces diastaticus*, *Saccharomyces norbensis*, *Saccharomyces kluyveri*, *Schizosaccharomyces pombe*, *Pichia pastoris*, *Pichia finlandica*, *Pichia trehalophila*, *Pichia kodamae*, *Pichia membranaefaciens*, *Pichia opuntiae*, *Pichia thermotolerans*, *Pichia salictaria*, *Pichia quercuum*, *Pichia pijperi*, *Pichia stipitis*, *Pichia methanolica*, *Pichia angusta*, *Kluyveromyces lactis*, *Candida albicans*, y *Yarrowia lipolytica*.

En algunas formas de realización de la invención, la célula hospedadora es de alga tal como, *Chlamydomonas* (por ejemplo, *C. reinhardtii*) y *Phormidium* (*P. sp.* ATCC29409).

En otras formas de realización, la célula hospedadora es una célula procariota. Las células procariotas adecuadas incluyen células de bacterias gram positivas, gram negativas y gram variables. La célula hospedadora puede ser una especie de, pero no limitada a, *Agrobacterium*, *Alicyclobacillus*, *Anabaena*, *Anacystis*, *Acinetobacter*, *Acidothermus*, *Arthrobacter*, *Azobacter*, *Bacillus*, *Bifidobacterium*, *Brevibacterium*, *Butyrivibrio*, *Buchnera*, *Campestris*, *Campylobacter*, *Clostridium*, *Corynebacterium*, *Chromatium*, *Coprococcus*, *Escherichia*, *Enterococcus*, *Enterobacter*, *Erwinia*, *Fusobacterium*, *Faecalibacterium*, *Francisella*, *Flavobacterium*, *Geobacillus*, *Haemophilus*, *Helicobacter*, *Klebsiella*, *Lactobacillus*, *Lactococcus*, *Ilyobacter*, *Micrococcus*, *Microbacterium*, *Mesorhizobium*, *Methylobacterium*, *Mycobacterium*, *Neisseria*, *Pantoea*, *Pseudomonas*, *Prochlorococcus*, *Rhodobacter*, *Rhodopseudomonas*, *Roseburia*, *Rhodospirillum*, *Rhodococcus*, *Scenedesmus*, *Streptomyces*, *Streptococcus*, *Syneccoccus*, *Saccharomonospora*, *Staphylococcus*, *Serratia*, *Salmonella*, *Shigella*, *Thermoanaerobacterium*, *Tropheryma*, *Tularensis*, *Temecula*, *Thermosynechococcus*, *Thermococcus*, *Ureaplasma*, *Xanthomonas*, *Xylella*, *Yersinia* y *Zymomonas*.

En algunas formas de realización, la célula hospedadora es una especie de *Agrobacterium*, *Acinetobacter*, *Azobacter*, *Bacillus*, *Bifidobacterium*, *Buchnera*, *Geobacillus*, *Campylobacter*, *Clostridium*, *Corynebacterium*, *Escherichia*, *Enterococcus*, *Erwinia*, *Flavobacterium*, *Lactobacillus*, *Lactococcus*, *Pantoea*, *Pseudomonas*, *Staphylococcus*, *Salmonella*, *Streptococcus*, *Streptomyces*, y *Zymomonas*.

En otras formas de realización más, la cepa hospedadora bacteriana no es patógena para los seres humanos. En algunas formas de realización, la cepa hospedadora bacteriana es una cepa industrial. Se conocen y son adecuadas en la presente invención numerosas cepas industriales bacterianas.

En algunas formas de realización de la invención, la célula hospedadora bacteriana es de especies de *Agrobacterium*, por ejemplo, *A. radiobacter*, *A. rhizogenes* y *A. rubi*. En algunas formas de realización de la invención, la célula hospedadora bacteriana es de especies de *Arthrobacter*, por ejemplo, *A. aureus*, *A. citreus*, *A. globiformis*, *A. hydrocarboglutamicus*, *A. mysorens*, *A. nicotianae*, *A. paraffineus*, *A. protophonniae*, *A. roseoparaffinus*, *A. sulfureus*, y *A. ureafaciens*. En algunas formas de realización de la invención, la célula hospedadora bacteriana es de especies de *Bacillus*, por ejemplo, *B. thuringiensis*, *B. anthracis*, *B. megaterium*, *B. subtilis*, *B. lentus*, *B. circulans*, *B. pumilus*, *B. lautus*, *B. coagulans*, *B. brevis*, *B. firmus*, *B. alkaophilus*, *B. licheniformis*, *B. clausii*, *B. stearothermophilus*, *B. halodurans* y *B. amyloliquefaciens*. En formas de realización particulares, la célula hospedadora será una cepa de *Bacillus* industrial incluida, pero no limitada a, *B. subtilis*, *B. pumilus*, *B. licheniformis*, *B. megaterium*, *B. clausii*, *B. stearothermophilus* y *B. amyloliquefaciens*. Algunas formas de realización preferentes de una célula hospedadora de *Bacillus* incluyen *B. subtilis*, *B. licheniformis*, *B. megaterium*, *B. stearothermophilus* y *B. amyloliquefaciens*. En algunas formas de realización, la célula hospedadora bacteriana es de especies de *Clostridium*, por ejemplo, *C. acetobutylicum*, *C. tetani* E88, *C. lituseburense*, *C. saccharobutylicum*, *C. perfringens*, *C. y beijerinckii*. En algunas formas de realización, la célula hospedadora bacteriana es de especies de *Corynebacterium*, por ejemplo, *C. glutamicum* y *C. acetoacidophilum*. En algunas formas de realización, la célula hospedadora bacteriana es de especies de *Escherichia*, por ejemplo, *E. coli*. En algunas formas de realización, la célula hospedadora bacteriana es de especies de *Erwinia*, por ejemplo, *E. uredovora*, *E. carotovora*, *E. ananas*, *E. herbicola*, *E. punctata*, y *E. terreus*. En algunas formas de realización, la célula hospedadora bacteriana es de especies de *Pantoea*, por ejemplo, *P. citrea* y *P. agglomerans*. En algunas formas de realización, la célula hospedadora bacteriana es de especies de *Pseudomonas*, por ejemplo, *P. putida*, *P. aeruginosa*, *P. mevalonii* y *P. sp.* D-01 10. En algunas formas de realización, la célula hospedadora bacteriana es de especies de *Streptococcus*, por ejemplo, *S. equisimilis*, *S. pyogenes* y *S. uberis*. En algunas formas de realización, la célula hospedadora bacteriana es de especies de *Streptomyces*, por ejemplo, *S. ambofaciens*, *S. achromogenes*, *S. avermitilis*, *S. coelicolor*, *S. aureofaciens*, *S. aureus*, *S. fungicidicus*, *S. griseus* y *S. lividans*. En algunas formas de realización, la célula hospedadora bacteriana es de especies de *Zymomonas*, por ejemplo, *Z. mobilis* y *Z. lipolytica*.

Las cepas que pueden utilizarse en la práctica de la invención, incluidas cepas procariotas y eucariotas, son fácilmente accesibles al público a partir de varias colecciones de cultivos tales como la American Type Culture Collection (ATCC), Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH (DSMZ), Centraalbureau Voor Schimmelcultures (CBS) y el Agricultural Research Service Patent Culture Collection, Northern Regional Research Center (NRRL).

Las células hospedadoras pueden modificarse genéticamente para que tengan características que mejoren la secreción de proteínas, la estabilidad de las proteínas u otras propiedades deseables para la expresión y/o secreción de una proteína. La modificación genética puede conseguirse mediante técnicas de ingeniería genética o

utilizando técnicas microbiológicas clásicas, tales como mutagénesis química o por UV y selección posterior. Puede utilizarse una combinación de técnicas de selección clásica y de modificación recombinante para producir el organismo de interés. Utilizando la tecnología recombinante, pueden introducirse, delecionarse, inhibirse o modificarse moléculas de ácido nucleico, de modo que dé como resultado un aumento de los rendimientos de β -glucosidasa dentro del organismo o en el cultivo. Por ejemplo, la desactivación génica de la función de Alp1 da como resultado una célula que no expresa la mayoría de o todas las celulasas. La desactivación génica de la función de pyr5 da como resultado una célula con un fenotipo deficiente en pirimidina.

Transformación

La introducción de un vector o constructo de ADN en una célula hospedadora puede efectuarse mediante transfección con fosfato de calcio, transfección mediada por DEAE-dextrano, electroporación, u otras técnicas comunes (véase Davis *et al.*, 1986, Basic Methods in Molecular Biology). Se describe un método para la transformación de células hospedadoras C1 en el Ejemplo 4, que se presenta más adelante. La transformación de células hospedadoras C1 es conocida en la técnica (véase, por ejemplo, el documento US 2008/0194005).

Condiciones de cultivo

Las células hospedadoras manipuladas pueden cultivarse en medios nutrientes convencionales modificados según resulte apropiado para activar promotores, seleccionar transformantes o amplificar el polinucleótido β -glucosidasa. Las condiciones de cultivo, tales como temperatura, pH y similares, son las utilizadas anteriormente con la célula hospedadora seleccionada para la expresión, y resultarán evidentes para los expertos en la materia. Como se ha señalado, se dispone de muchas referencias para el cultivo y la producción de muchas células, incluidas células de origen bacteriano, vegetal, animal (especialmente de mamíferos) y arqueobacteriano. Sambrook, Ausubel y Berger (todos *supra*), así como Freshney (1994) Culture of Animal Cells, a Manual of Basic Technique, tercera edición, Wiley-Liss, Nueva York y las referencias citadas en los mismos; Doyle y Griffiths (1997) Mammalian Cell Culture: Essential Techniques John Wiley and Sons, NY; Humason (1979) Animal Tissue Techniques, cuarta edición, W.H. Freeman and Company; y Ricciardelli, *et al.*, (1989) In vitro Cell Dev. Biol. 25:1016-1024. Para el cultivo y regeneración de células vegetales, Payne *et al.* (1992) Plant Cell and Tissue Culture in Liquid Systems John Wiley & Sons, Inc. Nueva York, NY; Gamborg y Phillips (eds.) (1995) Plant Cell, Tissue and Organ Culture; Fundamental Methods Springer Lab Manual, Springer-Verlag (Berlin Heidelberg Nueva York); Jones, ed. (1984) Plant Gene Transfer and Expression Protocols, Humana Press, Totowa, Nueva Jersey y Plant Molecular Biology (1993) R.R.D. Croy, Ed. Bios Scientific Publishers, Oxford, Reino Unido. ISBN 0 12 198370 6. Los medios de cultivo celular en general se establecen en Atlas y Parks (eds.) The Handbook of Microbiological Media (1993) CRC Press, Boca Raton, FL. Se encuentra información adicional para el cultivo celular en la literatura disponible en el mercado, tal como el Life Science Research Cell Culture Catalogue (1998) de Sigma-Aldrich, Inc (St Louis, MO) ("Sigma-LSRCCC") y, por ejemplo, The Plant Culture Catalogue y suplemento (1997) también de Sigma-Aldrich, Inc (St Louis, MO) ("Sigma-PCCS").

Las condiciones de cultivo para las células hospedadoras C1 son conocidas en la técnica y pueden ser determinadas fácilmente por un experto. Véanse, por ejemplo, los documentos US 2008/0194005, US 20030187243, WO 2008/073914 y WO 01/79507.

En algunas formas de realización, las células que expresan los polipéptidos β -glucosidasa de la invención se cultivan en condiciones de fermentación discontinua o continua. La fermentación discontinua clásica es un sistema cerrado, en el que las composiciones del medio se establecen al comienzo de la fermentación y no está sujeta a alternancias artificiales durante la fermentación. Una variación del sistema discontinuo es una fermentación alimentada que también se utiliza en la presente invención. En esta variación, el sustrato se añade en incrementos a medida que avanza la fermentación. Los sistemas semicontinuos son útiles cuando es probable que la represión catabólica inhiba el metabolismo de las células y cuando resulta deseable tener cantidades limitadas de sustrato en el medio. Las fermentaciones discontinua y alimentada son comunes y bien conocidas en la técnica. La fermentación continua es un sistema abierto en el que se añade continuamente un medio de fermentación definido a un biorreactor y se elimina simultáneamente una cantidad igual de medio acondicionado para el procesamiento. La fermentación continua mantiene generalmente los cultivos a una alta densidad constante en la que las células se encuentran principalmente en la fase de crecimiento logarítmico. Los sistemas de fermentación continua procuran mantener condiciones de crecimiento en estado estacionario. Los métodos para la modulación de nutrientes y factores de crecimiento para los procesos de fermentación continua así como las técnicas para maximizar la velocidad de formación de producto son bien conocidos en la técnica de la microbiología industrial.

También pueden emplearse sistemas de transcripción/traducción sin células para producir polipéptidos β -glucosidasa utilizando los polinucleótidos de la presente invención. Varios de tales sistemas están disponibles en el mercado. Se encuentra una guía general de protocolos de transcripción y traducción *in vitro* en Tymms (1995) In vitro Transcription and Translation Protocols: Methods in Molecular Biology, volumen 37, Garland Publishing, NY.

V. PRODUCCIÓN Y RECUPERACIÓN DE POLIPÉPTIDOS β -GLUCOSIDASA

La presente invención se refiere a un método para producir un polipéptido con actividad β -glucosidasa, comprendiendo el método proporcionar una célula hospedadora transformada con cualquiera de los polinucleótidos β -glucosidasa descritos de la presente invención; cultivar la célula hospedadora transformada en un medio de cultivo en condiciones en las que la célula hospedadora expresa el polipéptido β -glucosidasa codificado; y, opcionalmente, recuperar o aislar el polipéptido β -glucosidasa expresado, o recuperar o aislar el medio de cultivo que contiene el polipéptido β -glucosidasa expresado. El método proporciona adicionalmente de manera opcional lisar las células hospedadoras transformadas después de que hayan expresado el polipéptido β -glucosidasa codificado y, opcionalmente, recuperar o aislar el polipéptido β -glucosidasa expresado a partir del lisado celular. La presente invención proporciona adicionalmente un método para producir un polipéptido β -glucosidasa, comprendiendo dicho método cultivar una célula hospedadora transformada con un polinucleótido β -glucosidasa en condiciones adecuadas para la producción del polipéptido β -glucosidasa y recuperar el polipéptido β -glucosidasa. En una forma de realización adicional, la presente invención proporciona un método para sobreexpresar (es decir, producir) un polipéptido C1 β -glucosidasa que comprende: (a) proporcionar una célula hospedadora C1 recombinante que comprende un constructo de ácido nucleico, en el que el constructo de ácido nucleico comprende una secuencia polinucleotídica que codifica una C1 β -glucosidasa madura de la presente invención y el constructo de ácido nucleico también comprende opcionalmente una secuencia polinucleotídica que codifica un péptido señal en el extremo amino terminal de dicha β -glucosidasa madura, en el que la secuencia polinucleotídica que codifica la C1 β -glucosidasa madura y el péptido señal opcional está unida operativamente a un promotor heterólogo; y (b) cultivar la célula hospedadora en un medio de cultivo en condiciones en las que la célula hospedadora expresa el polipéptido β -glucosidasa codificado, en el que el nivel de expresión de β -glucosidasa de la célula hospedadora es al menos aproximadamente 2 veces superior al de C1 de tipo silvestre cultivada en las mismas condiciones. Puede medirse la expresión del polipéptido β -glucosidasa utilizando un ensayo de actividad β -glucosidasa con para-nitrofenil- β -D-glucopiranosido (pNPG), tal como el descrito en el Ejemplo 5. Puede medirse la expresión (es decir, la actividad) de la β -glucosidasa secretada. Como alternativa, puede determinarse la expresión de la actividad β -glucosidasa total en el cultivo. El péptido señal empleado en este método puede ser cualquier péptido señal heterólogo conocido en la técnica o puede ser el péptido señal de C1 β -glucosidasa de tipo silvestre (SEQ ID NO: 6). En algunas formas de realización, el nivel de sobreexpresión es al menos aproximadamente 5 veces, 10 veces, 12 veces, 15 veces, 20 veces, 25 veces, 30 veces ó 35 veces superior al de la expresión de β -glucosidasa del C1 de tipo silvestre.

Por lo general, la recuperación o aislamiento del polipéptido β -glucosidasa se da a partir del medio de cultivo de la célula hospedadora, la célula hospedadora o ambos, utilizando técnicas de recuperación de proteínas que son bien conocidas en la técnica, incluidas las descritas en el presente documento. Las células se recolectan por lo general por centrifugación, se rompen mediante medios físicos o químicos, y el extracto bruto resultante puede conservarse para una purificación adicional. Las células microbianas empleadas en la expresión de proteínas pueden romperse mediante cualquier método conveniente, incluidos ciclos de congelación-descongelación, sonicación, rotura mecánica o uso de agentes de lisis celular, u otros métodos, que son bien conocidos para los expertos en la materia.

El polipéptido resultante puede recuperarse/aislarse y opcionalmente purificarse mediante cualquiera de varios métodos conocidos en la técnica. Por ejemplo, el polipéptido puede aislarse del medio nutritivo mediante procedimientos convencionales incluidos, pero no limitados a, centrifugación, filtración, extracción, secado por pulverización, evaporación, cromatografía (por ejemplo, de intercambio iónico, de afinidad, de interacción hidrófoba, cromatofocalización, y de exclusión por tamaño), o precipitación. Pueden utilizarse etapas de renaturalización de proteínas, como se desee, para completar la configuración de la proteína madura. Finalmente, puede emplearse la cromatografía de líquidos de alto rendimiento (HPLC) en las etapas finales de purificación. La purificación de BGL1 se describe en la publicación de patente de EE.UU. US 2007/0238155. Además de las referencias citadas anteriormente, se conocen en la técnica diversos métodos de purificación, incluidos, por ejemplo, los expuestos en Sandana (1997) *Bioseparation of Proteins*, Academic Press, Inc.; Bollag *et al.* (1996) *Protein Methods*, segunda edición, Wiley-Liss, Nueva York; Walker (1996) *The Protein Protocols Handbook*, Humana Press, NJ; Harris y Angal (1990) *Protein Purification Applications: A Practical Approach*, IRL Press en Oxford, Oxford, Inglaterra; Harris y Angal, *Protein Purification Methods: A Practical Approach*, IRL Press en Oxford, Oxford, Inglaterra; Scopes (1993) *Protein Purification: Principles and Practice*, 3ª edición, Springer Verlag, Nueva York; Janson y Ryden (1998) *Protein Purification: Principles, High Resolution Methods and Applications*, segunda edición, Wiley-VCH, Nueva York; y Walker (1998) *Protein Protocols* en CD-ROM, Humana Press, NJ. En algunos casos, la proteína purificada puede purificarse casi a homogeneidad o puede constituir al menos un 20%, al menos un 30% o al menos un 50% de la proteína en la composición.

Pueden utilizarse métodos inmunológicos para purificar los polipéptidos β -glucosidasa. En un enfoque, un anticuerpo generado contra los polipéptidos β -glucosidasa (por ejemplo, contra un polipéptido que comprende la SEQ ID NO: 5) que utiliza métodos convencionales, se inmoviliza sobre perlas, se mezcla con medios de cultivo celular en condiciones en las que la β -glucosidasa se una, y se hace precipitar. En un enfoque relacionado, se utiliza inmunocromatografía.

Tal como se ha señalado, en algunas formas de realización la β -glucosidasa se expresa como una proteína de fusión que incluye una porción que no es enzima. En algunas formas de realización, la secuencia de β -glucosidasa se fusiona a un dominio que facilita la purificación. Tal como se utiliza en el presente documento, la

expresión "dominio que facilita la purificación" se refiere a un dominio que interviene en la purificación del polipéptido al que se fusiona. Los dominios de purificación adecuados incluyen péptidos quelantes de metales, módulos de histidina-triptófano que permiten la purificación sobre metales inmovilizados, una secuencia que se une a glutatión (por ejemplo, GST), un marcador hemaglutinina (HA) (que corresponde a un epítipo derivado de la proteína hemaglutinina de la gripe; Wilson *et al.* (1984) Cell, 37: 767), secuencias proteicas de unión a maltosa, el epítipo FLAG utilizado en el sistema de purificación por extensión/afinidad FLAGS (Immunex Corp, Seattle, Wash.), y similares. La inclusión de una secuencia conectora polipeptídica escindible por proteasas entre el dominio de purificación y el polipéptido HHDH es útil para facilitar la purificación. Un vector de expresión contemplado para su uso en las composiciones y los métodos descritos en el presente documento proporciona la expresión de una proteína de fusión que comprende un polipéptido de la invención fusionado a una región de polihistidina, separados por un sitio de escisión de enteroquinasa. Los residuos de histidina facilitan la purificación en IMAC (cromatografía de afinidad por iones metálicos inmovilizados, como se describe en Porath *et al.* (1992) Protein Expression and Purification 3:263-281) mientras que el sitio de escisión de enteroquinasa proporciona un medio para separar el polipéptido HHDH de la proteína de fusión. También pueden utilizarse vectores pGEX (Promega, Madison, Wis.) para expresar polipéptidos extraños como proteínas de fusión con la glutatión S-transferasa (GST). En general, tales proteínas de fusión son solubles y pueden purificarse fácilmente a partir de células lisadas por adsorción a perlas de ligando-agarosa (por ejemplo, glutatión-agarosa en el caso de fusiones con GST), seguido de elución en presencia de ligando libre.

VI. MÉTODOS DE USO DE POLIPÉPTIDOS β -GLUCOSIDASA Y CÉLULAS QUE EXPRESAN POLIPÉPTIDOS β -GLUCOSIDASA

Tal como se ha descrito anteriormente, los polipéptidos β -glucosidasa de la presente invención pueden utilizarse para catalizar la hidrólisis de un dímero de azúcar con la liberación del monómero de azúcar correspondiente, por ejemplo, la conversión de celobiosa con la liberación de glucosa. Por lo tanto, la presente invención proporciona un método de producción de glucosa, (a) proporcionando una celobiosa; y (b) poniendo en contacto la celobiosa con un polipéptido β -glucosidasa de la invención en condiciones suficientes para formar una mezcla de reacción para convertir la celobiosa en glucosa. El polipéptido β -glucosidasa puede utilizarse en tales métodos, ya sea en forma aislada o como parte de una composición, tal como cualquiera de los descritos en el presente documento. El polipéptido β -glucosidasa también puede proporcionarse en medios de cultivo celular o en un lisado celular. Por ejemplo, después de producir el polipéptido β -glucosidasa mediante cultivo de una célula hospedadora transformada con un vector o polinucleótido β -glucosidasa de la presente invención, no es necesario aislar la β -glucosidasa del medio de cultivo (es decir, si la β -glucosidasa es secretada en el medio de cultivo) o lisado celular (es decir, si la β -glucosidasa no es secretada en el medio de cultivo) o utilizarla en forma purificada para que sea útil en otros métodos de utilización del polipéptido β -glucosidasa. Cualquier composición, medio de cultivo celular o lisado celular que contenga un polipéptido β -glucosidasa de la presente invención puede ser adecuado para su uso en métodos que utilizan una β -glucosidasa. Por lo tanto, la presente invención proporciona adicionalmente un método de producción de glucosa: (a) proporcionando una celobiosa; y (b) poniendo en contacto la celobiosa con un medio de cultivo o lisado celular o composición que comprende un polipéptido β -glucosidasa de la presente invención en condiciones suficientes para formar una mezcla de reacción para convertir la celobiosa en glucosa.

La presente invención proporciona adicionalmente composiciones que son útiles para la conversión enzimática de la celobiosa en glucosa. Por ejemplo, pueden combinarse uno o más polipéptidos β -glucosidasa de la presente invención con otra enzima y/o un agente que modifique las propiedades de manejo de material a granel o procesabilidad adicional de la(s) β -glucosidasa(s) (por ejemplo, un agente fluidificante, agua, tampón, un tensioactivo, y similares), o que mejore la eficacia de la conversión de la celobiosa en glucosa, tal como se describe con más detalle más adelante. La otra enzima puede ser una β -glucosidasa diferente u otra enzima celular.

Mezclas de celulasas

Por ejemplo, en algunas formas de realización, la β -glucosidasa se combina con otras celulasas para formar una mezcla de celulasas. La mezcla de celulasas puede incluir celulasas seleccionados de entre celulasas CBH, EG y BG (por ejemplo, celulasas de *Trichoderma reesei* (por ejemplo, celulasas C2730 de *Trichoderma reesei* ATCC N° 25921, Sigma-Aldrich, Inc.), *Acidothermus cellulolyticus*, *Thermobifida fusca*, *Humicola grisea* y *Chrysosporium* sp.). Las enzimas de la mezcla de celulasas trabajan conjuntamente, lo que da como resultado la desrecristalización y la hidrólisis de la celulosa a partir de un sustrato de biomasa para producir azúcares solubles, tales como, pero no limitados a, la glucosa (Véase Brigham *et al.*, 1995, en Handbook on Bioethanol (C. Wyman ed.) págs. 119-141, Taylor y Francis, Washington DC).

Las mezclas de celulasas para una hidrólisis enzimática eficaz de la celulosa son conocidas (véase, por ejemplo, Viikari *et al.*, 2007, "Thermostable enzymes in lignocellulose hydrolysis" Adv Biochem Eng Biotechnol 108:121-45, y las publicaciones de patentes de EE.UU. US 2009/0061484, US 2008/0057541 y US 2009/0209009 de logen Energy Corp.). En algunas formas de realización, las mezclas de enzimas naturales o recombinantes purificadas se combinan con materia prima celulósica o un producto de hidrólisis de la celulosa. Como alternativa, o además, puede utilizarse una o más células que producen celulasas naturales o recombinantes.

Otros componentes de las composiciones de β -glucosidasa

Los polipéptidos β -glucosidasa de la presente invención pueden utilizarse en combinación con otros ingredientes opcionales, tales como un tampón, un tensioactivo y/o un agente limpiador. Puede utilizarse un tampón con un polipéptido β -glucosidasa de la presente invención (opcionalmente en combinación con otras celulasas, incluida otra β -glucosidasa) para mantener un pH deseado dentro de la solución en la que se emplea la β -glucosidasa. La concentración exacta de tampón empleado dependerá de varios factores que el experto en la técnica puede determinar. Los tampones adecuados son bien conocidos en la técnica. Puede utilizarse adicionalmente un tensioactivo en combinación con las celulasas de la presente invención. Los tensioactivos adecuados incluyen cualquier tensioactivo compatible con la β -glucosidasa y otras celulasas opcionales que se estén utilizando. Los tensioactivos ejemplares incluyen un tensioactivo aniónico, un tensioactivo no iónico y tensioactivos anfólicos.

Los tensioactivos aniónicos adecuados incluyen, pero no se limitan a, alquilbencenosulfonatos ramificados o lineales; sulfatos de alquil éter o alquenil éter con grupos alquilo o grupos alquenilo lineales o ramificados; sulfatos de alquilo o alquenilo; olefinsulfonatos; alcanosulfonatos, y similares. Los contraiones adecuados para los tensioactivos aniónicos incluyen, por ejemplo, iones de metales alcalinos, tales como sodio y potasio; iones de metales alcalinotérreos, tales como calcio y magnesio; ion amonio; y alcanolaminas que tienen de 1 a 3 grupos alcohol con un número de átomos de carbono de 2 ó 3. Los tensioactivos anfólicos adecuados para su uso en la práctica de la presente invención incluyen, por ejemplo, sulfonatos de sales de amonio cuaternario, tensioactivos anfólicos de tipo betaína, y similares. Los tensioactivos no iónicos adecuados incluyen generalmente éteres de polioxialquileo, así como alcanolamidas de ácidos grasos superiores o aductos de óxido de alquileo de las mismas, monoésteres de ácidos grasos de glicerina, y similares. También pueden emplearse mezclas de tensioactivos como se conoce en la técnica.

Producción de azúcares fermentables a partir de biomasa celulósica

Los polipéptidos β -glucosidasa de la presente invención, así como cualquier composición, medio de cultivo o lisado celular que comprende tales polipéptidos β -glucosidasa, pueden utilizarse en la producción de monosacáridos, disacáridos u oligómeros de un mono o disacárido como materia prima química o de fermentación a partir de biomasa. Tal como se utiliza en el presente documento, el término "biomasa" se refiere a material biológico vivo o muerto que contiene un sustrato polisacárido, tal como, por ejemplo, celulosa, almidón, y similares. Por lo tanto, la presente invención proporciona un método para convertir un sustrato de biomasa en un azúcar fermentable, comprendiendo el método poner en contacto un medio de cultivo o lisado celular que contiene un polipéptido β -glucosidasa según la invención, con el sustrato de biomasa en condiciones adecuadas para la producción del azúcar fermentable. La presente invención proporciona adicionalmente un método para convertir un sustrato de biomasa en un azúcar fermentable (a) pretratando un sustrato de celulosa para aumentar su susceptibilidad a la hidrólisis; (b) poniendo en contacto el sustrato de celulosa pretratado de la etapa (a) con una composición, medio de cultivo o lisado celular que contiene un polipéptido β -glucosidasa de la presente invención (y opcionalmente otras celulasas) en condiciones adecuadas para la producción del azúcar fermentable.

En algunas formas de realización, la biomasa incluye sustratos celulósicos, incluidos pero no limitados a, madera, pasta de madera, pasta de papel, rastrojo de maíz, fibra de maíz, arroz, desechos de la elaboración de pulpa y papel, plantas herbáceas o leñosas, pulpa vegetal o de fruta, grano de destilería, hierbas, cascarilla de arroz, paja de trigo, algodón, cáñamo, lino, sisal, mazorcas de maíz, bagazo de caña de azúcar, pasto varilla, y mezclas de los mismos. La biomasa puede pretratarse opcionalmente para aumentar la susceptibilidad de la celulosa a la hidrólisis utilizando métodos conocidos en la técnica tales como pretratamientos químicos, físicos y biológicos (por ejemplo, explosión de vapor, reducción a pulpa, desfibrado, hidrólisis ácida, exposición a disolventes, y similares, así como combinaciones de los mismos). En algunas formas de realización, la biomasa comprende plantas transgénicas que expresan enzimas ligninasa y/o celulasa que degradan la lignina y la celulosa. Véase, por ejemplo, el documento US 20080104724. La biomasa puede incluir celobiosa y/o puede tratarse enzimáticamente para generar celobiosa para la conversión en un azúcar soluble (por ejemplo, glucosa).

En algunas formas de realización, pueden hacerse reaccionar el polipéptido β -glucosidasa y lisados celulares, medios de cultivo celular y composiciones que contienen polipéptido β -glucosidasa con la biomasa o biomasa pretratada a una temperatura en el intervalo comprendido entre aproximadamente 25°C y aproximadamente 100°C, entre aproximadamente 30°C y aproximadamente 90°C, entre aproximadamente 30°C y aproximadamente 80°C, entre aproximadamente 40°C y aproximadamente 80°C, y entre aproximadamente 35°C y aproximadamente 75°C. También pueden hacerse reaccionar la biomasa con los lisados celulares, medios de cultivo celular y composiciones que contienen polipéptido β -glucosidasa y polipéptidos 11-glucosidasa a una temperatura de aproximadamente 25°C, a aproximadamente 30°C, a aproximadamente 35°C, a aproximadamente 40°C, a aproximadamente 45°C, a aproximadamente 50°C, a aproximadamente 55°C, a aproximadamente 60°C, a aproximadamente 65°C, a aproximadamente 70°C, a aproximadamente 75°C, a aproximadamente 80°C, a aproximadamente 85°C, a aproximadamente 90°C, a aproximadamente 95°C y a aproximadamente 100°C. Además de las temperaturas descritas anteriormente, las condiciones adecuadas para la conversión de un sustrato de

biomasa en un azúcar fermentable que emplea un polipéptido β -glucosidasa de la presente invención (opcionalmente en una composición, medio de cultivo celular o lisado celular) incluyen llevar a cabo el proceso a un pH en un intervalo de aproximadamente pH 3,0 a aproximadamente 8,5, aproximadamente pH 3,5 a aproximadamente 8,5, aproximadamente pH 4,0 a aproximadamente 7,5, aproximadamente pH 4,0 a aproximadamente 7,0 y aproximadamente pH 4,0 a aproximadamente 6,5. Los expertos habituales en la materia comprenderán que los tiempos de reacción para convertir un sustrato de biomasa particular en un azúcar fermentable pueden variar, pero que puede determinarse fácilmente el tiempo de reacción óptimo. Los tiempos de reacción ejemplares pueden estar en el intervalo comprendido entre aproximadamente 1,0 y aproximadamente 240 horas, entre aproximadamente 5,0 y aproximadamente 180 horas, y entre aproximadamente 10,0 y aproximadamente 150 horas. Por ejemplo, el tiempo de incubación puede ser de al menos 1 hora, al menos 5 horas, al menos 10 horas, al menos 15 horas, al menos 25 horas, al menos 50 horas, al menos 100 horas, al menos 180, y similares.

La reacción de la β -glucosidasa con el sustrato de biomasa o sustrato de biomasa pretratado en estas condiciones puede dar como resultado la liberación de cantidades considerables de los azúcares solubles a partir del sustrato. Por ejemplo, puede disponerse de al menos un 20%, al menos un 30%, al menos un 40%, al menos un 50%, al menos un 60%, al menos un 70%, al menos un 80%, al menos un 90% o más de azúcar soluble, en comparación con la liberación de azúcar mediante la C1 de tipo silvestre. En alguna forma de realización, la cantidad de azúcar soluble disponible puede ser al menos 2 veces, al menos 3 veces, al menos 4 veces o al menos 5 veces superior a la disponible mediante la C1 de tipo silvestre en las mismas condiciones. En algunas formas de realización, los azúcares solubles comprenderán glucosa.

Los azúcares solubles producidos mediante los métodos de la presente invención pueden utilizarse para producir un alcohol (tal como, por ejemplo, etanol, butanol, y similares). Por consiguiente, la presente invención proporciona un método de producción de un alcohol, en el que el método comprende (a) proporcionar un azúcar fermentable producido utilizando un polipéptido β -glucosidasa de la presente invención en los métodos descritos anteriormente; (b) poner en contacto el azúcar fermentable con un microorganismo fermentador para producir el alcohol u otro producto metabólico; y (c) recuperar el alcohol u otro producto metabólico.

En algunas formas de realización, el polipéptido β -glucosidasa de la presente invención, o la composición, medio de cultivo celular o lisado celular que contiene el polipéptido β -glucosidasa puede utilizarse para catalizar la hidrólisis de un sustrato de biomasa hasta un azúcar fermentable en presencia de un microorganismo fermentador tal como una levadura (por ejemplo, *Saccharomyces* sp., tal como, por ejemplo, *S. cerevisiae*, *Zymomonas* sp., *E. coli*, *Pichia* sp., y similares) u otros microorganismos fermentadores de C5 ó C6 que son bien conocidos en la técnica, para producir un producto final tal como etanol. En este proceso de sacarificación y fermentación simultáneas (SSF) los azúcares fermentables (por ejemplo, glucosa y/o xilosa) se eliminan del sistema mediante el proceso de fermentación.

Los azúcares solubles producidos utilizando un polipéptido β -glucosidasa de la presente invención también pueden utilizarse en la producción de otros productos finales tal como, por ejemplo, acetona, un aminoácido (por ejemplo, glicina, lisina, y similares), un ácido orgánico (por ejemplo, ácido láctico, y similares), glicerol, un diol (por ejemplo, 1,3 propanodiol, butanodiol, y similares) y piensos.

Un experto en la materia comprenderá fácilmente que las composiciones de polipéptido β -glucosidasa de la presente invención pueden utilizarse en forma de solución acuosa o concentrado sólido. Cuando se emplean soluciones acuosas, la solución de β -glucosidasa puede diluirse fácilmente para permitir las concentraciones precisas. Un concentrado puede estar en cualquier forma reconocida en la técnica incluido, por ejemplo, líquidos, emulsiones, suspensiones, gel, pastas, gránulos, polvos, un aglomerado, un disco sólido, así como otras formas que son bien conocidas en la técnica. También pueden utilizarse otros materiales con o incluidos en la composición de β -glucosidasa de la presente invención como se desee, incluidas piedras, piedra pómez, cargas, disolventes, activadores enzimáticos y agentes anti-redeposición dependiendo del uso pretendido de la composición.

Además de utilizarse para la conversión de biomasa celulósica, los polipéptidos β -glucosidasa y las composiciones de los mismos también pueden utilizarse en la industria de alimentos y bebidas, por ejemplo, en el proceso de elaboración del vino para la liberación eficaz de los monoterpenoles (véase, por ejemplo, Yanai y Sato (1999) Am. J. Enol Eitic., 50: 231 - 235) y para la preparación de tofu enriquecido en glicona isoflavonas (véase, por ejemplo, Mase *et al.*, (2004). J. Appl. Glycosci., 51:211-216). Los polipéptidos β -glucosidasa de la presente invención también pueden emplearse en las composiciones de detergente para mejorar el rendimiento de limpieza (véase, por ejemplo, los documentos USP 7.244.605; US 5.648.263 y WO 2004/048592).

Los aspectos anteriores y otros aspectos de la invención pueden comprenderse mejor en relación a los siguientes ejemplos no limitativos.

VIII. EJEMPLOS

Ejemplo 1. Clonación de ADN genómico

Se aisló ADN genómico de células C1 mediante procedimientos convencionales. En resumen, se sembró un inóculo de hifas en un medio de crecimiento y se dejó crecer durante 72 horas a 35°C. La estera micelial se recogió por centrifugación, se lavó y se añadieron 50 µl de tampón de extracción de ADN (TRIS 200 mM pH 8,0; NaCl 250 mM; EDTA 125 mM; SDS al 0,5%). Se molió el micelio con una trituradora cónica, se volvió a extraer con 250 µl de tampón de extracción y se centrifugó la suspensión. Se transfirió el sobrenadante a un nuevo tubo que contenía 300 µl de isopropanol. Se recogió el ADN por centrifugación, se lavó dos veces con etanol al 70%, y se diluyó en 100 µl de agua.

Se clonó la región que incluía el gen *bgl1* de β-glucosidasa de cada una de tres cepas C1, incluidas las células de tipo silvestre y dos líneas celulares Alp1⁺ diferentes, y se secuenciaron. La clonación se llevó a cabo utilizando PCR con los cebadores internos y externos que se muestran en la Tabla 3. El producto de amplificación de los cebadores internos es la porción que codifica la proteína del gen *bgl1*, mientras que el cebador externo incluye las secuencias flanqueantes 5' y 3'.

TABLA 3

Cebadores internos:		
C1bgl_fwd1	ATGAAGGCTGCTGCGCTTTC	SEQ ID NO:22
C1bgl_rev1	TCATTAAGGAAGCTCAATCTTGAGATC	SEQ ID NO:23
Cebadores externos:		
C1_bgl_fwd2	TCTCTGCCGGTGCCATCAATCATCT	SEQ ID NO:24
C1_bgl_rev2	GCTCACCGGAAGTTGCCAAGTGCT	SEQ ID NO:25

La secuencia de *bgl1* se proporciona en la Figura 1 (secuencia superior). En particular, la secuencia genómica de C1 que se obtuvo difería de las secuencias publicadas en cuatro posiciones (correspondientes a los nucleótidos 1440, 1451, 1511 y 1521 de la SEQ ID NO: 1).

Ejemplo 2. Clonación de ADNc

Se clonó ADNc de *bgl1* de C1 y se secuenció como se describe a continuación.

Extracción de ARN a partir de medios de fermentación de C1

Se purificó ARN a partir de un cultivo de C1 de 3 días (cepa Alp1⁺) cultivado a 35°C con agitación a 250 rpm. Se transfirieron 500 µl-1.200 µl del caldo de cultivo a un mortero enfriado y se trituraron minuciosamente con la mano de mortero en presencia de N2 líquido. Se transfirió el polvo de células a un tubo Eppendorf y se añadieron a la muestra 1,500 µl de reactivo Trizol. Después de mezclar, se incubó durante 20 minutos a temperatura ambiente y se centrifugó a 12.000 g, 4°C, durante 15 minutos. Se transfirió el homogeneizado transparente a un tubo de centrifuga de 2 ml y se añadieron 300 µl de cloroformo, se mezcló y se centrifugó a 12.000 g 4°C durante 30 minutos. Se transfirió la fase acuosa superior a un nuevo tubo de 2 ml y se hizo precipitar por adición de 1.200 µl de EtOH al 96%. Se purificó adicionalmente la muestra utilizando el kit RNeasy de Qiagen utilizando las instrucciones del fabricante. Brevemente: se transfirieron 700 µl de muestra precipitada a la minicolumna de centrifugación del RNeasy y se centrifugó durante 20 segundos a 10.000 g a 20°C. Se añadieron 350 µl de tampón RW1 y se centrifugó durante 20 segundos a 10.000 g a 20°C. La solución de ADNasa se añadió directamente a la membrana de la columna de centrifugación del RNeasy, y se incubó a temperatura ambiente durante 25 minutos. Se añadieron 350 µl de tampón RW1 y se centrifugó durante 20 segundos a 10.000 g a 20°C. Se añadieron a la columna de centrifugación dos veces 500 µl de tampón RPE y se centrifugó durante 20 segundos a 10.000 g a 20°C. Se colocó la columna de centrifugación del RNeasy en un nuevo tubo de recogida de 2 ml y se centrifugó durante 1 minuto a 10.000 g a 20°C. A continuación, se colocó la columna de centrifugación del RNeasy en un nuevo tubo de recogida de 1,5 ml y se añadieron 30 µl de agua sin ARNasa directamente a la membrana de la columna de centrifugación y se centrifugó durante 1 minuto a 10.000 g a 20°C. La concentración de la solución de ARN total fue aproximadamente 1,9 µg/ul.

Síntesis de la primera hebra mediante transcriptasa inversa Superscript III

A un tubo de microcentrífuga sin nucleasa se añadieron los siguientes componentes:

1. 5 µl de oligo(dT) (400 ng/ul) (Qiagen Cat. N° 70DT01-1)
2. 1 µg de ARN total
3. 1 µl de dNTP 10 mM
4. Agua destilada estéril hasta 13 µl

Se calentó la mezcla a 65°C durante 5 minutos y se incubó en hielo durante 1 minuto y 4 µl de 5X tampón de primera hebra, 1 µl de DTT 0,1 M, 1,3 µl de inhibidor de ARNasa recombinante RNaseOUT™ (Cat. N° 10777-019) y 2 µl de transcriptasa inversa SuperScript™ III RT. Se incubó la mezcla a 50°C durante 60 minutos y se interrumpió la reacción mediante calentamiento a 70°C durante 15 minutos.

Clonación por PCR de ADNc de *bgl1*

Para clonar ADNc de longitud completa de *bgl1* a partir de la mezcla RT-PCR, se añadieron 2 µl de la mezcla RT-PCR, 10 µl de 5X tampón Phusion HF, 1 µl de dNTP (Σ10mM), 1 µl de cebador cdx09050 (10 µM), 1 µl de cebador cdx09052 (10 µM), 34,5 µl de agua MQ y 0,5 µl de polimerasa Phusion HF* (Finnzymes) y se mezclaron. Las condiciones de PCR fueron: desnaturalización inicial: 98°C-30 segundos, seguido de 35 ciclos de 98°C-10 segundos de desnaturalización y 72°C, 2 minutos 40 segundos de elongación, y un ciclo de extensión final de 72°C durante 5 minutos. El producto de PCR se purificó con columna de centrifugación EZ-10 según las instrucciones del fabricante y se clonó en pCRBlunt utilizando el kit de clonación StrataClone Ultra Blunt PCR (Stratagene) según las instrucciones del fabricante.

TABLA 4

Nombre del cebador	Secuencia (5'-3')	SEQ ID NO:
cdx09050	GGCTCATGAAGGCTGCTGCGCTTTCCTGC	26
cdx09052	GCCGAATTCTCAAGGAAGCTCAATCTTGAGATCC	27
TcbhC1bgl_R1	GCGTGTCTCAGAACCTCCTCAGAGAGGTTTCGTTTACTTACTTAT TATCAAGGAAGCTCAATCTTGAGATCC	28
PcbhC1bgl_F	GTCTTCAGATCAAGCAACTGTGTGCTGATCC TCTTCCGTCATGAAGGCTGCTGCGCTTTC	10

La Figura 2 muestra la secuencia de ADNc obtenida (véase también la Figura 1), junto con la secuencia proteica predicha. Tal como se ilustra en la Figura 3, la estructura del exón del gen *bgl1* se diferencia de la descrita anteriormente (véase la publicación de patente US 2007/0238155). La estructura del exón que se refleja en la secuencia de ADNc fue coherente con el análisis de la secuencia genómica correspondiente utilizando los algoritmos de predicción de genes Augustus v.2.03 y Genemark-ES v.2 (véase Stanke *et al.*, 2006 "AUGUSTUS: ab initio prediction of alternative transcripts" Nucleic Acids Res. 34 (publicación en el servidor Web): W435-9; Ter-Hovhannisyan *et al.*, 2008, Genome Res. 18:1979-1990). Como consecuencia de las diferencias en la secuencia de ADNc, se ha identificado que la secuencia de la proteína BGL1 se diferencia de la secuencia descrita anteriormente en el extremo amino terminal e internamente.

Puesto que BGL1 se secreta, se espera que tenga un péptido señal. SignalP (Bendtsen *et al.*, 2004, "Improved prediction of signal peptides: SignalP 3.0". J Mol Biol. 340(4):783-95) predice un sitio de escisión ("/") entre los aminoácidos 19 y 20 de la SEQ ID NO: 4.

Extremo N-terminal- MKAAALSCLFGSTLAVAGA // IESRKVHQKPLARSEPFYPS ... -extremo C-terminal [SEQ ID NO: 12]

Además, los resultados de los investigadores demuestran que BGL1 comprende ácido aspártico en la posición 358, glutamina en la posición 381 y ácido glutámico en la posición 385, mientras que la secuencia publicada proporciona, histidina, histidina y lisina, respectivamente, en las posiciones correspondientes.

Ejemplo 3. Construcción de vectores de expresión de *bgl1* con promotores de C1

Para producir la secuencia de *bgl1* bajo el control del promotor de C1 CBH1a, se clonó la secuencia genómica de *bgl1* en un vector que incluye una secuencia promotora de CBH1a (SEQ ID NO: 8, véase la Figura 4; véase también la publicación PCT WO 01/79507), secuencias reguladoras 3' que incluyen un terminador de la transcripción CBH1a, y un marcador de resistencia a ampicilina. Además, se clonó un casete génico de resistencia a fleomicina en el vector en 3' hasta el terminador. Mediante la técnica de clonación SLIC (Mamie *et al.*, 2007, Nature Methods 4:251- 56) se amplificó el ADN correspondiente a la porción de preproteína de la secuencia de *bgl1* utilizando los cebadores PcbhC1bgl_F y TcbhC1bgl_R1 (véase la tabla 4 en el Ejemplo 2, *supra*). El producto resultante se clonó en el vector digerido con PmlI/PacI en sentido 3' respecto al promotor de CBH1a para crear un vector de expresión que expresase el transcrito proteína BGL1 bajo el control del promotor de CBH1a.

Ejemplo 4. Método de transformación de C1

Preparación de protoplastos

Se inocularon células C1 (cepa Alp1[□]) en 100 ml de medio de crecimiento en un matraz Erlenmeyer de 500 ml utilizando 10⁶ esporas/ml. Se incubó el cultivo durante 48 horas a 35°C, 250 rpm. Para recolectar el micelio, se filtró el cultivo sobre un filtro de Myracloth estéril (Calbiochem) y se lavó con 100 ml de solución de 1700 mOsmol NaCl/CaCl₂ (NaCl 0,6 M, CaCl₂*H₂O 0,27 M). Se transfirió el micelio lavado a un tubo de 50 ml y se pesó. Se disolvió Caylasa (20 mg/gramo de micelio) en 1700 mOsmol NaCl/CaCl₂ y se esterilizó con UV durante 90 segundos. Se añadieron 3 ml de solución de Caylasa estéril al tubo que contenía los micelios lavados y se mezcló. Se añadieron al tubo otros 15 ml de solución de 1700 mOsmol NaCl/CaCl₂ y se mezcló. Se incubó la suspensión de micelio/Caylasa a 30°C, 70 rpm durante 2 horas. Se recolectaron los protoplastos por filtración a través de un filtro Myracloth estéril en un tubo de 50 ml estéril. Se añadieron al flujo que atraviesa la columna 25 ml de STC frío y centrifugó a 2.720 rpm durante 10 minutos a 4°C. Se resuspendió el sedimento en 50 ml de STC (sorbitol 1,2 M, CaCl₂*H₂O 50 mM, NaCl 35 mM, Tris-HCl 10 mM) y se centrifugó de nuevo. Después de las etapas de lavado, se resuspendió el sedimento en 1 ml de STC.

Transformación

Se pipetearon 6 µg de ADN de plásmido en la parte inferior de un tubo estéril de 15 ml y se añadieron 1 µl de ácido aurintricarboxílico y 100 µl de protoplastos. Se mezcló el contenido y se incubó el protoplasto con el ADN a temperatura ambiente durante 25 minutos. Se añadieron 1,7 ml de solución de PEG4000 (PEG4000 al 60% [polietilenglicol, peso molecular medio 4.000 daltons], CaCl₂*H₂O 50 mM, NaCl 35 mM, Tris-HCl 10 mM) y se mezclaron bien. Se mantuvo la solución a temperatura ambiente durante 20 minutos. Se llenó el tubo con STC, se mezcló y se centrifugó a 2.500 rpm durante 10 minutos a 4°C. Se tiró el STC y se resuspendió el sedimento en el STC restante y se sembró en placas de medio mínimo. Se incubaron las placas durante 5 días a 35°C. Se volvieron a sembrar en estrías las colonias y se comprobó la presencia del plásmido integrado. Se seleccionaron varios aislados y se ensayaron para la expresión de Bgl1.

Ejemplo 5. Ensayo de actividad β-glucosidasa

Se utilizó un ensayo basado en pNPG colorimétrico (p-nitrofenil-β-D-glucopiranosido) para medir la actividad β-glucosidasa. Se determinó la sobreexpresión de β-glucosidasa en base a la actividad de la proteína secretada con respecto a la C1 de tipo silvestre. Se cultivaron células no transformadas (por ejemplo, de tipo silvestre) o transformadas en placas de 96 pocillos, y se recogió el medio. El medio que contenía proteína secretada se diluyó 40 veces en tampón de Na-acetato 50 mM, pH 5. Se añadieron 20 µl de la mezcla de enzima diluida que contenía la enzima β-glucosidasa, a 80 µl de pNPG 2 mM (Sigma-Aldrich, Inc. St. Louis, MO), Na-acetato 50 mM pH5. Se incubaron las reacciones a pH 5, 50°C durante 10 minutos. Se inactivó la mezcla de reacción con 100 µl de solución pH 11 de carbonato de sodio 1 M. Se midió la absorbancia de la solución a 405 nm para determinar la conversión del pNPG en p-nitrofenol. La liberación de p-nitrofenol ($\epsilon = 17.700 \text{ M}^{-1} \text{ cm}^{-1}$) se mide a 405 nm para calcular la actividad β-glucosidasa.

Ejemplo 6. Sobreexpresión de BGL1

El vector de expresión descrito en el Ejemplo 3, que tenía una secuencia de *bgl1* bajo el control del promotor de C1 CBH1a, se co-transformó en células C1 (Alp1[□], cepa pyr5[□]) junto con un plásmido que expresaba *pyr*. Se seleccionaron los transformantes individuales y se cultivaron en placas de 96 pocillos. Se recogió el medio de los transformantes y de las células de control y se ensayó para la actividad β-glucosidasa utilizando el ensayo de pNPG. Se comparó la actividad pNPG de los transformantes con las células de control no transformadas (descrito anteriormente). Se midió la actividad como se describe en el Ejemplo 5. El mayor nivel de sobreexpresión observada fue aproximadamente una sobreexpresión 35 veces mayor, teniendo las 10 mayores expresoras una expresión 30-35 veces mayor que el tipo silvestre.

Aunque la presente invención se ha descrito con referencia a las formas de realización específicas de la misma, los expertos en la materia deben entender que pueden realizarse diversos cambios y que pueden sustituirse los equivalentes sin alejarse del alcance de la invención. Además, pueden realizarse muchas modificaciones para adaptar una situación, material, composición de materia, proceso, etapa de proceso o etapas particulares, para conseguir los beneficios proporcionados por la presente invención sin alejarse del alcance de la presente invención. Todas estas modificaciones pertenecen al alcance de las reivindicaciones adjuntas al presente documento.

LISTADO DE SECUENCIAS

<110> Codexis, Inc.

Clark, Louis

Baidyaroy, Dipnath

Szabo, Lorand

<120> C1 beta-glucosidasa recombinante para la producción de azúcares a partir de biomasa celulósica

<130> 026501-001810PC

<150> US 61/247,379

<151> 30-09-2009

<160> 31
<170> PatentIn versión 3.5
<210> 1
<211> 2981
5 <212> ADN
<213> *Myceliophthora thermophila*
<400> 1

10 atgaaggctg ctgcgctttc ctgcctcttc ggcagtaccc ttgccgttgc aggcgccatt
60
gaatcgagaa aggtatggac gggctttcgt caaagactcg ctccccgata aacttccctt
15 120
ttcatccaga ccaccccaac cctcccagtc ctgcttcgag cactatctct tcgggcagca
180
ccccaccac atccactcag attagcggcg acaccgttga ctgttgcaat ccgcaatcga
20 240
catgcaactt ccagccgcag cccaatggct gctcacgctt cccgcgaaag cctcacttgc
300
25 tgacaatcat cgtcaggttc accagaagcc cctcgcgaga tctgaacctt tttaccgtc
360
gccatggatg aatcccaacg ccgacggctg ggcggaggcc tatgccagg ccaagtcctt
30 420
tgtctcccaa atgactctgc tagagaaggt caacttgacc acgggagtcg ggttaagttt
480

35

40

45

50

55

60

65

gtcattttgt ccaggttaaca tgcaaattgg tctgctaaca ataacttacc gtagctgggg
540
5 ggctgagcag tgcgtcggcc aagtgggccc gateccctcg cttggacttc gcagtctgtg
600
catgcatgac tccctctctg gcatccgagg agccgactac aactcagcgt tccctctctg
10 660
ccagaccggt gctgctacct gggatcgcgg tctgatgtac cgtcgcggct acgcaatggg
720
15 ccaggaggcc aaaggcaagg gcatcaatgt ccttctcgga ccagtcgcgg gcccccttgg
780
ccgcatgccc gagggcggtc gtaactggga aggcttcgct ccggatcccg tccttaccgg
840
20 catcggcatg tccgagacga tcaagggcat tcaggatgct ggcgtcatcg cttgtgcgaa
900
gcactttatt ggaaacgagc agggtgagta gtcaaagacg ggccgtctcg gaccgcggc
25 960
ttcaagctgc tgactctgct gcagagcact tcagacaggt gccagaagcc cagggatacg
1020
30 gttacaacat cagcgaaacc ctctcctcca acattgacga caagaccatg cagcagctct
1080
acctttggcc gtttgccgat gccgtccggg ccggcgctcg ctctgtcatg tgctcgtacc
1140
35 agcagggtcaa caactcgtac gcctgccaga actcgaagct gctgaacgac ctctcaaga
1200
acgagcttgg gtttcagggc ttcgtcatga gcgactggca ggcacagcac actggcgag
40 1260
caagcgccgt ggctggtctc gatatgtcca tgccggggca caccagttc aacactggcg
1320
45 tcagtttctg gggcgccaat ctcaacctcg ccgtctctca cggcacagtc cctgcctacc
1380
gtctcgacga catggccatg cgcacatg cgcctctctt caaggtcacc aagaccaccg
1440
50 acctggaacc gatcaacttc tccttctgga ccgacgacac ttatggcccg atccactggg
1500

55

60

65

ccgccaagca gggctaccag gagattaatt cccacgttga cgtccgcgcc gaccacggca
1560

5 acctcatccg ggagattgcc gccaaaggta cgggtgctgct gaagaatacc ggctctctac
1620

10 ccctgaacaa gccaaagttc gtggccgtca tcggcgagga tgctgggtcg agccccaacg
1680

ggcccaacgg ctgcagcgac cgcggctgta acgaaggcac gctcgccatg ggctggggat
1740

15 ccggcacagc caactatccg tacctcgttt cccccgacgc cgcgctccag gcccgggcca
1800

tccaggacgg cacgaggtac gagagcgtcc tgtccaacta cgccgaggaa aagacaaagg
1860

20 ctctgggtct gcaggccaat gcaaccgcca tcgtcttcgt caatgccgac tcaggcgagg
1920

25 gctacatcaa cgtggacggg aacgagggcg accgtaagaa cctgactctc tggaacaacg
1980

gtgatactct ggtcaagaac gtctcgagct ggtgcagcaa caccatcgtc gtcattccact
2040

30 cggctggccc ggtcctcctg accgattggg acgacaaccc caacatcacg gccattctct
2100

35 gggctgggtct tccgggccag gagtcgggca actccatcac cgacgtgctt tacggcaagg
2160

tcaacccccg cgcgcgctcg cccttcactt ggggcaagac ccgcgaaagc tatggcgcg
2220

40 acgtcctgta caagccgaat aatggcaatg gtgcgcccc aacaggacttc accgagggcg
2280

45 tcttcacga ctaccgctac ttcgacaagg ttgaagatga ctcggtcac tacgagttcg
2340

gccacggcct gagctacacc accttcgagt acagcaacat ccgcgtcgtc aagtccaacg
2400

50 tcagcgagta ccggcccacg acgggcacca cggcccaggc ccgacggtt ggcaacttct
2460

ccaccgacct cgaggactat ctcttcccca aggacgagtt ccctacatc taccagtaca
2520

55

60

65

5 tctaccegta cctcaacacg accgaccccc ggagggcctc ggccgatccc cactacggcc
 2580
 agaccgccga ggagttcctc ccgccccacg ccaccgatga cgacccccag ccgctcctcc
 2640
 10 ggtcctcggg cggaaaactcc cccggcgga accgccagct gtacgacatt gtctacacaa
 2700
 tcacggccga catcacgaat acgggctccg ttgtaggcga ggaggtaccg cagctctacg
 2760
 15 tctcgtcggg cggccccgag gatcccaagg tgcagctgcg cgactttgac aggatgcgga
 2820
 tcgaacccgg cgagacgagg cagttcaccg gccgcctgac gcgcagagat ctgagcaact
 20 2880
 gggacgtcac ggtgcaggac tgggtcatca gcaggtatcc caagacggca tatgttggga
 2940
 25 ggagcagccg gaagttggat ctcaagattg agcttccttg a
 2981
 30 <210> 2
 <211> 2613
 <212> ADN
 <213> Secuencia artificial
 <220>
 35 <223> Constructo de ADN sintético de la secuencia de ADNc que codifica la C1 beta-glucosidasa de *Myceliophthora thermophila*
 <400> 2
 40 atgaaggctg ctgcgcttcc ctgcctcttc ggcagtaacc ttgccgttgc aggcgccatt
 60
 gaatcgagaa aggttcacca gaagccccctc gcgagatctg aaccttttta cccgtcgcca
 45 120
 tggatgaatc ccaacgccga cggctgggag gaggcctatg cccaggccaa gtcctttgtc
 180
 50 tcccaaatga ctctgctaga gaaggtcaac ttgaccacgg gactcggctg gggggctgag
 240
 cagtgcgtcg gccaaagtggg cgcgatccct cgccttggac ttgcagctct gtgcatgcat
 300
 55
 60
 65

gactccccctc tcggcatccg aggagccgac tacaactcag cgttccccctc tggccagacc
360
5 gttgctgcta cctgggatcg cggctctgatg taccgtcgcg gctacgcaat gggccaggag
420
gccaagggca agggcatcaa tgtccttctc ggaccagtcg ccggccccctc tggccgcatg
10 480
cccgagggcg gtcgtaactg ggaaggcttc gctccggatc ccgtccttac cggcatcggc
540
15 atgtccgaga cgatcaaggg cattcaggat gctggcgta tcgcttgtgc gaagcacttt
600
attggaaacg agcaggagca cttcagacag gtgccagaag ccaggggata cggttacaac
20 660
atcagcgaaa ccctctctc caacattgac gacaagacca tgcacgagct ctacctttgg
720
25 ccgtttgccg atgcgctcg ggccggcgtc ggctctgtca tgtgctcgta ccagcaggtc
780
aacaactcgt acgcctgcc gaactcgaag ctgctgaacg acctcctcaa gaacgagctt
840
30 gggtttcagg gcttcgtcat gagegactgg caggcacagc aactggcgc agcaagcgcc
900
gtggctggtc tcgatatgtc catgccgggc gacacccagt tcaacactgg cgtcagtttc
35 960
tgggggcgcca atctcaccct cgcgctctc aacggcacag tccctgccta ccgtctcgac
1020
40 gacatggcca tgcgcatcat ggccgccctc ttcaaggta ccaagaccac cgacctggaa
1080
ccgatcaact tctccttctg gaccgacgac acttatggcc cgatccactg ggccgccaag
1140
45 cagggtacc aggagattaa ttcccacgtt gacgtccgcg ccgaccacgg caacctcatc
1200
cgggagattg ccgccaaggg tacggtgctg ctgaagaata ccggtctctc acccctgaac
50 1260
aagccaaagt tcgtggccgt catcggcgag gatgctgggt cgagcccaa cgggccaac
1320

55

60

65

ggctgcagcg accgcggctg taacgaaggc acgctcgcca tgggctgggg atccggcaca
 1380
 5 gccaaactatc cgtacctcgt ttccccgac gccgcgctcc aggcccgggc catccaggac
 1440
 10 ggcacgaggt acgagagcgt cctgtccaac tacgccgagg aaaagacaaa ggctctggtc
 1500
 tcgcaggcca atgcaaccgc catcgtcttc gtcaatgccg actcaggcga gggctacatc
 1560
 15 aacgtggacg gtaacgaggg cgaccgtaag aacctgactc tctggaacaa cggtgatact
 1620
 ctggtcaaga acgtctcgag ctggtgcagc aacaccatcg tcgtcatcca ctcggtcggc
 1680
 20 ccggtcctcc tgaccgattg gtacgacaac cccaacatca cggccattct ctgggctggt
 1740
 25 ctccggggcc aggagtcggg caactccatc accgacgtgc ttacggcaa ggtcaacccc
 1800
 gccgcccgt cgcccttcac ttggggcaag acccgcgaaa gctatggcgc ggacgtcctg
 1860
 30 tacaagccga ataatggcaa tggcgcgccc caacaggact tcaccgaggg cgtcttcac
 1920
 gactaccgct acttcgacaa ggttgacgat gactcggtea tctacgagtt cggccacggc
 1980
 35 ctgagctaca ccaccttcga gtacagcaac atccgcgtcg tcaagtccaa cgtcagcgag
 2040
 40 taccggccca cgacgggcac cacggcccag gcccgcagct ttggcaactt ctccaccgac
 2100
 ctgaggact atctcttccc caaggacgag ttccctaca tctaccagta catctaccg
 2160
 45 tacctcaaca cgaccgaccc ccggagggcc tcggccgac cccactacgg ccagaccgcc
 2220
 50 gaggagtcc tcccgcacca cgccacgat gacgacccc agccgctcct ccggtcctcg
 2280
 ggcggaaact ccccgggcgg caaccgccag ctgtacgaca ttgtctacac aatcacggcc
 2340

gacatcacga atacgggctc cgttgtaggc gaggaggtagc cgcagctcta cgtctcgctg
2400
5 ggcgggtcccg aggatcccaa ggtgcagctg cgcgactttg acaggatgcg gatcgaaccc
2460
ggcgagacga ggcagttcac cggccgcctg acgcgcagag atctgagcaa ctgggacgtc
10 2520
acgggtgcagg actgggtcat cagcaggtat cccaagacgg catatggttg gaggagcagc
2580
15 cggaagttgg atctcaagat tgagcttccct tga
2613

20 <210> 3
<211> 851
<212> PRT
<213> *Myceliophthora thermophila*
<400> 3
25

Ile Glu Ser Arg Lys Val His Gln Lys Pro Leu Ala Arg Ser Glu Pro
1 5 10 15
30 Phe Tyr Pro Ser Pro Trp Met Asn Pro Asn Ala Asp Gly Trp Ala Glu
20 25 30
35 Ala Tyr Ala Gln Ala Lys Ser Phe Val Ser Gln Met Thr Leu Leu Glu
35 40 45
Lys Val Asn Leu Thr Thr Gly Val Gly Trp Gly Ala Glu Gln Cys Val
50 55 60
40 Gly Gln Val Gly Ala Ile Pro Arg Leu Gly Leu Arg Ser Leu Cys Met
65 70 75 80
His Asp Ser Pro Leu Gly Ile Arg Gly Ala Asp Tyr Asn Ser Ala Phe
45 85 90 95
Pro Ser Gly Gln Thr Val Ala Ala Thr Trp Asp Arg Gly Leu Met Tyr
100 105 110
50 Arg Arg Gly Tyr Ala Met Gly Gln Glu Ala Lys Gly Lys Gly Ile Asn
115 120 125
Val Leu Leu Gly Pro Val Ala Gly Pro Leu Gly Arg Met Pro Glu Gly
130 135 140
55 Gly Arg Asn Trp Glu Gly Phe Ala Pro Asp Pro Val Leu Thr Gly Ile
145 150 155 160

60

65

Gly Met Ser Glu Thr Ile Lys Gly Ile Gln Asp Ala Gly Val Ile Ala
 165 170 175
 5 Cys Ala Lys His Phe Ile Gly Asn Glu Gln Glu His Phe Arg Gln Val
 180 185 190
 Pro Glu Ala Gln Gly Tyr Gly Tyr Asn Ile Ser Glu Thr Leu Ser Ser
 10 195 200 205
 Asn Ile Asp Asp Lys Thr Met His Glu Leu Tyr Leu Trp Pro Phe Ala
 210 215 220
 15 Asp Ala Val Arg Ala Gly Val Gly Ser Val Met Cys Ser Tyr Gln Gln
 225 230 235 240
 Val Asn Asn Ser Tyr Ala Cys Gln Asn Ser Lys Leu Leu Asn Asp Leu
 20 245 250 255
 Leu Lys Asn Glu Leu Gly Phe Gln Gly Phe Val Met Ser Asp Trp Gln
 260 265 270
 25 Ala Gln His Thr Gly Ala Ala Ser Ala Val Ala Gly Leu Asp Met Ser
 275 280 285
 Met Pro Gly Asp Thr Gln Phe Asn Thr Gly Val Ser Phe Trp Gly Ala
 290 295 300
 30 Asn Leu Thr Leu Ala Val Leu Asn Gly Thr Val Pro Ala Tyr Arg Leu
 305 310 315 320
 35 Asp Asp Met Ala Met Arg Ile Met Ala Ala Leu Phe Lys Val Thr Lys
 325 330 335
 Thr Thr Asp Leu Glu Pro Ile Asn Phe Ser Phe Trp Thr Asp Asp Thr
 340 345 350
 40 Tyr Gly Pro Ile His Trp Ala Ala Lys Gln Gly Tyr Gln Glu Ile Asn
 355 360 365
 Ser His Val Asp Val Arg Ala Asp His Gly Asn Leu Ile Arg Glu Ile
 45 370 375 380
 Ala Ala Lys Gly Thr Val Leu Leu Lys Asn Thr Gly Ser Leu Pro Leu
 385 390 395 400
 50 Asn Lys Pro Lys Phe Val Ala Val Ile Gly Glu Asp Ala Gly Ser Ser
 405 410 415
 Pro Asn Gly Pro Asn Gly Cys Ser Asp Arg Gly Cys Asn Glu Gly Thr
 55 420 425 430
 Leu Ala Met Gly Trp Gly Ser Gly Thr Ala Asn Tyr Pro Tyr Leu Val

	435	440	445
5	Ser Pro Asp Ala Ala Leu Gln Ala Arg Ala Ile Gln Asp Gly Thr Arg		
	450	455	460
	Tyr Glu Ser Val Leu Ser Asn Tyr Ala Glu Glu Lys Thr Lys Ala Leu		
	465	470	475
10	Val Ser Gln Ala Asn Ala Thr Ala Ile Val Phe Val Asn Ala Asp Ser		
	485	490	495
	Gly Glu Gly Tyr Ile Asn Val Asp Gly Asn Glu Gly Asp Arg Lys Asn		
15	500	505	510
	Leu Thr Leu Trp Asn Asn Gly Asp Thr Leu Val Lys Asn Val Ser Ser		
	515	520	525
20	Trp Cys Ser Asn Thr Ile Val Val Ile His Ser Val Gly Pro Val Leu		
	530	535	540
	Leu Thr Asp Trp Tyr Asp Asn Pro Asn Ile Thr Ala Ile Leu Trp Ala		
25	545	550	555
	Gly Leu Pro Gly Gln Glu Ser Gly Asn Ser Ile Thr Asp Val Leu Tyr		
	565	570	575
	Gly Lys Val Asn Pro Ala Ala Arg Ser Pro Phe Thr Trp Gly Lys Thr		
30	580	585	590
	Arg Glu Ser Tyr Gly Ala Asp Val Leu Tyr Lys Pro Asn Asn Gly Asn		
	595	600	605
35	Gly Ala Pro Gln Gln Asp Phe Thr Glu Gly Val Phe Ile Asp Tyr Arg		
	610	615	620
	Tyr Phe Asp Lys Val Asp Asp Asp Ser Val Ile Tyr Glu Phe Gly His		
40	625	630	635
	Gly Leu Ser Tyr Thr Thr Phe Glu Tyr Ser Asn Ile Arg Val Val Lys		
	645	650	655
45	Ser Asn Val Ser Glu Tyr Arg Pro Thr Thr Gly Thr Thr Ala Gln Ala		
	660	665	670
	Pro Thr Phe Gly Asn Phe Ser Thr Asp Leu Glu Asp Tyr Leu Phe Pro		
	675	680	685
50	Lys Asp Glu Phe Pro Tyr Ile Tyr Gln Tyr Ile Tyr Pro Tyr Leu Asn		
	690	695	700
	Thr Thr Asp Pro Arg Arg Ala Ser Ala Asp Pro His Tyr Gly Gln Thr		
55	705	710	715
			720

60

65

Ala Glu Glu Phe Leu Pro Pro His Ala Thr Asp Asp Asp Pro Gln Pro
725 730 735
Leu Leu Arg Ser Ser Gly Gly Asn Ser Pro Gly Gly Asn Arg Gln Leu
740 745 750
Tyr Asp Ile Val Tyr Thr Ile Thr Ala Asp Ile Thr Asn Thr Gly Ser
755 760 765
Val Val Gly Glu Glu Val Pro Gln Leu Tyr Val Ser Leu Gly Gly Pro
770 775 780
Glu Asp Pro Lys Val Gln Leu Arg Asp Phe Asp Arg Met Arg Ile Glu
785 790 795 800
Pro Gly Glu Thr Arg Gln Phe Thr Gly Arg Leu Thr Arg Arg Asp Leu
805 810 815
Ser Asn Trp Asp Val Thr Val Gln Asp Trp Val Ile Ser Arg Tyr Pro
820 825 830
Lys Thr Ala Tyr Val Gly Arg Ser Ser Arg Lys Leu Asp Leu Lys Ile
835 840 845
Glu Leu Pro
850

<210> 4
<211> 870
<212> PRT
<213> *Myceliophthora thermophila*
<400> 4

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

	Leu	Cys	Met	His	Asp	Ser	Pro	Leu	Gly	Ile	Arg	Gly	Ala	Asp	Tyr	Asn	
							100					105				110	
5	Ser	Ala	Phe	Pro	Ser	Gly	Gln	Thr	Val	Ala	Ala	Thr	Trp	Asp	Arg	Gly	
							115					120				125	
10	Leu	Met	Tyr	Arg	Arg	Gly	Tyr	Ala	Met	Gly	Gln	Glu	Ala	Lys	Gly	Lys	
							130					135				140	
	Gly	Ile	Asn	Val	Leu	Leu	Gly	Pro	Val	Ala	Gly	Pro	Leu	Gly	Arg	Met	
							145					150			155		160
15	Pro	Glu	Gly	Gly	Arg	Asn	Trp	Glu	Gly	Phe	Ala	Pro	Asp	Pro	Val	Leu	
							165					170				175	
	Thr	Gly	Ile	Gly	Met	Ser	Glu	Thr	Ile	Lys	Gly	Ile	Gln	Asp	Ala	Gly	
20							180					185				190	
	Val	Ile	Ala	Cys	Ala	Lys	His	Phe	Ile	Gly	Asn	Glu	Gln	Glu	His	Phe	
							195					200				205	
25	Arg	Gln	Val	Pro	Glu	Ala	Gln	Gly	Tyr	Gly	Tyr	Asn	Ile	Ser	Glu	Thr	
							210					215				220	
	Leu	Ser	Ser	Asn	Ile	Asp	Asp	Lys	Thr	Met	His	Glu	Leu	Tyr	Leu	Trp	
30							225					230			235		240
	Pro	Phe	Ala	Asp	Ala	Val	Arg	Ala	Gly	Val	Gly	Ser	Val	Met	Cys	Ser	
							245					250				255	
35	Tyr	Gln	Gln	Val	Asn	Asn	Ser	Tyr	Ala	Cys	Gln	Asn	Ser	Lys	Leu	Leu	
							260					265				270	
	Asn	Asp	Leu	Leu	Lys	Asn	Glu	Leu	Gly	Phe	Gln	Gly	Phe	Val	Met	Ser	
							275					280				285	
40	Asp	Trp	Gln	Ala	Gln	His	Thr	Gly	Ala	Ala	Ser	Ala	Val	Ala	Gly	Leu	
							290					295				300	
	Asp	Met	Ser	Met	Pro	Gly	Asp	Thr	Gln	Phe	Asn	Thr	Gly	Val	Ser	Phe	
45							305					310			315		320
	Trp	Gly	Ala	Asn	Leu	Thr	Leu	Ala	Val	Leu	Asn	Gly	Thr	Val	Pro	Ala	
							325					330				335	
50	Tyr	Arg	Leu	Asp	Asp	Met	Ala	Met	Arg	Ile	Met	Ala	Ala	Leu	Phe	Lys	
							340					345				350	
	Val	Thr	Lys	Thr	Thr	Asp	Leu	Glu	Pro	Ile	Asn	Phe	Ser	Phe	Trp	Thr	
55							355					360				365	
	Asp	Asp	Thr	Tyr	Gly	Pro	Ile	His	Trp	Ala	Ala	Lys	Gln	Gly	Tyr	Gln	

	370	375	380
5	Glu Ile Asn Ser His Val Asp Val Arg Ala Asp His Gly Asn Leu Ile		
	385	390	395 400
	Arg Glu Ile Ala Ala Lys Gly Thr Val Leu Leu Lys Asn Thr Gly Ser		
10	405	410	415
	Leu Pro Leu Asn Lys Pro Lys Phe Val Ala Val Ile Gly Glu Asp Ala		
	420	425	430
15	Gly Ser Ser Pro Asn Gly Pro Asn Gly Cys Ser Asp Arg Gly Cys Asn		
	435	440	445
	Glu Gly Thr Leu Ala Met Gly Trp Gly Ser Gly Thr Ala Asn Tyr Pro		
20	450	455	460
	Tyr Leu Val Ser Pro Asp Ala Ala Leu Gln Ala Arg Ala Ile Gln Asp		
	465	470	475 480
	Gly Thr Arg Tyr Glu Ser Val Leu Ser Asn Tyr Ala Glu Glu Lys Thr		
25	485	490	495
	Lys Ala Leu Val Ser Gln Ala Asn Ala Thr Ala Ile Val Phe Val Asn		
	500	505	510
30	Ala Asp Ser Gly Glu Gly Tyr Ile Asn Val Asp Gly Asn Glu Gly Asp		
	515	520	525
	Arg Lys Asn Leu Thr Leu Trp Asn Asn Gly Asp Thr Leu Val Lys Asn		
35	530	535	540
	Val Ser Ser Trp Cys Ser Asn Thr Ile Val Val Ile His Ser Val Gly		
	545	550	555 560
	Pro Val Leu Leu Thr Asp Trp Tyr Asp Asn Pro Asn Ile Thr Ala Ile		
40	565	570	575
	Leu Trp Ala Gly Leu Pro Gly Gln Glu Ser Gly Asn Ser Ile Thr Asp		
	580	585	590
45	Val Leu Tyr Gly Lys Val Asn Pro Ala Ala Arg Ser Pro Phe Thr Trp		
	595	600	605
	Gly Lys Thr Arg Glu Ser Tyr Gly Ala Asp Val Leu Tyr Lys Pro Asn		
50	610	615	620
	Asn Gly Asn Gly Ala Pro Gln Gln Asp Phe Thr Glu Gly Val Phe Ile		
	625	630	635 640
	Asp Tyr Arg Tyr Phe Asp Lys Val Asp Asp Asp Ser Val Ile Tyr Glu		
55	645	650	655

5 Phe Gly His Gly Leu Ser Tyr Thr Thr Phe Glu Tyr Ser Asn Ile Arg
 660 665 670
 Val Val Lys Ser Asn Val Ser Glu Tyr Arg Pro Thr Thr Gly Thr Thr
 675 680 685
 10 Ala Gln Ala Pro Thr Phe Gly Asn Phe Ser Thr Asp Leu Glu Asp Tyr
 690 695 700
 Leu Phe Pro Lys Asp Glu Phe Pro Tyr Ile Tyr Gln Tyr Ile Tyr Pro
 705 710 715 720
 15 Tyr Leu Asn Thr Thr Asp Pro Arg Arg Ala Ser Ala Asp Pro His Tyr
 725 730 735
 Gly Gln Thr Ala Glu Glu Phe Leu Pro Pro His Ala Thr Asp Asp Asp
 20 740 745 750
 Pro Gln Pro Leu Leu Arg Ser Ser Gly Gly Asn Ser Pro Gly Gly Asn
 755 760 765
 25 Arg Gln Leu Tyr Asp Ile Val Tyr Thr Ile Thr Ala Asp Ile Thr Asn
 770 775 780
 Thr Gly Ser Val Val Gly Glu Glu Val Pro Gln Leu Tyr Val Ser Leu
 785 790 795 800
 30 Gly Gly Pro Glu Asp Pro Lys Val Gln Leu Arg Asp Phe Asp Arg Met
 805 810 815
 Arg Ile Glu Pro Gly Glu Thr Arg Gln Phe Thr Gly Arg Leu Thr Arg
 35 820 825 830
 Arg Asp Leu Ser Asn Trp Asp Val Thr Val Gln Asp Trp Val Ile Ser
 835 840 845
 40 Arg Tyr Pro Lys Thr Ala Tyr Val Gly Arg Ser Ser Arg Lys Leu Asp
 850 855 860
 Leu Lys Ile Glu Leu Pro
 45 865 870

50 <210> 5
 <211> 5
 <212> PRT
 <213> Secuencia artificial
 <220>
 <223> Secuencia peptídica sintética
 55 <400> 5

60 Ile Glu Ser Arg Lys
 1 5

65 <210> 6
 <211> 19

<212> PRT
<213> Secuencia artificial
<220>
<223> Secuencia peptídica sintética
<400> 6

[illegible]

15	<210> 7
	<211> 24
	<212> PRT
20	<213> Secuencia artificial
	<220>
	<223> Secuencia peptídica sintética
	<400> 7

25 Met Lys Ala Ala Ala Leu Ser Cys Leu Phe Gly Ser Thr Leu Ala Val
1 5 10 15
30 Ala Gly Ala Ile Glu Ser Arg Lys
20

35 <210> 8
 <211> 1817
 <212> ADN
 <213> *Myceliophthora thermophila*
 <400> 8

gagctccacc gcggtggcgg ccgcggatct tacaaaaaaa aggtatccga tttggggaac
60
gtcgatgaaa gtattgcaaa agtgacgaga gttgcgcaac taactcgctg ccgaagaagc
120
tgcggaagaa agagaacacc gaaagtggaa taacggttac gatgtcctga cctcaaagtt
180

55

60

65

gaaaccagcc cttoctgctc tatttgggaa agcggcttgc ccttgaatgc gctgcactgt
240
5 ggcacgacta ccagtgatcg ggaggagcaa actaccctgg tccgttcctt ggtggggcgg
300
cactaggccc aacttagggg gatcggaggt cgatgccgcg gtcctcgttg gtctgggctc
10 360
ttctcatttc cgggtttgca cccccgttg cacctgctga tcgcccgcc aacccgatga
420
15 ggttgccccc agaccgacaa tcaccgcggc tgcattccca agtatattga agatggcacc
480
aggtaccggg ttttgcgctc cagtcgtttg gtgccaaatt tgggagtttt tgagcctcaa
540
20 gatctgggga aatcgacctc aacttcata caagttaaag tcgcacacac ggcgagttcc
600
acgaagagac acattttttt ctgaaggcct ctctccccgc acatcagaaa ccaccaata
25 660
ccaagactgc agaagccggg gtaagtgggc caccgggact acactaaaat gcggggagaa
720
30 gcgagatccg ttgcgaaggg aagggatggg gtgtgctgcg gctttctccg ctctcgtgcg
780
ccttttgctt gaatctagtg tacaccaggg taggctccga aggagtatct acggcagcgc
35 840
tgttcgtgct gcgttgagag tcagggcgga gacgagcagg cgacaggagc ctgcacccgg
900
cacttcggat cgcatttgcg cggagcgtca aatacgtctt tctgcgggtca tcagagagca
40 960
tcgtgaacca aggttcttcc gcagggcggc ctgggcttcg cagagtcgca ctggcgggac
1020
45 gccttccttg tcacccctga taacctgggt gccgcgcccc gactcctcca atgaggtgtg
1080
tggttgccct cgccgaccct tcagcaacct taatcgcttc catcgcacgg ctccacgtcc
50 1140
tcgaacgatg ccttcagtcc gtgcccggcc gtggcaacca taacgtgaca tcgccgcccc
1200

55

60

65

gcctactagc cgctatcgac cggtaggct tgtcaccgca gcgcccattc tccatcgggc
1260
5 ctctactctg atccacctca cccaccgcaa gcactagcga gcctcaccag agtgcaagcg
1320
acacgaccog cttggccctt cgtccttgac tatctcccag acctottgcc atcttgccga
10 1380
cgccgcccc ttttttttct cctccccctg ccggcaggte ggtggcccca gtcccagat
1440
15 ggcattgctc cgttggtccat gacgacccat cattcgatgg ctgactggca cactcgtctt
1500
gtttgagcat cgacggcccg cggcccgctt cccacggtag ggaacctcgt tgtacagtac
1560
20 ctctcgtaat gataccaac accggggccg agcgtctggga gggcggcggt cccgagaagc
1620
cggaaggcg gctggccggc tgaccttgt gacttggcga tggatgcggc catggagaat
25 1680
gtccgtccga agcgacgca caattagcct ggctaccatc gatataaatt gggtgattcc
1740
30 cagctcttga tgggcgtgtc ttctgcctgg cagccctcgt cttcagatca agcaactgtg
1800
tgctgaccc cttccgt
1817
35

<210> 9
<211> 2556
40 <212> ADN
<213> Secuencia artificial
<220>
<223> Constructo de ADN sintético de polinucleótido artificial que codifica la proteína C1 BGL1 que con codones optimizados para la expresión en *S. cerevisiae*
45 <400> 9

atagaaagta gaaaggtaga tcaaaaacca ttagctagat cagaaccatt ctacccttct
50 60

55

60

65

ccatggatga accctaattgc agatggatgg gcagaagcat atgctcaggg caagagtttt
120
5 gtctcccaga tgactctgtt ggaaaagggtt aatctgacaa caggagtagg atgggggtgca
180
gaacagtgtg tcggccaagt tgggtgctatc cctagattgg gtcttagaag tttgtgtatg
10 240
cacgattctc ccttaggtat aagaggcgct gactataact cagcattccc atccgggcaa
300
15 actgttgctg cgacatggga caggggtttg atgtatagaa ggggttatgc gatgggtcaa
360
gaggcaaagg gtaaaggaat taatgtattg ttggggccgg tggcggggcc actgggaaga
20 420
atgccagaag gtggaaggaa ctgggaagga ttgcggggcg acccagtgtc aacaggtata
480
25 ggtatgtccg aaacgatcaa aggcatacaa gatgcaggtg ttatcgctg tgccaagcat
540
tttattggta atgaacaaga gcattttcgt caagtgccag aggtcaagg ttatggttat
600
30 aatatttctg aaactttaag ttccaacatc gatgacaaaa ccatgcacga gttatactta
660
tgcccttttg cagacgtgtg gagagctggc gttggctctg ttatgtgtc ttatcagcaa
35 720
gttaataact cttacgctg tcaaaattcc aagttactaa atgacttatt gaagaacgaa
780
40 ctaggattcc aaggattcgt catgagcgat tggcaagcac agcatactgg tgctgcatcc
840
gctgtggcag gattagatat gtcaatgcca ggagatacac aatttaatac tggcgtagt
45 900
ttttgggtg caaacctaac tttagctgtt ctaaacggta cggtacctgc atatcgttta
960
50 gacgacatgg ccatgcgtat aatggctgtc ttattcaaag ttacaaaaac caccgattta
1020
gaaccaatta attttagttt ttggacagat gacacatatg gtccatataca ctgggctgct
1080

55

60

65

aagcaagggt accaagaaat aaatagtcac gttgacgtaa gagcggatca cggcaatctt
 1140
 5 atcagagaga tagcagcaaa gggaactgta ttgttgaaga atactgggtc attaccacta
 1200
 aacaaaccaa agtttgtcgc agtcattggt gaagatgctg gttcatcccc taatggacca
 10 1260
 aatggttgta gtgacagagg ctgcaatgaa ggcacgttgg caatgggctg gggctcaggg
 1320
 actgccatt acccctatct ggtctctccg gatgcggctt tacaggctag agcaatccag
 15 1380
 gatggtacta gatacgagag cgtcctaagt aactatgccg aagaaaagac taaggcctta
 1440
 20 gtcagtcaag ccaatgccac tgctatcgtt ttcgtaaacg cggattctgg cgaaggttat
 1500
 atcaatgttg atggtaatga aggtgacaga aagaatttaa ctttatggaa taacggcgac
 25 1560
 acattagtta aaaatgtatc aagttggtgt tccaatacta tcgtcgtgat acattctgtt
 1620
 30 ggtccagttt tactgacaga ctggtacgat aacccaaaca ttaccgccat tttatgggca
 1680
 ggtctgccag ggcaggaatc aggaaattcc attacggacg tactatacgg aaagggttaac
 35 1740
 ccagccgcca ggagcccttt cacatggggt aagacaagag agagctacgg agctgatgtt
 1800
 40 ctttataaac cgaacaacgg gaatggagcg ccacagcaag attttactga aggtgtgttc
 1860
 attgactata gatacttcga caaagttgac gatgactcag ttatatatga attcggtcac
 1920
 45 ggtctatctt atactacttt tgaatattca aatataagag tagtcaaaaag taatgtttct
 1980
 gaatataggc cgaccaccgg aacgacggct caagcgcta cttcggtaa tttttcaacg
 50 2040
 gatttagaag attatcttatt tcccaaagac gaatttccat acatctacca atacatatac
 2100

55

60

65

ccctatctga atactaccga tccaagaaga gcttctgccg atccacatta cgggcagact
2160
5 gccgaagagt tcttgccacc acacgctact gacgacgac ctcaacctct tctgagggtcc
2220
agtggcggaa attcacctgg tggtaatagg cagctgtatg atattgtgta tactataacg
10 2280
gctgatatta ctaatactgg tagcggttgg gggaagaag tgccgcaatt atatgtgtct
2340
15 ttagggtggc cggaagatcc taagggttcag ttaagagact ttgataggat gagaatagaa
2400
cctggagaaa ctaggcaatt tacaggtaga ttgacccgta gggatctgtc aaactgggat
20 2460
gtaacagtgc aagattgggt aatcagcagg taccggaaaa ctgcatacgt gggtagatct
2520
25 tcccgtaagt tagatttgaa aattgaattg ccataa
2556

30 <210> 10
<211> 60
<212> ADN
<213> Secuencia artificial
<220>
<223> Constructo de ADN sintético de la secuencia del cebador PcbhC1bgl_F
35 <400> 10

gtcttcagat caagcaactg tgtgctgac ctcttccgctc atgaaggctg ctgcgcttcc
40 60

45 <210> 11
<211> 18
<212> PRT
<213> Secuencia artificial
<220>
<223> Secuencia peptídica sintética
50 <400> 11

Met Gln Leu Pro Ala Ala Ala Gln Trp Leu Leu Thr Pro Ala Lys Ala
1 5 10 15
55 Ser Leu

60 <210> 12
<211> 39
<212> PRT
<213> Secuencia artificial
<220>
<223> Secuencia peptídica sintética del péptido señal de C1 BGL1
65 <400> 12

Met Lys Ala Ala Ala Leu Ser Cys Leu Phe Gly Ser Thr Leu Ala Val
1 5 10 15
5 Ala Gly Ala Ile Glu Ser Arg Lys Val His Gln Lys Pro Leu Ala Arg
20 25 30
10 Ser Glu Pro Phe Tyr Pro Ser
35

<210> 13
<211> 83
15 <212> ADN
<213> Secuencia artificial
<220>
<223> Constructo de ADN sintético
20 <400> 13

cggtgccatc aatcatctcg gttcgcgcga gctgcttctt tctgtgcagt gaacgctctc
60
25 aaactgcaac gacgctgtcc gac
83

<210> 14
30 <211> 60
<212> ADN
<213> Secuencia artificial
<220>
<223> Constructo de ADN sintético
35 <400> 14

cggtgccatc aatcatctcg gttcgcgcga gctgcttctt tctgtgcagt gaacgctctc
40 60

<210> 15
45 <211> 43
<212> ADN
<213> Secuencia artificial
<220>
<223> Constructo de ADN sintético
50 <400> 15

tctgtgcagt gaacgctctc aaactgcaac gacgctgtcc gac
55 43

<210> 16
60 <211> 40
<212> ADN
<213> Secuencia artificial
<220>
<223> Constructo de ADN sintético
65 <400> 16

5 **aatcatctcg gttegcgcga gctgcttctt tctgtgcagt**
 40

10 <210> 17
 <211> 30
 <212> ADN
 <213> Secuencia artificial
 <220>
 <223> Constructo de ADN sintético
 15 <400> 17

20 **aatcatctcg gttegcgcga gctgcttctt**
 30

25 <210> 18
 <211> 250
 <212> ADN
 <213> Secuencia artificial
 <220>
 <223> Constructo de ADN sintético de la secuencia del intrón1 del gen bgll para C1 beta-glucosidasa
 <220>
 30 <221> Intrón
 <222> (1)..(250)
 <223> Intrón 1 de C1 bgll
 <400> 18

35 **aggtatggac gggctttcgt caaagactcg ctccccgatc aacttcccct ttcattccaga**
 60
 ccaccccaac cctcccagtc ctgcttcgag cacgatctct tcgggcagca cccacccac
 40 **120**
 atccactcag attageggcg acaccgttga ctggtgcaat ccgcaatcga catgcaactt
 180
 45 **ccagccgcag cccaatggct gtcacgctt cccgcgaaag cctcacttgc tgacaatcat**
 240
 cgtcagggtc
 50 **250**

55 <210> 19
 <211> 73
 <212> ADN
 <213> Secuencia artificial
 <220>
 <223> Constructo de ADN sintético
 60 <400> 19

65

atgcaacttc cagccgcagc ccaatggctg ctcacgcttc ccgcgaaagc ctcacttgct
 60
 5 gacaatcatc gtc.
 73
 10
 <210> 20
 <211> 70
 <212> ADN
 <213> Secuencia artificial
 15 <220>
 <223> Constructo de ADN sintético
 <220>
 <221> exón
 <222> (1)..(70)
 20 <223> Primer exón de C1 beta-glucosidasa
 <400> 20
 25 atg aag gct gct gcg ctt tcc tgc ctc ttc ggc agt acc ctt gcc gtt
 48
 Met Lys Ala Ala Ala Leu Ser Cys Leu Phe Gly Ser Thr Leu Ala Val
 30 1 5 10 15
 gca ggc gcc att gaa tcg aga a
 70
 35 Ala Gly Ala Ile Glu Ser Arg
 20
 40 <210> 21
 <211> 240
 <212> ADN
 <213> Secuencia artificial
 <220>
 45 <223> Constructo de ADN sintético
 <400> 21
 50 atgaaggctg ctgcgctttc ctgcctcttc ggcagtaccc ttgccgttgc aggcgccatt
 60
 gaatcgagaa aggtatggac cagcatctct tcgggcagca cccacccac atccactcag
 120
 55 attagcggcg acaccgttga ctgttgcaat ccgcaatcga catgcaactt ccagccgcag
 180
 cccaatggct gctcacgctt ccgcgaaag ctcacttgct tgacaatcat cgtcaggctc
 60 240
 <210> 22
 <211> 20
 65 <212> ADN
 <213> Secuencia artificial

<220>
<223> Constructo de ADN sintético de la secuencia del cebador interno C1bgl_fwd1
<400> 22

5

atgaaggctg ctgcgctttc
20

10

<210> 23
<211> 27
<212> ADN
<213> Secuencia artificial

15

<220>
<223> Constructo de ADN sintético de la secuencia del cebador interno C1bgl_rev1
<400> 23

20

tcattaagga agctcaatct tgagatc
27

25

<210> 24
<211> 25
<212> ADN
<213> Secuencia artificial
<220>

30

<223> Constructo de ADN sintético de la secuencia del cebador externo C1bgl_fwd2
<400> 24

35

tctctgcggg tgccatcaat catct
25

40

<210> 25
<211> 24
<212> ADN
<213> Secuencia artificial
<220>

45

<223> Constructo de ADN sintético de la secuencia del cebador externo C1bgl_rev2
<400> 25

50

gctcaccgga acttgccaag tgct
24

55

<210> 26
<211> 29
<212> ADN
<213> Secuencia artificial
<220>

60

<223> Constructo de ADN sintético de la secuencia del cebador cdx09050
<400> 26

ggctcatgaa ggctgctgcg ctttcctgc
29

65

5 <210> 27
 <211> 34
 <212> ADN
 <213> Secuencia artificial
 <220>
 <223> Constructo de ADN sintético de la secuencia del cebador cdx09052
 <400> 27

10 **gccgaattct caaggaagct caatcttgag atcc**
34

15 <210> 28
 <211> 60
 <212> ADN
 <213> Secuencia artificial
 <220>
 20 <223> Constructo de ADN sintético de la secuencia del cebador TcbhC1bgl_R1
 <400> 28

25 **gtcttcagat caagcaactg tgtgctgac ctcttcgctc atgaaggctg ctgcgcttcc**
60

30 <210> 29
 <211> 4
 <212> PRT
 <213> Secuencia artificial
 <220>
 <223> Secuencia peptídica sintética
 35 <400> 29

40 **Glu Ser Arg Lys**
1

45 <210> 30
 <211> 6
 <212> PRT
 <213> Secuencia artificial
 <220>
 <223> Secuencia peptídica sintética
 50 <400> 30

55 **Ala Asp Asn His Arg Gln**
1 5

60 <210> 31
 <211> 24
 <212> PRT
 <213> Secuencia artificial
 <220>
 <223> Secuencia peptídica sintética
 65 <400> 31

ES 2 538 396 T3

	Met	Gln	Leu	Pro	Ala	Ala	Ala	Gln	Trp	Leu	Leu	Thr	Pro	Ala	Lys	Ala
	1				5					10					15	
5	Ser	Leu	Ala	Asp	Asn	His	Arg	Gln								
				20												
10																
15																
20																
25																
30																
35																
40																
45																
50																
55																
60																
65																

Reivindicaciones

1. Polinucleótido que codifica un polipéptido β -glucosidasa recombinante, en el que dicho polipéptido:

- 5 a) es al menos un 90% idéntico a la SEQ ID NO: 3 y comprende una secuencia amino terminal que se expone como IESRK (SEQ ID NO: 5), ESRK (SEQ ID NO: 29) o SRK; y/o es una proteína secretada derivada de una preproteína, en el que la preproteína es al menos un 90% idéntica a la SEQ ID NO: 4 y comprende una secuencia amino terminal que comprende la SEQ ID NO: 6 o NO: 7; y/o
- 10 b) es al menos un 90% idéntico a la SEQ ID NO: 3 y comprende al menos uno, al menos dos, o tres residuos de aminoácidos seleccionados de entre ácido aspártico en la posición 358 (358ASP), glutamina en la posición 381 (381 GLN) y ácido glutámico en la posición 385 (385GLU), en los que los números de posición de los aminoácidos corresponden a la SEQ ID NO: 4.

15 2. Polinucleótido según la reivindicación 1 en el que el polipéptido β -glucosidasa recombinante comprende la secuencia de la SEQ ID NO: 3 o la SEQ ID NO: 4.

3. Polinucleótido según la reivindicación 1 o la reivindicación 2, en el que el polinucleótido comprende la SEQ ID NO: 2 o la SEQ ID NO: 9.

20 4. Vector de expresión que comprende el polinucleótido según cualquiera de las reivindicaciones 1-3, unido operativamente a un promotor heterólogo.

25 5. Método de producción de un polipéptido β -glucosidasa secretado, que comprende cultivar una célula que comprende un polinucleótido según cualquiera de las reivindicaciones 1-3, unido operativamente a un promotor heterólogo, o un vector de expresión según la reivindicación 4, en condiciones en las que se produce β -glucosidasa.

30 6. Célula hospedadora recombinante que comprende un polinucleótido según cualquiera de las reivindicaciones 1-3 que codifica un polipéptido β -glucosidasa recombinante, en el que la secuencia codificante no comprende intrones, y en el que opcionalmente el polinucleótido está unido operativamente a un promotor heterólogo, y en el que opcionalmente la célula expresa al menos otra enzima celulasa recombinante.

7. Composición que comprende una célula que comprende una secuencia polinucleotídica que codifica un polipéptido β -glucosidasa recombinante unido operativamente a un promotor heterólogo, en el que dicho polipéptido:

- 35 a) es al menos un 90% idéntico a la SEQ ID NO: 3 y comprende una secuencia amino terminal que se expone como IESRK (SEQ ID NO: 5), ESRK (SEQ ID NO: 29) o SRK; y/o es una proteína secretada derivada de una preproteína, en el que la preproteína es al menos un 90% idéntica a la SEQ ID NO: 4 y comprende una secuencia amino terminal que comprende la SEQ ID NO: 6 o NO: 7; y/o
- 40 b) es al menos un 90% idéntico a la SEQ ID NO: 3 y comprende al menos uno, al menos dos, o tres residuos de aminoácidos seleccionados de entre ácido aspártico en la posición 358 (358ASP), glutamina en la posición 381 (381 GLN) y ácido glutámico en la posición 385 (385GLU), en los que los números de posición de los aminoácidos corresponden a la SEQ ID NO: 4, y una proteína β -glucosidasa secretada por la célula.

45 8. Método de producción de glucosa a partir de celobiosa, que comprende poner en contacto celobiosa con una célula recombinante que comprende una secuencia polinucleotídica que codifica un polipéptido β -glucosidasa recombinante, unida operativamente a un promotor heterólogo, en el que dicho polipéptido:

- 50 a) es al menos un 90% idéntico a la SEQ ID NO: 3 y comprende una secuencia amino terminal que se expone como IESRK (SEQ ID NO: 5), ESRK (SEQ ID NO: 29) o SRK; y/o es una proteína secretada derivada de una preproteína, en el que la preproteína es al menos un 90% idéntica a la SEQ ID NO: 4 y comprende una secuencia amino terminal que comprende la SEQ ID NO: 6 o NO: 7; y/o
- 55 b) es al menos un 90% idéntico a la SEQ ID NO: 3 y comprende al menos uno, al menos dos, o tres residuos de aminoácidos seleccionados de entre ácido aspártico en la posición 358 (358ASP), glutamina en la posición 381 (381 GLN) y ácido glutámico en la posición 385 (385GLU), en los que los números de posición de los aminoácidos corresponden a la SEQ ID NO: 4, en condiciones en las que la β -glucosidasa es expresada y secretada por la célula y dicha celobiosa es convertida enzimáticamente en glucosa por dicha β -glucosidasa.

60 9. Método según la reivindicación 5 u 8 en el que la célula es una célula de la cepa C1.

10. Método según la reivindicación 5 u 8 en el que la célula no es una célula de la cepa C1.

65 11. Vector de expresión según la reivindicación 4, método según la reivindicación 5 u 8, célula hospedadora recombinante según la reivindicación 6, o composición según la reivindicación 7, en los que el promotor heterólogo es el promotor de C1 celobiohidrolasa 1a (CBH1a).

12. Polinucleótido según cualquiera de las reivindicaciones 1-3, vector de expresión según la reivindicación 4 u 11, o método según cualquiera de las reivindicaciones 5, 8, 9, 10 y 11 en los que la secuencia polinucleotídica no codifica una o más de entre la SEQ ID NO: 11, la SEQ ID NO: 30 o la SEQ ID NO: 31.

5

13. Método según cualquiera de las reivindicaciones 5, 8, 9, 10 y 11, que comprende adicionalmente recuperar la β -glucosidasa a partir del medio en el que se cultiva la célula.

10

14. Método de conversión de un sustrato de biomasa en un azúcar fermentable, que comprende combinar la célula según la reivindicación 6 con el sustrato de biomasa en condiciones adecuadas para la producción del azúcar fermentable.

15

15. Método según la reivindicación 14 en el que el sustrato de biomasa es la celobiosa y el azúcar fermentable es la glucosa.

20

25

30

35

40

45

50

55

60

65

FIGURA 1: Polinucleótidos C1 β -glucosidasa

La secuencia superior es ADN genómico (SEQ ID NO: 1).

La secuencia inferior es ADNc (SEQ ID NO: 2).

Los huecos en la secuencia inferior muestran la posición de los intrones.

```

ATGAAGGCTGCTGCGCTTTCCTGCCTCTTCGGCAGTACCTTGCCGTTGCAGGGCCATTGAATCGAGAAAGGTATGGAC
ATGAAGGCTGCTGCGCTTTCCTGCCTCTTCGGCAGTACCTTGCCGTTGCAGGGCCATTGAATCGAGAA-----
12345678901234567890123456789012345678901234567890123456789012345678901234567890
1      2      3      4      5      6      7      8

GGGCTTTCGTCAAAGACTCGCTCCCGCATCAACTTCCCCTTTCATCCAGACCACCCACCCCTCCAGTCTGCTTCGAG
-----
12345678901234567890123456789012345678901234567890123456789012345678901234567890
9      10     11     12     13     14     15     1

CACGATCTCTTCGGGCGAGCACCCACCCACATCCACTCAGATTAGCGGCGACACCGTTGACTGTTGCAATCCGCAATCGA
-----
12345678901234567890123456789012345678901234567890123456789012345678901234567890
6      17     18     19     20     21     22     23     2

CATGCAACTTCCAGCCGCGAGCCCAATGGCTGCTCAGCTTCCGCGAAAGCCTCACTTGCTGACAATCATCGTCAGGTTT
-----
12345678901234567890123456789012345678901234567890123456789012345678901234567890
4      25     26     27     28     29     30     31     3

ACCAGAAGCCCTCGCGAGATCTGAACCTTTTACCCGTCGCCATGGATGAATCCCAACGCCGACGGCTGGCGGAGGCC
ACCAGAAGCCCTCGCGAGATCTGAACCTTTTACCCGTCGCCATGGATGAATCCCAACGCCGACGGCTGGCGGAGGCC
12345678901234567890123456789012345678901234567890123456789012345678901234567890
2      33     34     35     36     37     38     39     4

TATGCCCCAGGCCAAGTCCTTTGTCTCCCAAATGACTCTGCTAGAGAAGGTCAACTTGACCACGGGAGTCGGGTAAGTTT

```

TATGCCAGGCCAAGTCCTTTGTCTCCCAATGACTCTGCTAGAGAAAGGTCAACTTGACCACGGGAGTCG-----
 1234567890123456789012345678901234567890123456789012345678901234567890
 0 41 42 43 44 45 46 47 4

 GTCATTTGTCCAGGTAACATGCAAAATGGTTCTGTCTAACAATAACTTACCGTAGCTGGGGGGCTGAGCAGTCCGTCCGGCC
 -----GCTGGGGGGCTGAGCAGTCCGTCCGGCC
 1234567890123456789012345678901234567890123456789012345678901234567890
 8 49 50 51 52 53 54 55 5

 AAGTGGGGCGGATCCCTCGCCTTGGACTTCGACTTCGAGTCTGTGCATGCATGACTCCCTCTCGGCATCCGAGGAGCCGACTAC
 AAGTGGGGCGGATCCCTCGCCTTGGACTTCGACTTCGAGTCTGTGCATGCATGACTCCCTCTCGGCATCCGAGGAGCCGACTAC
 12345678901234567890123456789012345678901234567890123456789012345678901234567890
 6 57 58 59 60 61 62 63 6

 AACTCAGCGTTCCCTCTGGCCAGACCGTTGCTGTGCTACCTGGGATCGCGGTCTGATGTACCGTCGCGGTACGCAATGGG
 AACTCAGCGTTCCCTCTGGCCAGACCGTTGCTGTGCTACCTGGGATCGCGGTCTGATGTACCGTCGCGGTACGCAATGGG
 12345678901234567890123456789012345678901234567890123456789012345678901234567890
 4 65 66 67 68 69 70 71 7

 CCAGGAGGCCAAGGCAAGGGCATCAATGTCCTTCTCGGACCAGTCGCCGGCCCCCTTGGCCGCATGCCCGAGGGCGGTC
 CCAGGAGGCCAAGGCAAGGGCATCAATGTCCTTCTCGGACCAGTCGCCGGCCCCCTTGGCCGCATGCCCGAGGGCGGTC
 12345678901234567890123456789012345678901234567890123456789012345678901234567890
 2 73 74 75 76 77 78 79 8

 GTAAC TGGGAAGGCTTCGCTCCGGATCCCGTCCTTACCGGGCATCGGCATGTCCGAGACGATCAAGGGCATTCAGGATGCT
 GTAAC TGGGAAGGCTTCGCTCCGGATCCCGTCCTTACCGGGCATCGGCATGTCCGAGACGATCAAGGGCATTCAGGATGCT
 12345678901234567890123456789012345678901234567890123456789012345678901234567890
 0 81 82 83 84 85 86 87 8

 GCGGTCA TCGCTTGTGCGAAGCACTTTATTGGAAACGAGCAGGGTGAGTAGTCAAAGACGGGCCGTCTCGGACCCGCGGC
 GCGGTCA TCGCTTGTGCGAAGCACTTTATTGGAAACGAGCAG-----
 12345678901234567890123456789012345678901234567890123456789012345678901234567890
 8 89 90 91 92 93 94 95 9

TTCAAGCTGCTGACTCTGCTGCAGAGCACTTCAGACAGGTGCCAGAAAGCCAGGGATACGGTTACAACATCAGCGAAACC
 -----GAGCACTTCAGACAGGTGCCAGAAAGCCAGGGATACGGTTACAACATCAGCGAAACC
 1234567890123456789012345678901234567890123456789012345678901234567890
 6 97 98 99 100 101 102 103 1

 CTCTCCTCCAACATTGACGACAAGACCATGCACGAGCTCTACCTTTGGCCGTTTGCCGATGCCGTCCGGCCGGCGTCGG
 CTCTCCTCCAACATTGACGACAAGACCATGCACGAGCTCTACCTTTGGCCGTTTGCCGATGCCGTCCGGCCGGCGTCGG
 1234567890123456789012345678901234567890123456789012345678901234567890
 04 105 106 107 108 109 110 111 1

 CTCTGTCATGTGCTCGTACCAGCAGGTCAACAACCTCGTACGCCCTGCCAGAACTCGAAGCTGTGAACGACCTCCTCAAGA
 CTCTGTCATGTGCTCGTACCAGCAGGTCAACAACCTCGTACGCCCTGCCAGAACTCGAAGCTGTGAACGACCTCCTCAAGA
 1234567890123456789012345678901234567890123456789012345678901234567890
 12 113 114 115 116 117 118 119 1

 ACGAGCTTGGGTTTCAGGGCTTCGTCTATGAGCGACTGGCAGGCACAGCACACTGGCCGACGCAAGCCCGTGGTGGTCTC
 ACGAGCTTGGGTTTCAGGGCTTCGTCTATGAGCGACTGGCAGGCACAGCACACTGGCCGACGCAAGCCCGTGGTGGTCTC
 1234567890123456789012345678901234567890123456789012345678901234567890
 20 121 122 123 124 125 126 127 1

 GATATGTCCATGCCGGGCGACACCCAGTTCAACACTGGCGTCAGTTTCTGGGGCGCCAATCTCACCCCTCGCCGTCCTCAA
 GATATGTCCATGCCGGGCGACACCCAGTTCAACACTGGCGTCAGTTTCTGGGGCGCCAATCTCACCCCTCGCCGTCCTCAA
 1234567890123456789012345678901234567890123456789012345678901234567890
 28 129 130 131 132 133 134 135 1

 CGGCACAGTCCCTGCCCTACCGTCTCGACGACATGGCCATCGGCATCATGGCCGCCCTCTTCAAGGTCAACAGACCACCG
 CGGCACAGTCCCTGCCCTACCGTCTCGACGACATGGCCATCGGCATCATGGCCGCCCTCTTCAAGGTCAACAGACCACCG
 1234567890123456789012345678901234567890123456789012345678901234567890
 36 137 138 139 140 141 142 143 1

 ACCTGGAACCGATCAACTTCTCCTTCTGGACCGACGACACTTATGGCCCGATCCACTGGGCCGCCCAAGCAGGGGTACCCAG
 ACCTGGAACCGATCAACTTCTCCTTCTGGACCGACGACACTTATGGCCCGATCCACTGGGCCGCCCAAGCAGGGGTACCCAG

1234567890123456789012345678901234567890123456789012345678901234567890
 44 145 146 147 148 149 150 151 1

 GAGATTAAATCCACGTTGACGTCGCGCCGACACGGCAACCTCATCCGGGAGATTGCCGCCAAGGGTACGGTGCTGCT
 GAGATTAAATCCACGTTGACGTCGCGCCGACACGGCAACCTCATCCGGGAGATTGCCGCCAAGGGTACGGTGCTGCT
 12345678901234567890123456789012345678901234567890123456789012345678901234567890
 52 153 154 155 156 157 158 159 1

 GAAGAAATACCGGCTCTTACCCCTGAACAAGCCAAAGTTTCGTGGCCGTCATCCGGCGAGGATGCTGGTTCGAGCCCCAACG
 GAAGAAATACCGGCTCTTACCCCTGAACAAGCCAAAGTTTCGTGGCCGTCATCCGGCGAGGATGCTGGTTCGAGCCCCAACG
 12345678901234567890123456789012345678901234567890123456789012345678901234567890
 60 161 162 163 164 165 166 167 1

 GGCCCAACGGCTGCAGGACCGCGGCTGTAAAGGACAGCGCTCGCCATGGGCTGGGATCCGGCACAGCCAACTATCCG
 GGCCCAACGGCTGCAGGACCGCGGCTGTAAAGGACAGCGCTCGCCATGGGCTGGGATCCGGCACAGCCAACTATCCG
 12345678901234567890123456789012345678901234567890123456789012345678901234567890
 68 169 170 171 172 173 174 175 1

 TACCTCGTTTCCCCCGACGCGCGCTCCAGGCCCCGGCCATCCAGGACGGCACGAGGTACGAGAGCGTCTGTCCAACTA
 TACCTCGTTTCCCCCGACGCGCGCTCCAGGCCCCGGCCATCCAGGACGGCACGAGGTACGAGAGCGTCTGTCCAACTA
 12345678901234567890123456789012345678901234567890123456789012345678901234567890
 76 177 178 179 180 181 182 183 1

 CGCCGAGGAAAGACAAAGGCTCTGGTCTCGCAGGCCAATGCAACCGCCATCGTCTTCGTCAATGCCGACTCAGGCGAGG
 CGCCGAGGAAAGACAAAGGCTCTGGTCTCGCAGGCCAATGCAACCGCCATCGTCTTCGTCAATGCCGACTCAGGCGAGG
 12345678901234567890123456789012345678901234567890123456789012345678901234567890
 84 185 186 187 188 189 190 191 1

 GCTACATCAACGTGGACGGTAACGAGGGCGACCGTAAGAACCTGACTCTCTGGAAACACGGTGATACTCTGGTCAAGAAC
 GCTACATCAACGTGGACGGTAACGAGGGCGACCGTAAGAACCTGACTCTCTGGAAACACGGTGATACTCTGGTCAAGAAC
 12345678901234567890123456789012345678901234567890123456789012345678901234567890
 92 193 194 195 196 197 198 199 2

 GTCTCGAGCTGGTGCAGCAACACCATCGTCTGTCATCCACTCGGTGCGCCCGTCTCTGACCCGATTGGTACGACAACCC

GTCTCGAGCTGGTGCAGCAACACCATCGTCTGTCATCCACTCGGTGGGCGGGTCTCTCTGACCGATTGGTACGACAACCC
 12345678901234567890123456789012345678901234567890123456789012345678901234567890
 00 201 202 203 204 205 206 207 2

 CAACATCACGGCCATTCTCTGGGCTGGTCTTCCGGGCCAGGAGTCGGGCAACTCCATCACCGACGTGCTTTACGGCAAGG
 CAACATCACGGCCATTCTCTGGGCTGGTCTTCCGGGCCAGGAGTCGGGCAACTCCATCACCGACGTGCTTTACGGCAAGG
 12345678901234567890123456789012345678901234567890123456789012345678901234567890
 08 209 210 211 212 213 214 215 2

 TCAACCCCGCCGCTCGCCCTTCACTTGGGGCAAGACCCGCGAAAGCTATGGCGGGACGTCCTGTACAAAGCCGAAT
 TCAACCCCGCCGCTCGCCCTTCACTTGGGGCAAGACCCGCGAAAGCTATGGCGGGACGTCCTGTACAAAGCCGAAT
 12345678901234567890123456789012345678901234567890123456789012345678901234567890
 16 217 218 219 220 221 222 223 2

 AATGGCAATGGTGGCCCCAACAGGACTTCACCGAGGGCGTCTTCAATCGACTACCGTACTTCGACAAAGTTGACGATGA
 AATGGCAATGGTGGCCCCAACAGGACTTCACCGAGGGCGTCTTCAATCGACTACCGTACTTCGACAAAGTTGACGATGA
 12345678901234567890123456789012345678901234567890123456789012345678901234567890
 24 225 226 227 228 229 230 231 2

 CTCGGTCATCTACGAGTTCGGCCACGGCCTGAGCTACACCACTTCGAGTACAGCAACATCCGCGTCGTCAAGTCCAACG
 CTCGGTCATCTACGAGTTCGGCCACGGCCTGAGCTACACCACTTCGAGTACAGCAACATCCGCGTCGTCAAGTCCAACG
 12345678901234567890123456789012345678901234567890123456789012345678901234567890
 32 233 234 235 236 237 238 239 2

 TCAGCGAGTACCGGCCACGACGGGCACACGGCCACGGCCCGACGTTTGGCAACTTCTCCACCGACCTCGAGGACTAT
 TCAGCGAGTACCGGCCACGACGGGCACACGGCCCGACGGCCCGACGTTTGGCAACTTCTCCACCGACCTCGAGGACTAT
 12345678901234567890123456789012345678901234567890123456789012345678901234567890
 40 241 242 243 244 245 246 247 2

 CTCTTCCCCAAGGACGAGTTCCTTACATCTACAGTACATCTACCGTACCTCAACACGACCGACCCCGGAGGCGCTC
 CTCTTCCCCAAGGACGAGTTCCTTACATCTACAGTACATCTACCGTACCTCAACACGACCGACCCCGGAGGCGCTC
 12345678901234567890123456789012345678901234567890123456789012345678901234567890
 48 249 250 251 252 253 254 255 2

GGCCGATCCCACTACGGCCAGACCGCCGAGGAGTTCTCCGCCCCACGCCACCGATGACGACCCCCAGCCGCTCCTCC
 GGCCGATCCCCACTACGGCCAGACCGCCGAGGAGTTCTCCGCCCCACGCCACCGATGACGACCCCCAGCCGCTCCTCC
 1234567890123456789012345678901234567890123456789012345678901234567890
 56 257 258 259 260 261 262 263 264
 GGTCCTCGGGGGAAACTCCCCCGGGGCAACCGCCAGCTGTACGACATTGTCTACACAATCACGGCCGACATCACGAAT
 GGTCCTCGGGGGAAACTCCCCCGGGGCAACCGCCAGCTGTACGACATTGTCTACACAATCACGGCCGACATCACGAAT
 1234567890123456789012345678901234567890123456789012345678901234567890
 64 265 266 267 268 269 270 271 272
 ACGGGCTCCGTTGTAGGCGAGGAGTACCGCAGCTCTACGCTCGCTGGGCGGTCCCGAGGATCCCAAGTGCAGCTGCG
 ACGGGCTCCGTTGTAGGCGAGGAGTACCGCAGCTCTACGCTCGCTGGGCGGTCCCGAGGATCCCAAGTGCAGCTGCG
 1234567890123456789012345678901234567890123456789012345678901234567890
 72 273 274 275 276 277 278 279 280
 CGACTTTGACAGGATCGGATCGAACCCGGCGAGACGAGGCAGTTCAACCGCCGCTGACGCGCAGAGATCTGAGCAACT
 CGACTTTGACAGGATCGGATCGAACCCGGCGAGACGAGGCAGTTCAACCGCCGCTGACGCGCAGAGATCTGAGCAACT
 1234567890123456789012345678901234567890123456789012345678901234567890
 80 281 282 283 284 285 286 287 288
 GGGACGTACGGTGCAGGACTGGGTCTATCAGCAGGTATCCCAAGACGGCATAATGTTGGGAGGAGCAGCCGGAAGTTGGAT
 GGGACGTACGGTGCAGGACTGGGTCTATCAGCAGGTATCCCAAGACGGCATAATGTTGGGAGGAGCAGCCGGAAGTTGGAT
 1234567890123456789012345678901234567890123456789012345678901234567890
 88 289 290 291 292 293 294 295 296
 CTCAAGATTGAGCTTCCTTGA
 CTCAAGATTGAGCTTCCTTGA
 1234567890123456789012
 96 297 298

FIGURA 2: ADNc de BGL1 y secuencias proteicas (SEQ ID NO: 2 Y 4).

ATGAAGGCTGCTGGCTTTCCCTGCCTCTTTCGGCAGTACCCCTTGCCGTTGTCAGGGCCCATTTGAATCGAGAAAGGTTACACGAAAGCCCTC
 M K A A A L S C L F G S T L A V A G A I E S R K V H Q K P L
 123456789012345678901234567890123456789012345678901234567890123456789012345678901234567890
 1 2 3 4 5 6 7 8 9
 GCGAGATCTGAACCTTTTACCCGTCGCCATGGATGAATCCCAACGCCGACGGCTGGGGGAGGGCCCTATGCCCCAGGCCAAGTCCTTTGTC
 A R S E P F Y P S P W M N P N A D G W A E A Y A Q A K S F V
 123456789012345678901234567890123456789012345678901234567890123456789012345678901234567890
 10 11 12 13 14 15 16 17 18
 TCCCAATGACTCTGTAGAGAAGGTCAACTTGACCACGGGAGTCGGCTGGGGGCTGAGCAGTGCCTCGGCCCAAGTGGCGCGATCCCT
 S Q M T L L E K V N L T T G V G W G A E Q C V G Q V G A I P
 123456789012345678901234567890123456789012345678901234567890123456789012345678901234567890
 19 20 21 22 23 24 25 26 27
 CGCCTTGGACTTCGCAGTCTGTGCATGCTCCCTCTCGGCATCCGAGGAGCCGACTACAACTCAGCGTTCCCTCTGGCCAGACC
 R L G L R S L C M H D S P L G I R G A D Y N S A F P S G Q T
 123456789012345678901234567890123456789012345678901234567890123456789012345678901234567890
 28 29 30 31 32 33 34 35 36
 GTTGCTGTACCTGGGATCGCGGTCTGATGTACCGTTCGGGCTACGCAATGGCCAGGAGGCCAAAGGCAAGGCGATCAATGTCCTCTC
 V A A T W D R G L M Y R R G Y A M G Q E A K G K G I N V L L
 123456789012345678901234567890123456789012345678901234567890123456789012345678901234567890
 37 38 39 40 41 42 43 44 45
 GGACAGTCGCGCGCCCTTGGCCGCGATGCCCGAGGGCGGTCTGTAAGTGGGAGGCTTCGCTCCGATCCCGTCCCTACCGGCATCGGC
 G P V A G P L G R M P E G G R N W E G F A P D P V L T G I G
 123456789012345678901234567890123456789012345678901234567890123456789012345678901234567890
 46 47 48 49 50 51 52 53 54
 ATGTCGAGACGATCAAGGGCATTCAGGATGCTGGCGTCATCGCTTGTGCGAAGCACTTTATTGGAAACGAGCAGGAGCACTTCAGACAG
 M S E T I K G I Q D A G V I A C A K H F I G N E Q E H F R Q
 123456789012345678901234567890123456789012345678901234567890123456789012345678901234567890
 4 55 56 57 58 59 60 61 62 63

GTGCCAGAAGCCAGGGATACGGTTACAACATCAGCGAAACCCCTCTCCTCCAACATTGACGACAAGACCATGACGAGCTCTACCTTTGG
 V P E A Q G Y G Y N I S E T L S S N I D D K T M H E L Y L W
 12345678901234567890123456789012345678901234567890123456789012345678901234567890
 3 64 65 66 67 68 69 70 71 7
 CCGTTTGGCGATGCCGTCCGGGGCCGGCTCTGTCTCATGTGTCTCGTACCAGCAGGTCAACAACCTCGTACGCCCTGCCAGAACTCGAAG
 P F A D A V R A G V G S V M C S Y Q Q V N N S Y A C Q N S K
 12345678901234567890123456789012345678901234567890123456789012345678901234567890
 2 73 74 75 76 77 78 79 80 8
 CTGCTGAACGACCTCCTCAAGAACGAGCTTGGGTTTCAGGGCTTCGTCTATGAGCGACTGGCAGGCACAGCACACTGGCGCAGCAAGCGCC
 L L N D L L K N E L G F Q G F V M S D W Q A Q H T G A S A
 12345678901234567890123456789012345678901234567890123456789012345678901234567890
 1 82 83 84 85 86 87 88 89 9
 GTGGCTGGTCTCGATATGTCCATGCCGGGGCGACACCCAGTTCAACACTGGCGTCAGTTTCTGGGGCGCCAAATCTCACCCCTCGCCGTCCCTC
 V A G L D M S M P G D T Q F N T G V S F W G A N L T L A V L
 12345678901234567890123456789012345678901234567890123456789012345678901234567890
 0 91 92 93 94 95 96 97 98 9
 AACGGCACAGTCCCTGCCTACCGTCTCGACGACATGGCCATGCGCATATGGCCGCCCTCTTCAAGGTACCAAGACCCAGACCGACCTGGAA
 N G T V P A Y R L D D M A M R I M A A L F K V T K T T D L E
 12345678901234567890123456789012345678901234567890123456789012345678901234567890
 9 100 101 102 103 104 105 106 107 1
 CCGATCAACTTCTCCTTCTGGACCGACGACACTTATGGCCCGATCCACTGGCCCGCCAAAGCAGGGCTACCAGGAGATTAAATCCCACGTT
 P I N F S F W T D D T Y G P I H W A A K Q G Y Q E I N S H V
 12345678901234567890123456789012345678901234567890123456789012345678901234567890
 08 109 110 111 112 113 114 115 116 1
 GACGTCCGGCCGACCCAGCGCAACCTCATCCGGGAGATTGCCGCCCAAGGTACGGTGTGCTGAAGAATACCGGCTCTCTACCCCTGAAC
 D V R A D H G N L I R E I A A K G T V L L K N T G S L P L N
 12345678901234567890123456789012345678901234567890123456789012345678901234567890
 17 118 119 120 121 122 123 124 125 1
 AAGCCAAAGTTTCGTGGCCGTCATCGCGGAGGATCTGGGTTCGAGCCCCAACGGGCCCGCTGCAGCGACCGCGGCTGTAAAGGCG
 K P K F V A V I G E D A G S S P N G P N G C S D R G C N E G
 12345678901234567890123456789012345678901234567890123456789012345678901234567890
 26 127 128 129 130 131 132 133 134 1

ACGCTCGCCATGGGCTGGGATCCGGCACAGCCAACTATCCGTACCTCGTTTCCCCGACGCCGCTCCAGGCCCGGGCCATCCAGGAC
 T L A M G W G S G T A N Y P Y L V S P D A A L Q A R A I Q D
 12345678901234567890123456789012345678901234567890123456789012345678901234567890
 35 136 137 138 139 140 141 142 143 1
 GGCACGAGGTACGAGAGCGTCTGTCCAACTACGCCGAGGAAAGACAAAGGCTCTGTCTCGAGGCCAATGCAACCGCCATCGTCTTC
 G T R Y E S V L S N Y A E E K T K A L V S Q A N A T A I V F
 12345678901234567890123456789012345678901234567890123456789012345678901234567890
 44 145 146 147 148 149 150 151 152 1
 GTCAATGCCGACTCAGGCGAGGGCTACATCAACGTGGACGGTAACGAGGGCGACCGTAAGAACCCTGACTCTCTGGAACAACGGTGATACT
 V N A D S G E G Y I N V D G N E G D R K N L T L W N G D T
 12345678901234567890123456789012345678901234567890123456789012345678901234567890
 53 154 155 156 157 158 159 160 161 1
 CTGGTCAAGAACGTCTCGAGCTGGTGCAGCAACACCATCGTCGTCATCCACTCGGTGCGGCCGGTCTCTCCTGACCGATTGGTACGACAAC
 L V K N V S S W C S N T I V V I H S V G P V L L T D W Y D N
 12345678901234567890123456789012345678901234567890123456789012345678901234567890
 62 163 164 165 166 167 168 169 170 1
 CCCAACATCACGGCCATTCTCTGGGCTGGTCTTCCGGGCCAGGAGTCGGGCAACTCCATCACCGACGTGCTTTACGGCAAGTCAACCCC
 P N I T A I L W A G L P G Q E S G N S I T D V L Y G K V N P
 12345678901234567890123456789012345678901234567890123456789012345678901234567890
 71 172 173 174 175 176 177 178 179 1
 GCCGCCGCTCGCCCTTCACTTGGGCAAGACCCCGGAAGCTATGGCGCGGACGTCTGTACAAGCCGAATAATGGCAATGGTGCGCCC
 A A R S P F T W G K T R E S Y G A D V L Y K P N N G N G A P
 12345678901234567890123456789012345678901234567890123456789012345678901234567890
 80 181 182 183 184 185 186 187 188 1
 CAACAGGACTTCACCGAGGGCTTTCATCGACTACCGTACTTCGACAAGTTGACGATGACTCGGTCTATCTACGAGTTCGGCCACGCC
 Q Q D F T E G V F I D Y R Y F D K V D D S V I Y E F G H G
 12345678901234567890123456789012345678901234567890123456789012345678901234567890
 89 190 191 192 193 194 195 196 197 1
 CTGAGCTACACCACTTCGAGTACAGCAACATCCGGCTCGTCAAGTCCAACGTACGCGAGTACCGGCCCCACGACGGGCAACCGGCCCAG
 L S Y T T F E Y S N I R V V K S N V S E Y R P T T G T A Q
 12345678901234567890123456789012345678901234567890123456789012345678901234567890
 98 199 200 201 202 203 204 205 206 2

G C C C G A C G T T T G G C A A C T T C T C C A C C G A C C T C G A G G A C T A T C T C T T C C C A A G G A C G A G T T C C C C T A C A T C T A C C A G T A C A C C G
 A P T F G N F S T D L E D Y L F P K D E F P Y I Y Q Y I Y P
 1234567890123456789012345678901234567890123456789012345678901234567890
 07 208 209 210 211 212 213 214 215 2
 T A C C T C A A C A C G A C C C C G G A G G C C T C G G C C G A T C C C C A C T A C G G C C A G A C C G C C G A G G A G T T C C T C C G C C C A C G C C A C C G A T
 Y L N T T D P R A S A D P H Y G Q T A E E F L P P H A T D
 1234567890123456789012345678901234567890123456789012345678901234567890
 16 217 218 219 220 221 222 223 224 2
 G A C G A C C C C A G C C G T C C T C G G G C G G A A C T C C C C G G G C A A C C G C C A G C T G T A C G A C A T T G T C T A C A C A T C A C G G C C
 D D P Q P L L R S S G G N S P G G N R Q L Y D I V Y T I T A
 1234567890123456789012345678901234567890123456789012345678901234567890
 25 226 227 228 229 230 231 232 233 2
 G A C A T C A C G A A T A C G G C T C C G T T G T A G C G A G G A G T A C C G C A G C T C T A C G T C T C G C T G G G C C G T C C C G A G G A T C C C A A G G T G C A G C T G
 D I T N T G S V V G E E V P Q L Y V S L G G P E D P K V Q L
 1234567890123456789012345678901234567890123456789012345678901234567890
 34 235 236 237 238 239 240 241 242 2
 C G C G A C T T T G A C A G G A T G C G G A C C C G G C G A C G A G G C A G T T C A C C G G C C G C C T G A C C G C C A G A G A T C T G A G C A A C T G G G A C G T C
 R D F D R M R I E P G E T R Q F T G R L T R R D L S N W D V
 1234567890123456789012345678901234567890123456789012345678901234567890
 43 244 245 246 247 248 249 250 251 2
 A C G G T G C A G G A C T G G T C A T C A G C A G G T A T C C C A A G A C G G C A T A T G T T G G A G G A G C A G C C G G A A G T T G G A T C T C A A G A T T G A G C T T C C T
 T V Q D W V I S R Y P K T A Y V G R S S R K L D L K I E L P
 1234567890123456789012345678901234567890123456789012345678901234567890
 52 253 254 255 256 257 258 259 260 2
 tga
 *
 1234
 61

FIGURA 3: Estructura del exón del gen *bg/1* de C1

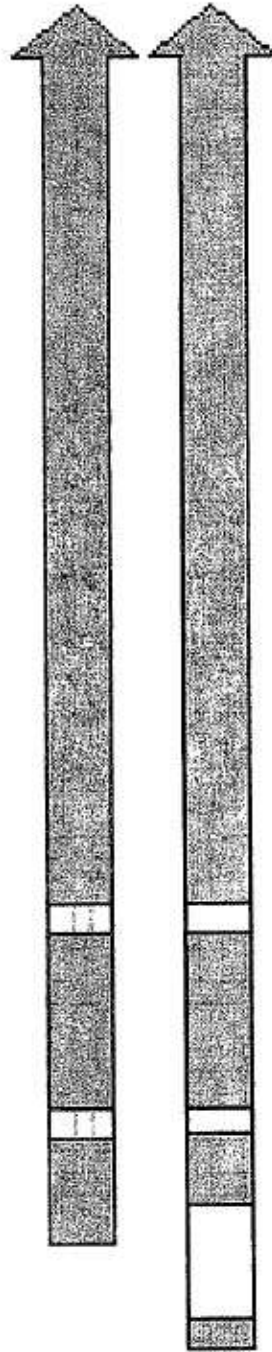


FIGURA 4: C1 CBH1a secuencia promotora (SEQ ID NO:8)

GAGCTCCACCGCGGTGGCGCGCGGATCTTACAAAAAAGGCTATCCGATTGGGAAACGTCGATAAAGTATTGCAAAAGTGACGAGAGTTGCGCAACTA
 ACTCGCTGCCGAAGAAGCTGCGGAAGAAAGAGAACACCGAAAGTGAATAACGTTACGGATGTCCTGACCTCAAAGTTGAAACACGACCTTCCGTGC
 TCTATTGGGAAAGCGGCTTGCCCTTGAATGCGCTGCACTGTGGCAGCACTACCAAGTATCGGGAGGAGCAAACTACCTGGTCCGTTCTTGGTG
 GGGCGCACTAGGCCCAACTTAGGGTGATCGGAGGTGATGCCGGGTCTCGTTGGTCTGGGCTCTTCTCATTTCCGGTTTGCACCCCCCGTTG
 CACCTGCTGATCGCCCCGCAACGCCGATGAGGTTGCGCCAGACCGACAATCACCGGGTGCATTTCCAAAGTATATTGAAGATGGCACCCAGGTAC
 CCGGTTTTCGCTCCAGTTCGTTTGGTGCCAAATTTTGGAGTTTGTAGCCTCAAGATCTGGGAAATCGACCTCAACTTCCATACAAGTTAAAGTC
 GCACACACGGCGAGTTCACGGAAGAGACACATTTTCTGAAGGCCCTCTCTCCCGCACATCAGAAACCAACCAAGACTGCAGAAAGCCG
 GGTAAAGTGGGCCACCGGACTACACTAAATGCGGGGAGAACGAGATCCGTTGCGAAGGGAAGGATGGGGTGTGCTGGGCTTCTCCGCTCT
 CGTGCCCTTTTGCTTGAATCTAGTGACACCAAGGTAGCTCCGAAGGAGTATCTACGGCAGCGCTGTTCGTGCTGCGTTGAGAGTCAGGGCGGA
 GACGAGCAGGACAGGACCTCGCACCGGCACTTCGGATCGCATTTGCGGGAGCGTCAAAATACGCTCTTCTGCGGTTCATCAGAGAGCATCGTGA
 ACCAAGTCTTCCGACGGCGGCTGGCTTCGCAGAGTCGCACCTCGCGGACGCTTCCGTGTCACCCCTGATAACCTGGCTGCCGCGCCGAG
 CTCTCCAAATGAGGTGTGTGGTTGCCCTCGCCGACCTTCAGCAACCTTAATCGCTTCCATCGCACGGCTCCACGCTCTCGAACGATGCCCTCAGT
 CCGTGCCCGCGGTGGCAACCATACGTGACATCGCCGCCAGCCTACTAGCCGCTATCGACCGGTAGGCTTGTACCCGAGCGGCCATTCTCCA
 TCGGGCTCTACTCTGATCCACCTCACCCACCGCAAGCACTAGCGAGCCTCACAGAGTGCAAGCGACACGACCGGCTTGGCCCTTCGTCTTGAC
 TATCTCCCAGACCTCTTGCCATCTTGCCGACGCCGCCCTTTTCTCTCCCTGCGGAGGTGGTGGCCCAAGTCCCGAGATGGCATT
 GCTCCGTTGTCCATGACGACCCATCATTCGATGGCTGACTGGCACACTCGTCTTGTTTGAGCATCGACGGCCCGCGGCCGCTCTCCACGGTACGG
 AACCTCGTTGTACAGTACCTCTCGTAATGATACCCAACACCGGGGCCGAGCGCTGGGAGGGCGGCTTCCCGAAGACCGGGAAGCGGCTGGCGG
 GCTGACCTTTGTGACTTGGCGATGATGCGGCCATGGAGAATGTCGTCGGAAGCGACGCGACAATTAGCCTGGTACCATCGATATAAATTGGGT
 GATCCCAGCTCTTGATGGGCGTGTCTTCTGCTTGGCAGCCCTCGTCTTCAGATCAAGCAACTGTGTGCTGATCTCTTCCGT

FIGURA 5: Secuencia polinucleotídica que codifica la proteína C1 β -glucosidasa (BGL1) que utiliza preferencia codónica para la expresión en *Saccharomyces cerevisiae* (SEQ ID NO:9).

ATAGAAAGTAGAAAGGTA
 I E S R K V
 CATCAAAAACCATTAGCTAGATCAGAACCATTTACCCCTTCTCCATGGATGAACCCTAATGCAGATGGATGGCAGAAAGCATATGCTCAG
 H Q K P L A R S E P F Y P S P W M N P N A D G W A E A Y A Q
 GCCAAGAGTTTGTCTCCAGATGACTCTGTGTGAAAAGGTTAATCTGACAACAGGAGTAGGATGGGTGCAGAACAGTGTCTCGGCCAA
 A K S F V S Q M T L L E K V N L T T G V G W G A E Q C V G Q
 GTTGGTGTATCCCTAGATTGGGTCTTAGAAGTTTGTGTATGCACGATTCTCCCTTAGGTATAAGAGCGCTGACTATACTCAGCATTC
 V G A I P R L G L R S L C M H D S P L G I R G A D Y N S A F
 CCATCCGGGCAAACTGTTGCTGCGACATGGACAGGGGTTTGATGTATAGAGGGGTTATGCGATGGGTCAAGAGGCAAGGGTAAGGA
 P S G Q T V A A T W D R G L M Y R R G Y A M G Q E A K G K G
 ATTAATGTATTGTTGGGGCCGGTGGCGGGCCACTGGGAAGAATGCCAGAAGGTGGGAAGGAATTCGGCAAGGATTCGGCCCCCGACCCAGTG
 I N V L L G P V A G P L G R M P E G G R N W E G F A P D P V
 CTAACAGGTATAGGTATGTCGAAACCATCAAAAGGCATACAAAGATGCAGGTGTTATCGCCTGTGCGAAGCATTTTATTGGTAATGAACAA
 L T G I G M S E T I K G I Q D A G V I A C A K H F I G N E Q
 GAGCATTTTCGTCAAAGTGCCAGAGGCTCAAGGTTATGGTTATAATATTTCTGAAACTTTAAAGTTCCAACATCGAIGACAAACCATGCAC
 E H F R Q V P E A Q G Y G Y N I S E T L S S N I D D K T M H
 GAGTATACTTATGGCCTTTTGACAGCGCTGTGACAGCTGGCGTCTGTATGTCTCTTATCAGCAAGTTAAATACTTTACGCC
 E L Y L W P F A D A V R A G V G S V M C S Y Q Q V N N S Y A
 TGTCAAAATCCAAAGTTACTAAATGACTTATTGAAGAACCAACTAGGATTCCAAGGATTCGTATGAGCGGATTGGCAAGCACAGCATACT
 C Q N S K L L N D L L K N E L G F Q G F V M S D W Q A Q H T

GGTGTGTCATCCGCTGTGGCAGGATTAGATATGTCATATGCCAGGAGATACACAATTTAATACTGGCGTTAGTTTTTGGGGTGC AAAACCTA
G A A S A V A G L D M S M P G D T Q F N T G V S F W G A N L

ACTTTAGCTGTTCTAAACGGTACGGTACCTGCGATATCGTTTAGACGACATGGCCATGCGTATAATGGCTGCTTTATTCAAAGTTACAAAA
T L A V L N G T V P A Y R L D D M A M R I M A A L F K V T K

ACCACCGATTTAGAACCAATTAATTTTAGTTTTTGGACAGATGACACATATGGTCTATACACTGGGCTGCTAAGCAAGGTACCAAGAA
T T D L E P I N F S F W T D D T Y G P I H W A A K Q G Y Q E

ATAAATAGTCACGTTGACGTAAGAGCGGATCACGGCAATCTTATCAGAGAGATAGCAGCAAGGAACTGTATTTGTTGAAGAATACTGGT
I N S H V D V R A D H G N L I R E I A A K G T V L L K N T G

TCATTACCACTAAACAAAGTTTGTGCGAGTCATTTGGTGAAGATGCTGGTTTCATCCCTAATGGACCAATGGTTGTAGTGACAGA
S L P L N K P K F V A V I G E D A G S S P N G P N G C S D R

GGCTGCAATGAAGGCACGTTGGCAATGGGCTGGGCTCAGGACTGCCAATTACCCCTATTGGTCTCTCCGGATGGGCTTTACAGGCT
G C N E G T L A M G W G S G T A N Y P Y L V S P D A A L Q A

AGAGCAATCCAGGATGGTACTAGATACGAGAGCGTCCTAAGTAACATGCGGAAGAAAGACTAAGGCCTTAGTCAGTCAAGCCAATGCC
R A I Q D G T R Y E S V L S N Y A E E K T K A L V S Q A N A

ACTGCTATCGTTTTTCGTAACGCGGATTCTGGCGAAGGTTATATCAATGTTGATGCTAATGAAGGTGACAGAAAGAAATTTAACTTTATGG
T A I V F V N A D S G E G Y I N V D G N E G D R K N L T L W

AATAACGGGACACATTAGTTAAAAATGTATCAAGTTGGTGTCCCAATACTATCGTGTGATACATTCTGTGTCAGTTTTTACTGACA
N N G D T L V K N V S S W C S N T I V V I H S V G P V L L T

GACTGGTACGATAACCCAAACATTACCGCCATTTTATGGCAGGCTCTCCAGGCGCAGGAATCAGGAAATTCATTACGGACGTACTATAC
D W Y D N P N I T A I L W A G L P G Q E S G N S I T D V L Y

GGAAAGGTTAACCCAGCCGCCAGCCCTTTTCACATGGGGTAAAGACAAGAGAGAGCTACGGAGCTGATGTTCTTTATAACCGAACAC
G K V N P A A R S P F T W G K T R E S Y G A D V L Y K P N N

GGGAAATGGAGCGCCACAGCAAGATTTTACTGAAGGTGTGTTTCATTTGACTATAGATACTTCGACAAAGTTGACGATGACTCAGTTATATAT
 G N G A P Q Q D F T E G V F I D Y R Y F D K V D D S V I Y
 GAAATTCGGTCACGGTCTATCTTATACTACTCTTTTGAATATTTCAAATATAAGAGTAGTCAAAAGTAATGTTTCTGAATATAGGCCGACCCACC
 E F G H G L S Y T T F E Y S N I R V V K S N V S E Y R P T T
 GGAACGACGGCTCAAGCGCCTACCTTCGGTAATTTTTCACCGGATTTAGAAGATTATTATTCCCAAGACGAATTTCCATACATCTAC
 G T T A Q A P T F G N F S T D L E D Y L F P K D E F P Y I Y
 CAATACATATACCCCTATCTGAATACTACCGATCCAAAGAGAGCTTCTGCCGATCCACATTACGGGCAGACTGCCGAAGAGTTCTTTGCCA
 Q Y I Y P Y L N T T D P R R A S A D P H Y G Q T A E E F L P
 CCACACGCTACTGACGACGATCCCTCAACCTCTTCTGAGGTCCAGTGGCGGAAATTCACCTGGTGGTAATAGGCAGCTGTATGATATTGTG
 P H A T D D D P Q P L L R S S G G N S P G G N R Q L Y D I V
 TATACTATAACGGCTGATATTACTAATACTGGTAGCGTTGTTGGTGAAGAAGTCCCGCAATTATATGTGTCTTTAGGTGGTCCGGAAGAT
 Y T I T A D I T N T G S V V G E E V P Q L Y V S L G G P E D
 CCTAAGGTTCAGTTAAGAGACTTTGATAGGATGAGAAATAGAACCCTGGAGAAACTAGGCAATTTACAGGTAGATTGACCCGTAGGGATCTG
 P K V Q L R D F D R M R I E P G E T R Q F T G R L T R R D L
 TCAAACCTGGGATGTAACAGTGAAGATTGGGTAATCAGCAGGTACCCGAAACTGCATACGTTGGGTAGATCTTCCCGTAAGTTAGATTG
 S N W D V T V Q D W V I S R Y P K T A Y V G R S S R K L D L
 AAAATTGAATTGCCATAA
 K I E L P *