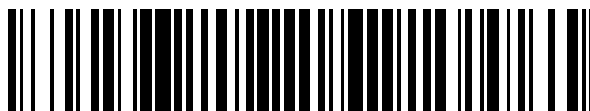


19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 623 288**

51 Int. Cl.:

C12N 9/30 (2006.01)
C12N 9/42 (2006.01)
C12P 19/02 (2006.01)
C12P 7/14 (2006.01)
C12P 19/14 (2006.01)
C12P 7/10 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **22.08.2011** **PCT/US2011/048700**

87 Fecha y número de publicación internacional: **23.02.2012** **WO12024698**

96 Fecha de presentación y número de la solicitud europea: **22.08.2011** **E 11818904 (2)**

97 Fecha y número de publicación de la concesión europea: **08.03.2017** **EP 2606131**

54 Título: **Uso de proteínas de la familia de glicólido hidrolasa 61 en el procesamiento de celulosa**

30 Prioridad:

20.08.2010 US 375788 P

45 Fecha de publicación y mención en BOPI de la traducción de la patente:
10.07.2017

73 Titular/es:

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

72 Inventor/es:

CAMPOPIANO, ONORATO;
TOROK, JANOS;
BAIDYAROY, DIPNATH;
CLARK, LOUIS;
RAO, KRIPA;
SZABO, LORAND y
YANG, JIE

74 Agente/Representante:

IZQUIERDO BLANCO, María Alicia

ES 2 623 288 T3

Aviso: En el plazo de nueve meses a contar desde la fecha de publicación en el Boletín Europeo de Patentes, de la mención de concesión de la patente europea, cualquier persona podrá oponerse ante la Oficina Europea de Patentes a la patente concedida. La oposición deberá formularse por escrito y estar motivada; sólo se considerará como formulada una vez que se haya realizado el pago de la tasa de oposición (art. 99.1 del Convenio sobre Concesión de Patentes Europeas).

DESCRIPCIÓN

Uso de proteínas de la familia de glicólido hidrolasa 61 en el procesamiento de celulosa**Campo de la invención**

5 La divulgación se refiere generalmente al campo de enzimas glicolíticas y su uso. Más específicamente, proporciona proteínas GH61 de *Myceliophthora thermophila*, y el uso de tales proteínas en la producción de azúcares fermentables y etanol de biomasa celulósica.

Antecedentes

10 La biomasa celulósica es un recurso renovable significativo para la generación de azúcares fermentables. Estos azúcares pueden usarse para fermentación y otros procesos metabólicos para producir biocombustibles, compuestos químicos y otros productos finales comercialmente valiosos. Mientras la fermentación de azúcares como glucosa a etanol es relativamente sencilla, la conversión eficiente de biomasa celulósica a azúcares fermentables supone un reto. Ladisch et al., 1983, Enzyme Microb. Technol. 5:82.

15 La conversión de biomasa celulósica a azúcares fermentables puede comenzar con pre-tratamientos químicos, mecánicos, enzimáticos u otros pre-tratamientos para aumentar la susceptibilidad de celulosa a hidrólisis. Tal pre-tratamiento puede ir seguido de conversión enzimática de celulosa a celobiosa, celo-oligosacáridos, glucosa y otros azúcares y polímeros de azúcar, usando enzimas que rompen la celulosa. Estas enzimas son referidas colectivamente como "celulasas" e incluyen endoglucanasas, β -glucosidasas y celobiohidrolasas.

20 WO 2009/033071 desvela SEQ ID NO: 44 que comparte el 100% de identidad con enzima GH61 *Myceliophthora thermophila* de SEQ ID: 2 como aquí se desvela.

25 WO 2009/085935 desvela dos enzimas GH61 de *Myceliophthora thermophila* (SEQ ID NO: 2 y 4) y mezclas enzimáticas.

Resumen de la invención

30 La invención proporciona una composición: a) una proteína GH61 recombinantemente producida que comprende una secuencia de aminoácido que es al menos 80% idéntica a SEQ ID NO: 2 o fragmentos de SEQ ID NO: 2 que tienen actividad GH61; y b) enzimas de celulasa que incluyen al menos una endoglucanasa (EG), al menos una β -glucosidasa (BGL), y al menos una celobiohidrolasa de Tipo 1 (CBH1), y al menos una celobiohidrolasa de Tipo 2 (CBH2).

35 La invención también proporciona un método para producir azúcares fermentables de biomasa celulósica que comprende poner en contacto un sustrato celulósico con una composición, medio de cultivo o lisado celular que contiene una proteína GH61 y enzimas de celulasa, donde: a) la proteína GH61 comprende una secuencia de aminoácido que es al menos 80% idéntica a SEQ ID NO: 2 o un fragmento del mismo biológicamente activo; y b) las enzimas de celulasa incluyen al menos una endoglucanasa (EG), al menos una β -glucosidasa (BGL), y al menos una celobiohidrolasa de Tipo 1 (CBH1), y al menos una celobiohidrolasa de Tipo 2 (CBH2).

40 La invención proporciona además un método para producir un producto final a partir de un sustrato celulósico que comprende: a) poner en contacto el sustrato celulósico con una composición de la invención por lo que los azúcares fermentables se producen a partir del sustrato; y b) poner en contacto los azúcares fermentables con un microorganismo en una fermentación para producir un producto final.

45 La invención proporciona un método para producir un producto final a partir de un sustrato celulósico, que comprende: a) obtener azúcares fermentables producidos de acuerdo con un método de la invención; y b) poner en contacto los azúcares fermentables con un microorganismo en una fermentación para producir un producto final.

Resumen de la divulgación

55 Aquí se desvelan proteínas recombinantes de la familia de Glicósido Hidrolasa 61 (GH61) obtenidas de *Myceliophthora thermophila*, y ácidos nucleicos que codifican tales proteínas. La divulgación también proporciona proteínas GH61 aisladas y purificadas de caldo de cultivo de *M. thermophila*. Estas proteínas pueden usarse para aumentar la producción de productos de reacciones donde un sustrato que contiene celulosa sufre sacarificación por una o una combinación de enzimas de celulasa, como endoglucanasas, β -glucosidasas y celobiohidrolasas. La adición o presencia de proteína GH61 recombinante o aislada puede aumentar la producción de enzimas de celulasa, al menos 20%, 30%, 50%, 70%, doble, triple o más.

60 La divulgación proporciona una composición que comprende una proteína GH61 aislada o recombinante, y un método para preparar tal composición. La proteína puede estar aislada de *M. thermophila*, o puede obtenerse a partir de producción recombinantes. Se proporcionan secuencias de aminoácido de veinticuatro proteínas GH61 *M.*

thermophila GH61 de longitud completa. Las proteínas GH61 desveladas incluyen proteínas que comprenden una secuencia de aminoácido que es al menos aproximadamente 60%, aproximadamente 70%, aproximadamente 80%, aproximadamente 85%, aproximadamente 90%, aproximadamente 92%, aproximadamente 93%, aproximadamente 94%, aproximadamente 95%, aproximadamente 96%, aproximadamente 97%, aproximadamente 98% o aproximadamente 99% idéntica a una cualquiera de las proteínas enumeradas (SEQ ID NOS: 1-30) o un fragmentos de la misma biológicamente activo. Un proteína GH61 ejemplar es GH61a (SEQ ID NO: 2). Otras proteínas GH61 ejemplares incluyen GH61o, GH61v, GH61x, GH61b y GH61e (SEQ ID Nos: 6, 13, 15, 16, 19, 30). Otras proteínas GH61 ejemplares incluyen GH61f, GH61v, GH61p, GH61g y GH61i (SEQ ID Nos: 7, 13, 20, 21, 23, 26).

En un aspecto de la divulgación, la composición preparada de acuerdo con el método de esta invención comprende una proteína GH61 aislada o recombinantes y una o más enzimas de celulasa enumeradas en esta divulgación, que incluyen aunque no se limitan a enzimas de celulasa seleccionadas de endoglucanasas (EG), β -glucosidasas (BGL), celobiohidrolasas de Tipo 1 (CBH1), y celobiohidrolasas de Tipo 2 (CBH2).

Otra realización de la divulgación es un método para producir un azúcar fermentable a partir de sustrato celulósico. Una suspensión que comprende el sustrato contacta con una composición que comprende una proteína GH61 para producir azúcares fermentables como glucosa y xilosa a partir del sustrato. La composición también puede contener una o más enzimas seleccionadas de proteínas de celulosa (endoglucanasas, β -glucosidasas, celobiohirolasas de Tipo 1, y celobiohirolasas de Tipo 2), esterases, xilanasas, hemicelulasas, lipasas, proteasas, amilasas y glucoamilasas. El sustrato puede derivarse de, por ejemplo, trigo, paja de trigo, sorgo, maíz, arroz, cebada, bagazo de caña de azúcar, hierbas, pasto varilla, grano de maíz, mazorcas de maíz, fibra de maíz o una combinación de los mismos, ejemplificado por la paja de maíz pre-tratada.

El azúcar fermentable puede recuperarse y usarse para producir un producto final como un alcohol (como etanol o butanol), un alcohol de azúcar (como sorbitol), un ácido orgánico (como ácido láctico, ácido acrílico, ácido acético, ácido succínico, ácido glutámico, ácido cítrico o ácido propinoico), un aminoácido (como glicina, lisina) o ácido aspártico, un ácido orgánico, un alcanos, un alqueno, un diol o glicerol.

Otra realización de la divulgación es un método para aumentar la producción de azúcares fermentables en una reacción de sacarificación por una o más enzimas de celulosa, realizando la reacción en presencia de una proteína GH61 como las referidas anteriormente.

En otro aspecto, la divulgación proporciona un método para hidrolizar un sustrato de celulosa. El sustrato contacta con una composición que comprende una o más proteínas recombinantes GH61, una o más glucosidasas (BGL) y una o más celobiohidrolasas (CBH). En algunas realizaciones de la divulgación, la composición enzimática está sustancialmente libre de endoglucanasa (EG). La hidrólisis puede dar como resultado una producción de glucosa que es al menos 20%, 30%, 50%, 70%, doble o triple o más que de la producción de la misma reacción realizada en ausencia de dicha proteína GH61.

Otra realización de la divulgación es un método para producir un producto final a partir de un sustrato celulósico. El sustrato contacta con una composición que contiene proteína GH61 como ya se ha referido bajo condiciones por las cuales se producen azúcares fermentables. Los azúcares fermentables después contactan con un microorganismo en una fermentación para producir un producto final como los enumerados arriba. Este método es adecuado para preparar un alcohol, particularmente etanol, donde el microorganismo es una levadura.

Otras realizaciones de la divulgación incluyen: a) las proteínas recombinantes GH61 ya referidas, opcionalmente producidas con un péptido heterólogo de señal; b) una secuencia de ácido nucleico que codifica la proteína GH61 que puede estar operativamente unida a un promotor heterólogo; c) una célula huésped que produce tales proteínas GH61 recombinantes, ejemplificadas por *M. thermophila*, levadura, una células huésped *Chaetomium*, *Thielavia*, *Acremonium*, *Myceliophthora*, *Aspergillus*, o *Trichoderma*.

La divulgación incorpora además el uso de una proteína GH61 recombinantemente producida en la producción de etanol. La proteína GH61 comprende una secuencia de aminoácido que es al menos 60%, aproximadamente 70%, aproximadamente 80%, aproximadamente 85%, aproximadamente 90%, aproximadamente 92%, aproximadamente 93%, aproximadamente 94%, aproximadamente 95%, aproximadamente 96%, aproximadamente 97%, aproximadamente 98% o aproximadamente 99% idéntica a una cualquiera de las proteínas enumeradas o un fragmento de las mismas que tenga actividad GH61.

Aquí se describe un ácido nucleico que codifica una proteína GH61. La divulgación también proporciona una célula que contiene una secuencia de ácido nucleico recombinantes que codifica una secuencia de proteína de esta invención (SEQ ID Nos: 1 a 30) operativamente unida a un promotor heterólogo. Las células huéspedes pueden ser (por ejemplo) células *M. thermophila* o células de levadura. La secuencia de ácido nucleico recombinante puede incluir SEQ ID NO. 31 o una cualquiera de SEQ ID Nos: 32 a 59. La célula también puede expresar al menos una proteína de celulasa recombinantes seleccionada de una endoglucanasa (EG), una β -glucosidasa (BGL), una celobiohidrolasa de Tipo 1 (CBH1), y/o una celobiohidrolasas de Tipo 2 (CBH2) y/o variantes de dichas proteínas de celulasa.

En un aspecto, la invención proporciona una composición que contiene una proteína GH61 (por ejemplo, SEQ ID NO: 2), una endoglucanasa (EG), una β -glucosidasa (BGL), una celobiohidrolasa de Tipo 1 (CBH1), y/o una celobiohidrolasa de Tipo 2 (CBH2), donde la masa combinada de la proteína GH61, EG, BGL, CHB1 y CBH2 es al menos aproximadamente 20%, 40%, 60%, 70%, 80%, 90%, 95% o sustancialmente todo el total de la proteína libre de célula en la composición. Una, dos, tres o las cuatro de las enzimas CBH1, CBH2, EG y BGL que pueden estar presentes en la composición pueden ser variantes derivadas de proteínas de celulasa que ocurren de manera natural. La composición también puede ser un caldo de cultivo de celulosa que contiene proteínas de celulasa.

Aquí se describe una proteína GH61 que comprende cualquiera de las SEQ ID Nos: 3 a 30 o un fragmento secretado de la misma. Opcionalmente, la proteína comprende el fragmento secretado y la secuencia peptídica de señal correspondiente, si está presente. En un aspecto relacionado la divulgación proporciona una variante de proteína GH61 con al menos aproximadamente 60%, al menos aproximadamente 70%, al menos aproximadamente 75%, al menos aproximadamente 80%, al menos aproximadamente 85%, al menos aproximadamente 90%, al menos aproximadamente 91%, al menos aproximadamente 92%, al menos aproximadamente 93%, al menos aproximadamente 94%, al menos aproximadamente 95%, al menos aproximadamente 96%, al menos aproximadamente 97%, al menos aproximadamente 98% o al menos aproximadamente 99% de identidad secuencial con un polipéptido o un fragmento biológicamente activo aquí proporcionado.

Aquí se describe una secuencia de ácido nucleico recombinante que codifica una proteína descrita anteriormente, que puede tener la secuencia del gen que ocurre de manera natural o un fragmento del mismo. La secuencia que codifica la proteína del ácido nucleico puede estar operativamente unida a una secuencia de señal heteróloga. Típicamente, la secuencia de ácido nucleico recombinante está operativamente unida a un promotor heterólogo. Se desvela una célula huésped que contiene un ácido nucleico recombinante de la divulgación. La célula puede expresar al menos una, dos, tres o cuatro proteínas recombinantes de celulosa seleccionadas de endoglucanasas (EG), β -glucosidasas (BGL), celobiohidrolasas de Tipo 1 (CBH1), y/o celobiohidrolasas de Tipo 2 (CBH2).

Aquí se describe una composición que contiene al menos una proteína aislada que comprende una secuencia seleccionada de la parte secretada de SEQ ID Nos. 1 a 30, y/o al menos un fragmento biológicamente activo de la misma. La composición también puede contener al menos una endoglucanasa (EG), una β -glucosidasa (BGL), una celobiohidrolasa de Tipo 1 (CBH1), y/o una celobiohidrolasa de Tipo 2 (CBH2), donde la masa combinada de la proteína GH61, EG, BGL, CHB1 y CBH2 es al menos aproximadamente 70% del total de la proteína libre de célula en la composición.

En un aspecto, la divulgación proporciona métodos para sacarificación (a) cultivando una célula de la invención bajo condiciones donde al menos una proteína GH61 se secreta en el caldo de cultivo, y (b) combinando el caldo y una biomasa celulósica bajo condiciones donde ocurre la sacarificación, donde (a) puede tener lugar antes o simultáneamente con (b).

En un aspecto, la divulgación proporciona un método para hidrolizar un almidón, que comprende poner el almidón en contacto con una composición que comprende una o más proteínas GH61 recombinantes y una o más amilasa(s).

Descripción de los dibujos

La FIGURA 1 está tomada de un experimento que usa proteína GH61 recombinantemente producida a partir de *Myceliophthora thermophila*. La proteína se probó para mejorar la actividad de celulasa usando enzimas de celulasa de un caldo de cultivo de *M. thermophila*. La FIGURA 1(A) muestra el porcentaje de producción mejorada de una reacción de sacarificación realizada en presencia de GH61a, en comparación con la misma reacción en ausencia de GH61a. En la FIGURA 1(B), los datos se trazan para mostrar la producción total de glucosa.

La FIGURA 2 está tomada de un experimento que usa GH61 que contiene fracciones aisladas de caldo de cultivo de *M. thermophila*. Las proteínas GH61, GH61f, GH61p y/o GH61a se combinaron con CBH1a y CBH2. Estos resultados demuestran que una mezcla de enzima que comprende estos componentes tiene suficiente actividad enzimática para la conversión de un sustrato de celulosa a glucosa.

La FIGURA 3 muestra el efecto de la proteína GH61a en la viscosidad de biomasa celulósica.

Descripción detallada

I. Introducción

Se determinó que el hongo filamentoso *Myceliophthora thermophila* produce proteínas GH61. Las proteínas GH61 aumenta la producción de azúcares fermentables cuando un sustrato que contiene celulosa sufre sacarificación por una o más enzimas de celulasa. Los azúcares fermentables producidos por sacarificación pueden

usarse, entre otros usos, en reacciones de fermentación para producir productos finales, como, pero sin limitar a, etanol.

Las proteínas GH61 pueden aislarse de células de *M. thermophila*. Las proteínas GH61 también pueden producirse recombinantemente expresando un ácido nucleico que codifica cualquiera de las secuencias de proteína GH61 proporcionadas en la divulgación, incluyendo secuencias de tipo salvaje y fragmentos de secuencias de tipo salvaje.

La preparación y uso de proteínas GH61, y las composiciones de la invención se describen con más detalle en las secciones a continuación.

II. Definiciones

A menos que se indique lo contrario, la práctica de la presente invención implica técnicas convencionales comúnmente usadas en biología molecular, fermentación, microbiología y campos relacionados, que son conocidas por aquellos expertos en la técnica. A menos que aquí se defina lo contrario, todos los términos técnicos y científicos aquí usados tiene el mismo significado como comúnmente los entiende un experto en la técnica a la que pertenece esta invención. Aunque pueden usarse algunos métodos y materiales similares o equivalentes a aquellos aquí descritos en la práctica o en las pruebas de la presente invención, se describen los métodos y materiales preferentes. De hecho, se pretende que la presente invención no esté limitada a la metodología, protocolos y reactivos particulares aquí descritos, ya que estos pueden variar dependiendo del contexto en el que se usen. Los títulos aquí provistos no son limitaciones de los varios aspectos o realizaciones.

Sin embargo, con el fin de facilitar la comprensión de la presente invención, más abajo se definen un número de términos. Los rangos numéricos incluyen los números que definen el rango. Así, cada rango numérico aquí desvelado pretende incluir cada rango numérico más estrecho que esté clasificado dentro de un rango numérico más amplio, como si tales rangos numéricos más estrechos estuvieran aquí expresamente escritos. También se pretende que cada limitación numérica máxima (o mínima) aquí desvelada incluye cada limitación numérica más baja (o más alta), como si tales limitaciones numéricas más bajas (o más altas) estuvieran aquí expresamente escritas.

Como aquí se usan, los términos “que comprende(n)” y sus cognados se usan en su sentido inclusivo (esto es, equivalentes al término “que incluye(n)” y sus correspondientes cognados).

Como aquí y en las reivindicaciones adjuntas se usa, el singular “uno”, “una”, “el”, “la” incluye la referencia plural a menos que el contexto claramente dicte lo contrario. Así, por ejemplo, la referencia a una “célula huésped” incluye una pluralidad de tales células huéspedes.

A menos que se indique lo contrario, los ácidos nucleicos se escriben de izquierda a derecha en una orientación de 5' a 3'; las secuencias de aminoácido se escriben de izquierda a derecha en orientación carboxi, respectivamente. Los títulos aquí proporcionados no son limitaciones de los varios aspectos o realizaciones que pueden tenerse por referencia a la especificación como un todo. Por consiguiente, los términos definidos más abajo se definen de manera más completa por referencia a la especificación como un todo.

Como aquí se usa, una “proteína GH61” es una proteína con actividad GH61 (o “favorecedor de celulasa”). Una “proteína GH61” o un polipéptido GH61 pueden tener una secuencia de una proteína que ocurre de manera natural (de tipo salvaje) o puede comprender variaciones relativas a una proteína de tipo salvaje.

Como aquí se usan, las secuencias de proteína GH61 mostradas en las Tablas 1 y 2 (SEQ ID Nos: 1-30) se consideran secuencias de tipo salvaje.

Una proteína tiene “actividad GH61” o “actividad favorecedora de celulosa” o es “biológicamente activa” si, cuando se incluye en una reacción de sacarificación (por ejemplo, se realiza usando una endogluconasa, una β -glucosidasa y celobiohidrolasas de Tipo 1 y Tipo 2) da como resultado una mayor cantidad (esto es, mayor producción) de uno o más azúcares solubles (por ejemplo, glucosa) que la reacción de sacarificación realizada bajo las mismas condiciones en ausencia de la proteína GH61. Una “variante biológicamente activa” es una variante o fragmento que conserva al menos algo (por ejemplo, al menos el 10%) de la actividad de GH61 de la proteína de tipo salvaje.

Como aquí se usa, una proteína GH61 “variante” (o polinucleótido que codifica una proteína GH61) es una proteína GH61 que comprende una o más modificaciones relativas a GH61 de tipo salvaje (o el polinucleótido de tipo salvaje que codifica GH61). Las modificaciones incluyen sustituciones, inserciones, eliminaciones y/o truncamientos de amino o carboxi de uno o más de los residuos aminoácidos (o de uno o más nucleótidos o codones en el polinucleótido). Una variante que comprende una eliminación relativa a la proteína de tipo salvaje puede ser referida como un “fragmento”.

Como aquí se usa, un “fragmento” de una proteína GH61 es (a) un polipéptido con una secuencia de tipo salvaje pero que comprende una eliminación relativa a SEQ ID Nos: 1-30 o (b) una variante de GH61 que comprende una eliminación relativa a un polipéptido de SEQ ID Nos: 1-30.

5 Una “sustitución” de aminoácido en una secuencia de proteína es una sustitución de un aminoácido sencillo por una secuencia con otro aminoácido.

10 Una sustitución de aminoácido puede ser una sustitución “conservadora”, en cuyo caso el aminoácido sustituido comparte una o más propiedades químicas con el aminoácido al que sustituye. Las propiedades compartidas incluyen las siguientes: aminoácidos básicos: arginina (R), lisina (K), histidina (H); aminoácidos ácidos: ácido glutámico (E) y ácido aspártico (D); aminoácidos polares no cargados: glutamina (Q) y asparagina (N); hidrofóbicos: leucina (L), isoleucina (I), valina (V); aminoácidos aromáticos: fenilalanina (F), triptófano (W), y tirosina (Y); aminoácidos que contienen sulfuro: cisteína (C), metionina (M); aminoácidos pequeños: glicina (G), alanina (A), serina (S), treonina (T), prolina (P), cisteína (C), y metionina (M).

15 El término “pre-proteína” tiene un significado estándar en la técnica y se refiere a un polipéptido que incluye una región de péptido de señal en terminal amino (o secuencia líder) unida. El péptido de señal se parte de la pre-proteína por una peptidasa de señal simultánea a la secreción de la proteína. La parte secretada de la proteína puede ser referida como proteína “madura” o proteína “secretada”. Así, una secuencia de aminoácido de una pre-proteína comprenderá una parte de péptido de señal y una parte secretada (madura).

20 Como aquí se usan, los términos “péptido de señal” tiene su significado habitual y se refiere a una secuencia de aminoácido enlazada con la terminal amino de un polipéptido, que dirige al polipéptido codificado a una secuencia secretora de célula.

25 La “sacarificación” se refiere a una reacción catalizada por enzima que da como resultado hidrólisis de un carbohidrato complejo a azúcares fermentables (por ejemplo, monosacáridos como glucosa o xilosa). Las enzimas pueden ser enzimas de celulosa, tales como endoglucanasas, β -glucosidasas y celobiohidrolasas de Tipo 1 y Tipo 2, y combinaciones de tales enzimas de celulosa. Las enzimas de celulosa pueden ser de un caldo de cultivo de un organismo de tipo salvaje o construido con celulasa que produzca enzimas de celulosa (por ejemplo, un hongo como *M. thermophila* o levadura).

30 La “hidrólisis” de celulosa u otro polisacárido (por ejemplo, almidón) ocurre cuando los enlaces glicosídicos entre al menos algunos de los monosacáridos adyacentes se hidrolizan, separando así los pares monómeros previamente unidos.

35 “Aumentar” la producción de un producto (como un azúcar fermentable) de una reacción ocurre cuando un componente particular presente durante una reacción (como una proteína GH61) provoca que se produzca más de un producto, en comparación con una reacción realizada bajo las mismas condiciones con el mismo sustrato y otros sustituyentes, pero en ausencia del componente de interés.

40 Los términos “mejorado/a” o “propiedades mejoradas”, como se usan en el contexto de descripción de las propiedades de una variante de GH61, se refieren a un polipéptido variante de GH61 que muestra una mejora en una propiedad o propiedades en comparación con GH61 de tipo salvaje (por ejemplo, SEQ ID NO: 2) o un polipéptido de referencia especificada. Las propiedades mejoradas pueden incluir, aunque no se limitan a, una mayor expresión de proteínas, una mayor termoactividad, una mayor termoestabilidad, una mayor actividad específica, una mayor especificidad de sustrato, una mayor resistencia al sustrato o a inhibición del producto final, una mayor estabilidad química, inhibición reducida por glucosa, una mayor resistencia a inhibidores (por ejemplo, ácido acético, lectinas, ácidos tánicos y compuestos fenólicos) y un perfil alterado pH/temperatura.

45 El término “celulasa” (o “enzima de celulasa”) se refiere en términos generales a enzimas que catalizan la hidrólisis de enlaces de celulosa β -1,4-glicosídicos a cadenas más cortas de celulosa, oligosacáridos, celobiosa y/o glucosa, por ejemplo, endoglucanasas, β -glucosidasas y celobiohidrolasas.

50 Como aquí se usa, el término “polinucleótido” se refiere a un polímero de deoxiribonucleótidos o ribonucleótidos en una forma de una cadena o doble cadena, y complementos de los mismos.

55 Como aquí se usa, un “gen” es una secuencia de ácido nucleico que codifica una proteína. El ácido nucleico puede o no tener intrones, puede o no ser recombinante y puede o no comprender además elementos que afectan a la transcripción o traslación. Para fines de esta descripción, una secuencia cADN que codifica una proteína puede ser referida algunas veces como un “gen”.

60 Los términos “ácido nucleico recombinantes” tienen su significado convencional. Un ácido nucleico recombinante, o equivalentemente, “polinucleótido”, es uno que se inserta en una localización heteróloga de tal manera que no esté asociado con secuencias de nucleótidos que normalmente flanquearían el ácido nucleico

cuando se encuentra en la naturaleza (por ejemplo, un ácido nucleico insertado en un vector o un genoma de un organismo heterólogo). De la misma manera, una secuencia de ácido nucleico que no aparece en la naturaleza, por ejemplo, una variante de un gen que ocurre de manera natural, es recombinante. Una célula que contiene un ácido nucleico recombinante, o proteína expresada *in vitro* o *in vivo* de un ácido nucleico recombinante también son recombinantes. Ejemplos de ácidos nucleicos recombinantes incluyen una secuencia de ADN que codifica una proteína que (i) esté operativamente unida a un promotor heterólogo y/o (ii) codifique un polipéptido de fusión con una secuencia de proteína y una secuencia de péptido de señal heterólogo.

Los ácidos nucleicos "hibridan" cuando se asocian, típicamente en solución. Los ácidos nucleicos hibridan debido a una variedad de fuerzas físico-químicas bien caracterizadas, como un enlace de hidrógeno, exclusión de disolvente, acumulación de base y similares. Como aquí se usan, los términos "condiciones de lavado con hibridación severa" en el contexto de experimentos de hibridación de ácido nucleico, como hibridación Southern y Northern, son dependientes de secuencia, y son diferentes bajo diferentes parámetros medioambientales. Una guía extensiva para la hibridación de ácidos nucleicos se encuentra en Tijssen, 1993 "Técnicas de laboratorio en bioquímica y biología molecular - Hibridación con sondas de ácido nucleico", Parte I, Capítulo 2 (Elsevier, Nueva York). Para polinucleótidos de al menos 100 nucleótidos de longitud, las condiciones de baja a muy alta severidad se definen de la siguiente manera: pre-hibridación e hibridación a 42 °C en 5xSSPE, 0,3% SDS, 200 µg/ml de ADN de esperma de salmón cortado y desnaturalizado y 25% formamida para severidades bajas, 35% formamida para severidades medias y medias-altas o 50% formamida para severidades alta y muy altas, siguiendo los procedimientos de ensayo Southern estándar. Para polinucleótidos de al menos 100 nucleótidos de longitud, el material transportador se lava finalmente tres veces cada 15 minutos usando 2xSSC, 0,2% SDS 50 °C (severidad baja), a 55 °C (severidad media), a 60 °C (severidad media-alta), a 65 °C (severidad alta) o a 70 °C (severidad muy alta).

Los términos "vector de expresión" se refieren a una molécula de ADN, lineal o circular, que comprende un segmento que codifica un polipéptido de la divulgación, y que está operativamente unido a segmentos adicionales que proporcionan su transcripción (por ejemplo, un promotor, una secuencia exterminadora de transcripción, potenciadores) y opcionalmente un marcador seleccionable.

Para fines de esta divulgación, un promotor es "heterólogo" a una secuencia de gen si el promotor no está asociado por naturaleza con el gen. Un péptido de señal es "heterólogo" a una secuencia de proteína cuando la secuencia de péptido de señal no está asociada con la proteína por naturaleza.

En relación con las secuencias reguladoras (por ejemplo, promotores), los términos "operativamente unido/a" se refieren a una configuración donde una secuencia reguladora está situada en una posición en relación con una secuencia codificadora de polipéptido de tal manera que la secuencia reguladora influya en la expresión del polipéptido. En relación con una secuencia de señal, los términos "operativamente unido/a" se refieren a una configuración donde la secuencia de señal que codifica un péptido de señal de amino-terminal se fusiona con el polipéptido, de tal manera que la expresión del gen produzca una pre-proteína.

Como aquí se usan, los términos "péptido", "polipéptido" y "proteína" se usan aquí intercambiabilmente para referirse a un polímero de residuos de aminoácido.

Como aquí se usa, el término "aminoácido" se refiere a aminoácidos que ocurren de manera natural y sintéticos, así como análogos de aminoácidos.

Los términos "biomasa", "sustrato de biomasa", "biomasa celulósica", "materia prima celulósica", "materia prima lignocelulósica" y "sustrato celulósico", se refieren todos a materiales que contienen celulosa. Por simplicidad, los términos "sustrato celulósico" aquí se usan para referirse a materiales que contienen celulosa que pueden funcionar con celulosas (con proteínas GH61 opcionalmente presentes), típicamente después de pre-tratamiento, para producir azúcares fermentables. Los ejemplos de sustratos celulósicos incluyen, aunque no se limitan a, biomasa como madera, pulpa de madera, pulpa de papel, fibra de maíz, grano de maíz, mazorcas de maíz, residuos de grano como cáscaras de maíz, rastrojos de maíz, hierbas, trigo, paja de trigo, cebada, paja de cebada, heno, paja de arroz; pasto varilla, papel sobrante, sobrante del procesamiento de papel y pulpa, plantas leñosas y herbáceas, pulpa de fruta y vegetales, grano de destilador, vaina de arroz, algodón, cáñamo, lino, sisal, bagazo de caña de azúcar, sorgo, soja, componentes obtenidos de moler granos, árboles, ramas, raíces, hojas, astillas de madera, serrín, matorrales y arbustos, verduras, frutas y flores y mezclas de los mismos. En algunas realizaciones, el material celulósico se "pre-trata", o trata usando métodos conocidos en la técnica, como pre-tratamiento químico (por ejemplo, pre-tratamiento con amoníaco, pre-tratamiento con ácido diluido, pre-tratamiento con álcali diluido o exposición a disolvente), pre-tratamiento físico (por ejemplo, exposición a rayos o radiación), pre-tratamiento mecánico (por ejemplo, pulverización o molienda) y pre-tratamiento biológico (por ejemplo, aplicación de microorganismos que solubilizan lignina) y combinaciones de ellos, para aumentar la susceptibilidad de celulosa a hidrólisis. Los sustratos celulósicos y sus procesos se describen con más detalle más abajo.

Un sustrato celulósico se “deriva de” una fuente específica (como maíz o trigo) mediante un proceso que comprende la obtención de la fuente o una parte física de la fuente (como una mazorca de maíz o paja de trigo), y después opcionalmente pre-tratar la fuente o parte.

Una secuencia de proteína de celulasa (esto es, una variante de celulosa) se “deriva de” una secuencia de celulasa cuando se produce introduciendo variaciones en una secuencia de tipo salvaje usando mutagénesis in vitro o métodos de evolución molecular. Típicamente, la secuencia de proteína será al menos 70%, aproximadamente 75%, aproximadamente 80%, aproximadamente 85%, aproximadamente 90%, aproximadamente 91%, aproximadamente 92%, aproximadamente 93%, aproximadamente 94%, aproximadamente 95% a la secuencia de tipo salvaje.

Se dice que un sustrato celulósico está “pre-tratado” cuando ha sido procesado mediante algunos medios físicos y/o químicos para facilitar la sacarificación.

Los “azúcares fermentables” se refieren a azúcares simples (monosacáridos, disacárido y oligosacáridos cortos) como, aunque sin limitar a, glucosa, xilosa, galactosa, arabinosa, manosa y sacarosa. El azúcar fermentable es cualquier azúcar que un microorganismo pueda utilizar o fermentar.

Los términos “transformar” o “transformación”, como se usa en referencia con una célula, significan que una célula tiene una secuencia de ácido nucleico no nativo integrado en su genoma o como un episoma (por ejemplo, plásmido) que se mantiene a través de múltiples generaciones.

El término “introducido”, como se usa en el contexto de inserción de secuencia de ácido nucleico en una célula, significa conjugado, transfectado, transducido o transformado (colectivamente “transformado”) o incorporado de otra manera en el genoma, o mantenido como un episoma en una célula.

Una célula “construida por celulosa” es una célula que comprende al menos una, al menos dos, al menos tres o al menos cuatro secuencias recombinantes que codifican una celulasa o variante de celulasa, y donde la expresión de la(s) celulasa(s) o variante(s) de celulasa se ha modificado en relación con la forma de tipo salvaje. La expresión de una celulasa se “modifica” cuando una variante de celulasa que ocurre de manera no natural se expresa o cuando una celulasa que ocurre de manera natural se sobreexpresa. Una manera de sobreexpresar una celulasa es unir operativamente un promotor fuerte (opcionalmente constitutivo) con la celulasa que codifica la secuencia. Otra manera de sobreexpresar una celulasa es aumentar el número de copia de un gen de celulasa heterólogo, variante o endógeno. La célula construida por celulasa puede ser una célula fúngica, como una célula de levadura o una célula fúngica filamentosa (por ejemplo, *Acidothermus cellulolyticus*, *Thermobifida fusca*, *Humicola grisea*, *Myceliophthora thermophila*, *Chaetomium thermophilum*, *Acremonium* sp., *Thielavia* sp, *Trichoderma reesei*, *Aspergillus* sp.). En algunas realizaciones la célula construida por celulasa es una célula *M. thermophila*.

El término “cultivar” se refiere a hacer crecer una población de células microbianas (por ejemplo, fúngica) bajo un medio de cultivo líquido o sólido. Algunas células cultivadas expresan o secretan proteínas, como por ejemplo proteínas GH61 o proteínas de celulasa. Cuando las células crecen en un medio líquido pueden secretarse proteínas al medio, lo que es referido mediante varios términos, incluyendo medio de cultivo celular, sobrenadante de cultivo celular y caldo de cultivo.

Como aquí se usa, el término “recuperación”, se refiere a la cosecha, aislamiento, recogida y separación de una proteína, azúcar, célula o producto final (por ejemplo, alcohol) de un caldo de cultivo o, alternativamente, medio de cultivo sólido.

Como aquí se usa, el término “aislado/a” se refiere a un ácido nucleico, polipéptido u otro componente que está parcialmente o completamente separado de componentes con los que normalmente se asocia por naturaleza (como otras proteínas, ácidos nucleicos o células).

Un producto que ha sido “purificado” es un producto que se ha enriquecido de la fuente de la que se obtiene al menos 10 veces, y opcionalmente al menos 100 veces, o al menos 1000 veces. Por ejemplo, una proteína que se obtiene de un caldo de cultivo donde representa el 0,1% de la proteína total puede purificarse para que sea el 1%, 2,5%, 10%, 25%, 50%, 80%, 95% o 100% (p/p) de la proteína en la preparación. Un producto que se ha “purificado parcialmente” se ha enriquecido de la fuente de la que se produce al menos 2 veces, al menos 5 veces, pero aún contiene al menos el 50%, y algunas veces al menos el 90% (p/p) de componentes no relacionados diferentes al disolvente.

“Fraccionar” un producto líquido como un caldo de cultivo significa aplicar un proceso de separación (como precipitación con sal, cromatografía de columna, exclusión por tamaño y filtración) o una combinación de tales procesos para obtener una solución donde una proteína desea (como una proteína GH61, una enzima de celulasa o una combinación de ellas) tenga un porcentaje mayor de proteína total en la solución que el producto líquido inicial.

Una “suspensión” es una solución acuosa donde se dispersan uno o más componentes sólido, como sustrato celulósico.

Como aquí se usa una “composición” que comprende una o más proteínas puede ser, por ejemplo, una composición libre de células que comprende la(s) proteína(s), un lisado celular, un caldo de cultivo que comprende la(s) proteína(s), por ejemplo, en forma secretada; una célula que comprende la(s) proteína(s), como una célula recombinante que expresa la(s) proteína(s); una mezcla de dos o más poblaciones celulares que expresan diferentes proteínas (por ejemplo, una célula que expresa una proteína recombinantes GH61 y una segunda célula que expresa proteína de celulasa).

Los términos “composición libre de células” se refieren a una composición que contiene proteínas donde las células y los restos celulares se han retirado, como una mezcla de celulasa purificada o un caldo de incubación celular que contiene proteína secretada, del cual se han retirado células y material insoluble.

Los términos “identidad porcentual”, “% identidad”, “idéntico porcentual” y “% idéntico” se usan intercambiamente para referirse a una comparación de dos secuencias óptimamente alineadas sobre una ventana de comparación. La ventana de comparación puede incluir adiciones o eliminaciones en cualquier secuencia para optimizar la alineación. El porcentaje de identidad es el número de posiciones que son idénticas entre las secuencias, dividido entre el número total de posiciones en la ventana de comparación (incluyendo posiciones donde una de las secuencias tiene un espacio). Por ejemplo, una proteína con una secuencia de aminoácido que coincide en la posición 310 con una secuencia de GH61a (que tiene 323 aminoácidos en la forma secretada) tendrá 310/323 = 95,9% identidad con la referencia. Similarmente, una variante de proteína que tiene 300 residuos (esto es, menos que la longitud completa) y coincide la secuencia de referencia en 280 posiciones tendrá 280/300 = 93,3% identidad. Mientras la alineación y la clasificación óptimas pueden realizarse manualmente, el proceso puede facilitarse usando un algoritmo de alineación implementado por ordenador. Los ejemplos son los algoritmos BLAST y BLAST 2.0, descritos en Altschul et al., 1990, J. Mol. Biol. 215: 403-410 y Altschul et al., 1977, Nucleic Acids Res. 3389-3402. Alternativamente, el grado de identidad entre dos secuencias de aminoácido puede determinarse usando el algoritmo Needleman-Wunsch (Needleman y Wunsch, 1970, J. Mol. Biol. 48: 443-53), que se ha implementado en el programa Needle del paquete EMBOSS (EMBOSS: The European Molecular Biology Open Software Suite, Rice et al, 2000, Trends in Genetics 16: 276-77), preferentemente la versión 3.0 o posterior. Los parámetros opcionales usados son penalización de espacio abierto de 10, penalización de extensión de espacio de 0,5, y la matriz de sustitución EBLOSUM62 (versión EMBOSS de BLOSUM62). La producción de “identidad más larga” con etiqueta Needle (obtenida usando la opción the-nobrief) se usa como la identidad porcentual y se calcula de la siguiente manera: (Residuos Idénticos x 100) / (Longitud de Alineación – Número Total de Espacios en Alineación).

III. Identificación de genes GH61

Se identificaron veinticuatro proteínas GH61 endógenas a *Myceliophthora thermophila* como se ha descrito en el Ejemplo 1. Véase la TABLA 1 y TABLA 2 más abajo. Como se muestra en los Ejemplos 2 y 3, una proteína particular GH61 de *M. thermophila* con la designación GH61a (SEQ ID NO: 2) se mostró para mejorar las reacciones de sacarificación en presencia de celulasa. Similarmente, pueden usarse otras proteínas GH61 (SEQ ID Nos. 3 a 30) para mejorar la actividad de celulasa.

La TABLA 1 proporciona la secuencia de una pre-proteína GH61 (SEQ ID NO: 1), que muestra el péptido de señal nativo predicho subrayado, y la forma secretada (madura) predicha (SEQ ID NO: 2). La secuencia ID NO: 2 tiene 323 aminoácidos de longitud. Se contempla que ciertas variantes de proteína GH61a comprenderán residuos 11-323 de SEQ ID NO: 2 (incluyendo, aunque sin limitar a, fragmentos truncados amino-terminal), o tendrán al menos: aproximadamente 70%, aproximadamente 75%, aproximadamente 80%, aproximadamente 85%, aproximadamente 90%, aproximadamente 91%, aproximadamente 92%, aproximadamente 93%, aproximadamente 94%, aproximadamente 95%, aproximadamente 96%, aproximadamente 97%, aproximadamente 98% o aproximadamente 99% de identidad de secuencia con residuos 11-323 de SEQ ID NO: 2. En algunas realizaciones, la variante de proteína GH61 tendrá al menos 90% identidad con residuos 11-323 de SEQ ID NO: 2 y tendrá 315 residuos de longitud.

La TABLA 2 proporciona 28 secuencias de pre-proteína GH61, con el péptido de señal nativa predicha subrayada. No se predice que GH61t y GH61n tengan péptidos de señal. Excepto donde se especifique lo contrario o se aclare del contexto, la referencia a una o más secuencias enumeradas en la TABLA 2 se inserta para abarcar una o más de la forma de pre-proteína y la forma secretada (esto es, una proteína que comprende la secuencia entera, y una proteína que no incluye la secuencia subrayada). También se contempla que ciertas proteínas GH61 comprendan residuos 11 con la terminal C de las secuencias mostradas en la Tabla 2 (incluyendo, aunque sin limitar a, fragmentos truncados de amino-terminal), o tendrán al menos: aproximadamente 70%, aproximadamente 75%, aproximadamente 80%, aproximadamente 85%, aproximadamente 90%, aproximadamente 91%, aproximadamente 92%, aproximadamente 93%, aproximadamente 94%, aproximadamente 95%, aproximadamente 96%, aproximadamente 97%, aproximadamente 98% o aproximadamente 99% de identidad de secuencia con residuos 11 para terminal C de estas secuencias. En la Tabla 2 pueden usarse las siguientes denominaciones: B = SEQ ID NO. de proteína de longitud completa (incluyendo péptido de señal); C = Péptido de señal (Subrayado); D = Proteína

madura (No subrayado); E = Parte de D que comienza en residuo 11 de la parte secretada (esto es, entre el espacio y la terminal C, una parte truncada de amino terminal).

Los límites de péptido de señal en la TABLA 1 y TABLA 2 se predicen en base al análisis de la secuencia primaria. Es posible que el sitio real de división difiera del sitio predicho hasta en varios residuos (por ejemplo, 1-20, por ejemplo hasta 1, 2, 3, 4, 5, 6, 7, 8, 9 o 10). La presencia de un péptido de señal heterólogo puede también afectar al sitio de la división. Una división de péptido de señal puede determinarse expresando la secuencia de longitud completa e identificando los residuos de terminal amino de la proteína GH61 secretada. Se espera que la adición o eliminación de varios residuos en la terminal amino de la proteína madura tenga poco o ningún efecto en la actividad de la proteína secretada.

TABLA 1

Glicósido hidrolasa 61 proteína GH61a		
SEQ ID NO:	Designación	
1	GH61a	Pe-Proteína: <u>MSKASALLAGLTGAALVAA</u> HGHVSHIVVN GVYYRNYDPTTDWYQPNPPTVIGWTAADQDN GFVEPNSEFGTPDIICHKSATPGGGHATVAAGDKINIVWTPPEWESHIGPVIDYLAACNGDCE TVDKSSLRWFKIDGAGYDKAAGRWAADALRANGNSWLVOIPSDLKAGNYVLRHEIIALHGAQ SPNGAQAYPQCINLRVTGGGSNLPSGVAGTSLYKATDPGILFNPNVSSPDYTVPGPALIAGA ASSIAQSTSVATATGTATVPGGGGANPTATTTAATSAAPSTTLRTTTTSAQTAPPSSGDVQ TKYGQCGGNGWTGPTVCAPGSSCSVLNEWYSQCL
2	GH61a	Proteína Madura: <u>MSKASALLAGLTGAALVAA</u> HGHVSHIVVN GVYYRNYDPTTDWYQPNPPTVIGWTAADQDN GFVEPNSEFGTPDIICHKSATPGGGHATVAAGDKINIVWTPPEWESHIGPVIDYLAACNGDCE TVDKSSLRWFKIDGAGYDKAAGRWAADALRANGNSWLVOIPSDLKAGNYVLRHEIIALHGAQ SPNGAQAYPQCINLRVTGGGSNLPSGVAGTSLYKATDPGILFNPNVSSPDYTVPGPALIAGA ASSIAQSTSVATATGTATVPGGGGANPTATTTAATSAAPSTTLRTTTTSAQTAPPSSGDVQ TKYGQCGGNGWTGPTVCAPGSSCSVLNEWYSQCL

TABLA 2

Glicósido hidrolasa 61 proteína GH61a

B	Designación	Secuencia (Secuencia de Péptido de Señal Putativa)
3	GH61l	<u>MFSLKFFILAGGLAVL</u> TEA HIRLVSPAPF TNPDQGPSPLLEAGSDYPCHNGNGGGYQGT PTQMAKGSQQLAFQGSVAHGGGSCQVSITYDENPTAQSSFKVIHSIQGGC PARAETIPDC SAQNIACNIKPDNAQMDTPDKYEFTIPEDLPSGKATLAWTWINTIGNREFYMACAPVEIT GDGGSESALAALPDMVIANIPISIGGTCAEEGKYEYPNPGKSVETIPGWTDLVPLQGECC AASGVSGSGGNASSATPAAGAAPTAVRGRRTWNA
4	GH61m	<u>MKLATLLAALT</u> LGVA DQLSVGSRKFG VYEHIRKNTNYSNPVTDLSDTNLRNCGVGGSGT STTVLDVKAGDSFTFFSDVAVYHQGPISLCVDRTSAESMDGREPDMRCRTGSQAGYLAVTD YDGSQDCFKIYDWGPTFNGGQASWPTNSYEYSILKCIIRDGEYLLRIQSLAIHNPGALPQF YISCAQVNVTTGGGTVTPRSRRPILIIYFNHFSYIVPGPAVFKC
5	GH61n	MTKNAQSKQG VENPTSGLDIRCYTSQTAAANVVTPAGSTIHYISTQQINHPGPTQYYLAKV PPGSSAKTFDGSAGVWFKISTTMTPTVDSNKQMFPGQNTYETSNTTIPANTPDGEYLLRVK QIALHMASQPNKVQFYLAQTQKITGGRNGTPSPLVALPGAYKSTDPGILVDIYSMPESY QPPGPPVWRG
6	GH61o	<u>MKPFSLVALATA</u> VSG HAIFQRVSVN GQDQGLKGVRAPSSNSPIQNVNDANMACNANIV YHDNTIIVKIPAGARVGAWWQHVIIGGPQGANPDNPAAASHKGPIQVYLAKVDNAATASPSG LKWFVAERGLNNGVWAYLMRVELLALHSASSPGGAQFYMGCAQIEVTGSGTNSGSDVFSF PGAYSANDPGILLSIYDSSGKPNNGGRSYIPGPRPISCSGSGGGGNNGGDGGDDNNGGGN NNGGGSVPLYGQCGGIGYTGPPTCAQGTCKVSN EYYSQCLP
7	GH61p	<u>MKLTSSLAVLAA</u> AGAAQ HYTFPRAGTG GSLSGEWEVVRMTENHYSHGPVTDVTSPEMTC YQSGVQAGAPQTVQVKAGSQFTFSVDPSIGHPGPLQFYMAKVPVSGQTAAATFDGTGAVWFKIY QDGPNGLTGDSITWPSAGKTEVSVTIPSCIEDGEYLLRVEHTPLPTAPAAQNRARSSPSA AYKATDPGILFQLYWIPT EYINPGPAPVSC
8	GH61q	<u>MPPPRLSTLLPL</u> LALIAPTALG HSHLGYIIN GEVYQGFDPREPEQANSPLRVGWSTGAI DDGFVAPANYSSPDIIICHIEGASPPAHAPVRAGDRVHVQWNGWPLGHVGPVLSYLPACGGL EGSESGCAGVDKROLRWTKVDDSLPAMEL
9	GH61 r	<u>MRSTLAGALAA</u> IAAQKVAG HATFQQLWHG SSCVRLPASNSPVTNVGSRDFVCNAGTRPV SGKCPVKAGGTVTIEMHQQPGDRSCNNEAIGGAHWGPVQVYLTKVQDAATADGSTGWFKIF SDSWSKKPGGNLGGDDNWGTRDLNACCGKMD

(continua)

10	GH61s	MLLLTLATLVTLARHVS HARLFRVSVD GKDQGDGLNKYIRSPATNDPVRDLSSAAIV CNTQGSKAAPDFVRAAGDKLTFLWAHDNPDDPDYVLDPSHKGAILTYVAAYPSGDPTGP IWSKLAEEGFTGGQWATIKMIDNGGKVDVTLPEALAPGKYLRQELLALHRADFACDDPAH PNRGAESYPNCVQVEVSGSGDKKPDQNFDFNKGYTCDNKGLHFKEYIGQDSQYVAPGPRPW NGS
11	GH61t	MFTSLCITDH WRTLSSHSGPVMNYLAHCTNDCKSFKGDSGNVWVKIEQLAYNPSANPPW ASDLLREHGAKWKVTIPPSLVPGEYLLRHEILGLHVAGTVMGAQFYPGCTQIRVTEGGSTQ LPSGIALPGAYGPQDEGILVDLWRVNQGVNYTAPGGPVWSEAWDTEFGGSNTTECATMLD DLDDYMAANDEWIGWTA
12	GH61u	MKLSAAIAVLAAALAEG HYTFPSIANT ADWQYVRITTNFQSNGPVTDVNSDQIRCYERN PGTGAPGIYNTAGTTINYNKSSISHPGPMFYIAKVPAGQSAATWDGKAWWSKIHQEM PHFGTSLTWDSNGRTSMPVTIPRCLQDGEYLLRAEHIALHSAGSPGGAQFYISCAQLSVTG GSGTWNPRNKVSFPGAYKATDPGILINIYYPVPTSYPAGPPVDTC
13	GH61v	MYRTLGSIALLAGGAAAHG AVTSYNIAGK DYPGYSGFAPTQQDVIQWQWPDYNPVLSAS DPKLRCNGGTGAALYAEAAPGDTITATWAQWTHSQGPILVWMYKCPGDFSSCDGSGAGWFK IDEAGFHGDGTTVFLDTETPSGWDIAKLVGGNKSWSKIPDGLAPGNYLVRHEIALHQAN NPQFYPECAQIKVTGSGTAEPAAASYKAAIPGYCQQSDPNISFNINDHSLPQEYKIPGPPVF KGTASAKARAFQA
14	GH61w	MLTTTFALLTAALGVSA HYTLPRVGTG SDWQHVRADNWQNGFVGDVNSEQIRCFQAT PAGAQDVYTVQAGSTVYHANPSIYHPGPMQFYLARVPDQDVKSWTGEAVWFKVYEEQP QFGAQLTWPSNGKSSFEVPIPCIRAGNYLLRAEHIALHVAQSOGGAQFYISCAQLQVTGG GSTEPSQKVSFPGAYKSTDPGILININYPVPTSYPQNPAPVFR
15	GH61x	MKVLAPLILAGAASA HTIFSSLEV GVNQIGQGVRVPSYNGPIEDVTSNSIACNGPPN PTTPTNKVITVRAGETVTAVWRYMLSTTGSAENDIMDSHKGPTMAYLKKVDNATTDGSGV GGWFKIQEDGLTNGVWGTERVINGQGRHNIKIPECIAPGOYLLRAEMLALHGSANYPGAQF YMECAQLNIVGGTGSKTPTSVSFPAYKGTDPGVKINIYWPVTSYQIPGPGVFTC
16	GH61b	MKLSLFSVLATALTVEGHA IFQKVSVNGA DQGSALTGLRAPNNPNVQNVNSQDMICGQS GSTNTIIEVKAGDRIGAWYQHVIGGAQFPNDPDNPIAKSHKGPVMAYLAKVDNAATASKT GLKWFKIWEDTFNPSTKTWGVNDLINNNGWVYFNLQCIADGNYLRLVEVLALHSAYSQGG AQFYQSCAQINVSOGGSFTASTVSFPGAYSASDPGILINIYGATGQPDNNGQPYTAPGPA PISC
17	GH61c	MALQLLASLALLSVPALAHGGLA NYTVGDTWYR GYDPNLPPETQLNQTMQIRQWATID PVFTVSEPYLACNNPGAPPPSYIPIRAGDKITAVYWYWLHAIGPMSVWLARCGDTPAADCR DVDVNRVWGFKIWEGGLLEGNLAEGWYQKDFQRWDGSPSLWPVTIPKGLKSGTYIIRHE ILSLHVALKPQFYPECAHLNITGGGDLPEETLVRFPGVYKEDDPSIFIDVYSEENART DYTVPGGPIWEG
18	GH61d	MKALSLLAAAGAVSA HTIFVQLEAD GTRYPVSYGIRDPTYDGPITDVTSDNVACNGGPN PTTPSSDVITVTAAGTVKAIWRHTLQSGPDDVMDASHKGPTLAYIKKVGDKATKDSGVGGGW FKIQEDGYNNGWGTSTVISNGGEHYIDIPACIPEGQYLLRAEMIALHAAGSPGGAQLYME CAQINIVGSGSVPSSTVSFPGAYSPNDPGLLINIYSMPSSSYTIPGPPVFKC

19	GH61e	MKSSTPALFAAGLLAQHAAA HSIFQQASSG STDFTLCTRMPPNNSPVTSVTSGDMTCK VGGTKGVSGFCEVNAGDEFTVEMHAQPGDRSCANEIAGGNHFGPVLIIYMSKVDDASTADGS GDWFKVDEFGYDASTKTWGTDKLNCNGKRTFNIPSHIPAGDYLVRAEALHTANQPGGA QFYMSCYQVRISSGEGGQLPAGVKIPGAYSANDPGILVDIWGNDFNDPPGHSARHAIIS SSSNNSGAKMTKKIQEPTITSVTDLPTDEAKWIALQKISYVDQGTARTYEPASRKTRSPR V
20	GH61f	MKSFTLTTLAALAGNAAA HATFQALWVD GVDYGAQCARLPASNSPVTDTVTSNAIRCNAV PSPARGKCPVKAGSTVTVMHQQPGDRSCSSEIAGGAHYGPVMVYMSKVSDAASADGSSGW FKVFEDGWAKNPSGGSGDDDYWGTDKLNCCGKMNVKIPADLPAGDYLLRAEALHTAGS AGGAQFYMTCYQLTVTSGGSASPTVSFPGAYKATDPGILVNIHAPLSGYTVPGPAVYSGG STKKAGSACTGCESTCAVSGGPTATVSQSPGSTATAPGGGGGCTVQKYQQCGGQYTGCT NCASGSTCSAVSPPYYSQCV
21	GH61g	MKGLLGAAALSLAVSDVSA HYIFQQLTTG GVKHAVYQYIRKNTNYSNPVTDLTSDNLR NVGATGAGTDTVTRAGDSFTTTDTVPYHQGPTSIYMSKAPGSASDYDGGGWFKIKDWA DYTATIPECIPPGDYLLRIQQLGIHNPWPAGIPQFYISCAQITVTGGGSANPGPTVSIPGA FKETDPGYTVNIYNNFNHYTVPGPAVFTCNGSGGNNGGSPVTTTTTTTTTRPSTSTAQSQ PSSSPTSPSSCTVAKWGQCGGQYSGCTVCAAGSTCQKTNDYYSQCL
22	GH61h	MSSFTSKGLLSALMGAATVA AHGHVTNIVI NGVSYQNFDPFTHPYMQNPPTVVGWTSN TDNGFVGPEFSFSPDIICHKSATNAGGHAVVAAGDKVFIQWDTWPESHGHPVIDYLADCGD AGCEKVDKTTLKFFKISESGLLDGTNAPGKWASDTLIANNNSWLVIQIPNIAAGNYLVRHE IIALHSAGQQNGAQNYPQCFLQVTGSGTQKPSGVLGTELKATDAGILANIYTSPTYQI PGPATISGASAVQOTTSAITASASAITGSATAAPTATTTTAAATTTTATAGSGATATPST GGSPSSAQPAPTTAAATSSPARPTRCAGLKKRRRHARDVKVAL
23	GH61i	MKTALALVSAALVAHAH YVDHATIGGK DYQFYQPYQDPYMGDNKPDVRSRIPGNGPV EDVNSIDLQCHAGAEPKALHAPAAAGSTVTLVWTLWPDSDHVPVITYMARCPDTGCQDWSP GTKPWVFKIKEGGREGTSNTPLMTAPSAITYTIPSCLSKGYLLVRHEIALHSAWQYPGAQ FYPGCHQLQVTGGGSTVPSTNLVSFPGAYKGSDDPGITYDAYKAQPYTIPGPAVFTC
24	GH61j	MRYFLQAAAAFAVNSAAG HYIFQOFATG GSKYPPWKYIRRTNPDWLQNGPVTDLSS TDLRCNVGGQVSNGTETITLNAAGDEFSFILDTPVYHAGPTSLYMSKAPGAVADYDGGGAWF KIYDWGPGSGTWTLSGTYTQRIKCI PDGEYLLRIQIGLHNPGAAAPQFYISCAQVKVVDG GSTNPTPTAQIPGAFHSNDPGLTVNIYNDPLTNVYVPGPRVSHW
25	GH61k	MHPSLLFTLGLASVLVPLSSA HTTFTTLFVN DVNQDGTICIRMAKKGNVATHPLAGGLD SEDMACGRDQEPVAFTCPAPAGAKLTLEFRMWADASQSGSIDPSHLGVMAIYLLKKVSDMK SDAAAGPGWFKIWDQGYDLAACKWATEKLIDNGLLSVNLPTGLPTGYLLARQEITLQNV TNDRPEPQFYVGCAQLYVEGTSDSPIPSDKTVSIPGHISDPADPGLTFNVTGDASTYKPP GPEVYFPTTTTTSSSSGSDNKGARRQTPDDKQADGLVPADCLVKANWCAALPPYT DEAGCWAHAEDCNKQLDACYTSAPPSGSKGCKVWEQVCTVVSQKCEAGDFKGPPLGKEL GEGIDEPIPGGKLPPAVNAGENGNGHGGGGGDDGDDNDNEAGAGAASTPTFAAPGAATPQP NSEARRRREAHWRRLSAE
26	GH61p2	MKLTSSLAVLAAAGAAQ HYTFPRAGTG GSLSGEWVVRMTENHSHGVPVTDVTSPEMTC YQSGVQGAQPTVQVKAGSQFTFSVDPSIGHPGPLQFYMAKVPSPGQTAATFDGTGAVWFKIY QDGPNGLGTDSITWPSAGKTEVSVTIPSCIEDGEYLLRVEHIALHSASSVGAQFYIACAQ LSVTGGSGTLNTGSLVSLPGAYKATDPGILFQLYWPIPTEYINPGPAPVSC
27	GH61q2	MPPRLSTLLPLLALIAPTALG HSHLGYIIING EVYQGFDPREQANSPLRVGWSTGAI DDGFVAPANYSSPDIICHIEGASPPAHAPVRAGDRVHVQWKRLAARTRGAGAVVPGALRRA GGVREVDLSLPADELVGAAGGAGGEDDGGSGSDGSGSGSGSRVGVPGQRWATDVLIAANNS WQVEIPRGLRDGPYVLRHEIVALHYAAEPGGAQNYPLCVNLWVEGGDGSMELDHFDATQFY RPDPPGILLNVTAGLSYAVPGPTLAAGATPVYPYAQQNISSARADGTPVIVTRSTETVPFT AAPTPAETAEEKGRYDDQTRTKDLNERFFYSSRPEQKRLTATSRRELVDHRTYLSVAVC ADFGAHKAETNHEALRGNGKHGGVSE
28	GH61r2	MRSTLAGALAAIAAQKVAG HATFQQLWHG SSCVRLPASNSPVTNVGSRDFVCNAGTRPV SGKCPVKAGGTVTIEMHQQPGDRSCNNEIAGGAHWGPVQVYLTKVQDAATADGSTGWFKIF SDSWSKKPGGNSGDDDNWGTDLNACCCKMDVAIPADIASGDYLLRAEALHTAGQAGGA QFYMSCYQMTVEGGSGTANPPTVKFPGAYSANDPGILVNIHAPLSSTAPGPAVYAGGTIR EAGSACTGCAQTCKVGSSPSAVAPGSGAGNGGGFQPR
29	GH61t2	MNYLAHCTND DCKSFKGDSGNVWKIEQLAYNPSANPPWASDLLREHGAKWKVTIPPSLV PGEYLLRHEILGLHVAGTVMGAQFYPGCTQIRVTEGGSTQLPSGIALPGAYGPPQDEGILVD LWRVNGQGVNYTAPGGPVWSEAWDTEFGGSNTTECATMLDLDLDYMAANDPCCTDQNGFG SLEPGSKAAGGSPSLYDVTVLVPVLQKKVPTKLQWSPASVNGDELTERP
30	GH61e2	MKSSTPALFAAGLLAQHAAA HSIFQQASSG STDFTLCTRMPPNNSPVTSVTSGDMTCN VGGTKGVSGFCEVNAGDEFTVEMHAQPGDRSCANEIAGGNHFGPVLIIYMSKVDDASTADGS GDWFKVDEFGYDASTKTWGTDKLNCNGKRTFNIPSHIPAGDYLVRAEALHTANQPGGA QFYMSCYQVRISSGEGGQLPAGVKIPGAYSANDPGILVDIWGNDFNEYVYIPGPPVIDSSYF

IV. Ácidos nucleicos recombinantes y proteínas

60 En un aspecto, la divulgación proporciona ácidos nucleicos recombinantes que comprenden secuencias de proteína expuestas en la TABLA 1 o TABLA 2 y variantes (por ejemplo, variantes biológicamente activas) de las mismas. La divulgación también proporciona vectores de expresión que contienen los ácidos nucleicos recombinantes, células que comprenden un ácido nucleico recombinante o vector, proteínas recombinantes producidas por tales células y métodos para usar las células y proteínas.

En un aspecto, la divulgación proporciona una secuencia de ácido nucleico recombinantes que codifica una pre-proteína que comprende SEQ ID NO: 1-30, la proteína madura (secretada) codificada en SEQ ID NO: 1-30 o un fragmento truncado amino terminal de SEQ ID NO: 1-30.

5 En un aspecto, la divulgación proporciona una proteína GH61 recombinante, aislada o purificada que tiene una secuencia que comprende SEQ ID NO: 1-30, la proteína madura (secretada) codificada en SEQ ID NO: 1-30 o un fragmento truncado amino terminal de SEQ ID NO: 1-30.

10 En un aspecto, la divulgación proporciona una secuencia de ácido nucleico recombinantes que codifica una proteína GH61 con al menos aproximadamente 70%, al menos aproximadamente 75%, aproximadamente 80%, aproximadamente 85%, aproximadamente 90%, aproximadamente 91%, aproximadamente 92%, aproximadamente 93%, aproximadamente 94%, aproximadamente 95%, aproximadamente 96%, aproximadamente 97%, aproximadamente 98% o aproximadamente 99% de identidad de secuencia con una secuencia que comprende SEQ ID NO: 1-30, la proteína madura (secretada) codificada en SEQ ID NO: 1-30 o un fragmento truncado amino terminal de SEQ ID NO: 1-30. En un aspecto relacionado, la divulgación proporciona una proteína GH61 recombinante, aislada o purificada que tiene una secuencia como se ha expuesto anteriormente. En algunos casos, una sustitución conservadora de aminoácido puede ser preferente sobre otros tipos de sustituciones.

20 Las proteínas GH61 pueden ser proteínas endógenas aisladas de células fúngicas (por ejemplo, *M. thermophila*) o pueden producirse recombinantemente.

25 Como se analiza con más detalle más abajo, las proteínas GH61 pueden comprender un péptido de señal endógeno o heterólogo. Como se analiza con más detalle más abajo, un ácido nucleico que codifica una proteína GH61 puede unirse operativamente a un promotor, como un promotor heterólogo.

En una realización la divulgación proporciona una secuencia de ácido nucleico recombinante que comprende una secuencia seleccionada de SEQ ID NOS: 31-59.

30 La secuencia de ácido nucleico usada para expresar una proteína GH61 puede ser una secuencia nativa obtenida de *M. thermophila* que codifica la respectiva proteína, o una parte de la misma que codifica la proteína madura. GH61 ejemplares que codifican secuencias de *M. thermophila* se muestran en SEQ ID NOS: 31 a 59. Alternativamente, las secuencias de ácido nucleico que codifican una proteína específica pueden designarse por referencia al código genético. En algunas realizaciones, se usa una secuencia que puede estar optimizada con codón para una célula huésped diferente a *M. thermophila* (por ejemplo, células de levadura). Véase GCG CodonPreference, Genetics Computer Group Wisconsin Package; Codon W, John Peder, Universidad de Nottingham; McInerney, J. O., 1998, Bioinformatics 14: 372-73, en cuyo caso la secuencia de ácido nucleico es diferente a la secuencia que ocurre de manera natural.

40 En algunas realizaciones, las proteínas GH61 de la divulgación son biológicamente activas, esto es, tienen actividad GH61. La actividad GH61 puede medirse usando métodos conocidos en la técnica, incluyendo métodos descritos aquí más abajo. Los polipéptidos que carecen de actividad GH61 también pueden encontrar variedad de usos, incluyendo uso para la generación de anticuerpos para purificación de proteínas GH61. Excepto que se aclare del contexto, la referencia aquí a una proteína GH61 (incluyendo variantes) se refiere a proteínas con actividad GH61.

45 Es preferente que las proteínas GH61 que sean variantes de SEQ ID NOS: 1-30 tengan al menos 10% de la actividad de la misma cantidad molar de la proteína de tipo salvaje de la que se derivan. Pueden tener al menos 20%, al menos 30%, al menos 40%, al menos 50%, al menos 60%, al menos 70%, al menos 80%, al menos 90%, al menos 95% o al menos 99% de la actividad de la proteína de tipo salvaje.

50 Las proteínas GH61 que son de variantes de SEQ ID NOS: 1-30 pueden ser más cortas que la proteína de tipo salvaje en aproximadamente 2%, aproximadamente 3%, aproximadamente 4%, aproximadamente 5%, aproximadamente 6%, aproximadamente 7%, aproximadamente 8%, aproximadamente 9%, aproximadamente 10%, aproximadamente 15%, aproximadamente 20%, aproximadamente 25%, aproximadamente 30%, aproximadamente 35%, aproximadamente 40%, aproximadamente 45%, aproximadamente 50%, aproximadamente 60%, aproximadamente 70% o aproximadamente 80% en comparación con la secuencia referencia (tipo salvaje) y/o parte de una proteína de fusión en la que una parte de proteína GH61 se une a una o más secuencias. Estas variantes pueden tener al menos aproximadamente 70%, al menos aproximadamente 75%, aproximadamente 80%, aproximadamente 85%, aproximadamente 90%, aproximadamente 91%, aproximadamente 92%, aproximadamente 93%, aproximadamente 94%, aproximadamente 95%, aproximadamente 96%, aproximadamente 97%, aproximadamente 98% o aproximadamente 99% de identidad de secuencia con una secuencia que comprende SEQ ID NO: 1-30, la proteína madura (secretada) codificada en SEQ ID NO: 1-30 o un fragmento truncado amino terminal de SEQ ID NO: 1-30. Las proteínas GH61 pueden tener una eliminación interna o una eliminación en la terminal amino o carboxi en relación con una secuencia de tipo salvaje.

65

Las proteínas GH61 pueden comprender una eliminación en terminal amino y/o terminal carboxi y/o eliminación interna, pero donde el resto de secuencia de aminoácido es al menos aproximadamente 60%, 70%, 75%, 80%, 85%, 90%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% o 100% idéntica a las posiciones correspondientes en la secuencia con la que se está comparando (por ejemplo, una variante GH61 de longitud completa de la divulgación). En algunas realizaciones, el dominio de enlace de quitina, o dominio GH61 se retira. En algunas realizaciones, la proteína conservar sustancialmente todas la actividad del polipéptido de longitud completa.

Una proteína GH61 madura puede tener al menos 70%, 80%, 90% o 95% identidad de secuencia con una secuencia de tipo salvaje, y tiene sustancialmente longitud completa (al menos 90% de la longitud de la secuencia de tipo salvaje).

La divulgación también proporciona una secuencia de ácido nucleico GH61 recombinante y proteína expresada de la misma, donde la proteína tiene la secuencia de GH61f, GH61a, GH61v, GH61p, GH61g y GH61i (SEQ ID Nos: 2, 7, 13, 20, 21, 23, 26), o un fragmento secretada de la misma; así como variantes al menos aproximadamente 70%, al menos aproximadamente 75%, aproximadamente 80%, aproximadamente 85%, aproximadamente 90%, aproximadamente 91%, aproximadamente 92%, aproximadamente 93%, aproximadamente 94%, aproximadamente 95%, aproximadamente 96%, aproximadamente 97%, aproximadamente 98% o aproximadamente 99% idénticas a cualquiera de GH61f, GH61v, GH61p, GH61g y GH61i (SEQ ID Nos: 2, 13, 20, 21, 23, 26); o una cualquiera de GH61a, GH61o, GH61v, GH61x, GH61b y GH61e (SEQ ID Nos: 2, 6, 13, 15, 16, 19, 30), fragmentos y variantes de los mismas, y ácidos nucleicos que codifican tales proteínas. La divulgación también incluye vectores de expresión que comprenden una cualquiera de las secuencias de ácido nucleico anteriormente mencionadas, células que comprenden tales vectores de expresión y proteínas GH61 aisladas producidas por las células. Preferentemente, la proteína variante GH61 tiene actividad potenciadora de celulasa. La proteína que codifica las secuencias puede incluir un péptido de señal heterólogo y/o puede estar operativamente unido a un promotor heterólogo.

Sin pretender estar ligado a un particular mecanismo de acción, en presencia de GH61, la hidrólisis de polímeros de azúcar (por ejemplo, sustratos de celulosa) por las enzimas produce más producto durante un periodo de tiempo particular, procede más rápidamente o se completa más rápido cuando la proteína GH61 está presente, en comparación con una reacción similar bajo las mismas condicione donde la proteína GH61 está ausente.

Las proteínas GH61 de la divulgación que tienen actividad potenciadora de celulosa pueden identificarse usando métodos estándares para mapear la función en un polipéptido, como se conoce en la técnica. Por ejemplo, puede expresarse una variante truncada y después probarse en un ensayo de actividad de GH61. Pueden introducirse truncamientos adicionales hasta que se pierda la actividad, en cuyo punto podría identificarse la unidad funcional mínima de la proteína. Los fragmentos que contienen cualquier parte de la proteína por debajo del tamaño identificado serán típicamente funcionales, como lo serían las construcciones de fusión que contengan al menos el núcleo funcional de la proteína.

Para generar variantes biológicamente activas que incorporan uno o más cambios de aminoácido en una secuencia que codifica GH61 (cualquiera de SEQ ID Nos: 1 a 30), pueden introducirse sustituciones en la secuencia de proteína y la proteína expresada para la retención de actividad.

Puede usarse una estrategia de mutación aleatoria o semi-aleatoria para generar una gran colección de variantes activas. Los textos estándares PROTOCOLOS EN BIOLOGÍA MOLECULAR (Ausubel et al., eds.) y CLONACIÓN MOLECULAR: UN MANUAL DE LABORATORIO (Sambrook et al., eds.) describen técnicas que emplean mutagénesis química, mutagénesis de casete, degeneración de oligonucleótidos, preparación mutua de oligonucleótidos, mutagénesis con escaneo de enlazador, mutagénesis con escaneo de alanina y PCR propenso a error. Otros métodos eficientes incluyen las cepas mutantes de *E. coli* de Stratagene (Greener et al., Methods Mol. Biol. 57: 375, 1996) y la técnica de mezcla de ADN de Maxygen (Patten et al., Curr. Opin. Biotechnol. 8: 724, 1997; Harayama, Trends Biotechnol. 16:76, 1998; Patentes de Estados Unidos 5.605.793 y 6.132.970). Para aumentar la variación, puede usarse una tecnología que genera más cambios abruptos, como técnica de mezcla de ADN.

Pueden usarse kits disponibles comercialmente para obtener variantes, incluyendo el Sistema de Mutagénesis Dirigido al Sitio GeneTailor™ vendido por InVitrogen™ Life Technologies; el Kit de Mutagénesis Aleatorio BD Diversify™, vendido por BD Biosciences/Clontech; Template Generation System™, vendido por MJ Research Inc., la cepa mutante de *E. coli* XL1-Red™, vendido por Stratagene; y el kit de Mutagénesis Aleatoria GeneMorph®, también vendido por Stratagene. Al emplear cualquiera de estos sistemas junto con un ensayo adecuado de actividad GH61, pueden generarse y probarse variantes de una manera con un alto rendimiento.

Alternativamente o además de, el usuario puede emplear una estrategia de evolución dirigida. Véase, por ejemplo, patente de Estados Unidos 7.981.614; Métodos para generar polinucleótidos que tienen las características deseadas; US 2011/0034342: Método para generar una población optimizada y diversa de variantes; Patente de Estados Unidos 7.795.030: Métodos y composiciones para ingeniería celular y metabólica; Patente de Estados Unidos 7.647.184: Alto rendimiento evolutivo dirigido por mutagénesis racional; Patente de Estados Unidos 6.939.689: Re-montaje de ácido nucleico mediado por exonucleasa en evolución dirigida; y Patente de Estados

Unidos 6.773.900: Selección final en evolución dirigida. La mutagénesis puede realizarse de acuerdo con cualquiera de las técnicas conocidas en la técnica, incluyendo mutagénesis aleatoria y específica de sitio. La evolución dirigida puede realizarse con cualquiera de las técnicas conocidas en la técnica para investigar la producción de variantes incluyendo la mezcla.

Los métodos de mutagénesis y evolución dirigida son bien conocidos en la técnica. Véase las patentes de Estados Unidos Números 5.605.793, 5.830.721, 6.132.970, 6.420.175, 6.277.638, 6.365.408, 6.602.986, 7.288.375, 6.287.861, 6.297.053, 6.576.467, 6.444.468, 5.811.238, 6.117.679, 6.165.793, 6.180.406, 6.291.242, 6.995.017, 6.395.547, 6.506.602, 6.519.065, 6.506.603, 6.413.774, 6.573.098, 6.323.030, 6.344.356, 6.372.497, 7.868.138, 5.834.252, 5.928.905, 6.489.146, 6.096.548, 6.387.702, 6.391.552, 6.358.742, 6.482.647, 6.335.160, 6.653.072, 6.355.484, 6.03.344, 6.319.713, 6.613.514, 6.455.253, 6.579.678, 6.586.182, 6.406.855, 6.946.296, 7.534.564, 7.776.598, 5.837.458, 6.391.640, 6.309.883, 7.105.297, 7.795.030, 6.326.204, 6.251.674, 6.716.631, 6.528.311, 6.287.862, 6.335.198, 6.352.859, 6.379.964, 7.148.054, 7.629.170, 7.620.500, 6.365.377, 6.358.740, 6.406.910, 6.413.745, 6.436.675, 6.961.664, 7.430.477, 7.873.499, 7.702.464, 7.783.428, 7.747.391, 7.747.393, 7.751.986, 6.376.246, 6.426.224, 6.423.542, 6.479.652, 6.319.714, 6.521.453, 6.368.861, 7.421.347, 7.058.515, 7.024.312, 7.620.502, 7.853.410, 7.957.912, 7.904.249, y todos los homólogos relacionados que no son de Estados Unidos; Ling et al., Anal. Biochem., 254(2):157-78 [1997]; Dale et al., Meth. Mol. Biol., 57:369-74 [1996]; Smith, Ann. Rev. Genet., 19:423-462 [1985]; Botstein et al., Science, 229:1193-1201 [1985]; Carter, Biochem. J., 237:1-7 [1986]; Kramer et al., Cell, 38:879-887 [1984]; Wells et al., Gene, 34:315-323 [1985]; Minshull et al., Curr. Op. Chem. Biol., 3:284-290 [1999]; Christians et al., Nat. Biotechnol., 17:259-264 [1999]; Crameri et al., Nature, 391:288-291 [1998]; Crameri et al., Nat. Biotechnol., 15:436-438 [1997]; Zhang et al., Proc. Nat. Acad. Sci. U.S.A., 94:4504-4509 [1997]; Crameri et al., Nat. Biotechnol., 14:315-319 [1996]; Stemmer, Nature, 370:389-391 [1994]; Stemmer, Proc. Nat. Acad. Sci. USA, 91:10747-10751 [1994]; WO 95/22625; WO 97/0078; WO 97/35966; WO 98/27230; WO 00/42651; WO 01/75767; y WO 2009/152336.

En algunas realizaciones, una proteína GH61 de la divulgación tiene una secuencia de aminoácido que está codificada por un ácido nucleico que se hibrida bajo condiciones severas (esto es, severidad media alta, alta o muy alta) con el complemento de SEQ ID NO: 31-59 y comprende actividad GH61.

V. Ensayos de Actividad GH61

La actividad potenciadora de celulasa de proteínas GH61 de la divulgación puede determinarse usando cualquier ensayo adecuado de actividad GH61. Por ejemplo, se obtiene una proteína GH61 purificada o recombinante de esta divulgación y después se analiza para actividad GH61 combinándolo con enzimas de celulasa en una reacción de sacarificación, y se determina si hay un aumento en producción de glucosa, en comparación con la misma reacción de sacarificación realizada sin el GH61.

En una técnica, la actividad GH61 puede analizarse combinando sustrato celulósico con enzimas de celulasa (por ejemplo, 5-10 mg de peso total de enzimas de celulasa por gramo de sustrato) en presencia y ausencia de proteína GH61. En algunas realizaciones las enzimas de celulasa son un conjunto definido de enzimas recombinantes de celulasa de *M. thermophila*.

En otra técnica, se usa el caldo de un cultivo de *M. thermophila* de tipo salvaje (con o sin la complementación de la proteína GH61). La actividad GH61 se demuestra por una mayor producción de glucosa en presencia de GH61 exógeno (eso es, más allá de cualquier mejora resultante de GH61 endógeno en el caldo).

También es posible usar un caldo complementado con una o más enzimas purificadas.

Las enzimas adecuadas incluyen enzimas recombinantes aisladas clonadas de *M. thermophila*, como endoglucanasa (EG), β -glucosidasa (BGL); celobiodrilasa de Tipo 1 (CBH1) y/o celobiodrolasa de Tipo 2 (CBH2) en cualquier combinación adecuada para el sustrato elegido para producir un producto medible. Las enzimas de celulasa ejemplares que pueden usarse para analizar la actividad GH61 pueden tener secuencias de aminoácido seleccionadas de cualquiera de SEQ ID Nos: 61 a 68.

En un ensayo ejemplar para medir actividad GH61 de *M. thermophila* derivada de proteínas GH61 y proteínas variantes, las enzimas de celulasa usadas son *M. thermophila* BGL1 (SEQ ID NO:66; Badhan et al., Bioresour Technol. 2007 Feb; 98(3):504-10); *M. thermophila* CBH1 (SEQ ID NO:67; Park JI et al., Badhan et al., Bioresour Technol. 2007 Feb;98(3):504-10); and *M. thermophila* CBH2 (SEQ ID NO:68). En algunas realizaciones, también se usa endoglucanasa: *M. thermophila* EG2 (SEQ ID NO: 65; Rosgaard L. et al., Prog. 2006; 22(2):493-8; Badhan et al., supra).

Alternativamente, pueden usarse preparaciones comercialmente disponibles que comprenden una mezcla de enzimas de celulasa, como Laminex™ y Spezyme™ de Genencor International, Rohament™ de Rohm GmbH, y Celluzyme™, Cereflo™ y Ultraflo™ de Novozymes Inc.

Los ensayos con enzimas de celulosa se hacen típicamente a 50 °C, pero también pueden hacerse a 35, 45, 55, 60 o 65 °C. La proteína GH61 y las enzimas se combinan con el sustrato y se incuban para producir azúcares fermentables. Los azúcares después se recuperan y se cuantifican para producir glucosa. Un sustrato adecuado es paja de trigo (por ejemplo, paja de trigo pre-tratada). Otros sustratos celulósicos enumerados en esta divulgación pueden usarse como una alternativa, incluyendo rastrojo de maíz pre-tratado con ácido sulfúrico (véase la patente de Estados Unidos 7.868.227) y otros sustratos descritos en la Sección XII más abajo.

También puede usarse un método de ensayo que proporciona Harris et al., 2010, Biochemistry 49:3305-3316. En este método, el rastrojo de maíz se pre-trata con ácido sulfúrico, se lava, incuba con enzimas de celulosa y GH61 durante varios días, y después la producción de azúcares se cuantifica mediante refracción.

Otro método de ensayo se proporciona en la patente de Estados Unidos N° 7.868.227. En este método, el sustrato celulósico es PCS (rastrojo de maíz pre-tratado con calor y ácido sulfúrico diluido; WO 2005/074647); una mezcla de enzima de celulosa es Cellulcast®, una mezcla de enzimas de celulosa del hongo *Trichoderma reesei*, disponible en Sigma-Aldrich. La hidrólisis de PCS se realiza en un volumen total de reacción de 1,0 mL y una concentración de PCS de 50 mg/mL en 1 mM de sulfato de manganeso, 50 mM de tampón de acetato de sodio pH 5.0. La proteína de la prueba se combina con la mezcla de celulosa base en concentraciones relativa entre 0 y 100% de proteína total. La composición de proteína se incuba con PCS a 65 °C durante 7 días. La producción combinada de glucosa y celobiosa puede medirse mediante detección de índice de refracción.

La actividad GH61 se calcula como un incremento en la producción de glucosa a partir del sustrato por la(s) celulosa(s) en presencia de proteína GH61, en comparación con la misma mezcla de reacción en ausencia de proteína GH61. Típicamente, el incremento es dependiente de dosis en al menos un rango triple de concentraciones. La actividad GH61 puede expresarse como un grado de "sinergia" como se analiza en el Ejemplo 8.

La adición o presencia de proteína GH61 recombinante o aislada puede aumentar la producción del producto a partir de enzimas de celulosa, por ejemplo, al menos 1%, al menos 5%, al menos 10%, al menos 20%, 30%, 50%, 70%, doblar, triplicar o más.

VI. Expresión de proteínas GH61

El cultivo celular, genética recombinante, ingeniería de proteínas y técnicas de fermentación que pueden emplearse en la expresión, producción y uso de las proteínas GH61, composiciones y otros productos de esta divulgación son conocidos en la técnica. Por conveniencia, ciertos aspectos se describen brevemente más abajo. Aunque se describen principalmente en el contexto de expresión de proteínas GH61, se apreciará que los mismos métodos, células, etc., pueden usarse para expresar proteínas de celulosa y otras proteínas como, aunque sin limitar a, las descritas aquí en otro punto.

En algunas realizaciones de la divulgación, la construcción comprende además secuencias reguladoras, por ejemplo, un promotor, operativamente unido a la proteína que codifica la secuencia. Aquellos expertos en la técnica conocen grandes números de vectores y promotores adecuados.

Péptido de señal

En algunas realizaciones de la divulgación, una proteína GH61 puede incluir un péptido de señal, para que cuando se exprese en una célula huésped, la forma madura (por ejemplo, SEQ ID NO: 2) se secrete en un caldo de cultivo celular. La proteína GH61 (o variante) puede incluir su péptido de señal nativo correspondiente como se muestra en la TABLAS 1 y 2. Alternativamente, una secuencia de ácido nucleico recombinantes que codifica una proteína que comprende SEQ ID NO: 2, o la parte secretada de una cualquiera de SEQ ID Nos: 3 a 30, una parte truncada de amino terminal de una cualquiera de SEQ ID Nos: 2 a 30, o una variante de los mismos que tenga un péptido heterólogo de señal fusionado con terminal N.

Pueden usarse varios péptidos de señal, dependiendo de la célula huésped y otros factores. Los péptidos de señal útiles para células huéspedes fúngicas filamentosas incluyen los péptidos de señal obtenidos de TAKA amilasa de *Aspergillus oryzae*, amilasa neutral 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* (TrCBH2).

Los péptidos de señal útiles para células huéspedes bacterianas son los péptidos de señal obtenidos a partir de los genes para amilasa maltógena de *Bacillus* NCIB 11837, α -amilasa de *Bacillus stearothermophilus*, subtilisina de *Bacillus licheniformis*, β -lactamasa de *Bacillus licheniformis*, proteasas neutrales de *Bacillus stearothermophilus* (nprT, nprS, nprM), y prsA *Bacillus subtilis*. Más péptidos de señal se describen en Simonen and Palva, 1993, Microbiol Rev 57:109-137.

Los péptidos de señal útiles para células huéspedes de levadura también incluyen aquellos de los genes de alfa-factor de *Saccharomyces cerevisiae*, SUC2 invertasa de *Saccharomyces cerevisiae* (véase Taussig y Carlson,

1983, Nucleic Acids Res 11:1943-54; SwissPro, Nº de acceso. P00724), y otros. Romanos et al., 1992, Yeast 8:423-488. Las variantes de estos péptidos de señal y otros péptidos de señal son adecuados.

La divulgación también proporciona proteínas recombinantes que comprenden un péptido de señal mostrado en la Tabla 1 o Tabla 2 fusionado con terminal amino de una proteína heteróloga (esto es, una proteína con la que no se asocia por naturaleza, que puede ser una proteína diferente a una proteína GH61). Así, los péptidos de señal mostrados en la Tabla 1 y 2 pueden usarse para provocar secreción de una proteína heteróloga recombinantemente expresada en una célula huésped. En algunas realizaciones la célula huésped es *Myceliophthora thermophila*.

Promotores

Con el fin de obtener altos niveles de expresión en una célula particular a menudo es útil expresar la variante GH61 de la presente divulgación bajo el control de un promotor heterólogo. Una secuencia de promotor puede unirse operativamente a la región 5' de la secuencia que codifica GH61 usando métodos rutinarios.

Los ejemplos de promotores útiles para expresión de GH61s incluyen promotores de hongos. En algunas realizaciones, puede usarse una secuencia promotora que conduce la expresión de un gen diferente a un gen GH61 en una cepa fúngica. Como un ejemplo no limitativo, puede usarse un promotor fúngico de un gen que codifica una endoglucanasa. En algunas realizaciones, puede usarse una secuencia promotora que conduce la expresión de un gen GH61 en una cepa fúngica diferente a la cepa fúngica de la que la variante GH61 se derivó. Como un ejemplo no limitativo, si la variante GH61 se deriva de C1, puede usarse un promotor de *T. reesei* gen GH61 o un promotor como el descrito en WO 2010107303, como, aunque sin limitar a las secuencias identificadas como SEQ ID NO: 25, SEQ ID NO: 26, SEQ ID NO: 27, SEQ ID NO: 28 o SEQ ID NO: 29 en WO 2010107303.

Los ejemplos de otros promotores adecuados útiles para dirigir la transcripción de las construcciones de nucleótido de la presente divulgación en una célula huésped fúngica filamentosa son promotores obtenidos de genes para TAKA amilasa de *Aspergillus oryzae*, proteinasa aspártica de *Rhizomucor miehei*, alfa-amilasa neutral de *Aspergillus niger*, ácido estable alfa-amilasa 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 neutral de *Aspergillus niger* y triosa fosfato isomerasa de *Aspergillus oryzae*) promotes como *cbh1*, *cbh2*, *egl1*, *egl2*, *pepA*, *hfb1*, *hfb2*, *xyn1*, *amy*, y *glaA* (Nunberg et al., 1984, Mol. Cell Biol., 4:2306-2315, Boel et al., 1984, EMBO J. 3:1581-85 and EPA 137280), y promotores mutantes, truncados e híbridos de los mismos. En un huésped de levadura, los promotores útiles pueden ser de los genes de enolasa de *Saccharomyces cerevisiae* (*eno-1*), galactokinasa de *Saccharomyces cerevisiae* (*gal1*), alcohol dehidrogenasa/gliceraldehído-3-fosfato dehidrogenasa de *Saccharomyces cerevisiae* (*ADH2/GAP*), y 3-fosfoglicerato quinasa de *S. cerevisiae*. Otros promotores útiles para células huéspedes de levadura se describen en Romanos et al., 1992, Yeast 8:423-488. Los promotores asociados a producción de quitinasa en hongos también pueden usarse. Véase, por ejemplo, Blaiseau y Lafay, 1992, Gene 120243-248 (hongo filamentoso *Aphanocladium album*); Limon et al., 1995, Curr. Genet., 28:478-83 (*Trichoderma harzianum*).

Los promotores conocidos para expresión de control de genes en células procarióticas o eucarióticas o los virus y que pueden usarse en algunas realizaciones de la invención incluyen promotor SV40. Promotor *E. coli* lac o trp, promotor fago lambda P_L, promotor tac, promotor T7 y similares. En células huéspedes bacterianas, los promotores adecuados incluyen los promotores obtenidos de *E. coli* lac operon, gen agarasa de *Streptomyces coelicolor* (*dagA*), gen levansucrasa de *Bacillus subtilis* (*sacB*), gen α -amilasa de *Bacillus licheniformis* (*amyl*), gen amilasa maltogénico de *Bacillus stearothermophilus* (*amyM*), gen α -amilasa de *Bacillus amyloliquefaciens* (*amyQ*), genes de *Bacillus subtilis* *xylA* y *xyB* y gen procariótico β -lactamasa.

Pueden usarse cualquier otra secuencia promotora que conduzca la expresión en una célula huésped adecuada. Las secuencias promotoras adecuadas pueden identificarse usando métodos bien conocidos. En una técnica, una secuencia promotora putativa se une a una secuencia 5' que codifica una proteína reportera, la construcción se transfecta a la célula huésped (por ejemplo, C1) y el nivel de expresión de la reportera se mide. La expresión de la reportera puede determinarse midiendo, por ejemplo, niveles de mRNA de la secuencia reportera, una actividad enzimática de la proteína reportera o la cantidad de proteína reportera producida. Por ejemplo, la actividad promotora puede determinarse usando la proteína verde fluorescente como secuencia codificadora (Henriksen et al., 1999, *Microbiology* 145: 729-34) o un gen reportero lacZ (Punt et al., 1997, *Gene*, 197: 189-93). Los promotores funcionales pueden derivarse de secuencias promotoras que ocurren de manera natural mediante métodos de evolución dirigidos. Véase, por ejemplo, Wright et al., 1005, *Human Gene Therapy*, 16: 881-892.

Los promotores adicionales incluyen aquellos de *M. thermophila*, proporcionados en la solicitud de patente de Estados Unidos Nº 13/214.406 presentada el 22 de agosto de 2010, así como en WO 2010/107303.

Vectores

La presente invención hace uso de construcciones recombinantes que comprenden una secuencia que codifica GH61 como se ha descrito anteriormente. Las construcciones de ácido nucleico de la presente divulgación comprenden un vector, como un plásmido, un cósmido, un fago, un virus, un cromosoma artificial bacteriano (CAB), un cromosoma artificial de levadura (CAL) y similares, donde se ha insertado una secuencia de ácido nucleico de la divulgación. Los polinucleótidos de la presente divulgación pueden incorporarse en una cualquiera de una variedad de vectores de expresión adecuados para expresar un polipéptido. Los vectores adecuados incluyen secuencias de ADN de cromosomas, de no cromosomas y sintéticas, por ejemplo, derivados de SV40; plásmidos bacterianos, ADN fago; baculovirus; plásmidos de levadura; vectores derivados de combinaciones de plásmidos y ADN fago, ADN viral como virus vacuna, adenovirus, virus de viruela aviar, pseudorrabia, adenovirus, virus asociado con adeno, retrovirus y similares. Puede usarse cualquier vector que convierte material genético en una célula y, si se desea réplica, que sea replicable y viable en el huésped relevante.

La divulgación proporciona vectores de expresión para provocar que una proteína GH61 se produzca a partir de una célula huésped adecuada, que puede ser un hongo (por ejemplo, *M. thermophila* o levadura). Puede seleccionarse cualquier vecotr de, aunque sin limitarse a, derivados de vectores virales; plásmidos bacterianos, ADN fago; baculovirus, plásmidos de levadura; vectores derivados de combinaciones de plásmidos y ADN fago, y vectores lanzaderas recombinantes. El vector puede introducirse en una célula huésped con un polinucleótido que codifica GH61 para que esté operativamente unido a un promotor que está activo en la célula huésped. El vector se selecciona para expresar la proteína codificada, y puede replicar como un episoma en la célula huésped planeada, o integrarse en el genoma de la célula huésped.

En un aspecto particular la presente divulgación proporciona un vector de expresión que comprende un polinucleótido GH61 operativamente unido a un promotor heterólogo. Los vectores de expresión de la presente divulgación pueden usarse para transformar una célula huésped apropiada para permitir que el huésped exprese la proteína GH61. Los métodos para expresión recombinantes de proteínas en hongos y otros organismos son bien conocidos en la técnica, y un número de vectores de expresión están disponibles o pueden construirse usando métodos rutinarios. Véase, por ejemplo, Tkacz y Lange, 2004, ADVANCES IN FUNGAL BIOTECHNOLOGY FOR INDUSTRY, AGRICULTURE, AND MEDICINE, KLUWER ACADEMIC/PLENUM UBLISHERS. Nueva York; Zhu et al., 2009, Construcción de dos vectores de entrada para expression de gen en plásmido de hongos. 6:128-33; Kavanagh, K. 2005, FUNGI: BIOLOGY AND APPLICATIONS Wiley.

Células huéspedes

Las proteínas GH61 de la divulgación pueden expresarse en una célula huésped que comprende un ácido nucleico recombinantes que codifica la proteína GH61. La célula huésped puede también expresar otras proteínas de interés, particularmente una o más enzimas de celulasa que funcionan en concierto con la proteína GH61 en el proceso de sacarificación. En una realización la célula huésped es una célula construida con celulasa. Así, las enzimas de celulasa pueden expresarse endógenamente por la célula huésped, o pueden expresarse a partir de otros ácidos nucleicos.

En otra técnica, dos o más poblaciones de células huéspedes, cada una expresando una proteína diferente o un conjunto de proteínas (por ejemplo, una proteína GH61 y una celulasa) pueden cultivarse juntas. Las dos células huéspedes pueden ser iguales o pueden ser de diferente especie celular. Las células que expresan proteína GH61 y las células que expresan enzimas de celulasa pueden combinarse y cultivarse juntas para producir composiciones de esta invención que contienen tanto proteínas GH61 como enzimas de celulasa. Alternativamente, el caldo de cultivo de cada población celular puede recogerse por separado, fraccionarse opcionalmente para enriquecer las actividades respectivas y después mezclarse para producir la combinación deseada.

Las células huéspedes fúngicas incluyen, aunque no se limitan a *Ascomycota*, *Basidiomycota*, *Deuteromycota*, *Zygomycota*, *Fungi imperfecti*. En algunas realizaciones, las células huéspedes fúngicas preferentes son células de levadura, y células fúngicas filamentosas, incluyendo todas las formas filamentosas de la subdivisión *Eumycotina* y *Oomycota*. Hawksworth et al., en Ainsworth y Bisby's DICTIONARY OF THE FUNGI, 8ª edición, 1995, CAB International, University Press, Cambridge, UK. Los hongos filamentosos se caracterizan por un micelio vegetativo con una pared celular compuesta por quitina, celulosa y otros polisacáridos complejos, y son morfológicamente distintos de la levadura. *Trichoderma* también puede ser una Fuente de una o más celulasas para su uso con combinación con proteínas GH61.

La célula huésped puede ser de una especie de *Achlya*, *Acremonium*, *Aspergillus*, *Aureobasidium*, *Azospirillum*, *Bjerkandera*, *Cellulomonas*, *Cephalosporium*, *Ceriporiopsis*, *Chrysosporium*, *Clostridium*, *Coccidioides*, *Cochliobolus*, *Coprinus*, *Coriolus*, *Corynascus*, *Cryphonectria*, *Cryptococcus*, *Dictyostelium*, *Diplodia*, *Elizabethkingia*, *Endothia*, *Erwinia*, *Escherichia*, *Fusarium*, *Gibberella*, *Gliocladium*, *Gluconacetobacter*, *Humicola*, *Hypocrea*, *Kuraishia*, *Mucor*, *Myceliophthora*, *Neurospora*, *Nicotiana*, *Paenibacillus*, *Penicillium*, *Periconia*, *Phaeosphaeria*, *Phlebia*, *Piromyces*, *Podospora*, *Prevotella*, *Pyrularia*, *Rhizobium*, *Rhizomucor*, *Rhizopus*, *Ruminococcus*, *Saccharomycopsis*, *Salmonella*, *Schizophyllum*, *Scytalidium*, *Septoria*, *Sporotrichum*, *Streptomyces*, *Talaromyces*, *Thermoanaerobacter*, *Thermoascus*, *Thermotoga*, *Thielavia*, *Tolypocladium*, *Trametes*, *Trichoderma*, *Tropaeolum*,

Uromyces, *Verticillium*, *Volvariella*, *Wickerhamomyces*, o telemorfos o anamorfos, y sinónimos y equivalentes taxonómicos de los mismos.

Una célula huésped ejemplar es levadura. Ejemplos son *Candida*, *Hansenula*, *Saccharomyces*, *Schizosaccharomyces*, *Pichia*, *Kluyveromyces*, or *Yarrowia*. La levadura puede ser *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*, or *Yarrowia lipolytica*.

Una célula ejemplar puede ser *Myceliophthora thermophila*, algunas veces referida como "C1". Como aquí se usa, el término "C1" se refiere a *Myceliophthora thermophila*, incluyendo una cepa fúngica descrita por Garg (Véase, Garg, Mycopathol., 30: 3-4 [1966]). Como aquí se usa, "*Chrysosporium lucknowense*" incluye las cepas descritas en las patentes de Estados Unidos Números: 6.015.707, 5.811.381 y 6.573.086; Publicaciones de patentes de Estados Unidos Números: 2007/0238155, US 2008/0194005, US 2009/0099079; Publicaciones de patentes internacionales Números:

WO 2008/073914 y WO 98/15633, e incluyen, sin limitación *Chrysosporium lucknowense* Garg 27K, VKM-F 3500 D (Nº Acceso VKM F-3500-D), C1 cepa UV13-6 (Nº Acceso. VKM F-3632 D), C1 cepa NG7C-19 (Nº Acceso VKM F-3633 D), y C1 cepa UV18-25 (VKM F-3631 D), todos depositados en la Colección Rusa de Microorganismos de la Academia Rusa de Ciencias (VKM), Bakhurhina St. 8, Moscú, Rusia, 113184, y cualquier derivado de los mismos. Aunque inicialmente se describen como *Chrysosporium lucknowense*, C1 puede considerarse actualmente una cepa de *Myceliophthora thermophila*. Otras cepas de C1 incluyen células depositadas bajo número de acceso ATCC 44006, CBS (Centraalbureau voor Schimmelcultures) 122188, CBS 251.72, CBS 143.77, CBS 272.77, CBS122190, CBS122189, y VKM F-3500D. Los derivados C1 ejemplares incluyen organismos modificados donde uno o más genes o secuencias endógenas se han eliminado o modificado y/o uno o más genes o secuencias heterólogas se han introducido. Los derivados incluyen, aunque no se limitan a, UV18#100f Δalpl, UV18#100f Δpyr5 Δalpl, UV18#100.f Δalpl Δpep4 Δalp2, UV18#100.f Δpyr5 Δalpl Δpep4 Δalp2 y UV18#100.f Δpyr4 Δpyr5 Δalpl Δpep4 Δalp2, como se describe en WO 2008073914 y WO 2010107303.

En algunas realizaciones la célula huésped puede ser de la especie *Trichoderma*, como *T. longibrachiatum*, *T. viride*, *Hypocrea jecorina* or *T. reesei*, *T. koningii*, y *T. harzianum*. Alternativamente, la célula huésped puede ser de la especie, *Aspergillus*, como *A. awamori*, *A. funigatus*, *A. japonicus*, *A. nidulans*, *A. niger*, *A. aculeatus*, *A. foetidus*, *A. oryzae*, *A. sojae*, y *A. kawachi*. Alternativamente, la célula huésped es de la especie *Fusarium*, como *F. bactridioides*, *F. cerealis*, *F. crookwellense*, *F. culmorum*, *F. graminearum*, *F. graminum*, *F. oxysporum*, *F. roseum*, y *F. venenatum*.

La célula huésped también puede ser de la especie *Neurospora*, como *N. crassa*. Alternativamente, la célula huésped es de la especie *Humicola*, como *H. insolens*, *H. grisea*, and *H. lanuginosa*. Alternativamente, la célula huésped es de la especie *Mucor*, como *M. miehei* y *M. circinelloides*. La célula huésped puede ser de la especie *Rhizopus*, como *R. oryzae* y *R. niveus*. Alternativamente, la célula huésped es de la especie *Penicillium*, como *P. purpurogenum*, *P. chrysogenum*, y *P. verruculosum*.

Alternativamente, la célula huésped es de la especie *Thielavia*, como *T. terrestres*. Alternativamente, la célula huésped es de la especie *Tolypocladium*, como *T. inflatum* y *T. geodes*. Alternativamente, la célula huésped es de la especie *Trametes*, como *T. villosa* y *T. versicolor*. Alternativamente, la célula huésped es de la especie *Chrysosporium*, como *C. lucknowense*, *C. keratinophilum*, *C. tropicum*, *C. merdarium*, *C. inops*, *C. pannicola*, y *C. zonatum*. En una realización particular, la huésped es *C. lucknowense*.

Alternativamente, la célula huésped es un alga como *Chlamydomonas* (como *C. reinhardtii*) y *Phormidium* (P. sp. ATCC29409).

Alternativamente, la célula huésped es una célula procariótica. Las células procarióticas adecuadas incluyen células bacterianas Gram-positivas, Gram-negativas y Gram-variables. Ejemplos de células huéspedes bacterianas incluyen especies de *Bacillus* (como *B. subtilis*, *B. licheniformis*, *B. megaterium*, *B. stearothermophilus* y *B. amyloliquefaciens*), *Streptomyces* (por ejemplo, *S. ambofaciens*, *S. achromogenes*, *S. avermitilis*, *S. coelicolor*, *S. aureofaciens*, *S. aureus*, *S. fungicidicus*, *S. griseus*, y *S. lividans*), y *Streptococcus* (como *S. equisimilis*, *S. pyogenes*, y *S. uberis*).

Los ejemplos no limitativos de tipos de células en esta sección incluyen *Aspergillus aculeatus*, *Azospirillum irakense* KBC1, *Bacillus* sp. GL1, *Cellulomonas biazotea*, *Clostridium thermocellum*, *Thermoanaerobacter brockii*, *Coccidioides posadasii*, *Dictyostelium discoideum*, *Elizabethkingia meningoseptica*, *Erwinia chrysanthemi*, *Escherichia coli*, *Gluconacetobacter xylinus*, *Hypocrea jecorina*, *Kuraishia capsulata*, *Nicotiana tabacum*, *Paenibacillus* sp. C7, *Penicillium brasilianum*, *Periconia* sp. BCC 2871, *Phaeosphaeria avenaria*, *Prevotella albensis*, *Rhizobium leguminosarum*, *Rhizomucor miehei*, *Ruminococcus albus*, *Saccharomycopsis fibuligera*, *Salmonella*

typhimurium, *Septoria lycopersici*, *Streptomyces coelicolor*, *Talaromyces emersonii*, *Thermotoga maritima*, *Tropaeolum majus*, *Uromyces viciae-fabae*, y *Wickerhamomyces anomalus*.

Las cepas que pueden usarse en la práctica de la invención (tanto cepas procarióticas como eucarióticas) pueden obtenerse de cualquier fuente adecuada, incluyendo, aunque sin limitar a la Colección Americana de Cultivos Tipo (ATCC) u otros depósitos biológicos como Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH (DSMZ), Centraalbureau Voor Schimmelcultures (CBS), y la Colección de Cultivos Patentes para Servicio de Investigación Agrícola, Centro de Investigación Regional del Norte (NRRL).

Las células huéspedes pueden modificarse genéticamente para tener características que mejoran la manipulación genética, secreción de proteínas, estabilidad de proteínas u otras propiedades deseables para expresión o secreción de proteína. Por ejemplo, el bloqueo de la función Alp1 da como resultado una célula que es deficiente de proteasa. El bloqueo de la función pyr5 da como resultado una célula con un fenotipo deficiente de pirimidina. Las células huéspedes pueden modificarse para eliminar secuencias codificadoras de proteína celular endógena o de otra manera eliminar la expresión de una o más celulasas endógenas. La expresión de una o más celulasas endógenas no deseadas puede inhibirse para aumentar la proporción de celulasas de interés, por ejemplo, mediante mutagénesis química o UV y la selección posterior. La recombinación homóloga puede usarse para inducir modificaciones dirigidas a genes dirigiendo específicamente un gen in vivo para suprimir la expresión de la proteína codificada.

Transformación y Cultivo Celular

Los polinucleótidos de la divulgación, que codifican proteínas GH61, proteínas de celulasa u otras proteínas, pueden introducirse en células huéspedes para expresión. El polinucleótido puede introducirse en las células como un episoma auto-replicante (por ejemplo, vector de expresión), o puede integrarse establemente en el ADN de la célula huésped.

La introducción de un vector o una construcción de ADN en una célula huésped puede efectuarse mediante cualquier método adecuado, incluyendo aunque sin limitar a, transfección de fosfato de calcio, transfección mediada por DEAE-Dextrano, electroporación u otras técnicas comunes (Véase Davis et al., 1986, MÉTODOS BÁSICOS EN BIOLOGÍA MOLECULAR; Sambrook et al (2001) Clonación Molecular: Un Manual de Laboratorio, 3ª ed., Cold Spring Harbor Laboratory Press, Nueva York; "Guía para Genética de Levadura y Biología Molecular", C. Guthrie y G. Fink, Eds., Métodos en Enzimología 350 (Academic Press, San Diego, 2002). En algunas realizaciones, el polinucleótido que se introduce en la célula huésped permanece en el genoma o en el plásmido u otro vector establemente mantenido en la célula y es capaz de heredarse por la progenie. La transformación estable se realiza típicamente transformando la célula huésped con un vector de expresión que comprende el polinucleótido de interés junto con un gen marcador seleccionable (por ejemplo, un gen que confiera resistencia a un antibiótico). Solamente aquellas células huéspedes que han integrado secuencias de polinucleótido del vector de expresión en su genoma sobrevivirán a la selección con el marcador (por ejemplo, antibiótico). Estas células huéspedes establemente formadas pueden después propagarse de acuerdo con métodos conocidos en la técnica.

Las células huéspedes construidas pueden cultivarse en medios nutrientes convencionales modificados como sea apropiado para activar promotores, seleccionar transformantes o amplificar el polinucleótido GH61. Las referencias generales en las técnicas de cultivo celular y medios nutrientes incluyen MANIPULACIÓN GENÉTICA EN HONGOS, Benett, J. W., et al., Ed. Academic Press, 1985; MÁS MANIPULACIÓN GENÉTICA EN HONGOS, Benett, J. W., et al., Ed. Academic Press, 1991; Y EL MANUAL DE MEDIOS MICROBIOLÓGICOS, CRC Press, Boca Ratón, FL, 1993. Las condiciones de cultivo para células huéspedes C1 se describen en US 2008/0194005, US 2003/0187243, WO 2008/073914 y WO 01/79507. Las condiciones de cultivo, como temperatura, pH y similares, son aquellas previamente usadas con la célula huésped seleccionada para expresión, y será aparente para aquellos expertos en la técnica. Como se ha anotado, hay disponibles muchas referencias que describen el cultivo y producción de muchas células, incluyendo células de bacterias, plantas y animales (especialmente mamíferos) y de arqueobacterias. Atlas y Parks (eds.). El Manual de Medios Microbiológicos (1993) CRC Press, Boca Ratón, FL. Se encuentra más información para cultivo celular en bibliografía comercial disponible como Catálogo de Cultivo Celular para Investigación de Ciencia Viva (1998) de Sigma-Aldrich, Inc (St. Louis, MO) ("Sigma-LSRCCC") y, por ejemplo, El Catálogo de Cultivos de Plantas y Complementos (1997) también de Sigma-Aldrich, Inc (St. Louis, MO) ("Sigma-PCCS").

Enriquecimiento y Purificación de Proteínas

Un polipéptido expresado puede recuperarse de células o caldo. Opcionalmente, una proteína puede enriquecerse (por ejemplo, purificarse o purificarse parcialmente) usando métodos bien conocidos en la técnica. Por ejemplo, el polipéptido puede aislarse del medio nutriente mediante procedimientos convencionales incluyendo, aunque sin limitar, centrifugación, filtración, extracción, secado con pulverización, evaporación, cromatografía (por ejemplo, cromatografía de intercambio de iones, enlace de fase sólida, afinidad, interacción hidrofóbica, cromatografía de exclusión de tamaño) y/o filtración o precipitación. Las etapas de plegar la proteína pueden

usarse, como se desea, para finalizar la configuración de la proteína madura. Finalmente, cromatografía líquida de alto rendimiento (HPLC) puede emplearse en las etapas finales de purificación. Véase, por ejemplo, Parry et al., 2001, *Biochem. J.* 353: 117, y Hong et al., 2007, *Appl. Microbiol. Biotechnol.* 73: 1331. Otros métodos de purificación bien conocidos en la técnica incluyen aquellos expuestos en Sandana (1997) *Bioseparación de Proteína*, Academic Press, Inc.; Bollag et al. (1996) *Métodos de Proteínas*, 2ª Edición, Wiley-Liss, NY; Walker (1996) *El Manual de Protocolos de Proteínas*, Humana Press, NJ; Harris y Angal (1990) *Aplicaciones de Purificación de Proteínas: Una Técnica Práctica*, IRL Press en Oxford, Oxford, Inglaterra; Scopes (1993) *Purificación de Proteínas: Principios y Práctica*, 3ª Edición, Springer Verlag, NY; Janson y Ryden (1998) *Purificación de Proteínas: Principios, Métodos y Aplicaciones de Alta Resolución*, Segunda Edición, Wiley-VCH, NY; y Walker (1998) *Protocolos de Proteínas en CD-ROM*, Humana Press, NJ; PURIFICACIÓN DE PROTEÍNAS: PRINCIPIOS, MÉTODOS DE ALTA RESOLUCIÓN Y APLICACIONES, J. C. Janson (Ed.), Wiley 2011; EXPRESIÓN Y PURIFICACIÓN DE PROTEÍNAS DE ALTO RENDIMIENTO: MÉTODOS Y PROTOCOLOS, S. A., Doyle (Ed.) Humana Press 2009.

Técnicas generales

Los polinucleótidos que codifican proteínas GH61 y otras proteínas pueden prepararse, por ejemplo, mediante síntesis química usando el método clásico de fosforamidita descrito en Beaucage, et al., 1981, *Tetrahedron Letters*, 22: 1859-69 o el método descrito por Matthes, et al., 1984, *EMBO J.* 3:801-05. Los oligonucleótidos de hasta aproximadamente 40 bases se sintetizan individualmente, después se unen (por ejemplo, mediante métodos de ligación enzimática o química, o métodos mediados por polimerasa) para formar esencialmente cualquier secuencia continua deseada.

Los textos generales que describen técnicas biológicas moleculares que incluyen el uso de vectores, promotores, métodos de amplificación in vitro que incluyen reacción en cadena de la polimerasa (PCR) y la reacción en cadena de la ligasa (LCR) son Berger y Kimmel, GUÍA PARA TÉCNICAS DE CLONACIÓN MOLECULAR, MÉTODOS EN ENZIMOLOGÍA volumen 152 Academic Press, Inc., San Diego, CA (Berger); Sambrook et al., CLONACIÓN MOLECULAR - UN MANUAL DE LABORATORIO (2ª Ed.), Vol. 1-3, Cold Spring Harbor Laboratory, Cold Spring Harbor, Nueva York, 1989 y PROTOCOLOS ACTUALES EN BIOLOGÍA MOLECULAR, F. M. Ausubel et al., eds. Current Protocols (como suplemento a lo largo de 2009).

VII. Purificación de proteínas GH61 endógenas de caldo de cultivo

Como una alternativa a la expresión recombinante de proteínas HG61 de esta divulgación, las proteínas GH61 secretadas pueden fraccionarse del caldo de cultivo de *Myceliophthora thermophila* que produce y secreta una o más proteínas endógenas con actividad GH61. De la misma manera, GH61 endógeno no secretado puede recuperarse mediante lisis de células de *M. thermophila*.

Las proteínas GH61 de esta divulgación pueden obtenerse a partir de células que expresan proteínas GH61 usando técnicas estándares de separación de proteínas, como las descritas anteriormente, y siguiendo la actividad GH61 durante el fraccionamiento con un ensayo adecuado de actividad GH61.

Como se ilustra en los Ejemplos, cuando se aísla la proteína de caldo de cultivo de *M. thermophila*, una combinación efectiva es cromatografía en un grupo fenilo que presente resina, seguido de cromatografía de intercambio de anión. Como resultado de las técnicas de separación, la actividad específica de la proteína GH61 (la actividad observada en un ensayo de actividad por unidad total de proteína) puede aumentar aproximadamente 10, aproximadamente 25, aproximadamente 100, aproximadamente 250, aproximadamente 1000 veces o más.

Una vez que la actividad GH61 se ha fraccionado de una fuente adecuada, las fracciones pueden recombinarse entre sí y/o con proteínas GH61 recombinantes en cualquier combinación (véase Ejemplos). Tales fracciones o combinaciones pueden después usarse para promover la actividad de una o más celulasas, como aquí se describe. Las proteínas GH61 purificadas o recombinantes de esta divulgación, y combinaciones de ellas, pueden provocar un aumento en el índice de actividad de celulasa para conversión de biomasa celulósica u otro sustrato a azúcares fermentables en al menos aproximadamente 1%, 5%, 10%, 15%, 20%, al menos aproximadamente 25%, al menos aproximadamente 30%, al menos aproximadamente 35%, al menos aproximadamente 40%, al menos aproximadamente 45%, al menos aproximadamente 50%, al menos aproximadamente 55%, al menos aproximadamente 60%, al menos aproximadamente 65%, al menos aproximadamente 70%, al menos aproximadamente 75%, al menos aproximadamente 80%, al menos aproximadamente duplicar, al menos aproximadamente cuadruplicar o más.

Al usar tales técnicas de separación de proteínas en combinación con un ensayo de actividad GH61, se ha determinado que las fracciones de proteína que tienen datos de secuencia completa o parcial correspondientes a GH61f, GH61a, GH61v, GH61p, GH61g, y GH61i (SEQ ID NOs: 2, 7, 13, 20, 21, 23, 26) tienen la habilidad para mejorar la actividad de celulasa de acuerdo con su clasificación como una proteína GH61 (Véase Ejemplos).

VIII: Celulasas

Las proteínas GH61 de esta divulgación son útiles para aumentar la producción de azúcares fermentables en una reacción de sacarificación con una o más enzimas de celulasa. La proteína GH61 y las enzimas de celulasa pueden producirse en la misma célula o en células diferentes. En cada caso, las enzimas de celulasa pueden expresarse a partir de una región codificadora recombinante o a partir de un gen constitutivo. Las enzimas de celulasa pueden proporcionarse en forma de un caldo de cultivo o sobrenadante, o purificarse hasta el punto que se desee.

Las celulasas para uso en la presente invención pueden derivarse de cualquier organismo que produzca celulasas, y pueden expresarse en, para ilustración pero sin limitación, cualquier célula aquí descrita. En algunas realizaciones las celulasas se derivan de y/o se expresan en un hongo filamentoso (por ejemplo, especie *Myceliophthora*, *Aspergillus Azospirillum*, y *Trichoderma*) o célula de levadura. Para ilustración, pueden usarse las celulasas derivadas de cualquiera de las siguientes células. Por ejemplo, muchos hongos (incluyendo pero sin limitar a especies de *Thielavia*, *Humicola*, *Chaetomium*, *Neurospora*, *Chaetomidium*, *Botryosphaeria*, *Trichophaea*, *Aspergillus*, *Schizophyllum*, *Agaricus*, *Sporotrichium*, *Corynascus*, *Myceliophthora*, *Acremonium*, *Thermoascus*, *Alternaria*, *Botryotinia*, *Phanerochaete*, *Claviceps*, *Cochliobolus*, *Cryphonectria*, *Emericella*, *Fusarium*, *Gibberella*, *Hypocrea*, *Irpex*, *Magnaporthe*, *Nectria*, *Neosartorya*, *Penicillium*, *Phanerochaete*, *Pleurotus*, *Podospora*, *Polyporus*, *Sclerotinia*, *Sordaria*, *Talaromyces*, *Trichoderma*, y *Volvariella*. Por ejemplo, *Acremonium thermophilum*; *Agaricus bisporus*; *Alternaria alternate*; *Aspergillus aculeatus*; *Aspergillus clavatus*; *Aspergillus flavus*; *Aspergillus fumigatus*; *Aspergillus nidulans*; *Aspergillus niger*; *Aspergillus oryzae*; *Aspergillus terreus*; *Botryotinia fuckeliana*; *Chaetomium thermophilum*; *Phanerochaete Chrysosporium*; *Claviceps purpurea*; *Cochliobolus carbonum*; *Cryphonectria parasitica*; *Emericella nidulans*; *Fusarium oxysporum*; *Fusarium poae*; *Fusarium venenatum*; *Gibberella avenacea*; *Gibberella pulicaris*; *Gibberella zeae*; *Humicola grisea*; *Hypocrea koningii*; *Hypocrea lixii*; *Hypocrea virens*; *Irpex lacteus*; *Magnaporthe grisea*; *Nectria haematococca*; *Neosartorya fischeri*; *Neurospora crassa*; *Penicillium chrysogenum*; *Penicillium decumbens*; *Penicillium funiculosum*; *Penicillium janthinellum*; *Penicillium marneffeii*; *Penicillium occitanis*; *Penicillium oxalicum*; *Phanerochaete chrysosporium*; *Pleurotus sp. 'Florida'*; *Podospora anserine*; *Polyporus arcularius*; *Sclerotinia sclerotiorum*; *Sordaria macrospora*; *Talaromyces emersonii*; *Talaromyces stipitatus*; *Thermoascus aurantiacus*; *Trichoderma sp.*; *Trichoderma viride*; *Trichoderma reesei*; *Volvariella volvacea*. En una realización la célula es *M. thermophila*.

Endoglucanasa (EG)

La divulgación proporciona una célula que expresa una proteína GH61 en combinación con una endoglucanasa recombinantes. Los términos "endoglucanasa" o "EG" se refieren a un grupo de enzimas de celulasa clasificadas como E.C. 3.2.1.4. Estas enzimas catalizan la hidrólisis de enlaces glicosídicos internos β -1,4.

Por ejemplo, la célula puede contener una secuencia recombinante de polinucleótido que codifica la proteína EG. La secuencia de polinucleótido EG puede estar operativamente unida a un promotor heterólogo y/o la secuencia de polipéptido EG comprende una señal de secuencia. La proteína EG puede expresarse como una pre-proteína, que se secreta de la célula con la pérdida simultánea del péptido de señal.

EG puede comprender una endoglucanasa endógena de *M. thermophila* como EG2a de *M. thermophila* (véase WO 2007/109441) o una variante de la misma. EG puede ser de *S. avermitilis*, con una secuencia expuesta en el acceso GenBank NP_821730, o una variante como la descrita en US 2010/0267089 A1. EG puede ser un *Thermoascus aurantiacus* EG, o un EG endógeno de una bacteria, una levadura, o un hongo filamentoso diferente de *M. thermophila*. De hecho, se contempla que cualquier EG adecuado encontrará uso en combinación con las proteínas GH61 aquí proporcionadas. No se pretende que la presente invención esté limitada a ninguna EG específico.

β -Glucosidasa (BGL)

La divulgación proporciona una célula que expresa una proteína GH61 en combinación con una β -glucosidasa recombinante. Los términos " β -glucosidasa", "celobiodasa" o "BGL" se refieren a un grupo de enzimas de celulasa clasificados como E.C. 3.2.1.21. Estas enzimas hidrolizan celobiosa a glucosa.

Por ejemplo, la célula puede contener una secuencia de polinucleótido recombinantes que codifica la proteína BGL, donde la secuencia de polinucleótido está operativamente unida a un promotor heterólogo y/o secuencia de señal. La proteína BGL puede expresarse como una pre-proteína, que se secreta de la célula con pérdida simultánea del péptido de señal.

BGL puede ser una *M. thermophila* BGL1 o una variante de la misma. La BGL1 puede comprender la secuencia expuesta en SEQ ID NO: 60 o SEQ ID NO: 66, o es una variante de las mismas, o una variante descrita en US 2011/0129881 A1. Alternativamente, la BGL es de *Thermoascus aurantiacus* (TaBGL), que tiene una secuencia expuesta como SEQ ID NO: 61, o es una variante de la misma, o una variante como las descritas en US 2011/0124058 A1.

Alternativamente, la BGL es de *Azospirillum irakense* (CelA), que tiene una secuencia expuesta como SEQ ID NO: 62, o una variante de la misma, o una variante descrita en US 2011/0114744 A1. Alternativamente, BGL se describe en la TABLA 1 de la solicitud PCT N° PCT/US2010/038902. Alternativamente, la BGL es una BGL endógena de una bacteria, una levadura, un hongo filamentosos diferente a *M. thermophila*. También se contempla el uso de variantes de BGLs que ocurren de manera natural. De hecho, se contempla que cualquier BGL adecuada encontrará uso en combinación con las proteínas GH61 aquí proporcionadas. No se pretende que la presente invención se limite a ninguna BGL específica.

Celobiohidrolasa de Tipo 1 y Tipo 2

La divulgación proporciona una célula que expresa una proteína GH61 en combinación con una celobiodrilasa recombinante de Tipo 1. Los términos "celobiohidrolasa", "exoglucanasa", "exo-celobiodrilasa" o "CBH" se refieren a un grupo de enzimas de celulasa clasificadas como E.C. 3.2.1.91. Las celobiohidrolasas de Tipo 1 (CBH2) hidrolizan celobiosa en un proceso del extremo reductor las cadenas de celulosa. Las celobiodrilasas de Tipo 2 (CBH3) hidrolizan celobiosa en un proceso del extremo no reductor las cadenas de celulosa.

Por ejemplo, la célula puede contener una secuencia de polinucleótido recombinante que codifica la proteína CBH, donde la secuencia de polinucleótido está operativamente unida a un promotor heterólogo y/o secuencia de péptido de señal. La proteína CBH puede expresarse como una pre-proteína, que se secreta de la célula con pérdida simultánea del péptido de señal.

La célula puede ser una célula *M. thermophila*, y puede ser una celobiohidrolasa endógena, como CBH1a, que tiene una secuencia expuesta en SEQ ID No. 63 o 67 o una variante de la misma. Alternativamente, CBH1 es una CBH1 endógena de una bacteria, una levadura, un hongo filamentosos diferente a *M. thermophila* o una variante de CBH1s que ocurren de manera natural. De hecho, se contempla que cualquier CBH adecuada encontrará uso en combinación con las proteínas GH61 aquí proporcionadas. No se pretende que la presente invención se limite a ninguna CBH específica.

IX.: Composiciones libres de células donde la proteína GH61 se combina con enzimas de celulasa

En un aspecto, la divulgación proporciona una composición que comprende al menos una proteína GH61 aquí descrita (por ejemplo, que comprende una secuencia de SEQ ID NO. 1-30, que comprende una parte secretada de la proteína GH61, que comprende una parte truncada de terminal amino de la proteína GH61, y variantes biológicamente activas de las mismas), en combinación con al menos una, al menos dos, al menos tres o más celulasas seleccionadas de EGs, BGLs, CBH1s y/o CBH2s, donde la masa combinada de EG, BGL, CBH1 y/o CBH2 es al menos aproximadamente 50%, al menos aproximadamente 60% o al menos aproximadamente 70% de la proteína total libre de célula en la composición. La proteína GH61 (ya sea en caldo o en forma parcialmente purificada) puede combinarse con celulasas de *M. thermophila* o de otros organismos productores de celulasa (incluyendo, por ejemplo, organismos enumerados más abajo).

En algunas composiciones de la invención, la proteína GH61 comprende SEQ ID NO: 2; y (a) la CBH1 es una variante CBH1a de *M. thermophila* con al menos aproximadamente 80%, al menos aproximadamente 85%, algunas veces al menos aproximadamente 90%, y algunas veces al menos aproximadamente 95% de identidad de secuencia con SEQ ID NO: 63 o 67; y/o (b) la CBH2 es una variante CBH2b de *M. thermophila* con al menos aproximadamente 80%, al menos aproximadamente 85%, algunas veces al menos aproximadamente 90%, y algunas veces al menos aproximadamente 95% de identidad de secuencia con SEQ ID NO: 64 o 68; y/o (c) la BGL es una variante BGL1 de *M. thermophila* con al menos aproximadamente 80%, al menos aproximadamente 85%, algunas veces al menos aproximadamente 90%, y algunas veces al menos aproximadamente 95% de identidad de secuencia con SEQ ID NO: 60 o 66.

La composición también puede ser un medio de cultivo celular (esto es, caldo de cultivo) que contiene GH61 recombinante secretado y proteínas de celulasa. Tale medios pueden producirse cultivando células recombinantes descritas aquí anteriormente bajo condiciones donde una combinación de enzimas (por ejemplo, proteínas GH61, EG, CBH y/o BGL) se expresa y secreta. El medio de cultivo celular puede estar esencialmente libre de células, por ejemplo, retirándolas mediante centrifugación o filtración. Una composición para degradar celulosa puede producirse cultivando células recombinantes descritas anteriormente bajo condiciones donde las enzimas (por ejemplo, proteínas GH61, EG, CBH y/o BGL) se expresan y secretan, opcionalmente retirando las células del medio, y opcionalmente enriqueciendo el medio para aumentar la concentración de proteínas.

X. Uso de proteínas GH61 en reacciones de sacarificación

Las reacciones de sacarificación pueden realizarse exponiendo un sustrato celulósico (por ejemplo, biomasa pre-tratada) a una proteína GH61 y celulasa, que funcionan en concierto para hidrolizar una celulosa y producir azúcares fermentables.

Típicamente, las celulasas incluyen al menos una endoglucanasa (EG), al menos una β -glucosidasa (BGL), al menos una celobiohidrolasa de Tipo 1 (CBH1), y/o al menos una celobiohidrolasa de Tipo 2 (CBH2).

Las células y composiciones de la invención (incluyendo el caldo de cultivo o lisados celulares) pueden usarse en la producción de azúcares fermentables de biomasa celulósica. El sustrato de biomasa puede convertirse a un azúcar fermentable (a) pre-tratando opcionalmente un sustrato celulósico para aumentar su susceptibilidad a hidrólisis; (b) contactando el sustrato celulósico opcionalmente pre-tratado de la etapa (a) con una composición, medio de cultivo o lisado celular que contenga una proteína GH61 y celulasas bajo condiciones adecuadas para la producción de celobiosa y azúcares fermentables (por ejemplo, glucosa).

En una realización, para realizar una reacción de sacarificación, cada una de las proteínas GH61 y enzimas de celulasa referidas anteriormente pueden purificarse parcialmente o sustancialmente, y las proteínas purificadas se combinan con el sustrato celulósico. En otra realización las varias proteínas individuales se expresan recombinantemente en diferentes células, y los medios que contienen las proteínas secretadas se añaden a la biomasa.

Las composiciones pueden reaccionar con el sustrato a una temperatura en el rango de aproximadamente 25° C a aproximadamente 110 °C, de aproximadamente 30° C a aproximadamente 90 °C, de aproximadamente 30° C a aproximadamente 80 °C, de aproximadamente 40° C a aproximadamente 80 °C, de aproximadamente 35° C a aproximadamente 75 °C, de aproximadamente 55° C a aproximadamente 100 °C o a aproximadamente 90 °C. El proceso puede realizarse a un pH en un rango de desde aproximadamente pH 3,0 a aproximadamente 8,5, desde aproximadamente pH 3,5 a aproximadamente 8,5, desde aproximadamente pH 4,0 a aproximadamente 7,5, desde aproximadamente pH 4,0 a aproximadamente 7,0 y desde aproximadamente pH 4,0 a aproximadamente 6,5. Los tiempos de reacción para convertir un sustrato particular de biomasa en un azúcar fermentable pueden variar pero el tiempo óptimo de reacción puede determinarse fácilmente. Los tiempos ejemplares de reacción pueden estar en el rango de desde aproximadamente 1 a aproximadamente 240 horas, desde aproximadamente 5 a aproximadamente 180 horas y desde aproximadamente 10 a aproximadamente 150 horas. Por ejemplo, el tiempo de incubación puede ser al menos 1 hora, al menos 5 h, al menos 10 h, al menos 15 h, al menos 25 h, al menos 50 h, al menos 100 h o al menos 180 h.

En algunas realizaciones, los polipéptidos de GH61 de la presente divulgación se usan en combinación con otros ingredientes opcionales como al menos un tampón o un surfactante. En algunas realizaciones, se usa al menos un tampón con el polipéptido de GH61 de la presente divulgación (opcionalmente combinado con otras enzimas) para mantener un pH deseado en la solución donde se emplea GH61. Los tampones adecuados son bien conocidos en la técnica. En algunas realizaciones, se usa al menos un surfactante con la GH61 de la presente divulgación. Los cationes (por ejemplo, Cu^{++} , Mn^{++} , Co^{++} , Mg^{++} y Ca^{++} en concentraciones de 0,001 a 50 mM, de 5 μM a 1 mM, de 10-50 μM a 10-20 μM) pueden incluirse en la reacción.

Las combinaciones ejemplares de la proteína GH61 y celulasas incluyen: proteína GH61 con una o más endoglucanasas (EG); (BGL); proteína GH61 con una o más 1 (CBH1); o proteína GH61 con una o más celobiohidrolasas de Tipo 2 (CBH2). Otras combinaciones son proteína GH61 con EG y BGL; proteína GH61 con EG y CBH1; proteína GH61 con EG y CBH2; proteína GH61 con BGL y CBH1; proteína GH61 con BGL y CBH2, o proteína GH61 con CBH1 y CBH2. Otras combinaciones son proteína GH61 con EG, BGL, y CBH1; proteína GH61 con EG, BGL, y CBH2; proteína GH61 con EG, CBH1, CBH2; proteína GH61 con BGL, CBH1, y CBH2; y proteína GH61 con todos de EG, BGL, CBH1, y CBH2. Otras enzimas enumeradas en esta divulgación pueden incluirse en una o más de estas combinaciones.

En algunas realizaciones de la divulgación, la mezcla de enzimas comprende una GH61 aislada como aquí se proporciona y al menos una o más de una celobiodrolasa de tipo 1 como CBH1a, una CBH2b aislada, una endoglucanasa aislada (EG) como una endoglucanasa de tipo 2 (EG2), o una endoglucanasa de tipo 2 (EG2), y/o una β -glucosidasa aislada (BGL). En algunas realizaciones, al menos 5%, al menos 10%, al menos 15%, al menos 20%, al menos 25%, al menos 30%, al menos 35%, al menos 40%, al menos 45% o al menos 50% de la mezcla de enzimas es GH61. En algunas realizaciones, la mezcla de la encima comprende además una celobiodrolasa tipo 1a (por ejemplo, CBH2a), y GH61, donde las enzimas juntas comprenden al menos 25%, al menos 30%, al menos 35%, al menos 40%, al menos 45%, al menos 50%, al menos 55%, al menos 60%, al menos 65%, al menos 70%, al menos 75% o al menos 80% de la mezcla de la enzima. En algunas realizaciones, la mezcla de la enzima comprende además una β -glucosidasa (BGL), GH61, CBH2b, donde las tres enzimas juntas comprenden al menos 30%, al menos 35%, al menos 40%, al menos 45%, al menos 50%, al menos 55%, al menos 60%, al menos 65%, al menos 70%, al menos 75%, al menos 80% o al menos 85% de la mezcla de enzimas. En algunas realizaciones, la mezcla de la enzima comprende además una endoglucanasa (EG), GH61, CBH2b, CBH1a, BGL donde las cinco enzimas juntas comprenden al menos 35%, al menos 40%, al menos 45%, al menos 50%, al menos 55%, al menos 60%, al menos 65%, al menos 70%, al menos 75%, al menos 80%, al menos 85% o al menos 90% de la mezcla de la enzima. En algunas realizaciones, la mezcla de enzimas comprende GH61, CBH2b, CBH1a, BGL y al menos una EG, en cualquier proporción adecuada para la reacción deseada.

En algunas realizaciones, la composición de la mezcla de enzimas comprende celulasas aisladas en las siguientes proporciones por peso (donde el peso total de las celulasas es 100%); aproximadamente 20%-10% de BGL, aproximadamente 30%-25% de CBH1a, aproximadamente 10%-30% de GH61, aproximadamente 20%-10% de EG1b, y aproximadamente 20%-25% de CBH2b. En algunas realizaciones, la mezcla de enzimas comprende celulasas aislada en las siguientes proporciones por peso: aproximadamente 20%-10% de GH61, aproximadamente 25%-15% de BGL, aproximadamente 20%-30% de CBH1a, aproximadamente 10%-15% de EG, y aproximadamente 25%-30% de CBH2b. En algunas realizaciones, la mezcla de enzimas comprende celulasas aislada en las siguientes proporciones por peso: aproximadamente 30%-20% de GH61, aproximadamente 15%-10% de BGL, aproximadamente 25%-10% de CBH1a, aproximadamente 25%-10% de CBH2b, aproximadamente 15%-10% de EG. En algunas realizaciones, la mezcla de enzimas comprende celulasas aislada en las siguientes proporciones por peso: aproximadamente 40%-30% de GH61, aproximadamente 15%-10% de BGL, aproximadamente 20%-10% de CBH1a, aproximadamente 20%-10% de CBH2b, y aproximadamente 15%-10% de EG.

En algunas realizaciones, la mezcla de enzimas comprende celulasas aislada en las siguientes proporciones por peso: aproximadamente 50%-40% de GH61, aproximadamente 15%-10% de BGL, aproximadamente 20%-10% de CBH1a, aproximadamente 15%-10% de CBH2b, y aproximadamente 10%-5% de EG. En algunas realizaciones, la mezcla de enzimas comprende celulasas aislada en las siguientes proporciones por peso: aproximadamente 10%-15% de GH61, aproximadamente 20%-25% de BGL, aproximadamente 30%-20% de CBH1a, aproximadamente 15%-5% de EG, y aproximadamente 25%-35% de CBH2b. En algunas realizaciones, la mezcla de enzimas comprende celulasas aislada en las siguientes proporciones por peso: aproximadamente 15%-5% de GH61, aproximadamente 15%-10% de BGL, aproximadamente 45%-30% de CBH1a, aproximadamente 25%-5% de EG1b, y aproximadamente 40%-10% de CBH2b. En algunas realizaciones, la mezcla de enzimas comprende celulasas aislada en las siguientes proporciones por peso: aproximadamente 10% de GH61, aproximadamente 15% de BGL, aproximadamente 40% de CBH1a, aproximadamente 25% de EG, y aproximadamente 10% de CBH2b.

En algunas realizaciones, el componente de enzima comprende más de una enzima CBH2b, CBH1a, EG, BGL y/o GH61 (por ejemplo, 2, 3 ó 4 enzimas diferentes) en cualquier combinación adecuada. En algunas realizaciones, una composición de mezcla de enzimas de la invención comprende además al menos una proteína y/o enzima adicional. En algunas realizaciones, las composiciones de mezcla de enzimas de la presente invención comprenden además al menos un enzima adicional diferente a GH61, BGL, CBH1a, GH61 y/o CBH. En algunas realizaciones, las composiciones de mezcla de enzimas de la invención comprenden además una celulasa adicional, diferente a la variante de GH61, BGL, CBH1a, GH61 y/o CBH. En algunas realizaciones, el polipéptido GH61 de la invención está también presente en mezclas con enzimas de no-celulasa que degradan celulosa, hemicelulosa, pectina y/o lignocelulosa y/o enzimas descritas aquí más abajo.

Realizaciones ejemplares de *M. thermophila*

Para ilustración y no limitación, se proporcionan las siguientes realizaciones ejemplares:

Una realización de la divulgación es una célula huésped que es una célula *M. thermophila* que expresa una proteína recombinante que comprende SEQ ID NO: 1-30, que comprende una parte secretada de la proteína GH61, que comprende una parte truncada de terminal amino de la proteína GH61 y variantes biológicamente activas de las mismas. En algunos casos, la célula expresa una GH61 seleccionada de SEQ ID Nos: 3 a 12 o la proteína secretada correspondiente, o una GH61 seleccionada de SEQ ID Nos: 13 a 256 o la proteínas secretada correspondiente.

La divulgación proporciona una célula que comprende una secuencia de ácido nucleico recombinante que codifica una proteína GH61. En algunos aspectos, la divulgación proporciona una célula que comprende una secuencia de ácido nucleico recombinantes que codifica una proteína con al menos aproximadamente 70%, al menos aproximadamente 75%, al menos aproximadamente 80%, al menos aproximadamente 85%, al menos aproximadamente 90%, al menos aproximadamente 91%, al menos aproximadamente 92%, al menos aproximadamente 93%, al menos aproximadamente 94%, al menos aproximadamente 95%, al menos aproximadamente 96%, al menos aproximadamente 97%, al menos aproximadamente 98% o al menos aproximadamente 99% de identidad de secuencia con SEQ ID NO: 2, o al menos aproximadamente 70%, al menos aproximadamente 75%, al menos aproximadamente 80%, al menos aproximadamente 85%, al menos aproximadamente 90%, al menos aproximadamente 91%, al menos aproximadamente 92%, al menos aproximadamente 93%, al menos aproximadamente 94%, al menos aproximadamente 95%, al menos aproximadamente 96%, al menos aproximadamente 97%, al menos aproximadamente 98% o al menos aproximadamente 99% de identidad de secuencia con la parte secretadas de una cualquiera de las SEQ ID Nos: 1 a 30, que comprenden una parte secretada de la proteína GH61 (por ejemplo, SEQ ID NO: 2), que comprenden una parte truncada de terminal amino de la proteína GH61, y variantes biológicamente activas de las mismas. La secuencia de ácido nucleico recombinante puede codificar una proteína con al menos aproximadamente 70%, al menos aproximadamente 75%, al menos aproximadamente 80%, al menos aproximadamente 85%, al menos aproximadamente 90%, al menos aproximadamente 91%, al menos aproximadamente 92%, al menos aproximadamente 93%, al menos aproximadamente 94%, al menos aproximadamente 95%, al menos

aproximadamente 96%, al menos aproximadamente 97%, al menos aproximadamente 98% o al menos aproximadamente 99% de identidad de secuencia con una proteína GH61 enumerada en la TABLA 1 o TABLA 2.

El ácido nucleico puede comprender la secuencia de nucleótido mostrada en cualquiera de SEQ ID Nos: 31 a 59, o un fragmento de la misma, o un ácido nucleico que se hibrida con SEQ ID Nos: 31-59 (o la secuencia exactamente complementaria) bajo condiciones severas (esto es, severidad media alta, alta o muy alta), y que codifica un polipéptido con actividad GH61. Alternativamente, el ácido nucleico puede codificar un polinucleótido que es al menos aproximadamente 70%, al menos aproximadamente 80%, al menos aproximadamente 90% o al menos aproximadamente 95% o al menos aproximadamente 99%, idéntico a cualquiera de tales secuencias o fragmentos, donde el ácido nucleico codifica y puede expresarse para proporcionar un polipéptido con actividad GH61. Opcionalmente, tales secuencias de ácido nucleico pueden optimizarse con codón para expresión en una especie particular, como una levadura, como se describe en otro punto en esta divulgación.

En una realización de la divulgación, una célula huésped expresa al menos una GH61 recombinante que comprende una cualquiera de SEQ ID Nos: 1 a 30, y/O al menos una proteína GH61 recombinante que tiene al menos aproximadamente 70%, al menos aproximadamente 75%, al menos aproximadamente 80%, al menos aproximadamente 85%, al menos aproximadamente 90%, al menos aproximadamente 91%, al menos aproximadamente 92%, al menos aproximadamente 93%, al menos aproximadamente 94%, al menos aproximadamente 95%, al menos aproximadamente 96%, al menos aproximadamente 97%, al menos aproximadamente 98% o al menos aproximadamente 99% de identidad de secuencia con SEQ ID Nos: 1 a 25; y también expresa:

a) una proteína EG recombinante con al menos aproximadamente 70%, al menos aproximadamente 75%, al menos aproximadamente 80%, al menos aproximadamente 85%, algunas veces al menos aproximadamente 90% y algunas veces al menos aproximadamente 95% de identidad de secuencia con *M. thermophila* EG2a (SEQ ID NO: 65); y/o

b) una proteína CBH1a recombinante con al menos aproximadamente 70%, al menos aproximadamente 75%, al menos aproximadamente 80%, al menos aproximadamente 85%, algunas veces al menos aproximadamente 90% y algunas veces al menos aproximadamente 95% de identidad de secuencia con SEQ ID NO: 63 o 67; y/o

c) una proteína CBH2b recombinante con al menos aproximadamente 70%, al menos aproximadamente 75%, al menos aproximadamente 80%, al menos aproximadamente 85%, algunas veces al menos aproximadamente 90% y algunas veces al menos aproximadamente 95% de identidad de secuencia con SEQ ID NO: 64 o 68; y/o

d) una proteína BGL recombinante con al menos aproximadamente 70%, al menos aproximadamente 75%, al menos aproximadamente 80%, al menos aproximadamente 85%, algunas veces al menos aproximadamente 90% y algunas veces al menos aproximadamente 95% de identidad de secuencia con SEQ ID NO: 60, 61, 62 o 66.

En ciertas realizaciones de la divulgación, la célula expresa al menos una, al menos dos, al menos tres (por ejemplo, b-d) o las cuatro de (a), (b), (c) y (d).

XI. Sacarificación en ausencia de EG exógeno

En un aspecto, la divulgación proporciona un método para hidrolizar un sustrato celulósico que comprende combinar una proteína GH61 con enzimas de β -glucosidasa (BGL) y celobiodrolasa (CBH), en una composición sustancialmente libre de endoglucanasa (EG). Se apreciará que la actividad de tipo EG a la que contribuyó una proteína GH61 no es considerada una endoglucanasa. Como se ilustra en el ejemplo más abajo, una reacción de sacarificación puede realizarse en presencia de una proteína GH61, una β -glucosidasa, y una o más enzimas de celobiodrolasa, sin EG recombinante o sin EG añadido, o sustancialmente libre de EG. Bien celobiodrolasa de Tipo 1 o de Tipo 2 o ambas pueden estar presentes. En ausencia de EG, GH61 puede aumentar la producción de una reacción de sacarificación por BGL y CBH sencilla en más de 1,5 o 1,7 veces.

Se dice que una reacción está "sustancialmente libre" de endoglucanasa si (a) no haya actividad de endoglucanasa detectable en la reacción, o (b) la cantidad de esa enzima EG presente sea inferior a 2%, a menudo inferior a 1%, a menudo inferior a 0,5%, a menudo inferior a 0,2% y a menudo inferior a 0,1% (p/p) de la cantidad de BGL presente, o (c) la cantidad de esa enzima EG presente sea inferior a 2%, a menudo inferior a 1%, a menudo inferior a 0,5%, a menudo inferior a 0,2% y a menudo inferior a 0,1% (p/p) de la cantidad de CBH presente, o (d) la cantidad de esa enzima EG presente sea inferior a 2%, a menudo inferior a 1%, a menudo inferior a 0,5%, a menudo inferior a 0,2% y a menudo inferior a 0,1% (p/p) de la cantidad de GH61 presente, o (e) la cantidad de esa enzima EG presente sea inferior a 2%, a menudo inferior a 1%, a menudo inferior a 0,5%, a menudo inferior a 0,2% y a menudo inferior a 0,1% (p/p) de la cantidad de celulosa total presente.

XII. Composiciones que comprenden otras enzimas

Las enzimas adicionales que pueden actuar en concierto para hidrolizar un sustrato celulósico (tal como celulosa o un sustrato que contiene almidón) en el proceso de sacarificación pueden incluirse en las composiciones

o incorporarse en los métodos de esta invención. Tales enzimas incluyen, aunque no se limitan a, xilanasas, hemicelulasas, amilasas, esterasas y celulasas, α -glucosidasas, aminopeptidasas, carbohidrasas, carboxipeptidasas, catalasas, quitinasas, cutinasas, ciclodextrina glicosiltransferasas, deoxirribonucleasas, α -galactosidasas, β -galactosidasas, glucoamilasas, glucocerebrosidasas, invertasas, lacasas, lipasas, manosidasas, mutanasas, oxidasas, enzimas pectinolíticas, peroxidasas, fosfolipasas, fitasas, polifenoloxidasas, ribonucleasas y trans-glutaminasas, así como otras celulosas (por ejemplo, celobiodrolasas de tipo 1 y tipo 2, endoglucanasas y β -glucosidasas). Las mezclas de celulasa para hidrólisis enzimática eficiente de celulosa son bien conocidos (Véase, por ejemplo, Viikari et al., 2007, Adv. Biochem. Eng. Biotechnol., 108: 121-45; y publicaciones de patentes de Estados Unidos 2009/0061484; US 2008/0057541; y US 2009/0209009). En algunas realizaciones, las mezclas de enzimas purificadas que ocurren de manera natural o recombinantes se combinan con materia prima celulósica o un producto de hidrólisis de celulosa. En algunas realizaciones, una o más poblaciones, cada una produciendo una o más celulosas que ocurren de manera natural o recombinantes, se combinan con materia prima celulósica o un producto de hidrólisis de celulosa.

En algunas realizaciones adicionales, la presente divulgación proporciona al menos una GH61 y al menos una endoxilanasas. Las endoxilanasas (EC 3.2.1.8) catalizan la endohidrólisis de 1,4- β -D-enlaces xilosídicos en xilanos. Esta enzima también puede referirse como endo-1,4- β -D-xilanasas o 1,4- β -D-xilano xilanolhidrolasa. En algunas realizaciones, una alternativa es EC 3.2.1.136, o glucuronoarabinosilano endoxilanasas, una enzima que es capaz de hidrolizar 2,4 enlaces xilosídicos en glucuronoarabinosilanos.

En algunas realizaciones adicionales, la presente divulgación proporciona la menos una GH61 y al menos una β -xilosidasa. β -xilosidasas (EC 3.2.1.37) catalizan la hidrólisis de 1,4- β -D-xilanos, para retirar los residuos sucesivos de D-xilosa de las terminales no reductoras. Esta enzima puede también referirse como xilano, 1,4- β -xilosidasa, 1,4- β -D-xilano xilohidrolasa, exo 1,4- β -xilosidasas o xilobiasa.

En algunas realizaciones adicionales, la presente divulgación proporciona al menos una GH61 y al menos una α -L-arabinofuranosidasa. α -L-arabinofuranosidasas (EC 3.2.1.55) catalizan la hidrólisis de residuos α -L-arabinofuranosidasas no reductoras de terminal. La enzima actúa en α -L-arabinofuranosidasas, α -L-arabinans que contiene (1,3)-y/o (1,5)-uniones, arabinosilanos y arabinogalactanos. α -L-arabinofuranosidasa es también conocida como arabinosidasa, α -arabinosidasa, α -L-arabinosidasa, α -arabinofuranosidasa, arabinofuranosidasa, polisacárido α -L-arabinofuranosidasa, α -L-arabinofuranosidasa hidrolasa, L-arabinosidasa y α -L-arabinananasas.

En algunas realizaciones adicionales, la presente divulgación proporciona al menos una GH61 y al menos una α -glucuronidasa. α -glucuronidasa (EC 3.2.1.139) catalizan la hidrólisis de una α -D-glucuronidasa a D-glucuronato y un alcohol.

En algunas realizaciones adicionales, la presente divulgación proporciona al menos una GH61 y al menos una acetilxilanesterasa. Acetilxilanesterasas (EC 3.1.1.72) catalizan la hidrólisis de grupo de acetilo de xilano polimérico, xilosa acetilada, glucosa acetilada, acetato α -naftilo, y acetato p-nitrofenilo.

En algunas realizaciones adicionales, la presente divulgación proporciona al menos una GH61 y al menos un feruloil esterasa. Ferulil esterasas (EC 3.1.1.73) tienen actividad 4-hidroxi-3-metoxycinnamoyl-azúcar hidrolasa (EC 3.1.1.73) que cataliza la hidrólisis del grupo 4-hidroxi-3-metoxycinnamoyl (feruloil) a partir de un azúcar esterificado, que es normalmente arabinosa en sustratos "naturales", para producir ferulato (4-hidroxi-3-metoxycinnamato). Feruloil esterasa es también conocido como esterasa de ácido ferúlico, esterasa de hidroxycinnamoyl, FAE-III, cinnamoyl éster hidrolasa, FAEA, cinnAE, FAE-I o FAE-II.

En algunas realizaciones adicionales, la presente divulgación proporciona al menos una GH61 y al menos una coumaroil esterasa. Coumaroil esterasas (EC 3.1.1.73) catalizan una reacción de la forma: coumaroil-sacárido + H₂O coumarato + sacárido. En algunas realizaciones, el sacárido es un oligosacárido o un polisacárido. Esta enzima también puede ser referida como trans-4-coumaroil esterasa, trans-p-coumaroil esterasa, p-coumaroil esterasa o ácido p-coumárico esterasa. La enzima también se clasifica en EC 3.1.1.73 de manera que también puede ser referida como feruloil esterasa.

En algunas realizaciones adicionales, la presente divulgación proporciona al menos una GH61 y al menos una α -galactosidasa. α -galactosidasas (EC 3.2.1.222) catalizan la hidrólisis de residuos de β -D-galactosa no reductora de terminal en β -D-galactosidos. En algunas realizaciones, el polipéptido es también capaz de hidrolizar α -L-arabinosidos. Esta enzima también puede ser referida como exo-(1 $\rightarrow$ 4)- β -D-galactanasas o lactasa.

En algunas realizaciones adicionales, la presente divulgación proporciona al menos una GH61 y al menos una beta-mananasas. Beta-mananasas (EC 3.2.1.78) catalizan la hidrólisis aleatoria de enlaces 1,4- β -D manosídicos en mananos, galactomananos y glucomanos. Esta enzima también puede ser referida como manano endo-1,4- β -manosidasa o endo1,4-mananasas.

En algunas realizaciones adicionales, la presente divulgación proporciona al menos una GH61 y al menos una beta-manosidasa. Beta-manosidasas (EC 3.2.1.25) catalizan la hidrólisis de residuos β -D-manosa no reductores de terminal en β -D-manosidos. Esta enzima también puede ser referida como mananansa o manasa.

5 En algunas realizaciones adicionales, la presente divulgación proporciona al menos una GH61 y al menos una glucoamilasa. Glucoamilasas (EC 3.2.1.3) catalizan la liberación de D-glucosa de extremos no reductores de moléculas oligo- y poli-sacáridos. La glucoamilasa es a menudo considerada un tipo de amilasa conocido como amilo-glucosidasa.

10 En algunas realizaciones adicionales, la presente divulgación proporciona al menos una GH61 y al menos una amilasa. Amilasas (EC 3.2.1.1) son enzimas divisoras de almidón que degradan el almidón y compuestos relacionados hidrolizando los enlaces α -1,4 y/o α -1-6 glucosídicos de una manera que actúa endo- o exo-. Las amilasas incluyen α -amilasas (EC 3.2.1.1); β -amilasas (3.2.1.2), amilo-amilasas (EC 3.2.1.3), α -glucosidasas (EC 3.2.1.20), pululanastas (EC 3.2.1.41) e isoamilasas (EC 3.2.1.68). En algunas realizaciones, la amilasa es α -amilasa.

15 En algunas realizaciones una o más enzimas que degradan pectina se incluyen en las mezclas de enzimas que comprenden GH61 de la presente divulgación. Una pectinasa cataliza la hidrólisis de pectina en unidades más pequeñas como oligosacárido o sacáridos monoméricos. En algunas realizaciones, las mezclas de enzimas comprenden cualquier pectinasa, por ejemplo una edo-poligalacturonasa, una pectina metil esterasa, una endo-galactanasa, una pectina acetil esterasa, una endo-pectina liasa, pectato liasa, alfa ramosidasa, exo-galacturonasa, una exo-poligalacturoanto liasa, una ramnogalacturonano hirolasa, una ramnogalacturonano liasa, una ramnogalacturonano acetil esterasa, una ramnogalacturonano galacturonohidrolasa y/o una xilogalacturonasa.

20

En algunas realizaciones, la presente divulgación proporciona al menos una GH61 y al menos una endo-poligalacturonasa. Endo-poligalacturonasas (EC 3.2.1.15) catalizan la hidrólisis aleatoria de enlaces 1,4, α -D-galactosidurónicos en pectato y otros galacturonanos. Esta enzima puede también referirse como poligalacturonasa pectina depolimerasa, pectinasa, endopoligalacturonasa, pectolasa, pectina hidrolasa, pectina poligalacturonasa, poli- α -1,4-galacturonido glicanohidrolasa, endogalacturonasa; endo-D-galacturonasa o poli (1,4- α -D-galacturonido) glicanohidrolasa.

25

30 En algunas realizaciones adicionales, la presente divulgación proporciona al menos una GH61 y al menos una pectina metil esteras. Pectin metil esterases (EC 3.1.1.11) catalizan la reacción: pectina + n H₂O = n metanol + pectato. La enzima también puede ser conocida como pectinesterasa, pectina, demetoxilasa, pectina metoxilasa, pectina metilesterasa, pectasa, pectinoesterasa o pectina pectilhidrolasa.

35 En algunas realizaciones adicionales, la presente divulgación proporciona al menos una endo-galactanasa. Endo-galactanasas (EC 3.2.1.89) catalizan la endohidrólisis de enlaces 1,4- β -D-galactosídicos en arabinogalactanos. La enzima también puede ser conocida como arabinogalactano endo-1,4- β -galactosidasa, endo-1,4- β -galactanasa, galactanasa, arabinogalactanasa o arabinogalactano 4- β -D-galactanohidrolasa.

40 En algunas realizaciones adicionales, la presente divulgación proporciona al menos una GH61 y al menos una pectina acetil esterasa. Pectina acetil esterasa catalizan la desacetilación de los grupos de acetilo en los grupos hidroxilos de residuos GalUA de pectina.

45 En algunas realizaciones adicionales, la presente divulgación proporciona al menos una GH61 y al menos una endo-pectina liasa. Endo-pectina liasas (EC 4.2.2.10) catalizan la división eliminativa de (1 $\rightarrow$ 4)- α -D-galacturonano metil éster para dar oligosacáridos con grupos 4-deoxi-6-O-metil- α -D-galacto-4-enuronosil en sus extremos no reductores. La enzima también puede ser conocida como pectina liasa, pectina trans-eliminasa; endo-pectina liasa, transeliminasa polimetilgalacturónica, pectina metiltranseliminasa, pectoliasa, PL, PNL o PMGL o (1 $\rightarrow$ 4)-6-O-metil- α -D-galacturonano liasa.

50

En algunas realizaciones adicionales, la presente divulgación proporciona al menos una GH61 y al menos una pectato liasa. Pectato liasas (EC 4.2.2.2) catalizan la división alternativa de (1 $\rightarrow$ 4)- α -D-galacturonano para dar oligosacáridos con grupos 4-deoxi- α -D-galacto-4-enuronosil en sus extremos no reductores. La enzima también puede ser conocida como transeliminasa poligalacturónica, transeliminasa de ácido péctico, poligalacturonato liasa, endopectina metiltranseliminasa, pectato transeliminasa, endogalacturonato transeliminasa, liasa de ácido péctico, liasa péctica, ácido α -1,4-D-endopoligalacturonico liasa, PGA liasa, PPasa-N, ácido endo- α -1,4-poligalacturonico liasa, ácido poligalacturónico liasa, pectina trans-eliminasa, ácido poligalacturónico trans-eliminasa o (1 $\rightarrow$ 4)- α -D galacturonano liasa.

55

60 En algunas realizaciones adicionales, la presente divulgación proporciona al menos una GH61 y al menos una alfa-ramnosidasa. Alfa-ramnosidasas (EC 3.2.1.40) catalizan la hidrólisis de residuos α -L-ramnosa no reductores de terminal en un α -L-ramnosidasa N o α -L-ramnosidasa ramnoidrolasa.

65 En algunas realizaciones adicionales, la presente divulgación proporciona al menos una GH61 y al menos una exo-galacturonasa. Exo-galacturonasas (EC 3.2.1.82) hidrolizan ácido péctico del extremo no reductor, liberando

digalacturonato. La enzima también puede ser conocida como exo-poli- α -galacturonosidasa, exopoligalacturonosidasa o exopoligalacturanosidasa.

En algunas realizaciones adicionales, la presente divulgación proporciona al menos una GH61 y al menos una galacturano 1,4- α -galacturonidasa. Galacturano 1,4- α -galacturonidasas (EC 3.2.1.67) catalizan una reacción del siguiente tipo: $(1,4,\alpha\text{-D-galacturonido})_n + \text{H}_2\text{O} = (1,4,\alpha\text{-D-galacturonido})_{n-1} + \text{D-galacturonato}$. La enzima también puede ser conocida como poli [1- $\rightarrow$ 4) α -D-galacturonido] galacturonohidrolasa, exopoligalacturonasa, poli(galacturonato) hidrolasas, exo-D-galacturonasa, exo-D-galacturonosasa, exopololi-D-galacturonasa o poli (1,4, α -D-galacturonido) galacturonohidrolasa.

En algunas realizaciones adicionales, la presente divulgación proporciona al menos una GH61 y al menos una exopoligalacturonato liasa. Exopoligalacturonato liasas (EC 4.2.2.9) catalizan la división eliminativa de 4-(4-deoxi- α -D-galacto-4-enuronosil)-D-galacturonato del extremo reductor de pectato (esto es, pectina desterificada). Esta enzima también puede ser conocida como pectato disacárido-liaso, pectato exo-liaso, ácido exopéctico transeliminasa, exopéctato liasa, ácido exopoligalacturónico trans-eliminasa, PATE, exo-PATE, exo-PGL o liasa disacárido reductor de extremo de (1 $\rightarrow$ 4)- α -galacturonano.

En algunas realizaciones adicionales, la presente divulgación proporciona al menos una GH61 y al menos una ramnogalacturonanasa. Ramnogalacturonanasa hidrolizan el enlace entre ácido galactosilurónico y ramnopiranosilo de una manera endo en estructuras de ramnogalacturonano estrictamente alternativas, que consisten en el disacárido [(1,2- α -L-ramnoil-(1,4)- α -ácido galactosilurónico].

En algunas realizaciones adicionales, la presente divulgación proporciona al menos una GH61 y al menos una ramnogalacturonano liasa. Ramnogalacturonano liasas dividen enlaces α -L-Rhap-(1 $\rightarrow$ 4)- α -GalpA de una manera endo en ramnogalacturonano mediante eliminación de beta.

En algunas realizaciones adicionales, la presente divulgación proporciona al menos una GH61 y al menos una ramnogalacturonano acetil esterasa. Ramnogalacturonano acetil esterases catalizan la desacetilación de la red central de residuos alternantes de ramnosa y ácido galacturónico en ramnogalacturonano.

En algunas realizaciones adicionales, la presente divulgación proporciona al menos una GH61 y al menos una ramnogalacturonano galacturonohidrolasa. Ramnogalacturonano galacturonohidrolasas hidrolizan ácido galacturónico del extremo no reductor de estructuras de ramnogalacturonano estrictamente alternantes en una manera exo. Esta enzima también puede ser conocida como xilogalacturonano hidrolasa.

En algunas realizaciones adicionales, la presente divulgación proporciona al menos una GH61 y al menos una endo-arabinanasa. Endo-arabinasas (EC 3.2.1.99) catalizan endohidrólisis de enlaces 1,5- α -arabinofuranosídicos en 1,5-arabinanos. La enzima también puede ser conocida como endo-arabinasa, arabinana endo-1,5- α -L-arabinosidasas, endo-1,5- α -L-arabinanasa y endo- α -1,5-arabinasa; endo-arabinasa o 1,5- α -L-arabinana, 1,5- α -L-arabinanohidrolasa.

En algunas realizaciones adicionales, la presente divulgación proporciona al menos una GH61 y al menos una enzima que participa en la degradación de lignina en una mezcla de enzimas. La despolimerización de lignina enzimática puede realizarse mediante peroxidasas de lignina, peroxidasas de manganeso, lacasas y dehidrogenasas de celobiosa (CDH), a menudo trabajando en sinergia. Estas enzimas extracelulares a menudo son referidas como "encimas modificadoras del lignina" o "EMLs". Tres de estas enzimas comprenden dos peroxidasas que contienen hemoglobina glicosiladas: lignina peroxidasas (LiP); peroxidasa dependiente de Mn (MnP); y lacasa peroxidasa que contiene cobre (LCC):

En algunas realizaciones adicionales, la presente divulgación proporciona al menos una GH61 y al menos una lacasa. Las lacasas son enzimas de oxidasa que contienen cobre que se encuentran en muchas plantas, hongos y microorganismos. Las lacasas son enzimáticamente activas en fenoles y moléculas similares y realizan una oxidación de electrones. Las lacasas pueden ser poliméricas y la forma enzimáticamente activa puede ser un dímero o trímero.

En algunas realizaciones adicionales, la presente divulgación proporciona al menos una GH61 y al menos una peroxidasa dependiente de Mn. La actividad enzimática de peroxidasa dependiente de Mn (MnP) es dependiente de Mn^{2+} . Sin estar ligado a ninguna teoría, se ha sugerido que el papel principal de esta enzima es oxidar Mn^{2+} a Mn^{3+} (Véase, por ejemplo, Glenn et al., Arch. Biochem. Biophys., 251:688-696 [1986]). Posteriormente, los sustratos fenólicos se oxidan por el Mn^{3+} generado.

En algunas realizaciones adicionales, la presente divulgación proporciona al menos una GH61 y al menos una lignina peroxidasa. La lignina peroxidasa es una hemoglobina extracelular que cataliza la despolimerización oxidativa de soluciones diluidas de lignina polimérica in vitro. Algunos de los sustratos de LiP, más notablemente 3,4-dimetoxibenil alcohol (veratril alcohol, VA), son compuestos activos redox que han demostrado actuar como mediadores redox. VA es un metabolito secundario producido al mismo tiempo que LiP mediante cultivos

ligninolíticos de *P. chrysosporium* y sin estar ligado a ninguna teoría, se han propuesto para que funcionen como un mediador redox fisiológico en la oxidación LiP-catalizada de lignina in vivo (Véase, por ejemplo, Harvey, et al., FEBS Lett., 195-242-246 [1986])

5

En algunas realizaciones adicionales, la presente divulgación proporciona al menos una GH61 y al menos una proteasa, amilasa, glucoamilasa y/o una lipasa que participa en la degradación de celulosa.

10

Como aquí se usa, el término "proteasa" incluye enzimas que hidrolizan enlaces de péptido (peptidasas), así como enzimas que hidrolizan enlaces entre péptidos y otras fracciones, como azúcares (glicopeptidasas). Muchas proteasas se caracterizan bajo EC 3.4, y son adecuadas para su uso en la invención. Algunos tipos específicos de proteasas incluyen, cisteína proteasas que incluyen pepsina, papaína y serina proteasas que incluyen quimotripsinas, carboxipeptidasas y metaloendopeptidasas.

15

Como aquí se usa, el término "lipasa" incluye enzimas que hidrolizan lípidos, ácidos grasos y acilglicéridos, incluyendo fosfoglicéridos, lipoproteínas, diacilgliceroles y similares. En plantas, los lípidos se usan como componente estructural para limitar la pérdida de agua y la infección de patógenos. Estos lípidos incluyen ceras derivadas de ácidos grasos, así como cutina y suberina.

20

En algunas realizaciones adicionales, la presente divulgación proporciona al menos una GH61 y al menos una expansina o una proteína de tipo expansina, tal como swollenina (Véase, por ejemplo, Salheimo et al., Eur. J. Biochem., 269:42002-4211 [2002]) o una proteína de tipo swollenina. Las expansinas están implicadas en la liberación de la estructura de pared celular durante el crecimiento celular de la planta. Las expansinas se han propuesto para interrumpir el enlace de hidrógeno entre celulosa y otros polisacáridos de pared celular sin tener actividad hidrolíticas. De esta manera, se piensa que permiten el deslizamiento de fibras de celulosa y el aumento de la pared celular. Swollenina, una proteína de tipo expansina contiene un dominio de la Familia 1 del Módulo de Enlace de Carbohidrato N-terminal (CBD) y un dominio de tipo expansina C-terminal. En algunas realizaciones, una proteína de tipo expansina o una proteína de tipo swollenina comprende uno o ambos de dichos dominios y/o interrumpe la estructura de paredes celulares (como interrumpir la estructura de celulosa), opcionalmente sin producir cantidades detectables de azúcares reductores.

25

30

En algunas realizaciones adicionales, la presente divulgación proporciona al menos una GH61 y al menos un producto polipéptido de una proteína integrante de celulosa, andamio o una proteína de tipo andamio, por ejemplo CipA o CipC de *Clostridium thermocellum* o *Clostridium cellulositycum* respectivamente. Los andamios y las proteínas integrantes de celulosa son sub-unidades multi-funcionales que pueden organizar sub-unidades celulolíticas en un complejo multi-enzima. Esto se consigue mediante la interacción de dos clases complementarias de dominio (esto es, un dominio de cohesión en andamio y un dominio de acoplamiento molecular en cada unidad enzimática). La sub-unidad de andamio también tiene un módulo que se enlaza con celulosa que media la unión de la celulosa con su sustrato. Un andamio o proteína que integra celulosa para los fines de esta invención pueden comprender uno o ambos de estos dominios.

35

40

En algunas realizaciones, la presente divulgación proporciona una GH61 y al menos una proteína inducida por celulosa o proteína moduladora, por ejemplo como la codificada por el gen cip1 cip2 o genes similares de *Trichoderma reesei* (Véase, por ejemplo, Foreman et al., J. Biol. Chem. 278: 31988-31997 [2003]).

45

En algunas realizaciones adicionales, la presente divulgación proporciona al menos una GH61 y al menos un miembro de cada una de las clases de los polipéptidos descritos anteriormente, varios miembros de una clase de polipéptido, o una combinación de estas clases de polipéptidos para proporcionar mezclas de enzimas adecuadas para varios usos.

50

En algunas divulgaciones de la invención, la mezcla de enzimas comprende otros tipos de celulasas, seleccionadas aunque no limitadas a celobiohidrolasa, endoglucanasa, β -glucosidasa y proteína glicósido hidrolasa 61 (GH61). Estas enzimas pueden ser enzimas de tipo salvaje o recombinantes. En algunas realizaciones, la celobiohidrolasa es una celobiohidrolasa de tipo 1 (por ejemplo, *T. reesei* celobiohidrolasa I). En algunas realizaciones, la endoglucanasa comprende un dominio catalítico derivado del dominio catalítico de una endoglucanasa de *Streptomyces avermitilis* (Véase, por ejemplo, la publicación de solicitud de Patente de Estados Unidos Nº 2010/0267089). En algunas realizaciones, la al menos una celulasa se deriva de *Acidothermus cellulolyticus*, *Thermobifida fusca*, *Hemicola grisea*, *Myceliophthora thermophila*, *Chaetomium thermophilum*, *Acremonium* sp., *Thielavia* sp, *Trichoderma reesei*, *Aspergillus* sp., o una *Chrysosporium* sp. Las enzimas de celulasa de la mezcla de celulasa trabajan juntas dando como resultado una descristalización e hidrólisis de la celulosa de un sustrato de biomasa para producir azúcares fermentables, como, aunque sin limitar a, glucosa.

55

60

Algunas mezclas de celulasa para hidrólisis enzimática eficiente son conocidas (Véase, por ejemplo, Viikari et al., Adv. Biochem. Eng. Biotechnol., 108:121-45 [2007]; y publicaciones de solicitudes de patentes de Estados Unidos Números US 2009/0061484, US 2008/0057541 y US 2009/0209009). En algunas realizaciones, las mezclas de enzimas purificadas que ocurren de manera natural o recombinantes se combinan con materia prima celulósica o

65

un producto de hidrólisis de celulosa. Alternativamente o además de, una o más poblaciones celulares, cada una produciendo una o más celulasas que ocurren de manera natural o recombinantes, se combinan con materia prima celulósica o un producto de hidrólisis de celulosa.

En algunas realizaciones, la mezcla de enzimas comprende celulasas purificadas comercialmente disponibles. Las celulasas comerciales son conocidas y están disponibles (por ejemplo, celulasa C2730 de *Trichoderma reesei* ATCC No. 25921 disponible en Sigma-Aldrich, Inc.; y C9870 ACCELLERASE® 1500, disponible en Genencor).

XIII. Sustrato celulósico

Los sustratos celulósicos pueden derivarse de cualquier material que contenga celulosa, tal como biomasa derivada de plantas, animales o microorganismos, y pueden incluir residuos agrícolas, industriales o forestales, desechos industriales o municipales, y cultivos terrestres y acuáticos que crecen para fines energéticos. Los "sustratos celulósicos" abarcan ampliamente cualquier material vivo o muerto que contenga un sustrato de polisacárido, incluyendo aunque sin limitar a, celulosa, almidón, otras formas de polímeros de carbohidrato de cadena larga, y mezclas de tales fuentes. Puede o no montarse enteramente o principalmente a partir de glucosa o xilosa, y pueden también contener opcionalmente varios otros monómeros de pentosa o hexosa. La xilosa es una aldopentosa que contiene cinco átomos de carbono y un grupo aldehído. Es el precursor de hemicelulosa, y a menudo es un constituyente principal de biomasa.

Los sustratos celulósicos a menudo se proporcionan como materias primas de lignocelulosa que pueden procesarse antes de hidrólisis por celulasas. Como aquí se usan, los términos "materia prima de lignocelulosa" se refieren a cualquier tipo de biomasa de planta como, aunque sin limitar a, cosechas cultivada (por ejemplo, hierbas, incluyendo hierbas C4 como pasto varilla, espartina, centeno, miscanthus, hierba cinta o cualquier combinación de ellas), residuos que procesan azúcar, por ejemplo, aunque sin limitar a, bagazo (por ejemplo, bagazo de azúcar de caña, pulpa de remolacha o una combinación de ellas), residuos agrícolas (por ejemplo, rastrojo de soja, rastrojo de maíz, fibra de maíz, paja de arroz, paja de caña de azúcar, arroz, vainas de arroz, paja de cebada, mazorcas de maíz, paja de trigo, paja de canola, paja de avena, vainas de avena, fibra de maíz, cáñamo, linaza, sisal, algodón o cualquier combinación de ellos), pulpa de fruta, pulpa vegetal, granos de destilador, biomasa forestal (por ejemplo, madera, pulpa de madera, pulpa de papel, fibra de pulpa de madera reciclada, serrín, madera dura, como madera de álamo, madera blanda, o una combinación de ellos). Además, en algunas realizaciones, la materia prima de lignocelulosa comprende material celulósico de desecho y/o materiales forestales de desecho, incluyendo aunque sin limitar a, desechos del procesamiento de papel y pulpa, papel de periódico, cartón y similares. La biomasa también puede comprender plantas transgénicas que expresan ligninasa y/o enzimas de celulasa (US 2008/0104724 A1).

En algunas realizaciones, la materia prima lignocelulósica comprende una especie de fibra, mientras en algunas realizaciones alternativas, la materia prima celulósica comprende una mezcla de fibras que se originan de diferentes materias primas lignocelulósicas. En algunas otras realizaciones, la materia prima lignocelulósica comprende materia prima lignocelulósica fresca, materia prima lignocelulósica parcialmente seca, materia prima lignocelulósica completamente seca, y/o combinaciones de las mismas. En algunas realizaciones, la materia prima lignocelulósica comprende celulosa en una cantidad superior a aproximadamente 20%, más preferentemente superior a 30%, más preferentemente superior a 40% (p/p). Por ejemplo, en algunas realizaciones, el material lignocelulósico comprende desde aproximadamente 20% a aproximadamente 90% (p/p) de celulosa, o cualquier cantidad entre ellas, aunque en algunas realizaciones, el material lignocelulósico comprende menos de aproximadamente 19%, menos de aproximadamente 18%, menos de aproximadamente 17%, menos de aproximadamente 16%, menos de aproximadamente 15%, menos de aproximadamente 14%, menos de aproximadamente 13%, menos de aproximadamente 12%, menos de aproximadamente 11%, menos de aproximadamente 10%, menos de aproximadamente 9%, menos de aproximadamente 8%, menos de aproximadamente 7%, menos de aproximadamente 6%, menos de aproximadamente 5% de celulosa (p/p). Además, en algunas realizaciones, la materia prima lignocelulósica comprende lignina en una cantidad superior a aproximadamente 10%, más típicamente en una cantidad superior a aproximadamente 15% (p/p). En algunas realizaciones, la materia prima lignocelulósica comprende pequeñas cantidades de sacarosa, fructosa y/o almidón.

La materia prima lignocelulósica generalmente se somete a una reducción de tamaño mediante métodos que incluyen, aunque no se limitan a, molienda, pulverización, agitación, triturado, compresión/expansión u otros tipos de acción mecánica. La reducción de tamaño mediante acción mecánica puede realizarse mediante cualquier tipo de equipo adaptado para tal fin, por ejemplo, aunque sin limitar a, moledores con martillo, picadoras con cubo, prensas con rollos, refinadoras y licuadoras. En algunas realizaciones, al menos el 90% por peso de las partículas producidas a partir de la reducción de tamaño tienen longitudes inferiores a entre aproximadamente 1/16 y aproximadamente 4 (la medición puede ser por un volumen o una longitud media de peso). En algunas realizaciones, el equipo usado para reducir el tamaño de partícula es un molidor con martillo o una trituradora. Después de la reducción de tamaño, la materia prima se licua en agua, ya que esto facilita el bombeo de la materia prima. En algunas realizaciones, la materia prima lignocelulósica de tamaño de partícula inferior a aproximadamente 6 pulgadas no requiere reducción de tamaño. La biomasa puede opcionalmente pre-tratarse para aumentar la

susceptibilidad de la celulosa a hidrólisis mediante pre-tratamientos químicos, físicos y biológicos (como explosión de vapor, reducción a pulpa, pulverización, hidrólisis ácida, exposición a disolventes y similares, así como combinaciones de los mismos).

En algunas realizaciones, la materia prima se licua antes del pre-tratamiento. En algunas realizaciones, la consistencia del lodo de materia prima es entre aproximadamente 2% y aproximadamente 3% y más típicamente entre aproximadamente 4% y aproximadamente 5%. En algunas realizaciones, el lodo se somete a una operación de remojo en agua y/o ácido antes del pre-tratamiento. En algunas realizaciones, el lodo se desagua usando cualquier método adecuado para reducir el uso de vapor o químicos antes del pre-tratamiento. Ejemplos de dispositivos para desaguar incluyen, aunque no se limitan a, prensas presurizadas con tornillos (Véase, por ejemplo WO 2010/022511), filtros presurizados o extrusoras.

Como aquí se usan, los términos "materia prima lignocelulósica pre-tratada" y "lignocelulosa pre-tratada" se refieren a materias primas lignocelulósicas que se han sometido a procesos físicos y/o químicos para hacer la fibra más accesible y/o receptiva a las acciones de las enzima celulolíticas. Así, "materia prima lignocelulósica pre-tratada" es un ejemplo de un "sustrato celulósico pre-tratado". En algunas realizaciones, el pre-tratamiento se realiza para hidrolizar hemicelulosa y/o una parte de la misma presente en la materia lignocelulósica a pentosa monomérica y azúcares de hexosa (por ejemplo, xilosa, arabinosa, manosa, galactosa y/o cualquier combinación de las mismas). En algunas realizaciones, el pre-tratamiento se realiza para que ocurra casi una hidrólisis completa de la hemicelulosa y una pequeña cantidad de conversión de celulosa o glucosa. En algunas realizaciones, típicamente se usa una concentración ácida en el lodo acuoso de desde aproximadamente 0,2% (p/p) a aproximadamente 2% (p/p) o cualquier cantidad intermedia, para el tratamiento de la materia prima lignocelulósica. Cualquier ácido adecuado encuentra uso en estos métodos, incluyendo aunque sin limitar a, ácido hidrocórico, ácido nítrico y/o ácido sulfúrico. En algunas realizaciones, el ácido usado durante el pre-tratamiento es ácido sulfúrico. La explosión en vapor es un método de realizar el pre-tratamiento con ácido de materia prima (Véase, por ejemplo, patente de Estados Unidos N° 4.461.648). Otro método para pre-tratar el lodo de materia prima incluye un pre-tratamiento continuo (esto es, la materia prima lignocelulósica se bombea a través de un reactor continuamente). Estos métodos son bien conocidos por aquellos expertos en la técnica (Véase, por ejemplo, patente de Estados Unidos N° 7.754.457).

En algunas realizaciones, se usa álcali en el pre-tratamiento. A diferencia del pre-tratamiento con ácido, el pre-tratamiento con álcali puede no hidrolizar el componente de hemicelulosa de la materia prima. Más bien el álcali reacciona con grupos ácidos presentes en hemicelulosa para abrir la superficie del sustrato. En algunas realizaciones, la adición de álcali altera la estructura de cristal de la celulosa para que sea más susceptible para hidrólisis. Los ejemplos de álcali que encuentran uso en el pre-tratamiento incluyen, aunque no se limitan a, amoníaco, hidróxido de amonio, hidróxido de potasio e hidróxido de sodio. Un método de pre-tratamiento con álcali es explosión de suspensión de amoníaco, explosión de fibra de amoníaco o expansión de fibra de amoníaco (Proceso "AFEX"; Véase, por ejemplo, Patentes de Estados Unidos Números 5.171.592; 5.037.663; 4.600.590; 6.106.888; 4.356.196; 5.939.544; 6.176.176; 5.037.663 y 5.171.592). Durante este proceso, la materia prima lignocelulósica contacta con amoníaco o hidróxido de amonio en un recipiente de presión durante un tiempo suficiente que permita al amoníaco o hidróxido de amonio alterar la estructura de cristal de las fibras de celulosa. Después la presión se reduce rápidamente, lo que permite que el amoníaco destelle o caliente y explote la estructura de fibra de celulosa. En algunas realizaciones, el amoníaco destellado después se recupera usando métodos conocidos en la técnica. En un método alternativo, se utiliza un pre-tratamiento con amoníaco diluido. El método con el pre-tratamiento de amoníaco diluido utiliza más soluciones diluidas de amoníaco o hidróxido de amonio que AFEX (Véase, por ejemplo, WO 2009/045651 y US 2007/0031953). Este proceso de pre-tratamiento puede o no producir monosacáridos.

Un proceso adicional de pre-tratamiento para su uso en la presente invención incluye tratamiento químico de la materia prima con disolventes orgánicos, en métodos tales como aquellos que utilizan líquidos orgánicos en sistemas de pre-tratamiento (Véase, por ejemplo, patente de Estados Unidos N° 4.556.430). Estos métodos tienen la ventaja de que los líquidos de punto de ebullición bajo pueden recuperarse y reusarse fácilmente. Otros pre-tratamientos, como el proceso Organosolv™, también usa líquidos orgánicos (Véase, por ejemplo, patente de Estados Unidos N° 7.465.791). Someter la materia prima a agua presurizada también puede ser un método adecuado de pre-tratamiento (Véase, por ejemplo, Weil et al., (1997) Appl. Biochem. Biotechnol. 68(1-2): 21-40).

En algunas realizaciones, la materia prima lignocelulósica pre-tratada se procesa después del pre-tratamiento mediante cualquiera de varias etapas, como dilución con agua, lavado con agua, amortiguación, filtración o centrifugación o cualquier combinación de estos procesos, antes de hidrólisis enzimática, como será familiar para aquellos expertos en la técnica. El pre-tratamiento produce una composición de materia prima pre-tratada (por ejemplo, un "lodo de materia prima pre-tratada") que contiene un componente soluble que incluye azúcares resultantes de la hidrólisis de la hemicelulosa, opcionalmente ácido acético y otros inhibidores, y sólidos que incluyen materia prima y lignina no hidrolizada. En algunas realizaciones, los componentes solubles de la composición de materia prima pre-tratada se separan de los sólidos para producir una fracción soluble. En algunas realizaciones, la fracción soluble, incluyendo azúcares liberados durante el pre-tratamiento y otros componentes solubles (por ejemplo, inhibidores), se envía después a fermentación. Sin embargo, en algunas realizaciones donde la hemicelulosa no se hidroliza de manera efectiva durante el pre-tratamiento, se incluyen una o más etapas

adicionales (por ejemplo, una etapa adicional de hidrólisis y/o etapas de tratamiento enzimático y/o tratamiento adicional con álcali y/o ácido) para producir azúcares fermentables. En algunas realizaciones, la separación se realiza lavando la composición de materia prima pre-tratada con una solución acuosa para producir un chorro de lavado y un chorro de sólidos que comprende la materia prima pre-tratada no hidrolizada.

Alternativamente, el componente soluble se separa de los sólidos sometiendo a la composición de materia prima pre-tratada a una separación sólidos-líquido, usando cualquier método adecuado (por ejemplo, centrifugación, microfiltración, filtración de placas, filtración de flujo cruzado, filtración de presión, filtración de vacío, etc.). Opcionalmente, en algunas realizaciones, se incorpora una etapa de lavado en la separación sólidos-líquidos. En algunas realizaciones, los sólidos separados que contienen celulosa sufren después hidrólisis enzimática con enzimas de celulasa con el fin de convertir la celulosa en glucosa. En algunas realizaciones, la composición de materia prima pre-tratada se introduce en el proceso de fermentación sin separación de los sólidos contenidos. En algunas realizaciones, los sólidos no hidrolizados se someten a hidrólisis enzimática con enzimas de celulasa para convertir la celulosa en glucosa después del proceso de fermentación. La lignocelulosa pre-tratada se somete a hidrólisis enzimática con enzimas de celulasa.

El pre-tratamiento produce una composición de materia prima pre-tratada (por ejemplo, lodo de materia prima pre-tratada) que contiene un componente soluble que incluye los azúcares resultantes de la hidrólisis de hemicelulosa, opcionalmente ácido acético y otros inhibidores, y sólidos que incluyen materia prima y lignina no hidrolizada.

Los componentes solubles de la composición de materia prima pre-tratada pueden separarse de los sólidos para producir una fracción soluble para su uso en la reacción de sacarificación.

La separación puede realizarse lavando la composición de materia prima pre-tratada con una solución acuosa para producir un chorro de lavado, y un chorro de sólidos que comprende materia prima pre-tratada no hidrolizada. Alternativamente, el componente soluble se separa de los sólidos sometiendo a la composición de materia prima pre-tratada a una separación sólidos-líquido, usando métodos como centrifugación, microfiltración, filtración de placas, filtración de flujo cruzado, filtración de presión y/o filtración de vacío. Opcionalmente, puede incorporarse una etapa de lavado en la separación sólidos-líquidos. Los sólidos separados que contienen celulosa pueden después someterse a hidrólisis enzimática con enzimas de celulasa para conversión de celulosa a glucosa.

La lignocelulosa adecuadamente preparada puede someterse a hidrólisis enzimática usando una o más enzimas de celulasa en presencia de una o más proteínas GH61 o preparaciones de acuerdo con esta invención.

La hidrólisis de componentes de hemicelulosa y celulosa de una materia prima lignocelulosa produce un hidrolizado lignocelulósico que comprende xilosa y glucosa. Otros azúcares típicamente presentes incluyen galactosa, manosa, arabinosa, fucosa, ramnosa o una combinación de ellos. Con independencia de los medios para hidrolizar la materia prima lignocelulósica (hidrólisis ácida completa o pre-tratamiento químico con o sin posterior hidrólisis enzimática), la xilosa y glucosa forman generalmente una gran proporción de los azúcares presentes.

Si el hidrolizado lignocelulósico es un hidrolizado de hemicelulosa resultante del pre-tratamiento ácido, xilosa será el azúcar predominante y menores cantidades de glucosa estarán presentes, porque una cantidad modesta de hidrólisis de celulosa típicamente ocurre durante el pre-tratamiento. En este caso, la xilosa puede formar entre aproximadamente 50 y aproximadamente 90% de peso del contenido total de carbohidrato del hidrolizado de celulosa. Si el hidrolizado lignocelulósico resulta del pre-tratamiento secuencial y la hidrólisis enzimática de la materia prima lignocelulósica (esto es, sin una etapa de separación de sólidos después del pre-tratamiento), la xilosa puede formar entre aproximadamente 30 y aproximadamente 50% de peso del contenido total de carbohidrato. La cantidad relativa de xilosa presente en el hidrolizado lignocelulósico dependerá de la materia prima y el pre-tratamiento que se emplee.

Los componentes solubles del sustrato hidrolizado pueden separarse de los sólidos para producir una fracción soluble. La fracción soluble (incluyendo azúcares liberados durante hidrólisis, y algunas veces inhibidores) pueden después usarse para fermentación. Si la hemicelulosa no se hidroliza de manera efectiva durante el pre-tratamiento, puede ser deseable incluir una etapa adicional de hidrólisis o etapas con enzimas o un tratamiento adicional con álcali o ácido para producir azúcares fermentables.

XIV: Fermentación de azúcares

Los azúcares fermentables producidos en las reacciones de sacarificación usando proteínas GH61 de la divulgación pueden usarse para producir varios productos finales de interés.

En algunas realizaciones, los azúcares se usan en un proceso de fermentación para producir productos finales. El término "fermentación" se usa ampliamente para referirse al cultivo de un microorganismo o un cultivo de microorganismos que usan azúcares simples, tales como azúcares fermentables, como una fuente de energía para

obtener un producto deseado. En una realización diferente, la biomasa celulósica puede tratarse con una composición de esta invención para preparar comida para animales.

Los productos finales incluyen alcoholes (por ejemplo, etanol, butanol), acetona, aminoácidos (por ejemplo, glicina y lisina), ácidos orgánicos (por ejemplo, ácido láctico, ácido acético, ácido fórmico, ácido cítrico, ácido oxálico, ácido 3-hidroxipropinoico, ácido acrílico, ácido succínico, ácido málico, ácido fumárico o ácido úrico), glicerol, dioles (como 1,3 propanodiol o butanodiol), hidrocarburos con 1-20 átomos de carbono (por ejemplo, ésteres de cadena larga), alcoholes de azúcar (por ejemplo, xilitol), alcoholes grasos y β -lactamo y otros productos finales.

Puede usarse cualquier microorganismo para convertir azúcar en el hidrolizado de azúcar a etanol u otros productos de fermentación. Estos incluyen levaduras del género *Saccharomyces*, *Hansenula*, *Pichia*, *Kluyveromyces* y *Candida*. Pueden usarse levaduras comercialmente disponibles, tales como levadura Turbo, Ethanol Red® Safdistil®, Thermosacc®, Fermiol®, Fermivin® o Superstart™.

La levadura puede producirse genéticamente para azúcares de hexosa y pentosa a un producto final, incluyendo aunque sin limitar etanol. Alternativamente, la levadura puede ser una cepa que haya sido capaz de fermentar xilosa y glucosa mediante uno o más métodos no recombinantes, como evolución adaptiva o mutagénesis y selección aleatoria. Por ejemplo, la fermentación puede realizarse con levadura *Saccharomyces* recombinantes. La levadura recombinante es una cepa que ha sido capaz de fermentar xilosa mediante incorporación recombinantes de genes que codifican xilosa reductasa (XR) y xilitol dehidrogenasa (XDH) (Patentes de Estados Unidos 5.789.210, 5.866.382, 6.582.944 y 7.527.927 y EP 450 530), y/o genes que codifican una o más xilosa isomerasa (XI) (Patentes de Estados Unidos 6.475.768 y 7.622.284). Además, la cepa de levadura modificada puede sobreexpresar un gen endógeno o heterólogo que codifica xilquinasa (XK): otra levadura puede fermentar azúcares de hexosa y pentosa a al menos un producto final, incluyendo aunque sin limitar, etanol, como levadura de los géneros *Hansenula*, *Pichia*, *Kluyveromyces* y *Candida* (WO 2008/130603).

Un rango típico de temperatura para la fermentación de una azúcar a etanol usando *Saccharomyces* spp. está entre aproximadamente 25 °C y aproximadamente 37 °C, aunque la temperatura puede ser más alta (hasta 55 °C) si la levadura se ha modificado naturalmente o genéticamente para ser termoestable. El pH de una fermentación típica que emplea *Saccharomyces* spp. está entre aproximadamente 3 y aproximadamente 6, dependiendo del pH óptimo del microorganismo de fermentación. El azúcar hidrolizado también puede complementarse con nutrientes adicionales necesarios para el crecimiento y fermentación del microorganismo para fermentación. Por ejemplo, extracto de levadura, aminoácidos específicos, fosfato, fuentes de nitrógeno, sales, elementos de trazas y vitaminas (Verduyn et al., 1992, Yeast 8(7):501-170, Jørgensen, 2009, Appl Biochem Biotechnol, 153:44-57 and Zhao et al., 2009, Journal of Biotechnology, 139:55-60). Típicamente, la fermentación se realiza bajo condiciones anaeróbicas, aunque también pueden usarse condiciones aeróbicas o microaeróbicas.

Así, la invención proporciona procesos para producir un producto de fermentación, donde el método comprende: proporcionar las células huéspedes recombinantes como aquí se proporcionan, un medio de fermentación que comprende azúcares fermentables como glucosa y/o xilosa; y contactar el medio de fermentación con las células huéspedes fúngicas recombinantes bajo condiciones adecuadas para generar el producto de fermentación. En algunas realizaciones, los procesos comprenden además la etapa de recuperar el producto de fermentación. En algunas realizaciones adicionales, la etapa de fermentación se realiza bajo condiciones microaeróbicas o aeróbicas. En algunas realizaciones, la etapa de fermentación se realiza bajo condiciones anaeróbicas. En algunas realizaciones, el medio de fermentación comprende producto de un proceso de sacarificación.

Las proteínas GH61 y celulasas de la presente divulgación pueden utilizarse en cualquier método usado para generar alcoholes u otros biocombustibles a partir de celulosa, y no se limitan necesariamente a los aquí descritos. Dos métodos comúnmente empleados son el método de la sacarificación separada y fermentación (SHF) (véase, Wilke et al., Biotechnol. Bioengin. 6: 155-75 (1976)) o el método de sacarificación simultánea y fermentación (SSF) desvelado por ejemplo en las patentes de Estados Unidos N° 3.990.944 y 3.990.945.

El método SHF de sacarificación de la presente invención comprende las etapas de contactar una celulasa de proteína GH61 con una celulasa que contiene sustrato para romper enzimáticamente la celulasa en azúcares fermentables (por ejemplo, monosacáridos como glucosa), contactar los azúcares de fermentación con un microorganismo productor de alcohol para producir alcohol (por ejemplo, etanol o butanol) y recuperar el alcohol. En algunas realizaciones, puede usarse el método de bioprocesamiento consolidado (CBP), donde la producción de celulasa del huésped es simultánea a la sacarificación y fermentación, bien de un huésped o de un cultivo mezclado.

Además de los métodos SHF, puede usarse un método SSF. En algunos casos, los métodos SSF dan como resultado una mayor eficiencia de producción de alcohol que el que produce el método SHF (Drissen et al., Biocatalysis and Biotransformation 27: 27-35 (2009)). Una desventaja de SSF sobre SHF es que se necesitan temperaturas más altas para SSF que para SHF.

En una realización, la presente invención usa polipéptidos de celulasa que tienen mayor termoestabilidad que una celulasa de tipo salvaje.

EJEMPLOS

5

Ejemplo 1. Identificación de proteínas GH61 en *M. thermophila*

10

Se obtuvo la secuencia genómica de una cepa fúngica de tipo salvaje de *M. thermophila*. Se analizó el genoma entero para identificar y evaluar regiones codificadoras de proteína. Se seleccionaron veinticuatro proteínas endógenas para la cepa *M. thermophila* en base a factores que incluyen la presencia de motivos de secuencia de familia 61 de glicohidrolasa (GH61) (dominios Pfam). Los dominios Pfam se identificaron usando el algoritmo software "PFAM V.24", desarrollado por Wellcome Trust Sanger Institute (Henrissat et al., 1995, Proc Natl Acad Sci USA 92: 7090-94).

15

La TABLA 1 proporciona la secuencia de pre-proteína GH61 (SEQ ID NO: 1) y la forma secretada predicha (madura) (SEQ ID NO: 2). La proteína madura se designó "GH61a". La TABLA 2 proporciona la secuencia de otras pre-proteínas GH61, con el péptido de señal nativo predicho subrayado. Dos de las proteínas en la TABLA 2 no se predicen por tener péptidos de señal.

20

En la TABLA 3, más abajo, se muestra la numeración secuencial y el análisis de estructura de dominio de los polinucleótidos correspondientes.

TABLA 3: Numeración de secuencia y análisis de dominio

Designación de laboratorio	Proteína SEQ ID NO	Ácido nucleico SEQ ID NO	Proteína PFAM Dominio
GH61a	1	31	GH61—CBM_1
GH61l	3	32	Chitin_bind_3--GH61
GH61m	4	33	GH61
GH61n	5	34	GH61
GH61o	6	35	GH61—GH61-CBM_1
GH61p	7	36	GH61- GH61
GH61q	8	37	GH61
GH61r	9	38	GH61
GH61s	10	39	GH61
GH61t	11	40	GH61
GH61u	12	41	GH61
GH61v	13	42	GH61
GH61w	14	43	GH61
GH61x	15	44	GH61
GH61b	16	45	GH61
GH61c	17	46	GH61
GH61d	18	47	GH61
GH61e	19	48	GH61
GH61f	20	49	GH61—CBM_1
GH61g	21	50	GH61—CBM_1
GH61h	22	51	GH61
GH61i	23	52	GH61
GH61j	24	53	GH61
GH61k	25	54	GH61
GH61p2	26	55	GH61-- GH61
GH61q2	27	56	GH61
GH61r2	28	57	GH61
GH61t2	29	58	GH61
GH61e2	30	59	GH61

SEQ ID NO: 7 tiene un segundo dominio GH61 (GH61-GH61), SEQ ID Nos: 1, 20, 21 tienen la estructura GH61CBM1 (donde "CBM1" es un módulo 1 que se enlaza a carbohidrato), SEQ ID NO: 6 tiene la estructura GH61-GH61-CBM1, SEQ ID Nos: 4, 5, 8-19 y 22-25 tienen la estructura GH61, SEQ ID NO: 3 tiene la estructura Chitin_bind_3-GH61 (donde "Chitin_bind_3" es módulo 3 que se enlaza con quitina). SEQ ID Nos: 26-30 son secuencias alternativas correspondientes a los genes que codifican SEQ ID Nos: 7-9, 11 y 19, respectivamente.

Ejemplo 2. Expresión recombinantes de proteínas GH61

Entre las proteínas GH61 identificadas en el Ejemplo 1, se seleccionaron ciertos ejemplos para expresión en base a la estructura predicha y a aspectos funcionales, como si la proteína se secretaría de la célula, y su estructura de dominio.

Las seis proteínas GH61 enumeradas en la siguiente tabla se clonaron individualmente en un vector de expresión bajo el control de un promotor CHI (constitutivo a la célula diana) y se transformaron en una cepa de *M. thermophila* designada "CF-405" que se ha adaptado para ser deficiente en producción de celulasas endógenas.

TABLA 4: Proteínas GH61 expresadas recombinantemente		
	Designación de laboratorio	Proteína SEQ ID NO
	GH61a	2
	GH61o	6
	GH61v	13
	GH61x	15
	GH61b	16
	GH61e	19, 30

Las células transformadas que expresaban la proteína GH61 recombinantes se seleccionaron y se prepararon cultivos sembrados. Las células de progenie se cultivaron durante 5 días, y se recogió el caldo que contenía la proteína GH61 secretada ("caldo GH61").

Ejemplo 3. Actividad potenciadora de celulasa de proteína GH61

Se recogió el caldo de células que expresaban la proteína GH61 que comprendía SEQ ID NO: 2 (Ejemplo 2), y el nivel de GH61 recombinante se cuantificó mediante SDS-PAGE. Se realizaron ensayos de celulasa para determinar si la adición de la proteína GH61 mejoró la hidrólisis de material celulósico por las enzimas de celulasa.

El caldo de cultivo se recogió de un cultivo de una cepa de *M. thermophila* designado "CF-402" que sobreexpresa β -glucosidasa. Los ensayos de digestión de celulasa se realizaron usando el caldo con y sin GH61 añadido. Las reacciones se hicieron en 96 placas de pocillo profundo Costar en un volumen de reacción total de aproximadamente 80 microlitros. Las reacciones se realizaron a 55 °C usando preparaciones del sustrato que contenía celulosa de paja de trigo que se había pre-tratado bajo condiciones ácidas (de ahora en adelante referido como "paja de trigo pre-tratada"). Se añadieron 8,1 mg de proteína de caldo por gramo de sustrato (0,82%) a cada muestra. Además, las muestras tuvieron 0 (control), 0,22%, 0,44% o 0,66% proteína de caldo GH61 añadida (concentración total final de proteína de caldo 0,81%, 1,03%, 1,25% o 1,47%). En los pocillos de control sin GH61 añadido, se usó agua para ajustar el volumen para que la carga de sustrato fuera igual en todos los pocillos. En otros experimentos, se añadieron 6-10 mg de proteína de caldo por gramo de sustrato (0,6-1%) a cada muestra, a una concentración total final de proteína de caldo de 0,6-1,7%.

La FIGURA 1 muestra producción adicional de glucosa sobre el control (caldo sin GH61 añadida) después de 48 horas de incubación. En la FIGURA 1(A), se muestra el porcentaje de producción mejorada sobre el control. En la FIGURA 1(B), se trazan los datos para mostrar la producción total de glucosa.

Ejemplo 4. Sobreexpresión de GH61a en una cepa CXP

Construcción de genes GH61a recombinantes

Se aisló ADN genómico de la cepa *M. thermophila* designada "CF-409". Esta cepa produce endógenamente endoglucanasa, β -glucosidasas, celobiohidrolasa Tipo 1, y celobiohidrolasa Tipo 2. El procedimiento fue de la siguiente manera. El inóculo de las hijas se sembró en un medio de cultivo y se dejó crecer durante 72 horas a 35 °C. La alfombrilla micelar se recogió mediante centrifugación, se lavó y se añadieron 50 μ L de tampón de extracción de ADN (200 mM Tris pH 8.0; 250 mM NaCl; 125 mM EDTA; .5% SDS). Los micelios se molieron con un molidor cónico, se re-extrajeron con 250 μ L de tampón de extracción y la suspensión se centrifugó. El sobrenadante se transfirió a un tubo nuevo que contenía 300 μ L isopropanol. El ADN se recogió mediante centrifugación, se lavó dos veces con 70% etanol y se volvió a disolver en 100 μ L de agua desionizada.

La secuencia GH61a ADN se amplificó de células CF-409 usando cebadores PchiC1GH61a_F y TcbhC1GH61a_R. Se realizó reacción PCR usando la Polimerasa Phusion, a 98 °C durante 30", seguido de 35 ciclos de 98 °C durante 10", 72 °C durante 1' y la extensión final a 72 °C durante 5'. El producto resultante se clonó en el vector 3' pC1DX20PhR al promotor chi1 para crear un vector de expresión que expresó la transcripción de proteína GH61 bajo el control del promotor chi1 (pC1DX20PhR-GH61a) usando técnica de clonación In-fusion (kit de clonación In-Fusion Advantage™ PCR con potenciador clonador, Clontech Cat. No. 639617 de acuerdo con las instrucciones del fabricante).

PchiC1GH61a_F tactcttctccaccATGTCCAAGGCCTCTGCTCT SEQ ID NO:69

TcbhC1GH61a_R ggatccgaattcttaTTACAAACTGGGAGTACCA SEQ ID NO:70

5 Preparación de protoplasto para transformación CXP

Las células *M. thermophila* ("CF-405", una cepa eliminada Alp1) se inocularon en 100 mL de medio de crecimiento en un matraz Erlenmeyer de 500 mL usando 10^6 esporas/mL. El cultivo se incubó durante 24 horas a 35 °C, 250 rpm. Para cosechar el micelio, el cultivo se filtró sobre un filtro estéril Myracloth™ (Calbiochem) y se lavó con 100 mL solución 1700 mosmol NaCl/CaCl₂ (0.6 M NaCl, 0.27 M CaCl₂·H₂O). El micelio lavado se transfirió a un tubo de 50 mL y se pesó. Se disolvió cailasa (20 mg/gramo micelio) en 1700 mosmol NaCl/CaCl₂ y se esterilizó con UV durante 90 segundos. Se añadieron 3 mL de solución estéril Caylase en el micelio lavado que contenía el tubo y se mezcló. Se añadieron 15 mL adicionales de solución 1700 mosmol NaCl/CaCl₂ en el tubo y se mezcló.

La suspensión micelio/Caylase se incubó a 30 °C, 70 rpm durante 2 horas. Los protoplastos se cosecharon filtrándolos a través de un filtro estéril Myracloth en un tubo estéril de 50 mL. Se añadieron 25 mL de STC frío al flujo y giró a 2720 rpm durante 10 min a 4 °C. El gránulo se volvió a suspender en 50 mL sTC (1.2 M sorbitol, 50 mM CaCl₂·H₂O, 35 mM NaCl, 10 mM Tris-HCl) y se centrifugó de nuevo. Después de las etapas de lavado, el gránulo se volvió a suspender en 1 mL STC.

20 Transformación

En el fondo de un tubo estéril de 5 mL se colocó en una pipeta ADN plásmido y se añadieron 1 µL de ácido aointricarboxílico y 100 µL de protoplasto. El contenido se mezcló y el protoplasto con el ADN se incubó a temperatura ambiente durante 24 min. Se añadieron 1,7 mL de solución PEG4000 (60% PEG4000 [glicol de polietileno, peso medio molecular 4000 Daltons], 50 mM CaCl₂·H₂O, 35 mM NaCl, 10 mM Tris-HCl) y se mezclaron cuidadosamente. La solución se mantuvo a temperatura ambiente durante 20 min. El tubo se llenó con STC, se mezcló y centrifugó a 2500 rpm durante 10 min a 4 °C. STC se vertió y el gránulo se volvió a suspender en el resto de STC y se colocó en placas en placas de medio mínimo. Las placas se incubaron durante 5 días a 35 °C. Las colonias se volvieron a manchar y se comprobaron para presencia de plásmido integrado. Varias aislados se seleccionaron y probaron para la expresión de GH61a.

La transformación se realizó en cepas CF405. Los transformantes de GH61a se probaron para sobreexpresión de GH61 en SDS-PAGE para la presencia de la banda extra que se confirmó mediante análisis MS-MS.

Ejemplo 5. Purificación de proteínas GH61 de sobrenadante de celulasa

En este experimento, la actividad de proteína GH61 se fraccionó del sobrenadante del cultivo de una cepa de *M. thermophila* designada "CF-401". CF-401 se derivó de CDXF que tiene una eliminación de genes CDH1 y CDH2. Esta cepa produce endógenamente endoglucanasa, β-glucosidasas, celobiohidrolasa Tipo 1, y celobiohidrolasa Tipo 2.

Para prepara para cromatografía, el sobrenadante completo de celulasa de CF-401 se aclaró mediante centrifugación a 12.000 rpm durante 35 minutos. El sobrenadante se filtró después a través de una membrana PES™ de 0,22 µm, que después se concentró y se intercambió con tampón en un tampón 25 mM bis-tris, pH 5,7. Se añadió sulfato de amonio saturado en 25 mM bis-tris a una concentración final de ≈50 g/L proteína en 0,9 M sulfato de amonio y 20-25 mM bis-tris. Esta solución se cargó inmediatamente en una columna FF de Fenilo (High Sub) (una columna de flujo rápido empaquetada con Phenyl-Sepharose™) y se enjuagó con 0,9 M sulfato de amonio en 25 mM bis-tris hasta que el A₂₈₀ cayó cerca del punto de partida. La proteína se eluyó con un gradiente de 0,9 M sulfato de amonio en 25 mM bis-tris a 0 M sulfato de amonio en 25 mM bis-tris, sobre aproximadamente 10 volúmenes de columna. Las fracciones se recogieron (25 en este caso) de acuerdo con los picos del cromatograma (A₂₈₀). Para inhibir el crecimiento simultaneo, se añadió NaN₃ a todas las fracciones hasta una concentración final de 0,05%. Después de mantener un gel SDS-PAGE en cada fracción, las fracciones similares se combinaron para crear grupos con un mínimo conjunto de componentes. Aquí, de las 25 fracciones producidas, resultaron 11 grupos. Para preparar la siguiente columna, cada grupo se concentró hasta ≈150 mL usando una unidad de filtración de flujo tangencial equipada con membranas 50 kDa MWCO PES™. El volumen de la solución de proteína se devolvió después a 500 mL con DI agua, y se concentró de nuevo a ≈150 mL; esta etapa se repitió 5 x y la solución se sometió a intercambio de tampón de manera efectiva.

Cada uno de los 11 grupos derivados de CF-401 se fraccionó además con una columna Q (resina de intercambio iónico de amonio cuaternario). Después de la cuantificación de proteína total usando un ensayo BCA (ácido bicinonínico), se prepararon soluciones de 50 g/L proteína en 10 mM bis-tris pH 7,5 tampón para cada grupo. Después de su aplicación a la columna, la resina se lavó con 10 mM bis-tris, pH 7,5 hasta que A₂₈₀ cayó cerca del punto de partida. La proteína unida se eluyó después con 1 M NaCl en 10 mM Bis-Tris, pH 7,5 usando un gradiente gradual (5% durante 10 minutos), y se mantuvo en esa concentración hasta que el pico de proteína

comenzó a bajar continuamente. Para analizar, se mantuvo un gel SDS-PAGE en cada fracción, y las fracciones con composiciones similares se agruparon. Cada una de estos grupos de segunda fase después se desalaron/concentraron con filtración de flujo tangencial.

En total, 79 grupos de enzimas derivadas de CF-401 se prepararon en este ensayo. Para ayudar en la carga en el ensayo de sacarificación, cada proteína total para cada grupo se midió usando el ensayo de proteína BCA.

Ejemplo 6. Uso de proteínas GH61 para promover la sacarificación

Las enzimas GH61 fraccionadas de acuerdo con el ejemplo previo se probaron en sacarificación en presencia de sistemas de celulasa completa obtenidas a partir de una cepa de célula designada "CF-404". Estas células se derivaron de la cepa CF-401 y sobreexpresaron β -glucosidasa. Como un experimento de control, 0,21% (generalmente 0,1-4%) de CF-404 se añadió a 0,61% (generalmente 0,6-1,5%) de CF-401. Así, la carga de proteína total fue igual en todas las reacciones. Estas mezclas después se incubaron con 110 g/L glucano (paja de trigo pre-tratada) a 55 °C, pH 4,6-5 durante 53 horas. Al finalizar el experimento, las reacciones se enfriaron mediante la adición de 10 mM de ácido sulfúrico. Para análisis de glucosa, las muestras se analizaron usando un Agilent HPLC 1200 equipado con columna de exclusión iónico HPX-87H (300 mM x 7,8 mM) con 5 mM H₂SO₄ como una fase móvil a una velocidad de flujo de 0,6 ml/min a 65 °C. El tiempo de retención de la glucosa fue 9,1 minutos.

Un ejemplo que demuestra una producción mejorada de glucosa se muestra en la Tabla 5.

TABLA 5: Sacarificación mejorada usando fracciones de GH61		
0,21% GH61 complementando 0,61% CF-404	Glucosa producida (g/L)	SEQ ID NO correspondiente
GH61	>30	20
GH61	>35	2
GH61	>35	13
GH61	>30	7
GH61	>30	21, 26
GH61	>35	23
Control (CF-404)	28,9 $\pm$ 0,66	

La secuencia de proteína parcial se obtuvo a partir de espectrometría de masa y se comparó con las secuencias que codifican proteínas identificadas en la secuencia del genoma de *M. thermophila* (Ejemplo 1) usando el software de alineación BIFX disponible en Bioinformatics Organization, Hudson MA. La concordancia con las secuencias conocidas de *M. thermophila* se muestra en la tabla anterior.

Ejemplo 7. Combinación mínima de proteína para convertir celulosa en glucosa

Las enzimas de *M. thermophila*, CBH1 CBH2 se combinaron con varias combinaciones de las proteínas GH61; GH61a, GH61f y GH61p, y se ensayaron en varias concentraciones relativas para la habilidad para convertir glucano en glucosa. El sobrenadante de la cepa CF-401 (que comprende celulasas y proteínas GH61) también se ensayaron para comparación. Para 110 g/L glucano, las producciones máximas posibles de glucosa son aproximadamente de 135 a 145 g/L:

Las reacciones de sacarificación se realizaron en 110 g/L carga de glucano de paja de trigo pre-tratada en pH 5,0 a una temperatura de 55 °C a 950 rpm en un volumen total de 95 μ L en 96 placas de pocillos profundos de alto rendimiento (HTP). Se añadieron 81 g/L xilosa y 128 mM acetato a la paja de trigo pre-tratada. β -glucosidasa en exceso (en relación con glucano) también se complementó para liberar la inhibición del producto de celobiosa. La celulasa completa (caldo de células CP-401), así como enzimas individuales se caracterizaron por ensayos estándares BCA para cuantificación total de proteína. Las respuestas a dosis de las mezclas de enzimas se realizaron añadiendo proteína conocida total (calculada como peso de proteína añadida / peso glucano). Los niveles de proteína total probados fueron 0,73 y 3% (peso proteína añadida / peso glucano). Una respuesta a dosis se midió en pH 5,0 a 55 °C durante 72 horas. Se usaron las siguientes combinaciones de enzimas (en combinación con BGL1):

a. Cultivo sobrenadante CF-401 (Control; contiene las 4 enzimas + proteínas GH61)

- b. 50%(CBH1a+ CBH2b) + 50% GH61f
 c. 50%(CBH1a+ CBH2b) + 50% GH61p
 d. 50%(CBH1a+ CBH2b) + 25% GH61f + 25%GH61p
 e. 50%(CBH1a+ CBH2b) + 16.6% GH61a+16.6% GH61f+16.6% GH61p

Las reacciones se templaron a las 72 horas mediante la adición de 10 mM de ácido sulfúrico. Para análisis de glucosa, las muestras se analizaron usando un Agilent HPLC 1200 equipado con columna de exclusión iónica HPX-87H (300 mM x 7,8 mM) con 5 mM H₂SO₄ como una fase móvil a una velocidad de flujo de 0,6 mL/min a 65 °C. El tiempo de retención de la glucosa fue 9,1 minutos.

La FIGURA 2 muestra los resultados. Las enzimas de celulasa CBH1 y CBH2 se combinaron con varias proteínas GH61. El experimento control (línea de puntos) fue cultivo sobrenadante de células CF-401 que contiene una pluralidad de enzimas de celulasa y actividad endógena de GH61. La carga total de proteína se añadió en las proporciones especificadas en la figura.

Estos resultados establecen que una enzima mínima fijada es capaz de conseguir niveles de glucosa similares al control: específicamente, las celulasas CBH1 y CBH2, combinadas con una o más de GH61f, GH61p y GH61a. Todas las combinaciones generaron producciones de glucosa más altas que el caldo de cultivo de CP-401. Así, es posible conseguir las máximas conversiones teóricas usando un conjunto mínimo de enzimas. Esto también demuestra que la mezcla mínima de enzimas aquí usada tiene todos los componentes necesarios para la conversión completa de celulosa a glucosa. Sin embargo, se contempla que las combinaciones adicionales de enzimas encontrarán uso en los procesos de sacarificación.

Ejemplo 8. Sinergia de actividad de GH61 con otras enzimas derivadas de CXP

Este ejemplo describe una evaluación de GH61a para sinergia con otras enzimas derivadas de *M. thermophila* como CBH1a, CBH2b y EG2.

Las reacciones de sacarificación se realizaron en una carga de 110 g/L glucano de paja de trigo pre-tratada en pH 5,0 a una temperatura de 55 °C a 950 rpm en un volumen total de 95 µL en 96 placas de pocillos profundos de alto rendimiento (HTP). Se añadieron 81 g/L xilosa y 128 mM acetato a la paja de trigo pre-tratada. β-glucosidasa en exceso (p/p glucano) también se complementó para liberar la inhibición del producto de celobiosa. La celulasa completa así como enzimas individuales se caracterizaron por ensayos estándares BCA para cuantificación total de proteína.

TABLA 6: Sinergia de GH61a con enzimas derivadas de *M. thermophila* y mezclas de enzimas

Enzimas complementadas con GH61a	CBH1a%	CBH2b%	CH61a%	EG2%	Carga de proteína (peso total proteína añadida /peso glucano)	Producción de glucosa (g/L)	Grado de sinergia
CBH1a	0,39%	-	0,18%	-	0,57%	>15	>1,6
CBH2b	-	0,30%	0,18%	-	0,48%	>35	>1,3
EG2	-		0,18%	0,20%	0,38%	>25	>1,2
CBH1a+CBH2b	0,39%	0,30%	0,18%	-	0,87%	>75	>1,7
CBH2b+EG2	-	0,30%	0,18%	0,20%	0,68%	>40	>1,1
CBH1a+CBH2b	0,39%	0,30%	0,18%	0,20%	1,07%	>75	>1,4
+ EG2	0,68%	-	0,37%	0,20%	1,25%	>75	>1,7
	1,35%	1,20%	0,37%	0,20%	3,12%	>125	>1,6
	1,35%	1,20%	0,74%	0,20%	3,49%	>125	>1,5

Para un sistema de enzima que comprende dos enzimas A y B, el grado de sinergia se calculó mediante la siguiente fórmula:

$$\text{Grado de Sinergia} = \frac{\text{Producción de glucosa de la combinación de GH61 y enzimas de celulosa}}{\text{Producción de glucosa de GH61} + \text{Producción de glucosa de enzimas de celulosa}}$$

La producción de glucosa mostrada en la tabla es la producción obtenida de la combinación de GH61 y las enzimas. Esto se divide por la producción de glucosa medida por separado para GH61 y la mezcla de enzimas (no mostrada) para cuantificar la sinergia entre las dos.

Los resultados muestran que GH61a tiene sinergia con todos los sistemas de enzimas derivados de *M. thermophila* probados. Para la conversión completa de celulosa a glucosa, la presencia de GH61a es beneficiosa.

Ejemplo 9. Usar GH61a para reducir viscosidad de paja de trigo pre-tratada

GH61a purificado de *M. thermophila* se evaluó para determinar la función enzimática y para evaluar cualquier actividad de tipo endo-glucanasa para reducción en longitud de cadena de celulosa, permitiendo así una reducción en viscosidad.

GH61a sola o en combinación con EG2 se probaron para reducción de viscosidad en paja de trigo pre-tratada en carga de glucano de 75 g/L glucano y en pH 5,0, 50 °C. Las pruebas de reducción de viscosidad se realizaron (ciclo de 30 minutos a 80 RPM) en un viscosímetro RVA-super4 (Newport Scientific, Australia) en un peso total de 21 g.

La FIGURA 3 muestra los resultados. La adición de 0,02% GH61a en relación con glucano mostró aproximadamente 2-3% de reducción de viscosidad en pH 5, 50 °C. En comparación, la adición de 0,01% *M. thermophila* EG2 en relación con glucano mostró aproximadamente 19% de reducción de viscosidad bajo las mismas condiciones. Con una combinación de 0,02% GH61a y 0,01% EG2, la reducción total de viscosidad fue aproximadamente 21%.

SECUENCIAS

5

Secuencias de celulasa ejemplares

10

Precursor C1 BGL1: SEQ ID NO: 60

15

MKAAALSCFLGSTLAVAGAIESRKVHQKPLARSEPFYPSPMNPNADGWAEAYQAQKSFVSQMTLLEKVNLTGTVGWGAEQCV
GQVGAIPRLGLRSLCMHDSPLGIRGADYNSAFPSGQTVAAATWDRGLMYRRGYANGQEAQKGINVLLGPVAGPLGRMPGEGRN
WEGFADPDPVLTGIGMSETIKGIQDAGVACAKHFIGNEQEHFRQVPEAQGYGYNISLSSNIDDKTMHELWLWPFADAVRAG
VGSVMCSYQOVNNSYACQNSKLLNDLKNELGFGQFVMSDWAQHTGAASAVAGLDMSPGDTQFNTGVSWGANLTLAVLNG
TVPAYRLDDMAMRIMAALFKVTKTTDLEPIFNSFWTDDTYGPIHWAQKQGYQEIINSHVDVRAHDHNLIREIAAKGTVLLKNTG
SLPLNPKPFVAVIGEDAGSSPNGPNCSDRGCEGTAMGWGSGTANYPYLVSPDAAQARAIQDGTREYVLSNYAEKTKA
LYSQANATAIVFVNADSGEGYINVDGNEGDRKNLTLWNNGDTLVKNVSSWCSNTIVVHVSVPVLLTDWYDNPNTAILWAGL
PQGESGNSITDVLGYKYNPAARSPTWGKTRESYGADVLYKPNNGNGAPQDDTEGVFIDYRYFDKVDSDSVIYEFHGHSYT
TFEYSNIRVVKSVSEYRPTTGTAGAPTGFNFSTDLLEDYLFKDEFPYIYQYIYPYLNITDPRRASADPHYGTAEELPPI
ATDDDPQPLLRSSGGSNPGGNRLYDVIYITITADITNTGVSVEEVPLVYSLGGPEDPKVQLRDFDRMREIPEGETRQFTCRL
TRRDLNNDVTVQDWISRYPKTAYVGRSSRLDLKIELP

20

Precursor TaBGL (Thermoascus aurantiacus): SEQ ID NO: 61

25

30

MRLQWLELAVAAAATVASAKDDLAISPPFYPSPMMDNGEWAEEYRRADVFSQTLAEKVNLTGTVGMQEKCVGETG5IPR
LGFRGLCLQDSPLGVRFADYVSAPFAGVNVAAATWOKNLAYLRGKAMGEEHKGKGVQVQLGPVAGPLGRHPDGGRNWIEGFS
VLTGVLMAETIKGIQDAGVACAKHFIGNEMEHFRQASEAVGYGFDITESVSSNIDDKTLHELWLWPFADAVRAGVGSFMCY
NQVNNYSYCSNLYLLNKLKSELDFOGFVMSDWAHHSVGGAALAGLDMSPGDTAFGTGKSFWGTNLTIAVLNGTVPEWRVD
DMAVRIMAAFYKVRDRYQVPVNFDSWTKDEYGYEHALVGQNVVKVNDKVDVRAHDADITRQIGSASVVLKNDGGILPLTGYE
KFTGVFGEDAGSNRWGADGCSDRGCDNGLAMGWGSGTADFPYLVTPAQAIQNEILSKGKGLVSAVTDNGALDQMEQVASQAS
VSVFVNADSGEGYINVDGNEGDRKNLTLNKGGEVYIKTVAANCNNTIVVMHTVGPVLIIDEWYDNPNTIAVWAGLPGQESGN
SLVDVLYGRVSPGGKTPTWGKTRESYGAPLLTKPNNGKGAPODDTEGVFIDYRRFDKYNETPIYEFGGLSYITFEYSYDIY
VQPLNARPYTPASGSTKAAPTFGNISTDYADLYPEDIHKVPYIYPLNITDPAKSSGDPDYGMAEDYIPSGATDGSQPI
LPAGGAPGGNPGLYDEMYRVAIITNTGNVVGDEVPQLYVSLGGPDDPKVVLNFDRIITLHPGQQTMTWTTLTRRDISNWDPA
SQNWVTKYPKTVYIGSSSRKLHLQAPLPY

35

Precursor CeiA BGL (Azospirillum irakense): SEQ ID NO: 62:

40

MGALRLLSISIVALTCGGIHASTAIAQEGAAPAILHPEKWRPRPATQRLIDPAVEKRVDAALLKQLSVEEKVQVQIGDIGTI
TPEDLRKYPLGSLAGNSGPNGGDRAPPKWLDLADAFYRVSLKRPHTPIPVLFIDAVHGHGNGISATIFPHNIALGAT
HDELLRRIGIEVTAEMAATGIDWTFAPALSVVRDRWGRTEYEGFSEDPETVAAYSAIIEGVQKFGSKDFMAPGRIVASAK
HFLADGGTDQGRDQDARISEDELIRIHNAAGYPAIDAGVLTVMAFSFSSWQGIKHHGHKQLLTDVLKGQMGFNGFIVGDWAH
DQVPGCTKFCNPTSLIAGLDMYMAADSWKQLYENTLAQYKOGTIPMARLDDAVRRLRVKVLAGLFEKPAKDRPGLPGLLETL
GSPHRAVGREAVRKSLLVLLKNKGTLLPLSPKARVLVAGDGADNIGKSGGWTSWQGTGNRNDEFFGATSIILGGIRDAVADA
GGSVEFDVAGQYKTKPDVAIVVFGEPEYAEFQGDVETLEYQPDQKQDLALLKKLKDQGIPIVAVFLSGRPMWVNPENASDAF
VAAWLPGTEGGGVADVLFTDKAGKVQHDFAKGLSYSWPRTAAQTTVNRGDADYNPLFAYGYGLTYKDKSKVGTLPFEESGVPAE
ARQNAGYFRAGALRLPGRFL

45

C1 CBH1a: SEQ ID NO: 63

QNACTLTAENHPSLTWSKCTSGGCTSVQGSITIDANWRWTHRTDSATNCYEGNKWDTSYCSGDPSCASKCCIDGADYSSITYG
ITTSGNLSNLKFTVKQYSTNIGSRTYLMESDTKYQMFQLLGNEFTFDVDSNLGCGLNGALYFVSMADGGMSKYSGNKAGA
KYGTGYCDSQCPRLDFKINGEANVENWQSSNDANAGTGKYGSCSEMDVWEANNMAAFTPHPCPTVIGQSRCEGDSGGTYS
TDYVAGICDPDGGCDFNRYQGNKTFYKGMTVDTTKITTVTQFLKNSAGELSEIKRFYVQNGKVIPNSESTIPGVEGNSITQ
DWCDRQKAAFGDVTDFQDKGGMVQMGKALAGPMVLNMSIWDHAYNMLWLDSTWPIDGAGKPGAERGACPTTSGVPAEVEAEA
PNSNIFSNIRFGPIGSTVSGLPDGGSGNPNPPVSSSTPVSSSTTSSGSSGPTGGTGVAKHIEQCGGIGFTGPTQCESPYTC
TKLNDWYSQCL

~

Precursor C1 CBH2a: SEQ ID NO: 64

60

MKFVQSATLFAAATALAAPSRTTPQKPRQASAGCASAVTLDASTNVFQQYTLHPNNFYRAEVEAAAAEATSDSALAEKARKVAD
VGFTLWLDTIENIGRLEPALEDVPCENIVGLVIYDLPGRDCAAKASNGELKVGELDRYKTEYIDKIAEILKAHSNTAFALVIE
PDSLPLNLTNSDLQTCQASAGYREGVAYALKQLNLPNVVMIIDAGHGGWLGWDANLKPGAQELASVYKSAGSPSQVRGISTN
VAGVNAWDQEPGEFSDASDAQYNKCKEYINTFGAELKSAGMPNHAIIDTGRNGVTGLRDEWGDWCNVNGAGFGVRPTANT
GDELADAFVWVKPGGESDGTSDSSAARYDSFCGKPDFAKPSPEAGTWNQAYFEMLLKNANPSF

65

M. thermophila Endoglucanasa 2 (EG2): SEQ ID NO: 65

5
QSGPWQQCGGIGWQGSTDCVSGYHCYQNDWYSQCVGAASSTTLQTSTTSRPTATSTAPPSSTTSPSKGKLKWL
GSNESGAEFEGGNYPGLWGKHFIPTSAIQTLLINDGYNIFRIDFSMERLVPNQLTSSFDEGYLRNLTEVVNFY
10 TNAGKYAVLDPHNYGRYVGNVITDTNAGRRTFTWNLAKQFASNSLVIIDTNNEYNTMDQTLVLNLNQAAIDGIRA
AGATSQYIFVEGNAWSGAWSWNTTNTNMAALTDPOKNIKIVEMHQYLDSDSSGTHAECVSSNIGAQRVVGATQWL
RANGKLGVLGEFAGGANAVCQAVTGLLDHLQDNSDVLGALWAAAGPWGDMYSFEPPSGTGYVNYNSILKK
YLP

M. thermophila Beta-glucosidasa (BG): SEQ ID NO: 66

15 IESRKVHQKPLARSEPFYPSWMNPNADGWAEAYAQAQSFVSQMTLLEKVNLTGTVGWGAEQCVGVGAIPRLG
LRSLCMHDSPLGIRGADYNSAFPSGQTVAAWDRGLMYRRGYAMQEAQKGINVLLGPVAGPLGRMPEGGRNW
EGFADPDLVLTGIGMSETIKGIDAGVIAKHFIGNEQEHFRQVPEAQGYNYISETLSSNIDDKTMHELYLWLP
20 FADAVRAGVGSVMCSYQQVNSYACQNSKLLNDLLKNELGFGQFVMSDWQAQHTGAASAVAGLDMSPGDTQFN
TGVSFWGANLTLAVLNGTVPAYRLDDMAMRIMAALFKVTKTTDLEPINFSFWTDDTYGPIHWAQKQGYQEI
VDVRADHGNLIREIAAKGTVLLKNTGSLPLNPKFVAVIGEDAGSSPNGPNCSDRGCNEGTLAMGWGSGTANY
PYLVSPDAALQARAIQDGTREYVLSNYAEKTKALVSQANATAIVFVNADSGEGYINVDGNEGDRKNLTWNN
30 GDTLVKNVSSWCSNTIIVIHVSVPVLLTDWYDNPNTAILWAGLPGQESGNSITDVLVYGVNPAARSPTWGT
RESYGADVLYKPNNGGAPQDFTGEGVFDYRYFDKVDSDSVIYEFHGLSYTTFEYSNIRVVKSNVSEYRPTT
GTTAQAPTFGNFSTDLEDYLFPKDEFYIYQYIYPYLNTDPRRASADPHYGTAEFLPPHATDDDPQPLLR
25 SGGNSPGGNRLYDIVYITADITNTGTVGEEVPLVSLGGPEDPKVQLRDFDRMRIEPGETRQFTGRLTR
DLSNWDVTQDWISRYPKTAYVGRSSRLDLKIELP

M. thermophila Celobiohidrolasa Tipo 1a (Cbh1a): SEQ ID NO: 67

30 QNACTLTAENHPSLTWSKCTSGGSCTSVQGSITIDANWRWTHRTDSATNCYEGNKWDTSYCSDGPSCASKCCID
GADYSTYGTITSGNSLNLKFTKQYSTNIGSRTYLMESDTKYQMFQLLGNEFTFDVDVSNLGCGLNGALYFV
SMDADGGMSKYSGNKAGAKYGTGYCDSQCPDLKFINGEANVENWQSSNDANAGTGKYGCCSEMDVWEANM
35 AAFTPHPCTVIGQSRCEGDSCTGTYTDYAGICDPDGCDFNSYRQGNKTFYKGMTVDTTKKITVVTQFLKN
SAGELSEIKRFYVQNGKVPINSESTIPGVEGNSITQDWCDRQKAAFGDVTDFQDKGGMVQMGKALAGPMVLVMS
IWDDHAVNMLWLDSTWPIDGAGKPGAERGACPTTSGVPAEVEAEAPNSNVIFSNIRFGPIGISTVSGLPDGGSGN
PNPPVSSSTPVPSSSTTSSGSSGPTGGTGVAKHYEQCGGIGFTGPTQCESPYTCTKLNWYSQCL

M. thermophila Celobiohidrolasa Tipo 2b (CBH2b): SEQ ID NO: 68

40 APVIEERQNCGAVWTQCGNGWQGPCTCCASGSTCVAQNEWYSQCLPNSQVTSSTTPSSTSTSRSTSTSSSTTR
SGSSSSSTTPPVSSPVTISIPGGATSTASYSGNPFSGVRLFANDYYRSEVHNLAIPSMGTGLAAKASAVEVP
45 SFQWLDNRVITIDTLMVQTLQVRLNKGANPPYAAQLVVDLPDRDCAAAASNGEFSIANGGAANYRSYDAI
RKHIIEYSDIRIILVIEPDSMANMVTNMVAKCSNAASTYHELTVYALQKLNLPNVAMYLDAGHAGWLGPANI
QPAELFAGIYNDAGKPAAVRGLATNVANYNASIASAPSYTSPNPNYDEKHYIEAFSPLLNSAGFPARFIVDT
GRNGKQPTGQQWGDWCNVKGTGFGVRPTANTGHELVDAFVWVKPGGESDGTSDTSAARYDYHCGLSDALQPAP
EAGQWFQAYFEQLLTNANPPF

Secuencia codificadora de GH61a

50
atgtccaagg cctctgctct cctcgtctgc ctgacgggag cggccctcgt cgctgcacat 60
ggccacgtca gccacatcgt cgtcaacggc gtctactaca ggaactacga cccacagaca 120
gactggtacc agcccaaccc gccaacagtc atcggttggc cggcagccga tcaggataat 180
ggcttcgttg aacccaacag ctttggtcag ccagatatca tctgccacaa gagcgccacc 240
cccggcgggc gccacgtac cgttgctgcc ggagacaaga tcaacatcgt ctggaccccc 300
gagtggtccc aatccacat cggccccgtc attgactacc tagccgctg caacggtgac 360
tgcgagaccg tcgacaagtc gtcgctgcgc tggttcaaga ttgacggcgc cggctacgac 420
aaggcgccgc gccgctgggc cgcgacgct ctgcgcgccca acggcaacag ctggctcgtc 480
cagatcccg cggatctcaa ggcgggcaac tacgtctccc gccacgagat catcgccctc 540
cacggtgctc agagccccc cggcgccag gccacccgc agtgcatcaa cctccgctc 600
accggcgggc gcagcaacct gccacgggc gtcgcccggc cctcgtgta caaggcgacc 660
gacccgggca tcctcttcaa cccctacgct tcctccccgg attacaccgt ccccgcccg 720
gccctcattg ccggcgccgc cagctcgatc gccagagca cgtcggtcgc cactgcccac 780
ggcacggcca ccgttcccg cggcgccgc gccaaccta ccgccaccac caccgccc 840
acctccgccc ccccgagcac caccctgagg acgaccata cctcggccgc gcagactacc 900
gccccgccct ccggcgatgt gcagaccaag tacggccagt gtggtggcaa cggatggacg 960
ggcccgacgg tgtgcgcccc cggctcgagc tgcctccgtc tcaacgagtg gtactccag 1020
tgtttataa 1029

SEQ ID Nos: 32 a 59 aparecen en el Listado Secuencial formal para esta divulgación.

El uso de proteínas GH61 no pretende limitarse de ninguna manera a la teoría como su modo de acción. Teóricamente, la producción del producto puede aumentar por la acción de la proteína GH61 en el sustrato, por interacción de la proteína GH61 directamente con una o más enzimas de celulasa en una mezcla, reduciendo la viscosidad de la mezcla de reacción o mediante otro mecanismo. La actividad de GH61 puede seguirse de manera empírica usando métodos de ensayo descritos en esta divulgación, sin conocer el mecanismo de operación de la proteína GH61 que se está usando.

DERRAME DE SECUENCIA

[0297]

<110> Codexis, Inc.

Rao, Kripa

Clark, Louis

Campopiano, Onorato

Szabo, Lorand

Torok, Janos

Baidyaroy, Dipnath

<120> Uso de Proteínas de la Familia Glycoside Hydrolase GH61

Facilitar el procesamiento de la celulosa

<130> 90834-818317

<150> US 61/375,788

<151> 2010-08-20

<160> 70

<170> PatentIn version 3.5

<210> 1

<211> 342

<212> PRT

<213> Myceliophthora thermophila

<400> 1

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

Val Ala Ala His Gly His Val Ser His Ile Val Val Asn Gly Val Tyr
20 25 30

Tyr Arg Asn Tyr Asp Pro Thr Thr Asp Trp Tyr Gln Pro Asn Pro Pro
35 40 45

Thr Val Ile Gly Trp Thr Ala Ala Asp Gln Asp Asn Gly Phe Val Glu
50 55 60

Pro Asn Ser Phe Gly Thr Pro Asp Ile Ile Cys His Lys Ser Ala Thr
65 70 75 80

Pro Gly Gly Gly His Ala Thr Val Ala Ala Gly Asp Lys Ile Asn Ile
85 90 95

Val Trp Thr Pro Glu Trp Pro Glu Ser His Ile Gly Pro Val Ile Asp
100 105 110

Tyr Leu Ala Ala Cys Asn Gly Asp Cys Glu Thr Val Asp Lys Ser Ser
115 120 125

Leu Arg Trp Phe Lys Ile Asp Gly Ala Gly Tyr Asp Lys Ala Ala Gly
130 135 140

5

Arg Trp Ala Ala Asp Ala Leu Arg Ala Asn Gly Asn Ser Trp Leu Val
145 150 155 160

10

Gln Ile Pro Ser Asp Leu Lys Ala Gly Asn Tyr Val Leu Arg His Glu
165 170 175

15

Ile Ile Ala Leu His Gly Ala Gln Ser Pro Asn Gly Ala Gln Ala Tyr
180 185 190

Pro Gln Cys Ile Asn Leu Arg Val Thr Gly Gly Gly Ser Asn Leu Pro
195 200 205

20

Ser Gly Val Ala Gly Thr Ser Leu Tyr Lys Ala Thr Asp Pro Gly Ile
210 215 220

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

25

Ala Leu Ile Ala Gly Ala Ala Ser Ser Ile Ala Gln Ser Thr Ser Val
245 250 255

30

Ala Thr Ala Thr Gly Thr Ala Thr Val Pro Gly Gly Gly Gly Ala Asn
260 265 270

35

Pro Thr Ala Thr Thr Thr Ala Ala Thr Ser Ala Ala Pro Ser Thr Thr
275 280 285

40

Leu Arg Thr Thr Thr Thr Ser Ala Ala Gln Thr Thr Ala Pro Pro Ser
290 295 300

Gly Asp Val Gln Thr Lys Tyr Gly Gln Cys Gly Gly Asn Gly Trp Thr
305 310 315 320

45

Gly Pro Thr Val Cys Ala Pro Gly Ser Ser Cys Ser Val Leu Asn Glu
325 330 335

50

Trp Tyr Ser Gln Cys Leu
340

<210> 2

<211> 323

55

<212> PRT

<213> Myceliophthora thermophila

<400> 2

60

His Gly His Val Ser His Ile Val Val Asn Gly Val Tyr Tyr Arg Asn
1 5 10 15

65

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

```

260                265                270
5      Thr Thr Thr Ser Ala Ala Gln Thr Thr Ala Pro Pro Ser Gly Asp Val
      275                280                285

10     Gln Thr Lys Tyr Gly Gln Cys Gly Gly Asn Gly Trp Thr Gly Pro Thr
      290                295                300

15     Val Cys Ala Pro Gly Ser Ser Cys Ser Val Leu Asn Glu Trp Tyr Ser
      305                310                315                320

20     Gln Cys Leu

25     <210> 3
      <211> 278
      <212> PRT
      <213> Mycolophthora thermophila

      <400> 3

```

```

30          260          265          270
      Thr Thr Thr Ser Ala Ala Gln Thr Thr Ala Pro Pro Ser Gly Asp Val
      275          280          285
35
      Gln Thr Lys Tyr Gly Gln Cys Gly Gly Asn Gly Trp Thr Gly Pro Thr
      290          295          300
40
      Val Cys Ala Pro Gly Ser Ser Cys Ser Val Leu Asn Glu Trp Tyr Ser
      305          310          315          320
45
      Gln Cys Leu
50
<210> 3
<211> 278
<212> PRT
<213> Mycolophthora thermophila
55
<400> 3

```

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

	145		150		155		160									
5	Thr	Trp	Ile	Asn	Thr	Ile	Gly	Asn	Arg	Glu	Phe	Tyr	Met	Ala	Cys	Ala
				165						170					175	
10	Pro	Val	Glu	Ile	Thr	Gly	Asp	Gly	Gly	Ser	Glu	Ser	Ala	Leu	Ala	Ala
				180						185				190		
15	Leu	Pro	Asp	Met	Val	Ile	Ala	Asn	Ile	Pro	Ser	Ile	Gly	Gly	Thr	Cys
			195					200					205			
20	Ala	Thr	Glu	Glu	Gly	Lys	Tyr	Tyr	Glu	Tyr	Pro	Asn	Pro	Gly	Lys	Ser
			210				215					220				
25	Val	Glu	Thr	Ile	Pro	Gly	Trp	Thr	Asp	Leu	Val	Pro	Leu	Gln	Gly	Glu
	225					230					235					240
30	Cys	Gly	Ala	Ala	Ser	Gly	Val	Ser	Gly	Ser	Gly	Gly	Asn	Ala	Ser	Ser
					245					250					255	
35	Ala	Thr	Pro	Ala	Ala	Gly	Ala	Ala	Pro	Thr	Pro	Ala	Val	Arg	Gly	Arg
				260					265					270		
40	Arg	Pro	Thr	Trp	Asn	Ala										
				275												

<210> 4
 <211> 223
 <212> PRT
 <213> Mycellophthora thermophila
 <400> 4

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

«210» 5

«211» 192

«212» PRT

«213» Mycelophthora thermophila

<400> 5

Met Thr Lys Asn Ala Gln Ser Lys Gln Gly Val Glu Asn Pro Thr Ser
1 5 10 15

Gly Asp Ile Arg Cys Tyr Thr Ser Gln Thr Ala Ala Asn Val Val Thr
20 25 30

Val Pro Ala Gly Ser Thr Ile His Tyr Ile Ser Thr Gln Gln Ile Asn
35 40 45

His Pro Gly Pro Thr Gln Tyr Tyr Leu Ala Lys Val Pro Pro Gly Ser
50 55 60

Ser Ala Lys Thr Phe Asp Gly Ser Gly Ala Val Trp Phe Lys Ile Ser
65 70 75 80

Thr Thr Met Pro Thr Val Asp Ser Asn Lys Gln Met Phe Trp Pro Gly

85 90 95

Gln Asn Thr Tyr Glu Thr Ser Asn Thr Thr Ile Pro Ala Asn Thr Pro
100 105 110

Asp Gly Glu Tyr Leu Leu Arg Val Lys Gln Ile Ala Leu His Met Ala
115 120 125

Ser Gln Pro Asn Lys Val Gln Phe Tyr Leu Ala Cys Thr Gln Ile Lys
130 135 140

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

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

Met Lys Pro Glu Ser Tyr Gln Pro Pro Gly Pro Pro Val Trp Arg Gly
180 185 190

<210> 6

<211> 283

<212> PRT

<213> Mycellophthora thermophila

<400> 6

	Met Lys Pro Phe Ser Leu Val Ala Leu Ala Thr Ala Val Ser Gly His	
	1 5 10 15	
5	Ala Ile Phe Gln Arg Val Ser Val Asn Gly Gln Asp Gln Gly Gln Leu	
	20 25 30	
10	Lys Gly Val Arg Ala Pro Ser Ser Asn Ser Pro Ile Gln Asn Val Asn	
	35 40 45	
15	Asp Ala Asn Met Ala Cys Asn Ala Asn Ile Val Tyr His Asp Asn Thr	
	50 55 60	
20	Ile Ile Lys Val Pro Ala Gly Ala Arg Val Gly Ala Trp Trp Gln His	
	65 70 75 80	
25	Val Ile Gly Gly Pro Gln Gly Ala Asn Asp Pro Asp Asn Pro Ile Ala	
	85 90 95	
30	Ala Ser His Lys Gly Pro Ile Gln Val Tyr Leu Ala Lys Val Asp Asn	
	100 105 110	
35	Ala Ala Thr Ala Ser Pro Ser Gly Leu Lys Trp Phe Lys Val Ala Glu	
	115 120 125	
40	Arg Gly Leu Asn Asn Gly Val Trp Ala Tyr Leu Met Arg Val Glu Leu	
	130 135 140	
45	Leu Ala Leu His Ser Ala Ser Ser Pro Gly Gly Ala Gln Phe Tyr Met	
	145 150 155 160	
50	Gly Cys Ala Gln Ile Glu Val Thr Gly Ser Gly Thr Asn Ser Gly Ser	
	165 170 175	
55	Asp Phe Val Ser Phe Pro Gly Ala Tyr Ser Ala Asn Asp Pro Gly Ile	
	180 185 190	
60	Leu Leu Ser Ile Tyr Asp Ser Ser Gly Lys Pro Asn Asn Gly Gly Arg	
	195 200 205	
65	Ser Tyr Pro Ile Pro Gly Pro Arg Pro Ile Ser Cys Ser Gly Ser Gly	
	210 215 220	
	Gly Gly Gly Asn Asn Gly Gly Asp Gly Gly Asp Asp Asn Asn Gly Gly	
	225 230 235 240	
	Gly Asn Asn Asn Gly Gly Gly Ser Val Pro Leu Tyr Gly Gln Cys Gly	
	245 250 255	
	Gly Ile Gly Tyr Thr Gly Pro Thr Thr Cys Ala Gln Gly Thr Cys Lys	
	260 265 270	
	Val Ser Asn Glu Tyr Tyr Ser Gln Cys Leu Pro	
	275 280	
	<210> 7 <211> 212 <212> PRT <213> Mycelophthora thermophila <400> 7	

	Met	Lys	Leu	Thr	Ser	Ser	Leu	Ala	Val	Leu	Ala	Ala	Ala	Gly	Ala	Gln	
	1				5					10					15		
5	Ala	His	Tyr	Thr	Phe	Pro	Arg	Ala	Gly	Thr	Gly	Gly	Ser	Leu	Ser	Gly	
				20					25					30			
10	Glu	Trp	Glu	Val	Val	Arg	Met	Thr	Glu	Asn	His	Tyr	Ser	His	Gly	Pro	
			35					40					45				
15	Val	Thr	Asp	Val	Thr	Ser	Pro	Glu	Met	Thr	Cys	Tyr	Gln	Ser	Gly	Val	
		50						55					60				
20	Gln	Gly	Ala	Pro	Gln	Thr	Val	Gln	Val	Lys	Ala	Gly	Ser	Gln	Phe	Thr	
	65					70					75					80	
25	Phe	Ser	Val	Asp	Pro	Ser	Ile	Gly	His	Pro	Gly	Pro	Leu	Gln	Phe	Tyr	
					85					90					95		
30	Met	Ala	Lys	Val	Pro	Ser	Gly	Gln	Thr	Ala	Ala	Thr	Phe	Asp	Gly	Thr	
				100					105					110			
35	Gly	Ala	Val	Trp	Phe	Lys	Ile	Tyr	Gln	Asp	Gly	Pro	Asn	Gly	Leu	Gly	
			115					120					125				
40	Thr	Asp	Ser	Ile	Thr	Trp	Pro	Ser	Ala	Gly	Lys	Thr	Glu	Val	Ser	Val	
		130					135					140					
45	Thr	Ile	Pro	Ser	Cys	Ile	Glu	Asp	Gly	Glu	Tyr	Leu	Leu	Arg	Val	Glu	
	145					150					155					160	
50	His	Thr	Pro	Leu	Pro	Thr	Ala	Pro	Ala	Ala	Gln	Asn	Arg	Ala	Arg	Ser	
					165					170					175		
55	Ser	Pro	Ser	Pro	Ala	Ala	Tyr	Lys	Ala	Thr	Asp	Pro	Gly	Ile	Leu	Phe	
				180					185					190			
60	Gln	Leu	Tyr	Trp	Pro	Ile	Pro	Thr	Glu	Tyr	Ile	Asn	Pro	Gly	Pro	Ala	
			195					200					205				
65	Pro	Val	Ser	Cys													
				210													

<210> 8
 <211> 149
 <212> PRT
 <213> *Mycelophthora thermophila*

<400> 8

Met Pro Pro Pro Arg Leu Ser Thr Leu Leu Pro Leu Leu Ala Leu Ile
1 5 10 15

5

Ala Pro Thr Ala Leu Gly His Ser His Leu Gly Tyr Ile Ile Ile Asn
20 25 30

10

Gly Glu Val Tyr Gln Gly Phe Asp Pro Arg Pro Glu Gln Ala Asn Ser
35 40 45

15

Pro Leu Arg Val Gly Trp Ser Thr Gly Ala Ile Asp Asp Gly Phe Val

50

55

60

20

Ala Pro Ala Asn Tyr Ser Ser Pro Asp Ile Ile Cys His Ile Glu Gly
65 70 75 80

25

Ala Ser Pro Pro Ala His Ala Pro Val Arg Ala Gly Asp Arg Val His
85 90 95

30

Val Gln Trp Asn Gly Trp Pro Leu Gly His Val Gly Pro Val Leu Ser
100 105 110

Tyr Leu Ala Pro Cys Gly Gly Leu Glu Gly Ser Glu Ser Gly Cys Ala
115 120 125

35

Gly Val Asp Lys Arg Gln Leu Arg Trp Thr Lys Val Asp Asp Ser Leu
130 135 140

40

Pro Ala Met Glu Leu
145

<210> 9

45

<211> 151

<212> PRT

<213> Mycellophthora thermophila

50

<400> 9

55

60

65

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

<210> 10

<211> 245

<212> PRT

<213> Myceliophthora thermophila

<400> 10

	Met	Leu	Leu	Leu	Thr	Leu	Ala	Thr	Leu	Val	Thr	Leu	Leu	Ala	Arg	His
	1				5					10					15	
5	Val	Ser	Ala	His	Ala	Arg	Leu	Phe	Arg	Val	Ser	Val	Asp	Gly	Lys	Asp
				20					25					30		
10	Gln	Gly	Asp	Gly	Leu	Asn	Lys	Tyr	Ile	Arg	Ser	Pro	Ala	Thr	Asn	Asp
			35					40					45			
15	Pro	Val	Arg	Asp	Leu	Ser	Ser	Ala	Ala	Ile	Val	Cys	Asn	Thr	Gln	Gly
	50						55					60				
20	Ser	Lys	Ala	Ala	Pro	Asp	Phe	Val	Arg	Ala	Ala	Ala	Gly	Asp	Lys	Leu
	65					70					75					80
25	Thr	Phe	Leu	Trp	Ala	His	Asp	Asn	Pro	Asp	Asp	Pro	Val	Asp	Tyr	Val
					85					90					95	
30	Leu	Asp	Pro	Ser	His	Lys	Gly	Ala	Ile	Leu	Thr	Tyr	Val	Ala	Ala	Tyr
				100					105					110		
35	Pro	Ser	Gly	Asp	Pro	Thr	Gly	Pro	Ile	Trp	Ser	Lys	Leu	Ala	Glu	Glu
			115					120					125			
40	Gly	Phe	Thr	Gly	Gly	Gln	Trp	Ala	Thr	Ile	Lys	Met	Ile	Asp	Asn	Gly
		130					135					140				
45	Gly	Lys	Val	Asp	Val	Thr	Leu	Pro	Glu	Ala	Leu	Ala	Pro	Gly	Lys	Tyr
	145					150					155					160
50	Leu	Ile	Arg	Gln	Glu	Leu	Leu	Ala	Leu	His	Arg	Ala	Asp	Phe	Ala	Cys
					165					170					175	
55	Asp	Asp	Pro	Ala	His	Pro	Asn	Arg	Gly	Ala	Glu	Ser	Tyr	Pro	Asn	Cys
60																
65																

[illegible]

20	<210> 11 <211> 109 <212> PRT <213> <i>Mycosiphthons thermophila</i>
25	<400> 11

	Met	Phe	Thr	Ser	Leu	Cys	Ile	Thr	Asp	His	Trp	Arg	Thr	Leu	Ser	Ser	
	1				5					10					15		
5	His	Ser	Gly	Pro	Val	Met	Asn	Tyr	Leu	Ala	His	Cys	Thr	Asn	Asp	Asp	
				20					25					30			
10	Cys	Lys	Ser	Phe	Lys	Gly	Asp	Ser	Gly	Asn	Val	Trp	Val	Lys	Ile	Glu	
			35				40						45				
15	Gln	Leu	Ala	Tyr	Asn	Pro	Ser	Ala	Asn	Pro	Pro	Trp	Ala	Ser	Asp	Leu	
	50					55						60					
20	Leu	Arg	Glu	His	Gly	Ala	Lys	Trp	Lys	Val	Thr	Ile	Pro	Pro	Ser	Leu	
	65				70					75						80	
25	Val	Pro	Gly	Glu	Tyr	Leu	Leu	Arg	His	Glu	Ile	Leu	Gly	Leu	His	Val	
				85					90						95		
30	Ala	Gly	Thr	Val	Met	Gly	Ala	Gln	Phe	Tyr	Pro	Gly	Cys	Thr	Gln	Ile	
				100				105						110			
35	Arg	Val	Thr	Glu	Gly	Gly	Ser	Thr	Gln	Leu	Pro	Ser	Gly	Ile	Ala	Leu	
			115				120						125				
40	Pro	Gly	Ala	Tyr	Gly	Pro	Gln	Asp	Glu	Gly	Ile	Leu	Val	Asp	Leu	Trp	
	130					135						140					
45	Arg	Val	Asn	Gln	Gly	Gln	Val	Asn	Tyr	Thr	Ala	Pro	Gly	Gly	Pro	Val	
	145					150						155			160		
50	Trp	Ser	Glu	Ala	Trp	Asp	Thr	Glu	Phe	Gly	Gly	Ser	Asn	Thr	Thr	Glu	
				165					170					175			
55	Cys	Ala	Thr	Met	Leu	Asp	Asp	Leu	Leu	Asp	Tyr	Met	Ala	Ala	Asn	Asp	
				180				185					190				
60	Glu	Trp	Ile	Gly	Trp	Thr	Ala										
			195														

<210> 12
 <211> 227
 <212> PRT
 <213> Mycellophthora thermophila

<400> 12

	Met	Lys	Leu	Ser	Ala	Ala	Ile	Ala	Val	Leu	Ala	Ala	Ala	Leu	Ala	Glu	
	1				5					10				15			
5	Gly	His	Tyr	Thr	Phe	Pro	Ser	Ile	Ala	Asn	Thr	Ala	Asp	Trp	Gln	Tyr	
				20					25					30			
10	Val	Arg	Ile	Thr	Thr	Asn	Phe	Gln	Ser	Asn	Gly	Pro	Val	Thr	Asp	Val	
			35					40					45				
15	Asn	Ser	Asp	Gln	Ile	Arg	Cys	Tyr	Glu	Arg	Asn	Pro	Gly	Thr	Gly	Ala	
		50					55					60					
20	Pro	Gly	Ile	Tyr	Asn	Val	Thr	Ala	Gly	Thr	Thr	Ile	Asn	Tyr	Asn	Ala	
	65					70					75					80	
25	Lys	Ser	Ser	Ile	Ser	His	Pro	Gly	Pro	Met	Ala	Phe	Tyr	Ile	Ala	Lys	
					85					90					95		
30	Val	Pro	Ala	Gly	Gln	Ser	Ala	Ala	Thr	Trp	Asp	Gly	Lys	Gly	Ala	Val	
				100					105					110			
35	Trp	Ser	Lys	Ile	His	Gln	Glu	Met	Pro	His	Phe	Gly	Thr	Ser	Leu	Thr	
			115					120					125				
40	Trp	Asp	Ser	Asn	Gly	Arg	Thr	Ser	Met	Pro	Val	Thr	Ile	Pro	Arg	Cys	
		130					135					140					
45	Leu	Gln	Asp	Gly	Glu	Tyr	Leu	Leu	Arg	Ala	Gln	His	Ile	Ala	Leu	His	
	145					150					155					160	
50	Ser	Ala	Gly	Ser	Pro	Gly	Gly	Ala	Gln	Phe	Tyr	Ile	Ser	Cys	Ala	Gln	
55	Leu	Ser	Val	Thr	Gly	Gly	Ser	Gly	Thr	Trp	Asn	Pro	Arg	Asn	Lys	Val	
			180						185					190			
60	Ser	Phe	Pro	Gly	Ala	Tyr	Lys	Ala	Thr	Asp	Pro	Gly	Ile	Leu	Ile	Asn	
			195					200					205				
65	Ile	Tyr	Tyr	Pro	Val	Pro	Thr	Ser	Tyr	Thr	Pro	Ala	Gly	Pro	Pro	Val	
	210						215					220					
	Asp	Thr	Cys														
	225																

«210» 13
 «211» 258
 «212» PRT
 «213» *Mycelophthora thermophila*

5

«400» 13

10

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

15

Ala His Gly Ala Val Thr Ser Tyr Asn Ile Ala Gly Lys Asp Tyr Pro
 20 25 30

20

Gly Tyr Ser Gly Phe Ala Pro Thr Gly Gln Asp Val Ile Gln Trp Gln
 35 40 45

25

Trp Pro Asp Tyr Asn Pro Val Leu Ser Ala Ser Asp Pro Lys Leu Arg
 50 55 60

30

Cys Asn Gly Gly Thr Gly Ala Ala Leu Tyr Ala Glu Ala Ala Pro Gly
 65 70 75 80

35

Asp Thr Ile Thr Ala Thr Trp Ala Gln Trp Thr His Ser Gln Gly Pro
 85 90 95

40

Ile Leu Val Trp Met Tyr Lys Cys Pro Gly Asp Phe Ser Ser Cys Asp
 100 105 110

45

Gly Ser Gly Ala Gly Trp Phe Lys Ile Asp Glu Ala Gly Phe His Gly
 115 120 125

50

Asp Gly Thr Thr Val Phe Leu Asp Thr Glu Thr Pro Ser Gly Trp Asp
 130 135 140

55

Ile Ala Lys Leu Val Gly Gly Asn Lys Ser Trp Ser Ser Lys Ile Pro

60

65

	145		150		155		160
5	Asp Gly Leu Ala Pro Gly Asn Tyr Leu Val Arg His Glu Leu Ile Ala	165		170		175	
10	Leu His Gln Ala Asn Asn Pro Gln Phe Tyr Pro Glu Cys Ala Gln Ile	180		185		190	
15	Lys Val Thr Gly Ser Gly Thr Ala Glu Pro Ala Ala Ser Tyr Lys Ala	195		200		205	
20	Ala Ile Pro Gly Tyr Cys Gln Gln Ser Asp Pro Asn Ile Ser Phe Asn	210		215		220	
25	Ile Asn Asp His Ser Leu Pro Gln Glu Tyr Lys Ile Pro Gly Pro Pro	225		230		235	240
	Val Phe Lys Gly Thr Ala Ser Ala Lys Ala Arg Ala Phe Gln Ala	245		250		255	

<210> 14
 <211> 225
 <212> PRT
 <213> Mycellophthora thermophila
 <400> 14

ES 2 623 288 T3

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

<210> 15
 <211> 237
 <212> PRT
 <213> *Mycelophthora thermophila*

400 15

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

					100					105					110		
5		Ala	Thr	Thr	Asp	Ser	Gly	Val	Gly	Gly	Gly	Trp	Phe	Lys	Ile	Gln	Glu
				115					120					125			
10		Asp	Gly	Leu	Thr	Asn	Gly	Val	Trp	Gly	Thr	Glu	Arg	Val	Ile	Asn	Gly
			130					135					140				
15		Gln	Gly	Arg	His	Asn	Ile	Lys	Ile	Pro	Glu	Cys	Ile	Ala	Pro	Gly	Gln
		145					150					155					160
20		Tyr	Leu	Leu	Arg	Ala	Glu	Met	Leu	Ala	Leu	His	Gly	Ala	Ser	Asn	Tyr
						165					170					175	
25		Pro	Gly	Ala	Gln	Phe	Tyr	Met	Glu	Cys	Ala	Gln	Leu	Asn	Ile	Val	Gly
					180					185					190		
30		Gly	Thr	Gly	Ser	Lys	Thr	Pro	Ser	Thr	Val	Ser	Phe	Pro	Gly	Ala	Tyr
				195				200						205			
35		Lys	Gly	Thr	Asp	Pro	Gly	Val	Lys	Ile	Asn	Ile	Tyr	Trp	Pro	Pro	Val
			210					215					220				
40		Thr	Ser	Tyr	Gln	Ile	Pro	Gly	Pro	Gly	Val	Phe	Thr	Cys			
		225					230					235					
		«210» 16															
		«211» 246															
		«212» PRT															
		«213» Mycellophthora thermophila															
		«400» 16															
45																	
50																	
55																	
60																	
65																	

ES 2 623 288 T3

Met Lys Leu Ser Leu Phe Ser Val Leu Ala Thr Ala Leu Thr Val Glu
1 5 10 15

Gly His Ala Ile Phe Gln Lys Val Ser Val Asn Gly Ala Asp Gln Gly
20 25 30

Ser Leu Thr Gly Leu Arg Ala Pro Asn Asn Asn Asn Pro Val Gln Asn
35 40 45

Val Asn Ser Gln Asp Met Ile Cys Gly Gln Ser Gly Ser Thr Ser Asn
50 55 60

Thr Ile Ile Glu Val Lys Ala Gly Asp Arg Ile Gly Ala Trp Tyr Gln
65 70 75 80

His Val Ile Gly Gly Ala Gln Phe Pro Asn Asp Pro Asp Asn Pro Ile
85 90 95

Ala Lys Ser His Lys Gly Pro Val Met Ala Tyr Leu Ala Lys Val Asp
100 105 110

Asn Ala Ala Thr Ala Ser Lys Thr Gly Leu Lys Trp Phe Lys Ile Trp
115 120 125

Glu Asp Thr Phe Asn Pro Ser Thr Lys Thr Trp Gly Val Asp Asn Leu
130 135 140

Ile Asn Asn Asn Gly Trp Val Tyr Phe Asn Leu Pro Gln Cys Ile Ala
145 150 155 160

Asp Gly Asn Tyr Leu Leu Arg Val Glu Val Leu Ala Leu His Ser Ala
165 170 175

Tyr Ser Gln Gly Gln Ala Gln Phe Tyr Gln Ser Cys Ala Gln Ile Asn
180 185 190

Val Ser Gly Gly Gly Ser Phe Thr Pro Ala Ser Thr Val Ser Phe Pro
195 200 205

Gly Ala Tyr Ser Ala Ser Asp Pro Gly Ile Leu Ile Asn Ile Tyr Gly
210 215 220

Ala Thr Gly Gln Pro Asp Asn Asn Gly Gln Pro Tyr Thr Ala Pro Gly
225 230 235 240

Pro Ala Pro Ile Ser Cys
245

<210> 17
<211> 254
<212> PRT
<213> Mycelophthora thermophila

<400> 17

Met Ala Leu Gln Leu Leu Ala Ser Leu Ala Leu Leu Ser Val Pro Ala
 1 5 10 15
 5 Leu Ala His Gly Gly Leu Ala Asn Tyr Thr Val Gly Asp Thr Trp Tyr
 20 25 30
 10 Arg Gly Tyr Asp Pro Asn Leu Pro Pro Glu Thr Gln Leu Asn Gln Thr
 35 40 45
 15 Trp Met Ile Gln Arg Gln Trp Ala Thr Ile Asp Pro Val Phe Thr Val
 50 55 60
 20 Ser Glu Pro Tyr Leu Ala Cys Asn Asn Pro Gly Ala Pro Pro Pro Ser
 65 70 75 80
 25 Tyr Ile Pro Ile Arg Ala Gly Asp Lys Ile Thr Ala Val Tyr Trp Tyr
 85 90 95
 30 Trp Leu His Ala Ile Gly Pro Met Ser Val Trp Leu Ala Arg Cys Gly
 100 105 110
 35 Asp Thr Pro Ala Ala Asp Cys Arg Asp Val Asp Val Asn Arg Val Gly
 115 120 125
 40 Trp Phe Lys Ile Trp Gln Gly Gly Leu Leu Gln Gly Pro Asn Leu Ala
 130 135 140
 45 Glu Gly Leu Trp Tyr Gln Lys Asp Phe Gln Arg Trp Asp Gly Ser Pro
 145 150 155 160
 50 Ser Leu Trp Pro Val Thr Ile Pro Lys Gly Leu Lys Ser Gly Thr Tyr
 165 170 175
 55 Ile Ile Arg His Glu Ile Leu Ser Leu His Val Ala Leu Lys Pro Gln
 180 185 190
 60 Phe Tyr Pro Glu Cys Ala His Leu Asn Ile Thr Gly Gly Gly Asp Leu
 195 200 205
 65 Leu Pro Pro Glu Glu Thr Leu Val Arg Phe Pro Gly Val Tyr Lys Glu
 210 215 220
 Asp Asp Pro Ser Ile Phe Ile Asp Val Tyr Ser Glu Glu Asn Ala Asn
 225 230 235 240
 Arg Thr Asp Tyr Thr Val Pro Gly Gly Pro Ile Trp Glu Gly
 245 250
 <210> 18
 <211> 235
 <212> PRT
 <213> Mycelophthora thermophila

	Met	Lys	Ala	Leu	Ser	Leu	Leu	Ala	Ala	Ala	Gly	Ala	Val	Ser	Ala	His
	1				5					10					15	
5	Thr	Ile	Phe	Val	Gln	Leu	Glu	Ala	Asp	Gly	Thr	Arg	Tyr	Pro	Val	Ser
					20				25					30		
10	Tyr	Gly	Ile	Arg	Asp	Pro	Thr	Tyr	Asp	Gly	Pro	Ile	Thr	Asp	Val	Thr
			35					40					45			
15	Ser	Asn	Asp	Val	Ala	Cys	Asn	Gly	Gly	Pro	Asn	Pro	Thr	Thr	Pro	Ser
		50					55					60				
20	Ser	Asp	Val	Ile	Thr	Val	Thr	Ala	Gly	Thr	Thr	Val	Lys	Ala	Ile	Trp
	65					70					75					80
25	Arg	His	Thr	Leu	Gln	Ser	Gly	Pro	Asp	Asp	Val	Met	Asp	Ala	Ser	His
					85					90					95	
30	Lys	Gly	Pro	Thr	Leu	Ala	Tyr	Ile	Lys	Lys	Val	Gly	Asp	Ala	Thr	Lys
				100					105					110		
35	Asp	Ser	Gly	Val	Gly	Gly	Gly	Trp	Phe	Lys	Ile	Gln	Glu	Asp	Gly	Tyr
			115					120					125			
40	Asn	Asn	Gly	Gln	Trp	Gly	Thr	Ser	Thr	Val	Ile	Ser	Asn	Gly	Gly	Glu
		130					135					140				
45	His	Tyr	Ile	Asp	Ile	Pro	Ala	Cys	Ile	Pro	Glu	Gly	Gln	Tyr	Leu	Leu
	145					150					155					160
50	Arg	Ala	Glu	Met	Ile	Ala	Leu	His	Ala	Ala	Gly	Ser	Pro	Gly	Gly	Ala
				165						170					175	
55	Gln	Leu	Tyr	Met	Glu	Cys	Ala	Gln	Ile	Asn	Ile	Val	Gly	Gly	Ser	Gly
				180					185					190		
60	Ser	Val	Pro	Ser	Ser	Thr	Val	Ser	Phe	Pro	Gly	Ala	Tyr	Ser	Pro	Asn
			195					200					205			
65	Asp	Pro	Gly	Leu	Leu	Ile	Asn	Ile	Tyr	Ser	Met	Ser	Pro	Ser	Ser	Ser
		210					215					220				
70	Tyr	Thr	Ile	Pro	Gly	Pro	Pro	Val	Phe	Lys	Cys					
	225					230					235					
75	«210» 19															

<211> 304
 <212> PRT
 <213> Mycellophthora thermophila
 <400> 19

5

Met Lys Ser Ser Thr Pro Ala Leu Phe Ala Ala Gly Leu Leu Ala Gln

10

15

20

25

30

35

40

45

50

55

60

65

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

5
10
15
20
25
30
35
40
45
50
55
60
65

Gln	Glu	Pro	Thr	Ile	Thr	Ser	Val	Thr	Asp	Leu	Pro	Thr	Asp	Glu	Ala	260	265	270
Lys	Trp	Ile	Ala	Leu	Gln	Lys	Ile	Ser	Tyr	Val	Asp	Gln	Thr	Gly	Thr	275	280	285
Ala	Arg	Thr	Tyr	Glu	Pro	Ala	Ser	Arg	Lys	Thr	Arg	Ser	Pro	Arg	Val	290	295	300
<210> 20																		
<211> 323																		
<212> PRT																		
<213> Myceliophthora thermophila																		
<400> 20																		
Met	Lys	Ser	Phe	Thr	Leu	Thr	Thr	Leu	Ala	Ala	Leu	Ala	Gly	Asn	Ala	1	5	10
Ala	Ala	His	Ala	Thr	Phe	Gln	Ala	Leu	Trp	Val	Asp	Gly	Val	Asp	Tyr	20	25	30
Gly	Ala	Gln	Cys	Ala	Arg	Leu	Pro	Ala	Ser	Asn	Ser	Pro	Val	Thr	Asp	35	40	45
Val	Thr	Ser	Asn	Ala	Ile	Arg	Cys	Asn	Ala	Asn	Pro	Ser	Pro	Ala	Arg	50	55	60
Gly	Lys	Cys	Pro	Val	Lys	Ala	Gly	Ser	Thr	Val	Thr	Val	Glu	Met	His	65	70	75
Gln	Gln	Pro	Gly	Asp	Arg	Ser	Cys	Ser	Ser	Glu	Ala	Ile	Gly	Gly	Ala	85	90	95
His	Tyr	Gly	Pro	Val	Met	Val	Tyr	Met	Ser	Lys	Val	Ser	Asp	Ala	Ala	100	105	110
Ser	Ala	Asp	Gly	Ser	Ser	Gly	Trp	Phe	Lys	Val	Phe	Glu	Asp	Gly	Trp	115	120	125
Ala	Lys	Asn	Pro	Ser	Gly	Gly	Ser	Gly	Asp	Asp	Asp	Tyr	Trp	Gly	Thr	130	135	140
Lys	Asp	Leu	Asn	Ser	Cys	Cys	Gly	Lys	Met	Asn	Val	Lys	Ile	Pro	Ala	145	150	155
Asp	Leu	Pro	Ser	Gly	Asp	Tyr	Leu	Leu	Arg	Ala	Glu	Ala	Leu	Ala	Leu	165	170	175

ES 2 623 288 T3

	His Thr Ala Gly Ser Ala Gly Gly Ala Gln Phe Tyr Met Thr Cys Tyr	
	180	185 190
5	Gln Leu Thr Val Thr Gly Ser Gly Ser Ala Ser Pro Pro Thr Val Ser	
	195	200 205
10	Phe Pro Gly Ala Tyr Lys Ala Thr Asp Pro Gly Ile Leu Val Asn Ile	
	210	215 220
15	His Ala Pro Leu Ser Gly Tyr Thr Val Pro Gly Pro Ala Val Tyr Ser	
	225	230 235 240
20	Gly Gly Ser Thr Lys Lys Ala Gly Ser Ala Cys Thr Gly Cys Glu Ser	
	245	250 255
25	Thr Cys Ala Val Gly Ser Gly Pro Thr Ala Thr Val Ser Gln Ser Pro	
	260	265 270
30	Gly Ser Thr Ala Thr Ser Ala Pro Gly Gly Gly Gly Gly Cys Thr Val	
	275	280 285
35	Gln Lys Tyr Gln Gln Cys Gly Gly Gln Gly Tyr Thr Gly Cys Thr Asn	
	290	295 300
40	Cys Ala Ser Gly Ser Thr Cys Ser Ala Val Ser Pro Pro Tyr Tyr Ser	
	305	310 315 320
45	Gln Cys Val	
46	<210> 21	
47	<211> 289	
48	<212> PRT	
49	<213> Mycollophthora thermophila	
50	<400> 21	
55	Met Lys Gly Leu Leu Gly Ala Ala Ala Leu Ser Leu Ala Val Ser Asp	
	1	5 10 15
60	Val Ser Ala His Tyr Ile Phe Gln Gln Leu Thr Thr Gly Gly Val Lys	
	20	25 30
65	His Ala Val Tyr Gln Tyr Ile Arg Lys Asn Thr Asn Tyr Asn Ser Pro	
	35	40 45
70	Val Thr Asp Leu Thr Ser Asn Asp Leu Arg Cys Asn Val Gly Ala Thr	
	50	55 60

5 Gly Ala Gly Thr Asp Thr Val Thr Val Arg Ala Gly Asp Ser Phe Thr
65 70 75 80

Phe Thr Thr Asp Thr Pro Val Tyr His Gln Gly Pro Thr Ser Ile Tyr
85 90 95

10 Met Ser Lys Ala Pro Gly Ser Ala Ser Asp Tyr Asp Gly Ser Gly Gly
100 105 110

15 Trp Phe Lys Ile Lys Asp Trp Ala Asp Tyr Thr Ala Thr Ile Pro Gln
115 120 125

20 Cys Ile Pro Pro Gly Asp Tyr Leu Leu Arg Ile Gln Gln Leu Gly Ile
130 135 140

25 His Asn Pro Trp Pro Ala Gly Ile Pro Gln Phe Tyr Ile Ser Cys Ala
145 150 155 160

Gln Ile Thr Val Thr Gly Gly Gly Ser Ala Asn Pro Gly Pro Thr Val
165 170 175

30 Ser Ile Pro Gly Ala Phe Lys Gln Thr Asp Pro Gly Tyr Thr Val Asn
180 185 190

35 Ile Tyr Asn Asn Phe His Asn Tyr Thr Val Pro Gly Pro Ala Val Phe
195 200 205

40 Thr Cys Asn Gly Ser Gly Gly Asn Asn Gly Gly Gly Ser Asn Pro Val
210 215 220

45 Thr Thr Thr Thr Thr Thr Thr Thr Arg Pro Ser Thr Ser Thr Ala Gln
225 230 235 240

Ser Gln Pro Ser Ser Ser Pro Thr Ser Pro Ser Ser Cys Thr Val Ala
245 250 255

50 Lys Trp Gly Gln Cys Gly Gly Gln Gly Tyr Ser Gly Cys Thr Val Cys
260 265 270

55 Ala Ala Gly Ser Thr Cys Gln Lys Thr Asn Asp Tyr Tyr Ser Gln Cys
275 280 285

60 Leu

<210> 22
<211> 346
<212> PRT

65

«213» *Mycollophthora thermophila*

«400» 22

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

5 Gln Ile Pro Gly Pro Ala Ile Ile Ser Gly Ala Ser Ala Val Gln Gln
 245 250 255

 10 Thr Thr Ser Ala Ile Thr Ala Ser Ala Ser Ala Ile Thr Gly Ser Ala
 260 265 270

 15 Thr Ala Ala Pro Thr Ala Ala Thr Thr Thr Ala Ala Ala Ala Thr
 275 280 285

 20 Thr Thr Thr Thr Ala Gly Ser Gly Ala Thr Ala Thr Pro Ser Thr Gly
 290 295 300

 25 Gly Ser Pro Ser Ser Ala Gln Pro Ala Pro Thr Thr Ala Ala Ala Thr
 305 310 315 320

 30 Ser Ser Pro Ala Arg Pro Thr Arg Cys Ala Gly Leu Lys Lys Arg Arg
 325 330 335

 35 Arg His Ala Arg Asp Val Lys Val Ala Leu
 340 345

 40 <210> 23
 <211> 237
 <212> PRT
 <213> *Mycobacterium thermophilus*

 45 <400> 23

 50 Met Lys Thr Leu Ala Ala Leu Val Val Ser Ala Ala Leu Val Ala Ala
 1 5 10 15

 55 His Gly Tyr Val Asp His Ala Thr Ile Gly Gly Lys Asp Tyr Gln Phe
 20 25 30

 60 Tyr Gln Pro Tyr Gln Asp Pro Tyr Met Gly Asp Asn Lys Pro Asp Arg
 35 40 45

 65 Val Ser Arg Ser Ile Pro Gly Asn Gly Pro Val Glu Asp Val Asn Ser
 50 55 60

 70 Ile Asp Leu Gln Cys His Ala Gly Ala Glu Pro Ala Lys Leu His Ala
 65 70 75 80

 75 Pro Ala Ala Ala Gly Ser Thr Val Thr Leu Tyr Trp Thr Leu Trp Pro
 85 90 95

 80 Asp Ser His Val Gly Pro Val Ile Thr Tyr Met Ala Arg Cys Pro Asp
 100 105 110

Thr Gly Cys Gln Asp Trp Ser Pro Gly Thr Lys Pro Val Trp Phe Lys
 115 120 125
 Ile Lys Glu Gly Gly Arg Glu Gly Thr Ser Asn Thr Pro Leu Met Thr
 130 135 140
 Ala Pro Ser Ala Tyr Thr Tyr Thr Ile Pro Ser Cys Leu Lys Ser Gly
 145 150 155 160
 Tyr Tyr Leu Val Arg His Glu Ile Ile Ala Leu His Ser Ala Trp Gln
 165 170 175
 Tyr Pro Gly Ala Gln Phe Tyr Pro Gly Cys His Gln Leu Gln Val Thr
 180 185 190
 Gly Gly Gly Ser Thr Val Pro Ser Thr Asn Leu Val Ser Phe Pro Gly
 195 200 205
 Ala Tyr Lys Gly Ser Asp Pro Gly Ile Thr Tyr Asp Ala Tyr Lys Ala
 210 215 220
 Gln Pro Tyr Thr Ile Pro Gly Pro Ala Val Phe Thr Cys
 225 230 235
 <210> 24
 <211> 225
 <212> PRT
 <213> Mycelophthora thermophila
 <400> 24
 Met Arg Tyr Phe Leu Gln Leu Ala Ala Ala Ala Ala Phe Ala Val Asn
 1 5 10 15
 Ser Ala Ala Gly His Tyr Ile Phe Gln Gln Phe Ala Thr Gly Gly Ser
 20 25 30
 Lys Tyr Pro Pro Trp Lys Tyr Ile Arg Arg Asn Thr Asn Pro Asp Trp
 35 40 45
 Leu Gln Asn Gly Pro Val Thr Asp Leu Ser Ser Thr Asp Leu Arg Cys
 50 55 60
 Asn Val Gly Gly Gln Val Ser Asn Gly Thr Glu Thr Ile Thr Leu Asn
 65 70 75 80
 Ala Gly Asp Glu Phe Ser Phe Ile Leu Asp Thr Pro Val Tyr His Ala
 85 90 95

	Gly	Pro	Thr	Ser	Leu	Tyr	Met	Ser	Lys	Ala	Pro	Gly	Ala	Val	Ala	Asp	
				100					105					110			
5	Tyr	Asp	Gly	Gly	Gly	Ala	Trp	Phe	Lys	Ile	Tyr	Asp	Trp	Gly	Pro	Ser	
			115					120					125				
10	Gly	Thr	Ser	Trp	Thr	Leu	Ser	Gly	Thr	Tyr	Thr	Gln	Arg	Ile	Pro	Lys	
		130					135					140					
15	Cys	Ile	Pro	Asp	Gly	Glu	Tyr	Leu	Leu	Arg	Ile	Gln	Gln	Ile	Gly	Leu	
	145					150					155					160	
20	His	Asn	Pro	Gly	Ala	Ala	Pro	Gln	Phe	Tyr	Ile	Ser	Cys	Ala	Gln	Val	
				165					170						175		
25	Lys	Val	Val	Asp	Gly	Gly	Ser	Thr	Asn	Pro	Thr	Pro	Thr	Ala	Gln	Ile	
				180					185					190			
30	Pro	Gly	Ala	Phe	His	Ser	Asn	Asp	Pro	Gly	Leu	Thr	Val	Asn	Ile	Tyr	
		195						200					205				
35	Asn	Asp	Pro	Leu	Thr	Asn	Tyr	Val	Val	Pro	Gly	Pro	Arg	Val	Ser	His	
	210						215					220					
40	Trp																
	225																
	<210> 25																
	<211> 444																
	<212> PRT																
	<213> Mycolophthora thermophila																
	<400> 25																
45	Met	His	Pro	Ser	Leu	Leu	Phe	Thr	Leu	Gly	Leu	Ala	Ser	Val	Leu	Val	
	1				5					10					15		
50	Pro	Leu	Ser	Ser	Ala	His	Thr	Thr	Phe	Thr	Thr	Leu	Phe	Val	Asn	Asp	
				20					25					30			
55	Val	Asn	Gln	Gly	Asp	Gly	Thr	Cys	Ile	Arg	Met	Ala	Lys	Lys	Gly	Asn	
		35						40					45				
60	Val	Ala	Thr	His	Pro	Leu	Ala	Gly	Gly	Leu	Asp	Ser	Glu	Asp	Met	Ala	
	50						55					60					
65	Cys	Gly	Arg	Asp	Gly	Gln	Glu	Pro	Val	Ala	Phe	Thr	Cys	Pro	Ala	Pro	
	65					70					75					80	

	Ala Gly Ala Lys Leu Thr Leu Glu Phe Arg Met Trp Ala Asp Ala Ser	85	90	95
5	Gln Ser Gly Ser Ile Asp Pro Ser His Leu Gly Val Met Ala Ile Tyr	100	105	110
10	Leu Lys Lys Val Ser Asp Met Lys Ser Asp Ala Ala Ala Gly Pro Gly	115	120	125
15	Trp Phe Lys Ile Trp Asp Gln Gly Tyr Asp Leu Ala Ala Lys Lys Trp	130	135	140
20	Ala Thr Glu Lys Leu Ile Asp Asn Asn Gly Leu Leu Ser Val Asn Leu	145	150	155
25	Pro Thr Gly Leu Pro Thr Gly Tyr Tyr Leu Ala Arg Gln Glu Ile Ile	165	170	175
30	Thr Leu Gln Asn Val Thr Asn Asp Arg Pro Glu Pro Gln Phe Tyr Val	180	185	190
35	Gly Cys Ala Gln Leu Tyr Val Glu Gly Thr Ser Asp Ser Pro Ile Pro	195	200	205
40	Ser Asp Lys Thr Val Ser Ile Pro Gly His Ile Ser Asp Pro Ala Asp	210	215	220
45	Pro Gly Leu Thr Phe Asn Val Tyr Thr Gly Asp Ala Ser Thr Tyr Lys	225	230	235
50	Pro Pro Gly Pro Glu Val Tyr Phe Pro Thr Thr Thr Thr Thr Ser	245	250	255
55	Ser Ser Ser Ser Gly Ser Ser Asp Asn Lys Gly Ala Arg Arg Gln Gln	260	265	270
60	Thr Pro Asp Asp Lys Gln Ala Asp Gly Leu Val Pro Ala Asp Cys Leu	275	280	285
65	Val Lys Asn Ala Asn Trp Cys Ala Ala Ala Leu Pro Pro Tyr Thr Asp	290	295	300
	Gln Ala Gly Cys Trp Ala Ala Ala Glu Asp Cys Asn Lys Gln Leu Asp	305	310	315
	Ala Cys Tyr Thr Ser Ala Pro Pro Ser Gly Ser Lys Gly Cys Lys Val	325	330	335

ES 2 623 288 T3

Trp Glu Glu Gln Val Cys Thr Val Val Ser Gln Lys Cys Glu Ala Gly
 340 345 350
 Asp Phe Lys Gly Pro Pro Gln Leu Gly Lys Glu Leu Gly Glu Gly Ile
 355 360 365
 5 Asp Glu Pro Ile Pro Gly Gly Lys Leu Pro Pro Ala Val Asn Ala Gly
 370 375 380
 10 Glu Asn Gly Asn His Gly Gly Gly Gly Gly Asp Asp Gly Asp Asp Asp
 385 390 395 400
 15 Asn Asp Glu Ala Gly Ala Gly Ala Ala Ser Thr Pro Thr Phe Ala Ala
 405 410 415
 20 Pro Gly Ala Ala Lys Thr Pro Gln Pro Asn Ser Glu Arg Ala Arg Arg
 420 425 430
 25 Arg Glu Ala His Trp Arg Arg Leu Glu Ser Ala Glu
 435 440
 30 «210» 28
 «211» 232
 «212» PRT
 «213» Mycolophthora thermophila
 «400» 28
 35 Met Lys Leu Thr Ser Ser Leu Ala Val Leu Ala Ala Ala Gly Ala Gln
 1 5 10 15
 40 Ala His Tyr Thr Phe Pro Arg Ala Gly Thr Gly Gly Ser Leu Ser Gly
 20 25 30
 45 Glu Trp Glu Val Val Arg Met Thr Glu Asn His Tyr Ser His Gly Pro
 35 40 45
 50 Val Thr Asp Val Thr Ser Pro Glu Met Thr Cys Tyr Gln Ser Gly Val
 50 55 60
 55 Gln Gly Ala Pro Gln Thr Val Gln Val Lys Ala Gly Ser Gln Phe Thr
 65 70 75 80
 60 Phe Ser Val Asp Pro Ser Ile Gly His Pro Gly Pro Leu Gln Phe Tyr
 85 90 95
 65 Met Ala Lys Val Pro Ser Gly Gln Thr Ala Ala Thr Phe Asp Gly Thr
 100 105 110

ES 2 623 288 T3

	Gly	Ala	Val	Trp	Phe	Lys	Ile	Tyr	Gln	Asp	Gly	Pro	Asn	Gly	Leu	Gly	
			115					120					125				
	Thr	Asp	Ser	Ile	Thr	Trp	Pro	Ser	Ala	Gly	Lys	Thr	Glu	Val	Ser	Val	
		130					135					140					
5	Thr	Ile	Pro	Ser	Cys	Ile	Glu	Asp	Gly	Glu	Tyr	Leu	Leu	Arg	Val	Glu	
	145				150						155					160	
10	His	Ile	Ala	Leu	His	Ser	Ala	Ser	Ser	Val	Gly	Gly	Ala	Gln	Phe	Tyr	
					165					170					175		
15	Ile	Ala	Cys	Ala	Gln	Leu	Ser	Val	Thr	Gly	Gly	Ser	Gly	Thr	Leu	Asn	
			180						185					190			
20	Thr	Gly	Ser	Leu	Val	Ser	Leu	Pro	Gly	Ala	Tyr	Lys	Ala	Thr	Asp	Pro	
			195					200					205				
25	Gly	Ile	Leu	Phe	Gln	Leu	Tyr	Trp	Pro	Ile	Pro	Thr	Glu	Tyr	Ile	Asn	
	210						215					220					
30	Pro	Gly	Pro	Ala	Pro	Val	Ser	Cys									
	225					230											
35	<210> 27																
	<211> 302																
	<212> PRT																
	<213> Mycolophthora thermophila																
40	Met	Pro	Pro	Pro	Arg	Leu	Ser	Thr	Leu	Leu	Pro	Leu	Leu	Ala	Leu	Ile	
	1				5					10					15		
45	Ala	Pro	Thr	Ala	Leu	Gly	His	Ser	His	Leu	Gly	Tyr	Ile	Ile	Ile	Asn	
			20					25						30			
50	Gly	Glu	Val	Tyr	Gln	Gly	Phe	Asp	Pro	Arg	Pro	Glu	Gln	Ala	Asn	Ser	
		35					40					45					
55	Pro	Leu	Arg	Val	Gly	Trp	Ser	Thr	Gly	Ala	Ile	Asp	Asp	Gly	Phe	Val	
	50					55						60					
60	Ala	Pro	Ala	Asn	Tyr	Ser	Ser	Pro	Asp	Ile	Ile	Cys	His	Ile	Glu	Gly	
	65				70					75					80		
65	Ala	Ser	Pro	Pro	Ala	His	Ala	Pro	Val	Arg	Ala	Gly	Asp	Arg	Val	His	
					85				90						95		

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

Gln Ala Gly Gly Ala Gln Phe Tyr Met Ser Cys Tyr Gln Met Thr Val
180 185 190

Glu Gly Gly Ser Gly Thr Ala Asn Pro Pro Thr Val Lys Phe Pro Gly
195 200 205

Ala Tyr Ser Ala Asn Asp Pro Gly Ile Leu Val Asn Ile His Ala Pro
210 215 220

Leu Ser Ser Tyr Thr Ala Pro Gly Pro Ala Val Tyr Ala Gly Gly Thr
225 230 235 240

Ile Arg Glu Ala Gly Ser Ala Cys Thr Gly Cys Ala Gln Thr Cys Lys
245 250 255

Val Gly Ser Ser Pro Ser Ala Val Ala Pro Gly Ser Gly Ala Gly Asn
260 265 270

Gly Gly Gly Phe Gln Pro Arg
275

<210> 29

<211> 231

<212> PRT

<213> Mycelophthora thermophila

<400> 29

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

<211> 242
 <212> PRT
 <213> Mycollophthora thermophila
 <400> 30

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

ES 2 623 288 T3

	85	90	95
	Gly Gly Asn His Phe Gly Pro Val Leu Ile Tyr Met Ser Lys Val Asp		
	100	105	110
5	Asp Ala Ser Thr Ala Asp Gly Ser Gly Asp Trp Phe Lys Val Asp Glu		
	115	120	125
10	Phe Gly Tyr Asp Ala Ser Thr Lys Thr Trp Gly Thr Asp Lys Leu Asn		
	130	135	140
15	Glu Asn Cys Gly Lys Arg Thr Phe Asn Ile Pro Ser His Ile Pro Ala		
	145	150	155
20	Gly Asp Tyr Leu Val Arg Ala Glu Ala Ile Ala Leu His Thr Ala Asn		
	165	170	175
25	Gln Pro Gly Gly Ala Gln Phe Tyr Met Ser Cys Tyr Gln Val Arg Ile		
	180	185	190
30	Ser Gly Gly Glu Gly Gly Gln Leu Pro Ala Gly Val Lys Ile Pro Gly		
	195	200	205
35	Ala Tyr Ser Ala Asn Asp Pro Gly Ile Leu Val Asp Ile Trp Gly Asn		
	210	215	220
40	Asp Phe Asn Glu Tyr Val Ile Pro Gly Pro Pro Val Ile Asp Ser Ser		
	225	230	235
	Tyr Phe		

	atgtcaaagg cctctgctct octogctggc ctgacgggog oggcctctgt cgtgcacat	60
	ggccacgtca gccacatcgt cgtcaacggc gtctactaca ggaactaaga ccccaagaca	120
5	gactggtaac agcccaaccc gccaacagtc atcggctgga oggcagcaga tcaggataat	180
	ggcttcgttg aacccaacag ctttggcacg ccagatataa tctgcacaa gagcgccac	240
	ccggggggcg gccacgtac cgttgctgac ggagacaaga tcaacatcgt ctggacccc	300
10	gagtggcccg aatccacat oggcctcgtc attgactaac tagcgcctg caacggtgac	360
	tgcgagaccg tcgacaagtc gtcgctgogc ttggttcaaga ttgacgggoc oggctacgac	420
15	aaggcgcgcg gccctgggc ogcgacgt ctgcgcgca aggcacacag ctgctcgtc	480
	cagatccctg cggatctcaa ggcgggcaac taagtctctc gccacgagat catcgccctc	540
20	caaggctgctc agagcccaaa oggcggccag gctacccgc agtgcatcaa cctcgcgctc	600
	acggggggcg gcagcaact gccagggc gtgcgggca octcgtgta caaggcgacc	660
	gaacggggca tctcttcaa cccctacgtc tctccccgg attacacgt cccgggccc	720
25	gcctctattg cgggcgcgc cagctcgatc gccagagca cgtcggctgc cactgccacc	780
	ggcacgggca cgttcccg oggcggggc gccaaccta ccgccaccac caccgcgcc	840
30	acctcgcgc ccccgagac caccctgagg acgaccata cctcggcgc gcagactacc	900
	gcccgcctt cggcgatgt gcagaccaag tacggccagt gtggtggca cggatggacg	960
	ggccgaagg tctgcgccc ogctcagac tctccgtc tcaacgagt gtactccag	1020
35	tgtttgtaa	1029

<210> 32

<211> 837

<212> DNA

<213> Mycelophthora thermophila

<400> 32

	atgtttttctc tcaagtttctt tatetttggec ggtgggcttg ctgtctctcac cagaggctcac	60
5	ataagactag tgtogococg ccctttttacc aacctgaac agggocccag cccactcta	120
	gaggctggca gogactatcc ctgccacaa cgcactggg ggggttatca gggaaogcca	180
	acccagatgg caaagggttc taagcagcag ctagecttcc aggggtctgc cgttcctggg	240
10	ggtggctect gccagtgtc catcaactac gacgaaaa cagccgctca gagctecttc	300
	aaggtcattc actcgattca aggtggctgc cccgccagg ccgagacgat cccggattgc	360
15	agcgcacaaa atatcaacgc ctgcaatata aagcccgata atgccagat ggacaccccg	420
	gataagtatg agttcaogat ccgggaggat ctccccagt gcaaggccac cctogocctg	480
	acatggatca acactatcgg caacogcgag ttttatatgg catgogcccc ggttgagatc	540
20	acgggogacg ggggtagcga gtgggctctg gctgogctgc ccgacatggt cattgocaa c	600
	atcccgctca toggaggaac ctgcgcgacc gaggaggga agtactaca ataccacac	660
25	cccggtaagt cggtcgaac catccggggc tggaccgat tggttccct gcaaggcgaa	720
	tggggtctgc cctccggtgt ctgggctcc ggcggaaac ccagcagtgc taccctgccc	780
30	gcaggggccg ccccgactcc tctgtctcgc ggcgcgcgc ccacctggaa cgcctaa	837
	<210> 33	
	<211> 872	
	<212> DNA	
35	<213> Mycolophilona thermophila	
	<400> 33	
40	atgaagctcg ccaagctctt ogccgcctc acctcgggg tggccgacca gctcagcgtc	60
	gggtccagaa agtttgccgt gtacgagcac attcgcaaga acacgaacta caactcgccc	120
45	gttaccgacc tgtoggacac caacctgccc tgcacgtcg gggggggctc gggcaccagc	180
	acacccgtgc togaagtcaa ggccggagac togttcacct tcttcagcga cgttgccgtc	240
	taaccacagg ggccatctc gctgtgctg gacgggacca gtgcagagag catggatgga	300
50	ogggaaacgg acatgocctg ccgaactggc tcacaagctg gctacctggc ggtgactgac	360
	taagacgggt ccggtgactg tttcaagatc tatgactggg gacgacgtt caacgggggc	420
55	caggcgctgt ggcgcagcag gaattcgta gagtacagca tctcaagtg catcaggga c	480
	ggcgaatacc tactgoggat taagtccttg gccatcata acccaggtgc ctttcgcag	540
	ttctacatca gctgcgocca ggtgaatgtg accgggggag gcacogtcac cccgagatca	600
60	aggcgacoga tctgatcta tttcaacttc cactcgtata togtccctgg gcccgcagt	660
	ttcaagtgt ag	672
65	<210> 34	

<211> 579
<212> DNA
<213> *Mycobacterium thermophilum*

<400> 34

5 atgaccaaga atgagcagag caagcagggc gttgagaacc caacaagggg cgacatccgc 60
tgctaacacat cgcagacggc ggcacaacgtc gtgacccgtc cggccgggctc gacattccac 120
10 tacatctoga cccagcagat caacccacccc ggcocgactc agtactaacct ggccaaggta 180
ccccccgggt cgtgggcca gacctttgac gggtcgggg cgtcttggtt caagatctcg 240
15 accacgatgc ctacccgtga cagcacaacg cagatgttct ggcacgggca gaacacttat 300
gagacctcaa acacacccat tcccgccaa accccggagc ggcagtaact ccttcggctc 360
aagcagatcg cctccacat gggtctcag cccaacaagg tccagttcta ccttcgctgc 420
20 accagatca agatcacggc tggtcgcaac ggcacccca gcccgctggc cgcgctgcc 480
ggagcctaca agagcacga ccccgccatc ctggtcgaca tctactccat gaagcccgaa 540
25 tctaacagc ctcccgggc gcccgctctg cgggctaa 579

<210> 35
<211> 852
<212> DNA
<213> *Mycobacterium thermophilum*

<400> 35

35 atgaagccct ttacccctgt cgcctcggc actgcctga ggcgcatgc catcttcag 60
cgggtgtcgg tcaacgggca ggcacgggc cagctcaagg gggtcgggg gccgtcagc 120

40

45

50

55

60

65

	aactccccga	tccagaaagt	caacgatgcc	aacatggcct	gcaacgccaa	catttgtgtac	180
	caagacaaca	ccatcatcaa	ggtgcccgcg	ggagcccgcg	tgggcggtg	gtggcagcac	240
5	gtcatcgggg	ggcgcagggg	cgccaacgac	cgggacaaac	cgatgcgcgc	ctcccccag	300
	ggcccccacc	aggtctacct	ggccaaagtg	gacaaacggg	cgacggcgtc	gcgctgggga	360
10	ctcaagtggg	tcaaggtggc	cgagcgcggc	ctgaacaaag	ggtgtgtggc	ctacctgatg	420
	cgcgctgagc	tgtctgcacct	gcacagcgcc	togagcccg	gcggcgccca	gttctacatg	480
	ggctgtgcac	agatogaagt	cactggctcc	ggcaccaaact	cgggctccga	ctttgtctcg	540
15	ttccccggcg	cctaactggc	caacgacccg	ggcatcttgc	tgagcatcta	cgacagctcg	600
	ggcaagccca	acaatggcgg	gcgctcgtac	cagatccccg	gcgcgcgcgc	catctctgc	660
20	tccggcagcg	ggggcgggcg	caacaaacgg	ggcgacggcg	gcgacgacaa	caacggtggg	720
	ggcaacacaa	acggcgggcg	cagcgctccc	ctgtacgggc	agtgcggcg	cctgggctac	780
	acgggcgcga	ccacctgtgc	ccagggaaact	tgcagggtgt	cgaaacgaata	ctacagccag	840
25	tgcctccct	ag					852
	<210> 36						
	<211> 639						
30	<212> DNA						
	<213> Mycollophthora thermophila						
	<400> 36						
35	atgaagctca	cctcgtccct	cgtgtctctg	gcgctgcgcg	gcgcaccaggc	tcactatacc	60
	ttccctaggg	cgggcactgg	tgggttcgtc	tctggcgagt	gggaggtggg	cgcctgacc	120
40	gagaacacat	actgcacagg	ccgggtcaac	gatgtcaca	gcgcgcagat	gaactgctat	180
	cagtcggggg	tgcagggtgc	gcgcacgacc	gtccaggtca	aggcggggtc	ccaattcacc	240
45	ttcagcgtgg	atccctccat	cgccaccccc	ggccctctcc	agttctacat	ggctaagggtg	300
	ccgtcggggc	agacggcgcg	caactttgac	ggcccgggag	ccgtgtgggt	caagatctac	360
	caagaaggcc	cgaaacggct	cgccacccgc	agcattacct	ggccacggcg	cgccaaaacc	420
50	gaggtctcgg	tcaccatccc	cagctgcata	gaggatggcg	agtaacctgt	ccgggtcgag	480
	cacacccccc	tcctacagc	gcacgcagcg	caaacccgag	ctcgtctcgt	accatcccca	540
55	gctgcataca	aggccaccca	ccggggcctc	ctcttcacag	tctactggcc	catcccgacc	600
	gagtaacata	accccgccgc	ggcccccgtc	tcttgcata			639
	<210> 37						
	<211> 447						
60	<212> DNA						
	<213> Mycollophthora thermophila						
65							

«400» 37

atgcgcgcac caagactgag caacctctctt cccctctctag ccttaataagc ccccacccgc 60

5 ctggggcact cccactctgg gtacatcctc atcaacggcg aggtatacca aggattcgac 120

ccgcggccgg agcaggogaa ctgcgcgttg cgcgtgggct ggtcgacggg ggcaatcgac 180

10 gaagggttct tgggcgcggc caactactcg tgcgcgcgaa tcatctgcca catcgagggg 240

gccagccgcg cgggcgcgcg gcccgctcgg gggggcgacc gggcgccagt gcaatgggac 300

15 ggtcgccgcg tcggacacgt ggggcgcgtg ctgtcgtaac tggcgccctg cggcgggctg 360

gaggggtcgg agagcgggtg cggcggggtg gacaagcggc agctcggggtg gaccaagggtg 420

20 gaagactcgc tgcgggggat ggagctg 447

«210» 38

«211» 453

«212» DNA

«213» *Mycelophthora thermophila*

«400» 38

30 atgaggctga cattggccgg tgcctcgcca gccatcgctg ctacagaaagt agccggccac 60

gccacgtttc agcagctctg gccaggctcc tctgtgtcc gccttcgggc tagcaactca 120

35 cccgtcacca atgtgggaag cagagacttc gtctgcaacg ctggcaccgg ccccgctcgt 180

ggcagtgccc cgttggaagg tggcgggcac gtcaccatcg agatgcacca gccaccgggc 240

40 gaccgcagct gcaacaacga agccatcgga ggggcgcatt ggggcgccgt ccagggtgtac 300

ctgaccaagg ttcaggacgc cgcgacggcc gacggctcga cgggctggtt caagatcttc 360

45 tcgactcgt ggtccaagaa gcccgggggc aacttggggc acgacgacaa ctggggcacg 420

cgcgacctga acgctgctg cgggaagatg gac 453

«210» 39

«211» 738

«212» DNA

«213» *Mycelophthora thermophila*

«400» 39

	atgtctctctcc tcacctctagc cacactctgtc acctctctctgg cggcgccacgt ctctggctcac	60
	gcctcggtgtg tctcggtctc tgtctgacggg aaagaccagg ggcaggggt gaacaagtac	120
	atcgcctcgc cggcgaccaa cgaacctgtg cggcacctct cggcgccgcg catcgtgtgc	180
5	aaaccccagg ggtccaaagg cgcacctggc ttctgtcagg cggcgccgcg cgcacaagtg	240
	accttctctc gggcgccaga caacctggc gacctcggtg actaagtcct cgaacctgac	300
10	cacaagggcg ccatctctgac ctactctgcc gctacctct ccgggggacc gacctggccc	360
	atctggagca agcttgccga ggaaggattc acctgggggc agtggggcac catcaagatg	420
	atcgacaaag ggggcaaggt cgaagtgacg ctgcccaggg cctttggccc gggaaaagtac	480
15	ctgatccgcc aggagctgct ggcctctgac cgggcagact ttgcctgoga cgaacctggcc	540
	caaccccaacc gggcgccga gtctgacccc aactggtctc aggtggaggt gtccggccagc	600
20	ggcgacaaga agcgggacca gaactttgac ttcaacaagg gctatacctg cgataacaaa	660
	ggaactccact ttaagatcta catcggtcag gacagccagt atgtggcccc ggggcgcggg	720
	ccttggaatg ggagctga	738
25	<210> 40	
	<211> 600	
	<212> DNA	
30	<213> Mycolophthora thermophila	
	<400> 40	
	atgttcaactt cgttttgcat cacagatcat tggaggactc ttagcagcca ctctggggcca	60
35	gtcatgaaet atctcgccca ttgcacaaat gacgactgca agtctttcaa gggcgacagc	120
	ggcaagctct ggtcaagat cgaacagctc ggtacaaac cgtcagccaa cccccctgg	180
40	ggtcttgacc tctcctgtga gcaagggtgc aagtggaaag tgacgatccc gccacgtctt	240
	gtcccggggc aatatctgct ggggcacgag atcctggggt tgcaagtgc aggaacogtg	300
	atgggggccc agttctacc cggctgcacc cagatcaggg tcacogaagg cgggagcacg	360
45	cagctgcctt cgggtattgc gctcccaggc gcttacggcc cacaagacga ggttatcttg	420
	gtcgacttgt ggaggggtta ccagggccag gtcaactaca cggcgcttg aggaaccttt	480
50	tggagogaag cgtgggacac cgaatttggc gggccaaca cgacogagtg cgcacccatg	540
	ctcgacgacc tgcctgacta catgggggccc aacgacaggt ggatoggctg gacgggctag	600
55	<210> 41	
	<211> 684	
	<212> DNA	
	<213> Mycolophthora thermophila	
	<400> 41	
60		
65		

	atgaagctga gogctgocat cgcctgtctc ggggcggccc ttgcggaggg gaactatacc	60
5	ttccccagca togccaacac ggccgactgg caatatgtgc gcatcaagac caacttccag	120
	agcaacgggc cagtgaagga cgtcaactcg gaccagatcc ggtgctaaga ggcgaacccg	180
10	ggcaacgggc ccccgggcat ctacaacgtc aaggccggga caaccatcaa ctacaacgoc	240
	aagtgttoca ttccccaccc gggaccatg gcctttctaca ttgcaagggt tcocggcggc	300
	cagtggcgcg ccacctggga cggtaaggga gcctgtctgtt caagatoca ccaggagatg	360
15	ccgcactttg gcaccagcct cacttgggac tcaaacgggc gcacctccat gcccgtaac	420
	atcccccgct gtctgcagga cggcgagtat ctgttgcgtg cagagcacat tgcctctcac	480
20	agcgccggca gccccggggg cggccagttc tacatttctt gtgccagct ctacgtcaac	540
	gggggcagcg ggacctggaa ccccaggaa aaggtgttgt tcccggggc ctacaaggcc	600
	actgaacggg gcatctgat caacatctac taacccgtcc cgaactagcta cactcccgct	660
25	ggccccccg tgacacctg ctac	684
30	<210> 42 <211> 768 <212> DNA <213> Mycolophthora thermophila	
35	<400> 42	
	atgtaccgca cgtctgggtc cattgccctg ctgcgggggg gogctgcgcg ccacggcgcc	60
40	gtgaccagct acaacattgc gggcaaggac taacctggat actcgggctt ggcacctaac	120
	ggccaggatg tcatccagtg gcaatggccc gactataacc cgttctctgc cgcacagagc	180
45	cccagctcc gctgcacgg cggcaccggg ggggcgtgtt atgcggaggg ggcccccggc	240
	gacaccatca cggccacctg gggccagtgg aagcactccc agggcccgat cctggtgtgg	300
	atgtacaagt gccccgggca ctacagctcc tgcgacggct ccggcgcggg ttggttcaag	360
50	atogaagagg ccggcttcca cggcgacggc aagacggtct tcttogaac ccgagacccc	420
	tggggctggg acattgocaa gctggctggg ggcacaaagt cgtggagcag caagatccct	480
55	gaaggcctcg cccggggcaa ttacctgttc cgcacagaga tcatogccct gcaacaggcc	540
	aacacccgc aattctacc cagtgcgcc cagatcaagg tcaacggctc tggcaacgoc	600
	gagcccgccg cctctacaa ggcgcgacac cccgctact gccagcagag cgaacccaac	660
60	atttctttca acatcaacga ccactccctc ccgcaggagt acaagatccc cggccccccg	720
	gtcttcaagg gcacgcctc cgcacaggct cgcgctttcc aggcctaa	768
65	<210> 43	

<211> 678
 <212> DNA
 <213> Mycollophilora thermophila
 <400> 43

5	atgctgacaa caaccttgcg cctcctgacg ggcgtctctg ggcgcagcgc ccattatacc	60
	ctccccaggg tggggacggg ttccgactgg cagcacgtgc ggogggctga caactggcaa	120
10	aacaacggct tegtggcgga cgtcaactcg ggcagatca ggtgcttcca ggagaccct	180
	gcggggcgcc aagaagtcta cactgttcag gggggatcga ccgtgaacta ccacgcacac	240
	cccagtatct accaccccg gcccctgcag ttctacctgg ccggcgcttc ggacggacag	300
15	gaagtcaagt cgtggacggg cgagggtgac gtgtggttca aggtgtacga ggagcagcct	360
	caatttggeg ccagctgac ctggcctagc aacggcaaga gctcgttoga ggttccctac	420
20	cccagctgca ttggggcggg caactacctc ctccgcgtcg agcacatcgc cctgcacgtt	480
	gccccaaagc agggcgcgcg ccagttctac atctcgtcgg ccagctcca ggtcactggt	540
25	gggggcagca ccagaccttc tcagaagggt tcttcccggt gtgcctacaa gtccacggac	600
	ccgggcattc ttatcaacat caactacccc gtccctacct cgtaccagaa tcggggtcgg	660
30	gctgtcttcc gttgctaa	678

<210> 44
 <211> 714
 <212> DNA
 <213> Mycollophilora thermophila
 <400> 44

40

45

50

55

60

65

	atgaaggttc togogccoct gattctggcc ggtgcgcga ggcgccacac catcttctca	60
	tcocctogagg tggggggggt caaccagggc atcgggcagg gtgtccgggt gcctgtgtac	120
5	aaoggtcoga togaggaggt gaogtccaac togatcgctt gcaaggggcc ccccaaccog	180
	acgaagcoga ccaacaaggt catcagggtc cgggcggggc agacgggtgac ggcctgtctg	240
10	cgttacatgc tgagcaccac cggctcggcc cccaacgaca tcctggacag cagccacaag	300
	ggcccgacca tggcctacct caagaaggtc gacaacgcca ccaccgactc gggcgtcggc	360
	ggcggctggt tcaagatcca ggaggacggc cttaaccaag gctctgggg caccgagcgc	420
15	gtcatcaacg gacagggcgc ccacacatc agatccccc agtgcctgc ccccggcag	480
	tacctctccc ggcgcagat gcttgccctg caggagctt ccaactacc cggcgctcag	540
20	ttctacatgg agtgcgcga gctcaatata gtgggggcca ccggcagcaa gacgcctcc	600
	accgtcagct tccggggcgc ttacaagggt accgacccc gagtcaagat caacatctac	660
	tggcccccgc tcaccagcta ccagattccc ggcccccggc tgttcacctg ctca	714
25	<210> 45 <211> 741 <212> DNA <213> <i>Mycelophthora thermophila</i>	
30	<400> 45	
	atgaagctct cctcttttcc cgtcctggcc actgcctcca ccgtcagagg gcatgacatc	60
35	ttccagaagg tctcgtcaa cggagcggac cagggtctcc tcaccggcct ccgogctccc	120
	aaacaacaac accccgtgca gaatgtcaac agccaggaca tgatctggcg ccagtcggga	180
40	togaogtoga acactatcat cggagtcacg gcggcgata ggatcgggtc ctggtatcag	240
	cctgtcatcg ggggtgcaca gtcccccacc gaccagaca acccgattgc caagtcgac	300
	aagggccccc tcctggcta cctgcacaag gttgacaatg ccgcaacgcg cagcaagaag	360
45	ggcctgaagt ggttcaagat ttgggaggat acctttaatc ccagcaccac gacctggggt	420
	gtcgacaacc tcctcaaca caacggctgg gtgtacttca acctccgca gtgcctggcc	480
50	gaaggcaact acctctcctc cgtcagagtc ctgcctctgc actcggcta ctccagggc	540
	caggctcagt tctaccagtc ctgcgcccag atcaacgtat ccgggggggg ctcccttcag	600
	ccggcgtoga ctgtcagctt ccggggtgca tacagcgcca ggaacccgg tatcctgate	660
55	aacatctacg ggcgcacgg ccagcccagc aacaaggcc agccgtacac tgcacctggg	720
	cccgcgccca tctctgctg a	741
60	<210> 48 <211> 765	
65		

<212> DNA

<213> Mycellophthora thermophila

<400> 48

5 atggcoctcc agctcttggc gagcttggcc ctctctctcag tgcgggcoct tgcocacggc 60
 ggcttggcca actacacogt oggtgatact tggtaacagag gctacgaccc aaacctgcog 120
 ccggagacgc agctcaacca gaactggatg atccagoggc aatgggocac catcgacccc 180
 gtcttcacog tgtcggagcc gtacctggcc tgcacacacc cggggcgccg gcgcacctcg 240
 10 tacatcccca tccgcgcggg tgacaagatc accgcogtgt actggtactg gctgcacgcc 300
 atcgggcccc tgagcgtctg gctcgcggcg tgcggcgaca cggccgcggc cgactgcgcg 360
 15 gaogtgcag tcaacogggt oggtcgttc agatctggg agggcggcoct gctggagggc 420
 cccaaactgg ccgaggggct ctggtaccaa aaggacttcc agogctggga oggctccocg 480
 tccctctggc cgtcgcagat ccccaagggg ctcaagagcg ggacctacat catccggcac 540
 20 gagatcctgt cgttcacogt cgcctcacc cccagtttt acccgagtg tgogcatctg 600
 aatattactg ggggcggaga ctctctgcca ccgaagaga ctctggtggg gtttcggggg 660
 gtttacaaag aggcagatcc ctctatcttc atcgatgtct actcggagga gaacgcgaac 720
 25 oggacagatt ataoggttcc gggagggcca atctgggaag ggtga 765

<210> 47

<211> 708

<212> DNA

<213> Mycellophthora thermophila

<400> 47

35 atgaaggccc tctctctect tgggctgcc ggggcagttc ctgcgcatac catcttcgtc 60
 cagctogaag cagacggcac gaggtacccg gtttcgtacg ggatccggga cccaaactac 120
 40 gaogggcccc tcaacgagct cactccaac gaogttgctt gcaagggcgg tcgaaacccg 180
 acgaccccoct ccagcgagct catcacgtc aocgggggca ccacogtcaa ggcatctggg 240
 aggcacaccc tccaatccgg ccgggagat gtcatggag ccagocacca gggcccgacc 300
 45 ctggcctaca tcaagaaggc cggcgatgcc accaaggact cgggcgtcgg cggcggctgg 360
 ttcaagatcc agggggacgg ttacaacaa cggcagtggg gcaccagcac cgttatctcc 420
 50 aacggcgggc agcactacat tgacatcccg gctgcaccc ccgagggtea gtacctctc 480
 cggcgcgaga tgatcgcoct ccacggggcc gggcccccog cgggcgctca gctctacatg 540

55

60

65

	gaatgtgccc agatcaacat cgtcggcggc tccggctcgg tgcccagctc gacggtcagg	600
	ttccccggcg cgtatagccc caacgacccg ggtctctcca tcaacatcta ttccatgtcg	660
	ccctcagact cgtacaccat ccggggcccg ccggttttca agtgctag	708
5	<210> 48 <211> 915 <212> DNA <213> Mycollophthora thermophila	
10	<400> 48	
	atgaagtgtt ctaccccggc cttgttcggc gctgggctcc ttgctcagca tgctggggcc	60
15	cactccatct tccagcaggc gacgagcggc tggacgact ttgataagct gtgcaccogg	120
	atgcgcgcca acaatagccc cgtcactagt gtgaccagcg ggcacatgac ctgcaaaagtc	180
	gggggcaaca agggggtgtc cggcttctgc gagggtgaaag cggcgagcga gttcaagggt	240
20	gagatgcacg cgcagcccgg cgcagcctcg tgcgcacaag aggcctcgg cgggaaccac	300
	ttcggcccggt tctcatcta catgagcaag gtgcagcag cctccaccgc cgcggggtcc	360
25	ggcgactggt tcaagggtga caggttcggc taagacgcaa gcaccaagac ctggggcacc	420
	gacaagctca acgagaactg cggcaagcgc accttcaaca tccccagcca catccccgg	480
	ggcgactatc tggtcggggc cgaaggctatc ggcctacaca ctgccaaaca gccaggcggc	540
30	ggcgagttct acatgagctg ctatcaagtc aggatttccg gggggaagg gggccagctg	600
	cctgcoggag tcaagatccc gggcgcgtaa agtgccaaag acccgggcct ccttgctgac	660
35	atctggggta acgatttcaa cgaacctcca ggacactcgg ccggtcagc cctcatcctc	720
	atcagcagca gcagcaacaa cagcggcgac aagatgacca agaagatcca ggagccacc	780
	atcacatcgg tcacggacct cccaccgac gaggccaagt ggatcgcgct ccaaaagatc	840
40	tgttaagtgg accagacggg cagggcgagg acatacagc cggcgctcgg caagacggg	900
	tgcacaagag tctag	915
45	<210> 49 <211> 972 <212> DNA <213> Mycollophthora thermophila	
50	<400> 49	
55		
60		
65		

	atgaagttoct	toacootcanc	oactotggcc	gocotggctg	gcacogccgc	ogctcacgg	60
	acottocagg	cactotgggt	ogacggggtc	gactacggcg	ogcagtgtgc	cagtctgcc	120
5	gggtccaaet	ggcgggtcac	ogacgtgacc	tccacggcga	tccgttgcaa	ggccaaaccc	180
	tcgcccgcctc	ggggcaagtg	cccggtcaag	gcccgtctga	ccgttaagggt	cgagatgcac	240
10	cagcaacccg	gtgacccgctc	gtgcagcagc	gaggcgatcg	ggggggcgca	ctacggccccc	300
	gtgatgggtgt	acatgtocaa	gggtgtcggac	ggggcgctgg	cggaacgggtc	gtcgggctgg	360
15	ttcaagggtgt	tccaggacgg	ctggggcaca	aacccgtccg	gggggtcggg	cgacgacgac	420
	tactggggca	ccaaggacct	gaactcgtgc	tgcgggaaga	tgaacgtcaa	gaccccgcc	480
20	gaactgcctt	ggggcgacta	cctgctccgg	gcccaggccc	tccgcttgca	cacggccggc	540
	agcgccggcg	ggccccagtt	ctacatgacc	tgtaccagc	tcaccgtagc	cggtccggc	600
	agcgccagcc	ggccccacgt	ctccttcccg	ggcgccaca	aggccaccca	ccggggcctc	660
25	ctcgtcaaca	tccacggccc	gctgtccggc	tacaccgtag	ccggcccggc	cgtctactcg	720
	gggggtctca	ccaagaaggc	cgccagcgcc	tgcacgggt	gcaggtccac	ttggcgcgctc	780
30	ggctccggcc	ccacggccac	cgtctccag	tccccgggt	ccacggccac	ctggccccc	840
	ggggggggcg	ggggtgcac	cgtccagaa	taccagcagt	gggggggcca	gggtacacc	900
	ggctgcacca	actgcggctc	cggtccacc	tgcagcgagg	tctcgccgcc	ctactactcg	960
35	cagtgcgtct	aa					972

<210> 50

<211> 870

<212> DNA

<213> *Mycelophthora thermophila*

<400> 50

	atgaagggac tectoggegc egcgcgcctc tgcctggeg tcagogatgt ctgggccac	60
5	tacatctttc agcagctgac gaaggggggc gtcaggcaag ctgtgtacca gtacatcgc	120
	aagaacacca actataactc gcccgtagcc gatctgacgt ccaaogacct cgcctgcaat	180
10	gtgggtgcta cgggtgggg caacgatacc gtcacgggtg ggcggggga ttggttcacc	240
	ttcagcaccg atacgcgcgt ttaccaccag ggcgcgacct cgatctacat gtccaaggcc	300
	ccgggcagcg cgtccgacta cgaaggcagc ggcggctggt tcaagatcaa ggactgggct	360
15	gactacaccg ccacgattcc ggaatgtatt ccccccggcg actacctgct tgcctccag	420
	caactcggca tcacaaacc ttggcccggc ggcctccccc agttctacat ctcttggtgc	480
	cagatcaccg tgactggtgg cggcagtgcc aaccccggcc cgcagctctc catcccaggc	540
20	gccttcaagg agaccgacc ggcctacact gtcacatctt acaacaactt ccacaactac	600
	acgctccctg gcccgccgt cttaacctgc aacggtagcg gggcacaac ogggggcggc	660
25	tccacccag tcaccaccac caccaccacc accaccaggc cgtccaccag caccgccag	720
	tcccagccgt cgtcgagccc gaccagcccc tcagctgca ccgtcgcgaa gtggggccag	780
	tggggaggac agggttacag cggctgcacc gtgtgcggcg cggggtcgac ctgccagaag	840
30	accaacgact actacagcca gtgctttag	870

<210> 51

<211> 1041

<212> DNA

<213> Mycellophilus thermophila

<400> 51

	atgttcttct	tcacctccaa	gggtctcctt	tcgcctctca	tgggoggggc	aacggttgcc	60
5	gcacaaggta	agttcaccaa	catcgtcato	aacggggtct	cataccagaa	cttogaccca	120
	ttcacgcacc	cttatatgca	gaacctctcg	aagggttgctg	gctggacccg	gagcacacag	180
	gacaaagggt	tgttcggccc	cgagtccttc	tctagcccg	acatcatctg	ccacaagtc	240
10	gccaccaacg	ctgggggcca	tgcctcctc	ggggccggcg	ataaggctct	catccagtgg	300
	gacacctggc	cagagtcgca	ccacggctcg	gtcatcgact	atctcgccga	ctggggcgac	360
15	gggggctgcg	agaaggctga	caagaccacg	ctcaagttct	tcaagatcag	cgagtcgggc	420
	ctgtctgacg	gcactaacgc	cccgggcaag	tggggttcgg	acaagctgat	cgccacacac	480
	aactcgtggc	tggtcacgat	ccggcccaac	atcgcccgcg	gcactaagt	cctggcgccac	540
20	gagatcatcg	cctgcacacg	cgccggccag	cagaacggcg	cccagaacta	cctcagtgcc	600
	ttcaacctgc	aggtcacccg	ctccggcact	cagaagccct	ccggcgctct	cggcaccgag	660
25	ctctacaagg	ccaccgacgc	cgccatcctg	gccaacatct	acaactcgcc	cgtaacctac	720
	cagatccccg	gcacggccat	catctggggc	gcctccgcgg	tccagcagac	cacctggccc	780
	atcacggcct	ctgtctagcg	catcacgggc	tccgtctacg	ccggcgccac	ggctgcacac	840
30	accacggccg	cggcgccggc	caccactacc	accacggctg	gctccgggtg	taccggccacg	900
	ccctcgacgg	ggggtctctc	ttcttcggcc	cagcctgctc	ctaccacggc	tgcggtctac	960
35	tccagccctg	ctcgcccgac	ccgttgcgct	ggtctgaaga	agcgccgctg	ccacggccgt	1020
	gaagtcaagg	ttgcctctca	a				1041
40	<210> 52 <211> 714 <212> DNA <213> Mycollophilora thermophila						
45	<400> 52						
	atgaagacgc	tgcgcgcct	cgtggtctcg	gcgcgcctcg	tggcggcgca	cggtatggtt	60
50	gaacaagcca	cgatcgggtg	caaggattat	cagttctaac	agccgtacca	ggacccctac	120
	atgggggaca	acaagcccca	taggggtttc	cgctccatcc	cgggcaacgg	cccgtggag	180
	gaagtcaact	ccatcgacct	ccagtgcacc	gcgggtgcgg	aacgggcca	gtccacggcc	240
55	ccgcgcggcg	cgggtcgac	cgtgacgctc	tactggacc	tctggcccca	ctcccacgtc	300
	ggccccgtca	tcacctacat	ggctcgctgc	ccgacacgg	gctgcccagg	ctggtcccg	360
60	ggaactaagg	cgttttggtt	caagatcaag	gaaggcgccc	gtgagggcac	ctccaatacc	420
	ccgtctcatga	cggcccccctc	cgctacacc	tacacgctcc	cgtcctgctt	caagagcggc	480
65							

tactaacctcg tccgccacga gatacatgcc ctgcactcgg cctggcagta ccccgccgcc 540
 cagttctacc cgggctgcca ccagctccag gtcacccggg ggggtccac cgtgccctct 600
 accaacctgg tctccttccc cggcgctac aaggggagcg accccggcat cactacgac 660
 5 gcttacaagg cgcacctta caccatcct ggcccgcccg tgtttacctg ctga 714

<210> 53
 <211> 678
 10 <212> DNA
 <213> *Mycelophthora thermophila*

<400> 53
 15 atgagatact tectccagct cgtcgccgc ggggcctttg ccgtgaacag cggcgccggt 60
 cactacatct tocagcagtt ccgacccggc ggggtccagt acccgccctg gaagtacatc 120
 20 cggcgcaaca ccaaccgga ctggtgcag aacgggcggg tgaaggacct gtgtcgacc 180
 gacctgacct gcaacgtgg cgggcaggtc agcaacggga ccgagacct cacttgaac 240
 25 gccggcgagc agttcagctt catctcgac agcccgctct accatgcgg cccacctcg 300
 ctctacatgt ccaaggcgcc cggagctgtg gccgactaag accggcgccg ggcctggttc 360
 aagatctacg actggggtcc gtgggggacg agctggagct tgagtggcac gtacactcag 420
 30 agaattccca agtgcctcc tgacggcgag taactctctc gaatccagca gatggggctc 480
 cacaaccccg ggcgcgcgcc acagttctac atcagctgg ctcaagtcaa ggtctctgat 540
 35 gggggcagca ccaatccgac ccgacccgc cagattccgg gagccttcca cagcaacgac 600
 cctggcttga ctgtcaatat ctacaacgac cctctacca actacgtct cccgggacct 660
 40 agagtttgcg actggtag 678

<210> 54
 <211> 1335
 <212> DNA
 45 <213> *Mycelophthora thermophila*

<400> 54

50

55

60

65

5 atgcacccct cctttctttt caogettegg ctggcgagcg tgcttgtecc cctctogtct 60
gcacacacta ccttcaogac cctctctogtc aacgatgtca accaagggtga tggtaoctgc 120
10 attogcatgg cgaagaaggg caatgtogcc aoccatcttc togcaggogg totogactoc 180
gaagacatgg cctgtggtcg gcatggtcac gaacccgttg catttaogtg tocgggccca 240
gctggtgcca agttgactct cgaatttctc atgtgggccc atgcttcgca gtcgggatcg 300
15 atogatccat cccaccttgg cgtcatggcc atctacctca agaaggtttc cgacatgaaa 360
tctgacggcg ccgctggccc gggctggttc aagatttggg accaaggcta cgacttggcg 420
gccaagaagt gggccaccca gaagctcacc gacaacaaag gctctctgag cgtcaacctt 480
20 ccaacccgct taccacccgg ctactacctc gcccgccagg agatcaccac gctccaaaac 540
gttaccaatg acaggccaga gccccagttc tacgtoggtt ggcacacgct ctacgtcgag 600
25 ggcacctogg actcaccocat cccctoggac aagacgggtt ccattcccgg ccacatcagc 660
gaacccggcg aocccgggct gaccttcaac gtctacacgg gogaogcacc caoctacaag 720
ccgcccggcc ccgagggtta ctcccccacc accaccacca ccacctcttc ctctctctcc 780
30 ggaagcagcg acaacaaggg agccaggcgc cagcaaaccc ccgacgacca gcaggccgac 840
ggcctctgtc cagccgactg cctcgtcaag aacgcgaact ggtgcgcgc tgccctgccg 900
ccgtacaccg acagggccgg ctgctgggcc gccgcgagg actgcacaaa gaagctggac 960
35 gctgtctaca ccagcgcacc cccctcgggc agcaagggtt gcaagggtctg ggaaggagcag 1020
gtgtgcaccc togtctogca gaagtgcag gccggggatt tcaaggggccc ccgcagctc 1080
40 ggggaaggagc toggcgaggg gatcgatgag cctattccgg ggggaaagct gcccccggcg 1140
gtcaacggcg gagagaacgg gaatcatggc ggaggtggtg gtgatgatgg tgatgatgat 1200
aatgatgagg cgggggctcg ggcagcgtcg actccgactt ttgctgctcc tggtgggcc 1260
45 aagactcccc aaccaaaactc cgagagggcc cggcgccgtg aggcgcattg gcggcgactg 1320
gaatctgctg agtag 1335

50 <210> 55
<211> 606
<212> DNA
<213> *Mycobacterium thermophila*

55 <400> 55

60

65

	atgaagctca cctgtccct cgtgtctctg ggcgtgcgc ggcgccagga tcactatacc	60
5	ttccctaggg cgggcactcg tggttcgctc tctggcgagt gggaggtggt ccgcctgacc	120
	gagaaccatt actgcacgg ccgggtcacc gatgtcacc gcccgcagat gacctgctat	180
10	cagtcgggcg tgcagggtgc gcccagacc gtccaggta aggagggtc ccaattcacc	240
	ttcagcgtgg atccctccat cggccacccc ggcctctcc agttctacat ggctcaggtg	300
	ccgtcgggac agaaggcgc cacccttgac ggcacgggag ccgtgtggtt caagatctac	360
15	caagacggcc cgaacggcct cggcacggac agcattacct ggcgcagcgc cggcaaaccc	420
	gaggctctgg tcaccatccc cagctgcac gaggatggcg agtaacctgt ccgggtcgag	480
20	cacatcgcc tcacacggc cagcagcgtg ggcggcgccc agttctacat cgcctgcgcc	540
	cagctctccg tcacggcgcg ctccggcacc ctcaaacagg gctcgctcgt ctccctgccc	600
	ggcgctaca aggcaccca ccggggcacc ctcttcacgc tctactggcc catccgacc	660
25	gagtacatca acccggccc ggcgccgta tcttgc	696

30	<210> 56
	<211> 1176
	<212> DNA
	<213> <i>Mycelophthora thermophila</i>
35	<400> 56

ES 2 623 288 T3

	atgcggccac caagactgag caccctcctt cccctcctag ccttaatagc ccccaaccgc	60
	ctggggcaact cccacctcgg gtacatcato atcaagggcg aggtatacca aggattcgac	120
	ccggggccgg agcaggcgaa ctggccgttg cgggtgggtt ggtcgagggg ggcaatcgac	180
	gaagggttcg tgggcgcggc caactactcg tggcccgaca tcatctgcac catcgagggg	240
5	gccagcccg cgggcgcgc gcccgtcgg gggggcgacc gggcgacgt gcaatggaaa	300
	cggtcgccg ctgggacag tggggccgg gtgtgtgtac ctggcgccct ggggcgggt	360
10	ggaggggtcc gagagcgggt ggacgactcg ctggccggga tggagcttgt cggggcccg	420
	ggggggccgg ggggcgagga cgaaggcagc gccagcgag gcagcggcag cggcggcagc	480
	ggacgggtcg ggtgcccgg gcagcgctgg gccacgagc tgttgatcgc ggccaaacac	540
15	agctggcagg togagatccc ggcggggctg cgggaagggc cgtacgtgtt ggcacagag	600
	atcgctcgcc tgcactacgc ggcgagccc gggggcgcc agaactccc gctctgcgtc	660
20	aaactgtggg togagggcgg cgaaggcagc atggagctgg accacttcga cgcacccag	720
	ttctaccggc cgcagacc cggcatcctg ctcaacgtga cggccggcct ggcctcctac	780
25	gcgtgcgcgg gcccgagctt ggcgggggg ggcagccgg tgcgtacgc gcagcagaac	840
	atcagctcgg cagggcgcca tgggaacccc gtgattgtca ccaggagcac ggagacgggt	900
	cccttcaccc cgggaacccc gccagccgag accgcagag ccaaggggg gaggtatgat	960
30	gaccaaacc gaactaaaga cctaaatgaa cgtttctttt atagtaccc gccagacag	1020
	aagaggctga cagcgacctc aagaaggga ctagttgata atcgtaccc gtacctctcc	1080
35	gtagctgtct ggcagatctt cggcgctcat aaggcagcag aaaccaacca cgaagctttg	1140
	agaggcgcca ataagcacca tggcggtgtt tcagag	1176
40	<210> 57 <211> 837 <212> DNA <213> <i>Mycelophthora thermophila</i>	
45	<400> 57	
	atgaggtcga cattggccgg tgcctcgga gccatcgctg ctacagaaat agccggccac	60
50	gccacgtttc agcagctctg gcacggctcc tctgtgtcc gccttcgggc tagcaactca	120
	cccgtaacca atgtgggaag cagagacttc gtctgcaacg ctggcaccgg ccccgtaagt	180
	ggcaagtccc cgtgaaggc tggcgccacc gtcaccatcg agatgcacca gcaacccggc	240
55	gacgcagct gcaacaacga agccatcgga ggggcgcatt ggggccccgt ccaggtgtac	300
	ctgaccaagg ttcaggacgc cggagcgcc gacggctcga cgggctggtt caagatcttc	360
60		
65		

	tcogactcgt ggtccaagaa gcccgggggc aactcggggc adgacgacaa ctggggcaag	420
	cgagacctga agcctcgtg cgggaagatg gacgtggca tcccggcga catcgctcg	480
5	gggactacc tctcggggc cggggcgtg gccctgcaca cggcgggaca ggcgggggc	540
	gccagttct acatgagctg ctaccagatg aggttcgagg ggggtccgg gacggccaac	600
	cggcccaacg tcaagttcc gggcgctac agcgccaacg acccgggcat cctcgtcaac	660
10	atcaagccc ccctttccag ctacacggg ccgggcccgg cgtctacgc ggggggccc	720
	atcccgagg ccggctccg ctgcacggg tgcgcgcaga cctgcaagg cgggtcgtc	780
15	ccgagcgcc ttgccccgg cagcgggcg ggcacgggg gggggtcca acccca	837
	«210» 58	
	«211» 893	
20	«212» DNA	
	«213» <i>Mycelophthora thermophila</i>	
	«400» 58	
25	atgaactatc tgcgccattg caccaatgac gactgcaagt ctttcaaggg cgcagcggc	60
	aaagtctggg tcaagatcga gcagctcgcc tacaaccctg cagccaaccc ccctcgggg	120
30	tctgaactcc tccgtgagca cggtgccaag tggaaagtga cgtcccgcc cagtcttctc	180
	ccggcgcaat atctgctgag gcacgagatc ctgggggttc acgtcgcagg aacgtgatg	240
35	ggcgcccaat tctaccccg ctgcacccag atcagggtca ccgaaggcgg gacacgcag	300
	ctgcctcgg gtattcgct cccaggcgt tacggccac aagacgaggg tatcttggtc	360
	gacttctgga gggttaacca gggccaggtc aactacaagg cgcctggagg accgtttgg	420
40	agogaagcgt gggacacga gtttggcggg tccaaacga ccgagtcgc caccatgctc	480
	gacgaactgc tgcactacat gggggccaac gacgacccat gctgcacga ccgaaccag	540
45	ttggggagtc tgcagcggg gacgaaggcg gccggcggt cgcgagcct gtaagatacc	600
	gtcttggctc ccgttctcca gaagaaagt cgcacaaag tgcagtcgag cggacggcg	660
50	agcgtcaacg gggatgagtt gacagagagg ccc	693
	«210» 59	
	«211» 726	
55	«212» DNA	
	«213» <i>Mycelophthora thermophila</i>	
	«400» 59	
60		
65		

	atgaagtogt ctacccoggc ctgtttogcc gctgggctcc ttgttcagca tgcctggggcc	60
	cactccatct tccagcaggc gaggaggggc tggccgact ttgatacgt gtgcaccogg	120
5	atgcggccca acaatagccc cgtcactagt gtgaccagcg ggcacatgac ctgcaacgtc	180
	gggggcacca aggggggtgtc gggctttctgc gagggtgaaag cgggggagca gttcaagggt	240
	gagatgcacg cgcagccogg cgcacgctcg tgcgccaaag aggcctatcg cgggaaccac	300
10	ttgggcccgg tctccatcta catgagcaag gtgcagagcg cctccactgc cgaagggtcc	360
	gggactcgt tcaagggtga caggttcggc taagaagcaa gcaccaagac ctggggcacc	420
15	gacaagctca acgagaactg cggcaagcgc accttaaca tcccagcca catccccgg	480
	gggactata tggccgggc cagggtata gggtaacca ctgccaacca gccaggcggc	540
	gggcagttct acatgagctg ctatcaagtc aggatttcog gggggaagg gggccagctg	600
20	cctgcgggag tcaagatccc gggcggtac agtgccacag acccggcct ccttgtcgac	660
	atctggggta acgatttcaa cagttacgtt attccgggac ccccggtcat cgcagcagc	720
25	tacttc	726

<210> 60

<211> 870

<212> PRT

<213> Mycellophthora thermophila

<400> 60

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

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

ES 2 623 288 T3

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

ES 2 623 288 T3

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			
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
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
			740					745					750		
Pro	Gln	Pro	Leu	Leu	Arg	Ser	Ser	Gly	Gly	Asn	Ser	Pro	Gly	Gly	Asn
		755					760					765			
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
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
			820					825					830		
Arg	Asp	Leu	Ser	Asn	Trp	Asp	Val	Thr	Val	Gln	Asp	Trp	Val	Ile	Ser
		835				840						845			
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										
865					870										

<210> 61
 <211> 881
 <212> PRT
 <213> Thermosoccus aurantiacus

<400> 61

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

ES 2 623 288 T3

	Ser	Cys	Ser	Asn	Ser	Tyr	Leu	Leu	Asn	Lys	Leu	Leu	Lys	Ser	Glu	Leu
				260					265					270		
	Asp	Phe	Gln	Gly	Phe	Val	Met	Ser	Asp	Trp	Gly	Ala	His	His	Ser	Gly
			275					280					285			
5	Val	Gly	Ala	Ala	Leu	Ala	Gly	Leu	Asp	Met	Ser	Met	Pro	Gly	Asp	Thr
		290					295					300				
10	Ala	Phe	Gly	Thr	Gly	Lys	Ser	Phe	Trp	Gly	Thr	Asn	Leu	Thr	Ile	Ala
	305					310					315					320
15	Val	Leu	Asn	Gly	Thr	Val	Pro	Glu	Trp	Arg	Val	Asp	Asp	Met	Ala	Val
					325					330					335	
20	Arg	Ile	Met	Ala	Ala	Phe	Tyr	Lys	Val	Gly	Arg	Asp	Arg	Tyr	Gln	Val
			340						345					350		
25	Pro	Val	Asn	Phe	Asp	Ser	Trp	Thr	Lys	Asp	Glu	Tyr	Gly	Tyr	Glu	His
		355						360					365			
30	Ala	Leu	Val	Gly	Gln	Asn	Tyr	Val	Lys	Val	Asn	Asp	Lys	Val	Asp	Val
		370					375					380				
35	Arg	Ala	Asp	His	Ala	Asp	Ile	Ile	Arg	Gln	Ile	Gly	Ser	Ala	Ser	Val
	385					390					395					400
40	Val	Leu	Leu	Lys	Asn	Asp	Gly	Gly	Leu	Pro	Leu	Thr	Gly	Tyr	Glu	Lys
				405						410					415	
45	Phe	Thr	Gly	Val	Phe	Gly	Glu	Asp	Ala	Gly	Ser	Asn	Arg	Trp	Gly	Ala
			420						425					430		
50	Asp	Gly	Cys	Ser	Asp	Arg	Gly	Cys	Asp	Asn	Gly	Thr	Leu	Ala	Met	Gly
		435					440						445			
55	Trp	Gly	Ser	Gly	Thr	Ala	Asp	Phe	Pro	Tyr	Leu	Val	Thr	Pro	Glu	Gln
	450						455					460				
60	Ala	Ile	Gln	Asn	Glu	Ile	Leu	Ser	Lys	Gly	Lys	Gly	Leu	Val	Ser	Ala
	465				470						475					480
65	Val	Thr	Asp	Asn	Gly	Ala	Leu	Asp	Gln	Met	Glu	Gln	Val	Ala	Ser	Gln
				485					490						495	
70	Ala	Ser	Val	Ser	Ile	Val	Phe	Val	Asn	Ala	Asp	Ser	Gly	Glu	Gly	Tyr
			500						505					510		

Ile Asn Val Asp Gly Asn Glu Gly Asp Arg Lys Asn Leu Thr Leu Trp
 515 520 525
 Lys Gly Gly Glu Glu Val Ile Lys Thr Val Ala Ala Asn Cys Asn Asn
 530 535 540
 Thr Ile Val Val Met His Thr Val Gly Pro Val Leu Ile Asp Glu Trp
 545 550 555 560
 Tyr Asp Asn Pro Asn Val Thr Ala Ile Val Trp Ala Gly Leu Pro Gly
 565 570 575
 Gln Glu Ser Gly Asn Ser Leu Val Asp Val Leu Tyr Gly Arg Val Ser
 580 585 590
 Pro Gly Gly Lys Thr Pro Phe Thr Trp Gly Lys Thr Arg Glu Ser Tyr
 595 600 605
 Gly Ala Pro Leu Leu Thr Lys Pro Asn Asn Gly Lys Gly Ala Pro Gln
 610 615 620
 Asp Asp Phe Thr Glu Gly Val Phe Ile Asp Tyr Arg Arg Phe Asp Lys
 625 630 635 640
 Tyr Asn Glu Thr Pro Ile Tyr Glu Phe Gly Phe Gly Leu Ser Tyr Thr
 645 650 655
 Thr Phe Glu Tyr Ser Asp Ile Tyr Val Gln Pro Leu Asn Ala Arg Pro
 660 665 670
 Tyr Thr Pro Ala Ser Gly Ser Thr Lys Ala Ala Pro Thr Phe Gly Asn
 675 680 685
 Ile Ser Thr Asp Tyr Ala Asp Tyr Leu Tyr Pro Glu Asp Ile His Lys
 690 695 700
 Val Pro Leu Tyr Ile Tyr Pro Trp Leu Asn Thr Thr Asp Pro Lys Lys
 705 710 715 720
 Ser Ser Gly Asp Pro Asp Tyr Gly Met Lys Ala Glu Asp Tyr Ile Pro
 725 730 735
 Ser Gly Ala Thr Asp Gly Ser Pro Gln Pro Ile Leu Pro Ala Gly Gly
 740 745 750
 Ala Pro Gly Gly Asn Pro Gly Leu Tyr Asp Glu Met Tyr Arg Val Ser

	755	760	765
5	Ala Ile Ile Thr Asn Thr Gly Asn Val Val Gly Asp Glu Val Pro Gln 770 775 780		
10	Leu Tyr Val Ser Leu Gly Gly Pro Asp Asp Pro Lys Val Val Leu Arg 785 790 795 800		
15	Asn Phe Asp Arg Ile Thr Leu His Pro Gly Gln Gln Thr Met Trp Thr 805 810 815		
20	Thr Thr Leu Thr Arg Arg Asp Ile Ser Asn Trp Asp Pro Ala Ser Gln 820 825 830		
25	Asn Trp Val Val Thr Lys Tyr Pro Lys Thr Val Tyr Ile Gly Ser Ser 835 840 845		
30	Ser Arg Lys Leu His Leu Gln Ala Pro Leu Pro Pro Tyr 850 855 860		
	<210> 62 <211> 685 <212> PRT <213> Azospirillum lankense <400> 62		
35	Met Gly Ala Leu Arg Leu Leu Gly Ser Ile Ser Ile Val Ala Leu Thr 1 5 10 15		
40	Cys Gly Gly Ile His Ala Ser Thr Ala Ile Ala Gln Glu Gly Ala Ala 20 25 30		
45	Pro Ala Ala Ile Leu His Pro Glu Lys Trp Pro Arg Pro Ala Thr Gln 35 40 45		
50	Arg Leu Ile Asp Pro Ala Val Glu Lys Arg Val Asp Ala Leu Leu Lys 50 55 60		
55	Gln Leu Ser Val Glu Glu Lys Val Gly Gln Val Ile Gln Gly Asp Ile 65 70 75 80		
60	Gly Thr Ile Thr Pro Glu Asp Leu Arg Lys Tyr Pro Leu Gly Ser Ile 85 90 95		
65	Leu Ala Gly Gly Asn Ser Gly Pro Asn Gly Asp Asp Arg Ala Pro Pro 100 105 110		
	Lys Glu Trp Leu Asp Leu Ala Asp Ala Phe Tyr Arg Val Ser Leu Glu		

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

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

5 Thr Thr Val Asn Arg Gly Asp Ala Asp Tyr Asn Pro Leu Phe Ala Tyr
625 630 635 640

Gly Tyr Gly Leu Thr Tyr Lys Asp Lys Ser Lys Val Gly Thr Leu Pro
645 650 655

10 Glu Glu Ser Gly Val Pro Ala Glu Ala Arg Gln Asn Ala Gly Ile Tyr
660 665 670

15 Phe Arg Ala Gly Ala Leu Arg Leu Pro Gly Arg Phe Leu
675 680 685

20 <210> 63
<211> 509
<212> PRT
<213> Mycelophthora thermophila

25 <400> 63

Gln Asn Ala Cys Thr Leu Thr Ala Glu Asn His Pro Ser Leu Thr Trp
1 5 10 15

30 Ser Lys Cys Thr Ser Gly Gly Ser Cys Thr Ser Val Gln Gly Ser Ile
20 25 30

35 Thr Ile Asp Ala Asn Trp Arg Trp Thr His Arg Thr Asp Ser Ala Thr
35 40 45

40 Asn Cys Tyr Glu Gly Asn Lys Trp Asp Thr Ser Tyr Cys Ser Asp Gly
50 55 60

45 Pro Ser Cys Ala Ser Lys Cys Cys Ile Asp Gly Ala Asp Tyr Ser Ser
65 70 75 80

50 Thr Tyr Gly Ile Thr Thr Ser Gly Asn Ser Leu Asn Leu Lys Phe Val
85 90 95

55 Thr Lys Gly Gln Tyr Ser Thr Asn Ile Gly Ser Arg Thr Tyr Leu Met
100 105 110

60 Glu Ser Asp Thr Lys Tyr Gln Met Phe Gln Leu Leu Gly Asn Glu Phe
115 120 125

65 Thr Phe Asp Val Asp Val Ser Asn Leu Gly Cys Gly Leu Asn Gly Ala
130 135 140

Leu Tyr Phe Val Ser Met Asp Ala Asp Gly Gly Met Ser Lys Tyr Ser
145 150 155 160

ES 2 623 288 T3

5
10
15
20
25
30
35
40
45
50
55
60
65

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

ES 2 623 288 T3

Asn Ser Asn Val Ile Phe Ser Asn Ile Arg Phe Gly Pro Ile Gly Ser
420 425 430

Thr Val Ser Gly Leu Pro Asp Gly Gly Ser Gly Asn Pro Asn Pro Pro
435 440 445

Val Ser Ser Ser Thr Pro Val Pro Ser Ser Ser Thr Thr Ser Ser Gly
450 455 460

Ser Ser Gly Pro Thr Gly Gly Thr Gly Val Ala Lys His Tyr Glu Gln
465 470 475 480

Cys Gly Gly Ile Gly Phe Thr Gly Pro Thr Gln Cys Glu Ser Pro Tyr
485 490 495

Thr Cys Thr Lys Leu Asn Asp Trp Tyr Ser Gln Cys Leu
500 505

<210> 64

<211> 305

<212> PRT

<213> Mycelophthora thermophila

<400> 64

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

Ala Ala Pro Ser Arg Thr Thr Pro Gln Lys Pro Arg Gln Ala Ser Ala
20 25 30

Gly Cys Ala Ser Ala Val Thr Leu Asp Ala Ser Thr Asn Val Phe Gln
35 40 45

Gln Tyr Thr Leu His Pro Asn Asn Phe Tyr Arg Ala Glu Val Glu Ala
50 55 60

Ala Ala Glu Ala Ile Ser Asp Ser Ala Leu Ala Glu Lys Ala Arg Lys
65 70 75 80

Val Ala Asp Val Gly Thr Phe Leu Trp Leu Asp Thr Ile Glu Asn Ile
85 90 95

Gly Arg Leu Glu Pro Ala Leu Glu Asp Val Pro Cys Glu Asn Ile Val
100 105 110

Gly Leu Val Ile Tyr Asp Leu Pro Gly Arg Asp Cys Ala Ala Lys Ala
115 120 125

ES 2 623 288 T3

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

	370	375	380
	Glu Met Leu Leu Lys Asn Ala Asn Pro Ser Phe		
	385	390	395
5	<210> 65		
	<211> 373		
	<212> PRT		
	<213> Mycolophthora thermophila		
10	<400> 65		
	Gln Ser Gly Pro Trp Gln Gln Cys Gly Gly Ile Gly Trp Gln Gly Ser		
	1	5	10
15	Thr Asp Cys Val Ser Gly Tyr His Cys Val Tyr Gln Asn Asp Trp Tyr		
	20	25	30
20	Ser Gln Cys Val Pro Gly Ala Ala Ser Thr Thr Leu Gln Thr Ser Thr		
	35	40	45
25	Thr Ser Arg Pro Thr Ala Thr Ser Thr Ala Pro Pro Ser Ser Thr Thr		
	50	55	60
30	Ser Pro Ser Lys Gly Lys Leu Lys Trp Leu Gly Ser Asn Glu Ser Gly		
	65	70	75
	Ala Glu Phe Gly Glu Gly Asn Tyr Pro Gly Leu Trp Gly Lys His Phe		
	85	90	95
35	Ile Phe Pro Ser Thr Ser Ala Ile Gln Thr Leu Ile Asn Asp Gly Tyr		
	100	105	110
40	Asn Ile Phe Arg Ile Asp Phe Ser Met Glu Arg Leu Val Pro Asn Gln		
	115	120	125
45	Leu Thr Ser Ser Phe Asp Glu Gly Tyr Leu Arg Asn Leu Thr Glu Val		
	130	135	140
50	Val Asn Phe Val Thr Asn Ala Gly Lys Tyr Ala Val Leu Asp Pro His		
	145	150	155
	Asn Tyr Gly Arg Tyr Tyr Gly Asn Val Ile Thr Asp Thr Asn Ala Phe		
	165	170	175
55	Arg Thr Phe Trp Thr Asn Leu Ala Lys Gln Phe Ala Ser Asn Ser Leu		
	180	185	190
60	Val Ile Phe Asp Thr Asn Asn Glu Tyr Asn Thr Met Asp Gln Thr Leu		
65			

	195	200	205
5	Val Leu Asn Leu Asn Gln Ala Ala Ile Asp Gly Ile Arg Ala Ala Gly 210 215 220		
10	Ala Thr Ser Gln Tyr Ile Phe Val Glu Gly Asn Ala Trp Ser Gly Ala 225 230 235 240		
15	Trp Ser Trp Asn Thr Thr Asn Thr Asn Met Ala Ala Leu Thr Asp Pro 245 250 255		
20	Gln Asn Lys Ile Val Tyr Glu Met His Gln Tyr Leu Asp Ser Asp Ser 260 265 270		
25	Ser Gly Thr His Ala Glu Cys Val Ser Ser Asn Ile Gly Ala Gln Arg 275 280 285		
30	Val Val Gly Ala Thr Gln Trp Leu Arg Ala Asn Gly Lys Leu Gly Val 290 295 300		
35	Leu Gly Glu Phe Ala Gly Gly Ala Asn Ala Val Cys Gln Gln Ala Val 305 310 315 320		
40	Thr Gly Leu Leu Asp His Leu Gln Asp Asn Ser Asp Val Trp Leu Gly 325 330 335		
45	Ala Leu Trp Trp Ala Ala Gly Pro Trp Trp Gly Asp Tyr Met Tyr Ser 340 345 350		
50	Phe Glu Pro Pro Ser Gly Thr Gly Tyr Val Asn Tyr Asn Ser Ile Leu 355 360 365		
	Lys Lys Tyr Leu Pro 370		

<210> 66

<211> 851

<212> PRT

<213> Mycelophthora thermophila

<400> 66

ES 2 623 288 T3

Ile Glu Ser Arg Lys Val His Gln Lys Pro Leu Ala Arg Ser Glu Pro
1 5 10 15

Phe Tyr Pro Ser Pro Trp Met Asn Pro Asn Ala Asp Gly Trp Ala Glu
20 25 30

Ala Tyr Ala Gln Ala Lys Ser Phe Val Ser Gln Met Thr Leu Leu Glu

5

10

15

20

25

30

35

40

45

50

55

60

65

ES 2 623 288 T3

35

40

45

Lys Val Asn Leu Thr Thr Gly Val Gly Trp Gly Ala Glu Gln Cys Val
50 55 60

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
85 90 95

Pro Ser Gly Gln Thr Val Ala Ala Thr Trp Asp Arg Gly Leu Met Tyr
100 105 110

Arg Arg Gly Tyr Ala Met Gly Gln Gln 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

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

Gly Met Ser Glu Thr Ile Lys Gly Ile Gln Asp Ala Gly Val Ile Ala
165 170 175

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
195 200 205

Asn Ile Asp Asp Lys Thr Met His Glu Leu Tyr Leu Trp Pro Phe Ala
210 215 220

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
245 250 255

Leu Lys Asn Glu Leu Gly Phe Gln Gly Phe Val Met Ser Asp Trp Gln
260 265 270

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					
5	Asn	Leu	Thr	Leu	Ala	Val	Leu	Asn	Gly	Thr	Val	Pro	Ala	Tyr	Arg	Leu	
	305					310					315					320	
	Asp	Asp	Met	Ala	Met	Arg	Ile	Met	Ala	Ala	Leu	Phe	Lys	Val	Thr	Lys	
					325					330						335	
10	Thr	Thr	Asp	Leu	Glu	Pro	Ile	Asn	Phe	Ser	Phe	Trp	Thr	Asp	Asp	Thr	
				340					345					350			
15	Tyr	Gly	Pro	Ile	His	Trp	Ala	Ala	Lys	Gln	Gly	Tyr	Gln	Glu	Ile	Asn	
	355						360						365				
20	Ser	His	Val	Asp	Val	Arg	Ala	Asp	His	Gly	Asn	Leu	Ile	Arg	Glu	Ile	
	370						375				380						
25	Ala	Ala	Lys	Gly	Thr	Val	Leu	Leu	Lys	Asn	Thr	Gly	Ser	Leu	Pro	Leu	
	385					390					395					400	
30	Asn	Lys	Pro	Lys	Phe	Val	Ala	Val	Ile	Gly	Glu	Asp	Ala	Gly	Ser	Ser	
					405					410					415		
35	Pro	Asn	Gly	Pro	Asn	Gly	Cys	Ser	Asp	Arg	Gly	Cys	Asn	Glu	Gly	Thr	
			420					425						430			
40	Leu	Ala	Met	Gly	Trp	Gly	Ser	Gly	Thr	Ala	Asn	Tyr	Pro	Tyr	Leu	Val	
			435				440					445					
45	Ser	Pro	Asp	Ala	Ala	Leu	Gln	Ala	Arg	Ala	Ile	Gln	Asp	Gly	Thr	Arg	
	450						455					460					
50	Tyr	Glu	Ser	Val	Leu	Ser	Asn	Tyr	Ala	Glu	Glu	Lys	Thr	Lys	Ala	Leu	
	465					470					475					480	
55	Val	Ser	Gln	Ala	Asn	Ala	Thr	Ala	Ile	Val	Phe	Val	Asn	Ala	Asp	Ser	
					485					490					495		
60	Gly	Glu	Gly	Tyr	Ile	Asn	Val	Asp	Gly	Asn	Glu	Gly	Asp	Arg	Lys	Asn	
				500					505					510			
65	Leu	Thr	Leu	Trp	Asn	Asn	Gly	Asp	Thr	Leu	Val	Lys	Asn	Val	Ser	Ser	
			515					520					525				
70	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	
	545					550				555						560	
5	Gly	Leu	Pro	Gly	Gln	Gln	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	
10				580				585						590			
	Arg	Glu	Ser	Tyr	Gly	Ala	Asp	Val	Leu	Tyr	Lys	Pro	Asn	Asn	Gly	Asn	
			595					600					605				
15	Gly	Ala	Pro	Gln	Gln	Asp	Phe	Thr	Glu	Gly	Val	Phe	Ile	Asp	Tyr	Arg	
	610					615						620					
20	Tyr	Phe	Asp	Lys	Val	Asp	Asp	Asp	Ser	Val	Ile	Tyr	Glu	Phe	Gly	His	
	625					630					635					640	
25	Gly	Leu	Ser	Tyr	Thr	Thr	Phe	Glu	Tyr	Ser	Asn	Ile	Arg	Val	Val	Lys	
					645					650					655		
30	Ser	Asn	Val	Ser	Glu	Tyr	Arg	Pro	Thr	Thr	Gly	Thr	Thr	Ala	Gln	Ala	
			660						665					670			
35	Pro	Thr	Phe	Gly	Asn	Phe	Ser	Thr	Asp	Leu	Glu	Asp	Tyr	Leu	Phe	Pro	
			675					680					685				
40	Lys	Asp	Glu	Phe	Pro	Tyr	Ile	Tyr	Gln	Tyr	Ile	Tyr	Pro	Tyr	Leu	Asn	
	690					695						700					
45	Thr	Thr	Asp	Pro	Arg	Arg	Ala	Ser	Ala	Asp	Pro	His	Tyr	Gly	Gln	Thr	
	705					710				715						720	
50	Ala	Glu	Glu	Phe	Leu	Pro	Pro	His	Ala	Thr	Asp	Asp	Asp	Pro	Gln	Pro	
					725				730						735		
55	Leu	Leu	Arg	Ser	Ser	Gly	Gly	Asn	Ser	Pro	Gly	Gly	Asn	Arg	Gln	Leu	
				740				745						750			
60	Tyr	Asp	Ile	Val	Tyr	Thr	Ile	Thr	Ala	Asp	Ile	Thr	Asn	Thr	Gly	Ser	
			755					760					765				
65	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	

5 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
 10 Lys Thr Ala Tyr Val Gly Arg Ser Ser Arg Lys Leu Asp Leu Lys Ile
 835 840 845
 Glu Leu Pro
 850
 15
 <210> 67
 <211> 509
 <212> PRT
 20 <213> Mycolophthora thermophila
 <400> 67
 25 Gln Asn Ala Cys Thr Leu Thr Ala Glu Asn His Pro Ser Leu Thr Trp
 1 5 10 15
 Ser Lys Cys Thr Ser Gly Gly Ser Cys Thr Ser Val Gln Gly Ser Ile
 20 25 30
 Thr Ile Asp Ala Asn Trp Arg Trp Thr His Arg Thr Asp Ser Ala Thr
 35 40 45
 Asn Cys Tyr Glu Gly Asn Lys Trp Asp Thr Ser Tyr Cys Ser Asp Gly
 50 55 60
 40 Pro Ser Cys Ala Ser Lys Cys Cys Ile Asp Gly Ala Asp Tyr Ser Ser
 65 70 75 80
 Thr Tyr Gly Ile Thr Thr Ser Gly Asn Ser Leu Asn Leu Lys Phe Val
 85 90 95
 45 Thr Lys Gly Gln Tyr Ser Thr Asn Ile Gly Ser Arg Thr Tyr Leu Met
 100 105 110
 50 Glu Ser Asp Thr Lys Tyr Gln Met Phe Gln Leu Leu Gly Asn Glu Phe
 115 120 125
 55 Thr Phe Asp Val Asp Val Ser Asn Leu Gly Cys Gly Leu Asn Gly Ala
 130 135 140
 60 Leu Tyr Phe Val Ser Met Asp Ala Asp Gly Gly Met Ser Lys Tyr Ser
 145 150 155 160
 65

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

					405					410				415		
	Asn	Ser	Asn	Val	Ile	Phe	Ser	Asn	Ile	Arg	Phe	Gly	Pro	Ile	Gly	Ser
				420					425					430		
5																
	Thr	Val	Ser	Gly	Leu	Pro	Asp	Gly	Gly	Ser	Gly	Asn	Pro	Asn	Pro	Pro
				435				440					445			
10																
	Val	Ser	Ser	Ser	Thr	Pro	Val	Pro	Ser	Ser	Ser	Thr	Thr	Ser	Ser	Gly
				450			455					460				
15																
	Ser	Ser	Gly	Pro	Thr	Gly	Gly	Thr	Gly	Val	Ala	Lys	His	Tyr	Glu	Gln
	465					470					475					480
20																
	Cys	Gly	Gly	Ile	Gly	Phe	Thr	Gly	Pro	Thr	Gln	Cys	Glu	Ser	Pro	Tyr
				485						490						495
25																
	Thr	Cys	Thr	Lys	Leu	Asn	Asp	Trp	Tyr	Ser	Gln	Cys	Leu			
				500					505							
30																
	<210> 68															
	<211> 485															
	<212> PRT															
	<213> Mycelophthora thermophila															
35																
	Ala	Pro	Val	Ile	Glu	Glu	Arg	Gln	Asn	Cys	Gly	Ala	Val	Trp	Thr	Gln
	1				5					10				15		
40																
	Cys	Gly	Gly	Asn	Gly	Trp	Gln	Gly	Pro	Thr	Cys	Cys	Ala	Ser	Gly	Ser
				20					25					30		
45																
	Thr	Cys	Val	Ala	Gln	Asn	Glu	Trp	Tyr	Ser	Gln	Cys	Leu	Pro	Asn	Ser
			35				40					45				
50																
	Gln	Val	Thr	Ser	Ser	Thr	Thr	Pro	Ser	Ser	Thr	Ser	Thr	Ser	Gln	Arg
	50					55					60					
55																
	Ser	Thr	Ser	Thr	Ser	Ser	Ser	Thr	Thr	Arg	Ser	Gly	Ser	Ser	Ser	Ser
	65				70					75					80	
60																
	Ser	Ser	Thr	Thr	Pro	Pro	Pro	Val	Ser	Ser	Pro	Val	Thr	Ser	Ile	Pro
				85					90					95		
65																
	Gly	Gly	Ala	Thr	Ser	Thr	Ala	Ser	Tyr	Ser	Gly	Asn	Pro	Phe	Ser	Gly
			100					105					110			
70																
	Val	Arg	Leu	Phe	Ala	Asn	Asp	Tyr	Tyr	Arg	Ser	Glu	Val	His	Asn	Leu

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

	Asp Thr Gly Arg Asn Gly Lys Gln Pro Thr Gly Gln Gln Gln Trp Gly	
	370	375 380
5	Asp Trp Cys Asn Val Lys Gly Thr Gly Phe Gly Val Arg Pro Thr Ala	
	385	390 395 400
10	Asn Thr Gly His Glu Leu Val Asp Ala Phe Val Trp Val Lys Pro Gly	
		405 410 415
15	Gly Glu Ser Asp Gly Thr Ser Asp Thr Ser Ala Ala Arg Tyr Asp Tyr	
		420 425 430
20	His Cys Gly Leu Ser Asp Ala Leu Gln Pro Ala Pro Glu Ala Gly Gln	
		435 440 445
25	Trp Phe Gln Ala Tyr Phe Glu Gln Leu Leu Thr Asn Ala Asn Pro Pro	
		450 455 460
30	Phe	465
	<210> 60 <211> 35 <212> DNA <213> Artificial sequence	
35		
	<220> <223> Synthetic DNA forward primer	
40		
	<400> 60 taccctctt cccacctgtc ccaggccctt gctct	
45		35
	<210> 70 <211> 36 <212> DNA <213> Artificial sequence	
50		
	<220> <223> Synthetic DNA reverse primer	
55		
	<400> 70 ggatccgact tatattaca cccctggga gtaaca	
60		36
65		

Reivindicaciones

- 5 1. Una composición que comprende:
 - a) una proteína GH61 recombinantemente producida que comprende una secuencia de aminoácido que es al menos 80% idéntica a SEQ ID NO: 2 o fragmentos de SEQ ID NO: 2 que tienen actividad GH61; y
 - b) enzimas de celulasa que incluyen al menos una endoglucanasa (EG), al menos una β -glucosidasa (BGL), al menos una celobiohidrolasa de Tipo 1 (CBH1), y al menos una celobiohidrolasa de Tipo 2 (CBH2).
- 10 2. La composición de la reivindicación 1, donde la proteína GH61:
 - a) aumenta la producción de glucosa en al menos un 20% en una reacción donde la celulosa sufre sacarificación por una combinación de enzimas que comprende EG, BGL, CBH1 y CBH2, en comparación con la misma reacción en ausencia de la proteína GH61;
 - b) comprende una secuencia de aminoácido que es al menos 85%, 90%, 95%, 97% o 99% idéntica a SEQ ID NO: 2;
 - c) comprende SEQ ID NO: 2; o un fragmento de la misma que tiene actividad GH61.
- 15 3. La composición de la reivindicación 1 o 2, donde las enzimas de celulasa son:
 - a) recombinantes;
 - b) enzimas fúngicas, y/o
 - c) de *M. thermophila*.
- 20 4. La composición de una cualquiera de las reivindicaciones 1 a 3, donde la composición se obtiene:
 - a) expresando un gen que codifica la proteína GH61 en una población de célula huésped, y expresando al menos un gen que codifica una enzima de celulasa en otra población de célula huésped; o
 - b) expresando un gen que codifica la proteína GH61 y al menos un gen que codifica una enzima de celulasa en el mismo tipo de célula huésped.
- 25 5. La composición de la reivindicación 4, donde el gen codifica una proteína GH61 que comprende un péptido de señal heterólogo.
- 30 6. Un método para producir azúcares fermentables a partir de biomasa celulósica que comprende contactar un sustrato celulósico con una composición, medio de cultivo o lisado celular que contiene una proteína GH61 y enzimas de celulasa, donde:
 - a) la proteína GH61 comprende una secuencia de aminoácido que es al menos 89% idéntica a la SEQ ID NO: 2 o un fragmento biológicamente activo de la misma; y
 - b) las enzimas de celulasa incluyen al menos una endoglucanasa (EG), al menos una β -glucosidasa (BGL), al menos una celobiohidrolasa de Tipo 1 (CBH1), y al menos una celobiohidrolasa de Tipo 2 (CBH2).
- 35 7. El método de la reivindicación 6, donde:
 - a) antes de dicho contacto, el sustrato se pre-trata;
 - b), la composición, medio de cultivo o lisado celular comprende además una o más enzimas seleccionadas de esterasas, xilanasas, hemicelulasas, lipasas, proteasas, amilasas, glucoamilasas y combinaciones de ellas;
 - c) el sustrato celulósico se deriva de trigo, paja de trigo, sorgo, maíz, arroz, cebada, bagazo de caña de azúcar, hierbas, pasto varilla, grano de maíz, mazorcas de maíz, fibra de maíz o una combinación de los mismos, y donde el sustrato es opcionalmente paja de trigo derivada; y/o
 - d) el azúcar fermentable comprende glucosa.
- 40 8. El método de la reivindicación 6 o 7, donde la composición, medio de cultivo o lisado celular comprende GH61 aislado y al menos una o más de una endoglucanasa aislada (EG), una β -glucosidasa aislada (BGL), una celobiohidrolasa de Tipo 1 aislada (CBH1), y/o una celobiohidrolasa de Tipo 2 aislada (CBH2).
- 45 9. Un método para producir un producto final a partir de un sustrato celulósico, que comprende:
 - a) contactar el sustrato celulósico con una composición de acuerdo con cualquiera de las reivindicaciones 1 a 5, por lo que se producen azúcares fermentables a partir del sustrato; y
 - b) contactar los azúcares fermentables con un microorganismo en una fermentación para producir un producto final.
- 50 10. El método de la reivindicación 9, que es un proceso simultáneo de sacarificación y fermentación.
- 55 11. El método de la reivindicación 9, donde la sacarificación del sustrato celulósico y dicha fermentación ocurren consecutivamente.
- 60 12. Un método para producir un producto final a partir de un sustrato celulósico, que comprende:
- 65

- a) obtener azúcares fermentables producidos de acuerdo con el método de una cualquiera de las reivindicaciones 6-8; y
- b) contactar los azúcares fermentables con un microorganismo en una fermentación para producir un producto final.

5

13. El método de cualquiera de las reivindicaciones 9 a 12, donde el producto final es un alcohol.

14. El método de la reivindicación 13, donde el alcohol es etanol.

10

15. El método de cualquiera de las reivindicaciones 9 a 15, donde el microorganismo es una levadura.

Figura 1(A)

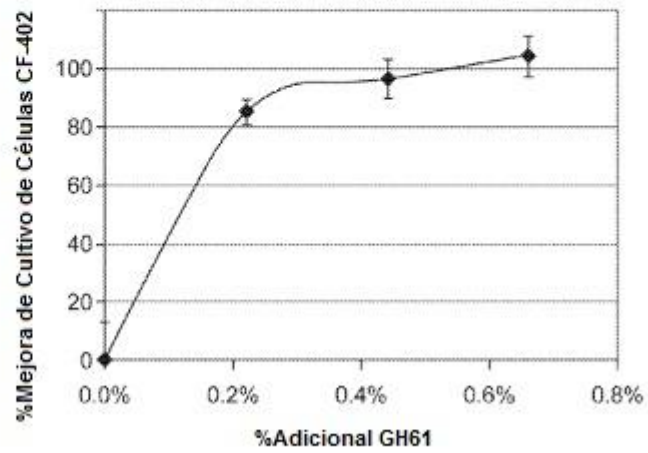


Figura 1(B)

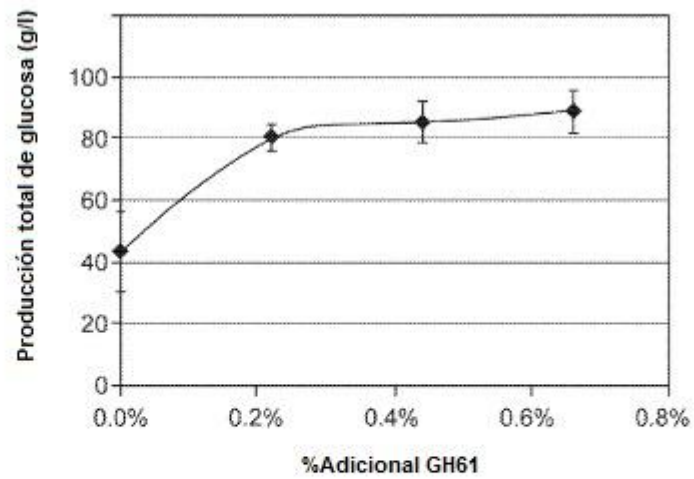
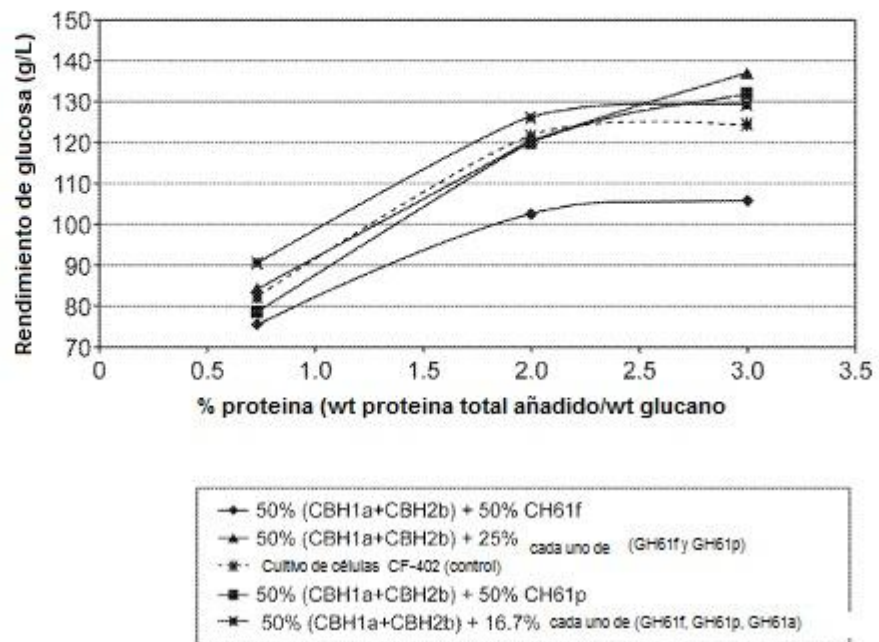


Figura 2**Figura 3**