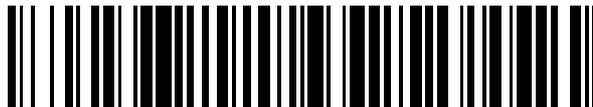


19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 426 954**

51 Int. Cl.:

**C12P 7/10** (2006.01)

**C12P 19/02** (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **10.07.2009 E 09802477 (1)**

97 Fecha y número de publicación de la concesión europea: **26.06.2013 EP 2307554**

54 Título: **Procedimiento para oxidar material lignocelulósico empleando una peroxidasa de Marasmius scorodonius**

30 Prioridad:

**29.07.2008 EP 08161376**

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

**28.10.2013**

73 Titular/es:

**DSM IP ASSETS B.V. (100.0%)  
Het Overloon 1  
6411 TE Heerlen, NL**

72 Inventor/es:

**ZORN, HOLGER;  
SZWEDA, RENATE;  
KUMAR, MANOJ y  
WILMS, JOHANNES**

74 Agente/Representante:

**LEHMANN NOVO, María Isabel**

**ES 2 426 954 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

Procedimiento para oxidar material lignocelulósico empleando una peroxidasa de *Marasmius scorodoni*

Campo de la invención

5 La presente invención se refiere a métodos para la oxidación de un material lignocelulósico que comprende un material de hidrato de carbono que no es almidón, para incrementar la susceptibilidad de tal material a la degradación enzimática, para la degradación de tal material y para producir un azúcar o azúcares a partir de tal material usando enzimas peroxidadas. La invención también se refiere al uso de enzimas peroxidadas en tales métodos.

Antecedentes de la invención

10 Los hidratos de carbono constituyen los compuestos orgánicos más abundantes en la tierra. Sin embargo, mucho de este hidrato de carbono está secuestrado en polímeros complejos que incluyen almidón (el principal hidrato de carbono de almacenamiento en semillas y granos), y una colección de hidratos de carbono y lignina conocida como lignocelulosa. Los componentes principales de hidratos de carbono de lignocelulosa son celulosa, hemicelulosa, y pectinas. Estos polímeros complejos se denominan a menudo de forma colectiva como lignocelulosa.

15 La bioconversión de biomasa lignocelulósica renovable en un azúcar fermentable que se fermenta subsiguientemente para producir alcohol (por ejemplo, etanol) como alternativa a combustibles líquidos ha atraído una intensa atención de los investigadores desde los años 1970, cuando la crisis del petróleo estalló debido a la disminución de la producción de petróleo por la OPEP. El etanol se ha usado ampliamente como una mezcla al 10% con gasolina en los Estados Unidos de América, o como combustible puro para vehículos en Brasil en las dos  
20 últimas décadas. Más recientemente, el uso de E85, una mezcla de etanol al 85% se ha implementado especialmente para aplicaciones de limpieza de la ciudad. La importancia del combustible bioetanol aumentará paralelamente al aumento de los precios del petróleo y el agotamiento gradual de sus fuentes. Adicionalmente, los azúcares fermentables están siendo usados para producir plásticos, polímeros y otros productos a base de fuentes biológicas, y se espera que esta industria crezca sustancialmente incrementando por lo tanto la demanda de  
25 azúcares fermentables de bajo coste abundantes que se puedan usar como materia prima en lugar de materias primas a base de petróleo.

El secuestro de tales grandes cantidades de hidratos de carbono en biomasa vegetal proporciona una fuente abundante de energía potencial en forma de azúcares, tanto azúcares de cinco carbonos como de seis carbonos, que se podrían utilizar para numerosos procesos industriales y agrícolas. Sin embargo, el enorme potencial  
30 energético de estos hidratos de carbono está actualmente siendo infrautilizado debido a que los azúcares están bloqueados en polímeros complejos, y por tanto no son fácilmente accesibles para la fermentación. Los métodos que generen azúcares a partir de biomasa vegetal proporcionarían materias primas abundantes, económicamente competitivas, para la fermentación en compuestos químicos, plásticos y combustibles.

A pesar de la investigación continua de las últimas décadas recientes por comprender la degradación enzimática de la biomasa lignocelulósica y la producción enzimática de celulasas, todavía sigue siendo deseable descubrir o  
35 manipular mediante ingeniería nuevas celulasas y hemicelulasas muy activas. También sería muy deseable construir composiciones enzimáticas muy eficientes capaces de llevar a cabo una degradación rápida y eficiente de materiales lignocelulósicos.

Además, típicamente se usan procedimientos quimiomecánicos para liberar la celulosa para la conversión  
40 subsiguiente en azúcares fermentables. Tales procedimientos quimiomecánicos requieren vasijas de reacción caras y son energéticamente costosas. Además, el pretratamiento químico que se produce a temperaturas elevadas y en condiciones extremas de pH no son compatibles con enzimas conocidas que degradan la celulosa. Adicionalmente, estas reacciones producen compuestos que se deben eliminar antes de que pueda transcurrir la fermentación. Como resultado, los procedimientos de pretratamiento químico actualmente se producen en vasijas de reacción  
45 distintas de la degradación de la celulosa, y se deben de producir antes de la degradación de la celulosa.

El documento US 5.865.898 A describe un método para oxidar un material lignocelulósico, método el cual comprende poner en contacto dicho material lignocelulósico con un agente oxidante seleccionado de oxígeno y gases que contienen oxígeno.

De este modo, son deseables los métodos que son más compatibles con el procedimiento de degradación de la  
50 celulosa, que no requieren temperatura y presión elevadas, que no generan productos de desecho tóxicos y que requieren menos energía.

Sumario de la invención

Se ha demostrado que las peroxidadas de *Marasmius scorodoni* pueden ayudar a las celulasas para liberar  
55 mayores cantidades de azúcar a partir de materia prima lignocelulósica que lo que sería el caso en ausencia de dichas peroxidadas.

Según la invención, se proporciona así un método para la oxidación de un material lignocelulósico que comprende un hidrato de carbono que no es almidón, método el cual comprende poner en contacto dicho material lignocelulósico que comprende un hidrato de carbono que no es almidón con un polipéptido que tiene actividad de peroxidasa y que es:

- 5 a. un polipéptido que comprende una secuencia de aminoácidos que tiene al menos 55% de homología con una secuencia de aminoácidos expuesta en uno cualquiera de: aminoácidos 21 a 513 de SEC NO: 2; aminoácidos 20 a 510 de SEC ID NO: 4; aminoácidos 1 a 493 de SEC ID NO 6; o aminoácidos 1 a 491 de SEC ID NO: 8, o
- 10 b. un polipéptido codificado por un polinucleótido que comprende una secuencia nucleotídica que tiene al menos 55% de homología con la secuencia nucleotídica expuesta en uno cualquiera de: nucleótidos 61 a 1539 de SEC ID NO: 1; nucleótidos 58 a 1530 de SEC ID NO: 3; nucleótidos 1 a 1479 de SEC ID NO: 5; o nucleótidos 1 a 1473 de SEC ID NO: 7,

para oxidar de ese modo dicho material lignocelulósico que comprende un hidrato de carbono que no es almidón.

La invención también proporciona:

- 15 - un método para incrementar la susceptibilidad de un material lignocelulósico que comprende un hidrato de carbono que no es almidón a la degradación enzimática, método el cual comprende poner en contacto dicho material lignocelulósico que comprende un hidrato de carbono que no es almidón con un polipéptido que tiene actividad de peroxidasa y que es:

20 a. un polipéptido que comprende una secuencia de aminoácidos que tiene al menos 55% de homología con una secuencia de aminoácidos expuesta en uno cualquiera de: aminoácidos 21 a 513 de SEC NO: 2; aminoácidos 20 a 510 de SEC ID NO: 4; aminoácidos 1 a 493 de SEC ID NO 6; o aminoácidos 1 a 491 de SEC ID NO: 8, o

25 b. un polipéptido codificado por un polinucleótido que comprende una secuencia nucleotídica que tiene al menos 55% de homología con la secuencia nucleotídica expuesta en uno cualquiera de: nucleótidos 61 a 1539 de SEC ID NO: 1; nucleótidos 58 a 1530 de SEC ID NO: 3; nucleótidos 1 a 1479 de SEC ID NO: 5; o nucleótidos 1 a 1473 de SEC ID NO: 7,

para incrementar de ese modo la susceptibilidad de un material lignocelulósico que comprende un hidrato de carbono que no es almidón a la degradación enzimática;

- 30 - un método para producir un azúcar o azúcares a partir de un material lignocelulósico que comprende un hidrato de carbono que no es almidón, método el cual comprende:

(i) oxidar dicho material lignocelulósico que comprende un hidrato de carbono que no es almidón, o incrementar la susceptibilidad de dicho material lignocelulósico que comprende un hidrato de carbono que no es almidón a la degradación enzimática usando un método como se describe anteriormente; y

35 (ii) poner en contacto el material lignocelulósico así modificado que comprende un hidrato de carbono que no es almidón con una celulosa y/o una hemicelulosa y/o una pectinasa,

para producir de ese modo un azúcar o azúcares a partir de un material lignocelulósico que comprende un hidrato de carbono que no es almidón; y

- 40 - un método para producir un azúcar o azúcares a partir de un material lignocelulósico que comprende un hidrato de carbono que no es almidón, método el cual comprende poner en contacto dicho material lignocelulósico que comprende un hidrato de carbono que no es almidón con:

(i) un polipéptido que tiene actividad de peroxidasa y que es:

45 a. un polipéptido que comprende una secuencia de aminoácidos que tiene al menos 55% de homología con una secuencia de aminoácidos expuesta en uno cualquiera de: aminoácidos 21 a 513 de SEC NO: 2; aminoácidos 20 a 510 de SEC ID NO: 4; aminoácidos 1 a 493 de SEC ID NO 6; o aminoácidos 1 a 491 de SEC ID NO: 8, o

b. un polipéptido codificado por un polinucleótido que comprende una secuencia nucleotídica que tiene al menos 55% de homología con la secuencia nucleotídica expuesta en uno cualquiera de: nucleótidos 61 a 1539 de SEC ID NO: 1; nucleótidos 58 a 1530 de SEC ID NO: 3; nucleótidos 1 a 1479 de SEC ID NO: 5; o nucleótidos 1 a 1473 de SEC ID NO: 7, y

50 (ii) una celulosa y/o una hemicelulosa y/o una pectinasa,

para producir de ese modo un azúcar o azúcares a partir de un material lignocelulósico que comprende un hidrato de carbono que no es almidón.

5 Los métodos de la invención como se exponen anteriormente se pueden llevar a cabo de tal manera que el polipéptido que tenga actividad de peroxidasa se ponga en contacto con el material lignocelulósico de hidrato de carbono que no es almidón durante, antes o después de una etapa de pretratamiento (o de hecho se puede usar por sí mismo como una etapa de pretratamiento). De este modo, el método de la invención puede ser un método de pretratamiento mejorado, ya que el uso del polipéptido que tiene actividad de peroxidasa durante, antes o después del pretratamiento puede permitir que se use menos compuestos químicos y/o menos energía en tal pretratamiento. 10 Las condiciones para tal pretratamiento pueden ser más compatibles con las condiciones en las que tiene lugar la degradación subsiguiente de un hidrato de carbono que no es almidón.

La invención proporciona además un método para la preparación de un producto de fermentación, método el cual comprende:

- 15 a. oxidar un material lignocelulósico de hidrato de carbono que no es almidón o producir un azúcar o azúcares a partir de un material lignocelulósico de hidrato de carbono que no es almidón usando un método como se describe anteriormente; y
- b. fermentar el material resultante,
- para preparar de ese modo un producto de fermentación; y

#### Breve descripción de los dibujos

20 La Figura 1 muestra la formación de productos de reacción parduzcos a partir de guayacol y MsP1 + H<sub>2</sub>O<sub>2</sub> (1) en comparación con una referencia inactivada por calor + H<sub>2</sub>O<sub>2</sub> (2).

La Figura 2 muestra el cromatograma de HPLC-DAD (195 nm) de extractos de lignina (organosolv) generados con MsP1 activa (-) o inactivada por calor (- - -) y H<sub>2</sub>O<sub>2</sub>.

La Figura 3 muestra el cromatograma de HPLC-ELSD de extractos de lignina (organosolv) generados con MsP1 activa (-) o inactivada por calor (- - -) y generación *in situ* de H<sub>2</sub>O<sub>2</sub>.

25 La Figura 4 muestra las actividades de lacasa y peroxidasa en cultivos sumergidos de *M. scorodoni* (- Peroxidasa ---- Lacasa); medio A, -medio B\*, medio C, medio D.

La Figura 5 muestra la electroforesis mediante IEF con tinción con plata y con ABTS, respectivamente. Línea 1: proteínas de referencia; línea 2-7 (teñidas con plata) y 8-13 (teñidas con ABTS): Medio A, días 2-7; línea 14-17 (teñidas con ABTS): Medio A-D en el día 11; línea 18; medio SNS en el día 11.

30 La Figura 6 muestra la inducción de actividades de lacasa y de peroxidasa en cultivos sumergidos de *M. scorodoni*. Medio SNS: Lacasa (---), peroxidasa (-----); SNS suplementado con lignina: Lacasa (~~~~), peroxidasa (-----); medio D suplementado con rastrojo de maíz: Lacasa (---), peroxidasa ( ).

La Figura 7 muestra actividad de β-glucosidasa en cultivos sumergidos de *M. scorodoni* suplementados con rastrojo de maíz; medio A, -medio B-, medio C, medio D.

35 La Figura 8 muestra actividad esterolítica en cultivos sumergidos de *M. scorodoni*.

La Figura 9 muestra electroforesis mediante IEF con tinción de la actividad para visualizar la actividad estereolítica; línea 1: medio B, día 7; línea 2: medio D, día 7; línea 3: medio B, día 10; línea 4: medio D, día 10; línea 5: medio B, día 14; línea 6: medio D, día 14.

40 La Figura 10 muestra la liberación de azúcares de glucosa a partir de rastrojo de maíz, expresado como porcentaje de glucosa de dm (= materia seca) de rastrojo de maíz, según se determina mediante ensayo de azúcar reductor.

#### Breve descripción del listado de secuencias

SEC ID NO: 1 expone la secuencia de ADNc de MsP1 procedente de *Marasmius scorodoni*.

45 SEC ID NO: 2 expone la secuencia de aminoácidos de MsP1 procedente de *Marasmius scorodoni*. Hay una secuencia de escisión del péptido señal en la posición 20/21. En consecuencia, los aminoácidos 21 a 513 representan la secuencia de MsP1 madura.

SEC ID NO: 3 expone la secuencia de ADNc de MsP2 procedente de *Marasmius scorodoni*.

SEC ID NO: 4 expone la secuencia de aminoácidos de MsP2 procedente de *Marasmius scorodoni*. Hay una secuencia de escisión del péptido señal en la posición 19/20. En consecuencia, los aminoácidos 20 a 510 representan la secuencia de MsP1 madura.

- 5 SEC ID NO: 5 expone la secuencia nucleotídica usada para expresar MsP1 procedente de *Marasmius scorodoni* en *A. niger*.

SEC ID NO: 6 expone la secuencia de aminoácidos de MsP1 procedente de *Marasmius scorodoni* expresada en *A. niger*.

- 10 SEC ID NO: 7 expone la secuencia nucleotídica usada para expresar MsP2 procedente de *Marasmius scorodoni* en *A. niger*.

SEC ID NO: 8 expone la secuencia de aminoácidos de MsP2 procedente de *Marasmius scorodoni* expresada en *A. niger*.

#### Descripción detallada de la invención

- 15 A lo largo de la presente memoria descriptiva y de las reivindicaciones que se acompañan, las palabras “comprender” e “incluir”, y variaciones tales como “comprende”, “que comprende”, “incluye” y “que incluye”, se han de interpretar inclusivamente. Esto es, estas palabras están destinadas a verbalizar la posible inclusión de otros elementos o entidades completas no citadas específicamente, cuando lo permita el contexto.

Los artículos “un” y “una” se usan aquí para referirse a uno o a más de uno (es decir, a uno o al menos uno) del objeto gramatical del artículo. A título de ejemplo, “un elemento” puede significar un elemento o más de un elemento.

- 20 La presente descripción se refiere al uso de polipéptidos que tienen actividad de peroxidasa. Estos polipéptidos se pueden usar para oxidar un material lignocelulósico que comprende un hidrato de carbono que no es almidón (también denominado aquí como material que no comprende almidón). Un material lignocelulósico que comprende un hidrato de carbono que no es almidón es aquel que comprende, consiste en o consiste sustancialmente en uno o más hidratos de carbono que no son almidón. Tal hidrato de carbono lignocelulósico que no es almidón puede ser  
25 una celulosa, una hemicelulosa o una pectina/sustancia péctica.

En este contexto, hidrato de carbono incluye todos los sacáridos, por ejemplo polisacáridos, oligosacáridos, disacáridos o monosacáridos.

- 30 Un polipéptido que tiene actividad de peroxidasa como se usa según la descripción puede modificar un material lignocelulósico que comprende un hidrato de carbono que no es almidón al modificar químicamente o modificar físicamente tal material lignocelulósico. La modificación química del material lignocelulósico que comprende un hidrato de carbono que no es almidón puede dar como resultado la degradación de tal material, por ejemplo mediante hidrólisis, oxidación u otra modificación química tal como mediante la acción de una liasa. Como alternativa, la modificación puede hacer al material lignocelulósico que comprende un hidrato de carbono que no es almidón más susceptible a la degradación enzimática, por ejemplo mediante una enzima que se sabe que degrada  
35 un hidrato de carbono que no es almidón (por ejemplo lignocelulosa), tal como una celulasa y/o una hemicelulasa y/o una pectinasa. Una celulasa puede comprender una o más de actividad de endocelulasa, de exocelulasa o de  $\beta$ -glucosidasa. Es decir, el material lignocelulósico que comprende un hidrato de carbono que no es almidón se puede modificar de manera que el componente de hidrato de carbono que no es almidón del material se hace más susceptible a la degradación mediante una enzima que puede degradar un hidrato de carbono que no es almidón.

- 40 En consecuencia, la invención se refiere a un método para la oxidación de un material lignocelulósico que comprende un hidrato de carbono que no es almidón, método el cual comprende poner en contacto dicho material lignocelulósico de hidrato de carbono que no es almidón con un polipéptido que tiene actividad de peroxidasa como se describe aquí.

- 45 Tal oxidación puede hacer al material lignocelulósico que comprende un material de hidrato de carbono que no es almidón más susceptible a la degradación enzimática (en particular, puede hacer al componente de hidrato de carbono que no es almidón del material más susceptible a la degradación enzimática). En consecuencia, la invención se refiere a un método para incrementar la susceptibilidad de un material lignocelulósico que comprende un hidrato de carbono que no es almidón a la degradación enzimática, método el cual comprende poner en contacto dicho material lignocelulósico que comprende un hidrato de carbono que no es almidón con un polipéptido que tiene  
50 actividad de peroxidasa como se describe aquí.

Por “más susceptible a la degradación enzimática” se indica que el material lignocelulósico que comprende hidrato de carbono que no es almidón (en particular el componente de hidrato de carbono que no es almidón) se degradará en un mayor grado mediante una enzima capaz de llevar a cabo tal degradación cuando se pone en contacto con un

polipéptido con actividad de peroxidasa como se describe aquí en comparación con la situación cuando no se pone en contacto con tal polipéptido.

5 Un método de la invención puede ser un método de pretratamiento. Es decir, se puede llevar a cabo un método de pretratamiento de un material lignocelulósico que comprende un hidrato de carbono que no es almidón, que implica el uso de un polipéptido que tiene actividad de peroxidasa. De este modo, se puede llevar a cabo un método de la invención de manera que el aporte energético y/o de compuestos químicos durante el pretratamiento pueda ser menor que el que sería el caso de otro modo.

10 De este modo, la invención proporciona un método para producir un azúcar o azúcares a partir de un material lignocelulósico que comprende un hidrato de carbono que no es almidón, método el cual comprende: poner en contacto dicho material lignocelulósico que comprende un hidrato de carbono que no es almidón con una peroxidasa como se describe aquí, es decir, oxidar dicho material lignocelulósico que comprende un hidrato de carbono que no es almidón o incrementar la susceptibilidad de tal material lignocelulósico a la degradación enzimática usando un método como se describe aquí; y poner en contacto el material lignocelulósico así modificado que comprende un hidrato de carbono que no es almidón con una enzima que degrada el material que comprende un hidrato de carbono que no es almidón. Tal enzima será típicamente aquella que es capaz de degradar el componente de hidrato de carbono que no es almidón del material, tal como una celulasa y/o una hemicelulasa y/o una pectinasa. De esa manera, se puede producir uno o más azúcares a partir de un material lignocelulósico que comprende un hidrato de carbono que no es almidón.

20 En los métodos de la invención, un polipéptido usado puede ser un polipéptido que tenga actividad de peroxidasa, por ejemplo un polipéptido aislado, y que comprenda una secuencia según una cualquiera de SEC ID NOs: 2, 4, 6 u 8 o un equivalente funcional o un fragmento de tal polipéptido.

25 SEC ID Nos: 6 y 8 exponen las secuencias de polipéptidos maduros (MsP1 y MsP2 respectivamente) que tienen actividad de peroxidasa a partir de *Marasmius scorodoni*. Sorprendentemente, MsP1 y MsP2 no son comparables a peroxidases de magnesio conocidas ni a peroxidases de lignina, pero muestran más homología con el grupo de peroxidases de "tipo DyP" (peroxidases decolorantes de tintes). También, MsP1 y MsP2 aparecen como dímeros. Véase Scheibner et al. Appl. Microbiol. Biotechnol. (2008) 77, 1241-1240, p. 1247. Sin embargo, SEC ID NOs: 2 y 4 muestran secuencias de longitud completa para MsP1 y MsP2, incluyendo las secuencias señal. Hay un sitio de escisión del péptido señal en la posición 20/21 de SEC ID NO: 2 y en la posición 19/20 de SEC ID NO: 4. En consecuencia, un polipéptido usado en la invención basado en SEC ID NO: 2 ó 4 comprenderá típicamente la secuencia expuesta en los aminoácidos 21 a 513 de SEC ID NO: 2 o en los aminoácidos 20 a 510 de SEC ID NO: 4, o será un equivalente funcional o fragmento de tal polipéptido.

35 Un polipéptido para uso en la invención, por ejemplo un polipéptido aislado, se puede obtener expresando un polinucleótido que comprende una secuencia como se expone en una cualquiera de SEC ID NOs: 1, 3, 5 ó 7 o un vector que comprende dicho polinucleótido en una célula hospedante apropiada, por ejemplo *Aspergillus niger*. Un polipéptido para uso en la invención puede ser un equivalente funcional o fragmento de tal polipéptido.

En vista del hecho de que SEC ID NOs: 1 y 3 comprenden porciones codificantes de la secuencia señal, un polipéptido de la descripción se puede codificar por un polinucleótido que comprende una secuencia como se expone en los nucleótidos 61 a 1539 de SEC ID NO: 1 o los nucleótidos 58 a 1530 de SEC ID NO: 3. Un polipéptido para uso en la invención puede ser un equivalente funcional o fragmento de tal polipéptido.

40 La expresión "polipéptido o polipéptidos que tienen actividad de peroxidasa" se define aquí y en lo sucesivo como un polipéptido de EC 1.11.1.- que será capaz típicamente de catalizar una reacción de la forma:



o



45 Un polipéptido que tiene actividad de peroxidasa puede tener peróxido de hidrógeno como su sustrato óptimo; como alternativa, un polipéptido que tiene actividad de peroxidasa puede ser más activo con un hidroperóxido orgánico tal como un peróxido de lípido. Un polipéptido que tiene actividad de peroxidasa puede contener un cofactor hemo en su sitio activo, o una cisteína o resto selenocisteínico activo redox.

50 El método de la invención puede hacer uso de un polipéptido purificado. Los polipéptidos descritos aquí y adecuados para uso en el método de la invención incluyen los polipéptidos codificados por los nucleótidos descritos aquí. Se prefiere especialmente un polipéptido que comprende la secuencia expuesta en los aminoácidos 21 a 513 de SEC ID NO: 2; los aminoácidos 20 a 510 de SEC ID NO: 4; los aminoácidos 1 a 493 de SEC ID NO 6; o los aminoácidos 1 a 491 de SEC ID NO: 8, o un equivalente funcional de cualquiera de esos polipéptidos.

55 En el método de la invención, también se pueden usar proteínas de fusión que comprenden un polipéptido descrito aquí.

Se describen aquí métodos para obtener los polipéptidos que se pueden usar en el método de la invención.

Se describen aquí polinucleótidos que comprenden una secuencia nucleotídica que se hibrida, preferiblemente en condiciones muy restrictivas, al complemento inverso de un polinucleótido que tiene la secuencia según uno cualquiera de: nucleótidos 61 a 1539 de SEC ID NO: 1; nucleótidos 58 a 1530 de SEC ID NO: 3; nucleótidos 1 a 1479 de SEC ID NO: 5; o nucleótidos 1 a 1473 de SEC ID NO: 7.

Tales ácidos nucleicos, que se pueden usar para proporcionar polipéptidos adecuados para uso en la invención, pueden comprender una secuencia que tiene al menos alrededor de 55%, preferiblemente al menos alrededor de 65%, más preferiblemente al menos alrededor de 70%, incluso más preferiblemente al menos alrededor de 75%, al menos alrededor de 80%, al menos alrededor de 85%, al menos alrededor de 90%, al menos alrededor de 95%, al menos alrededor de 96%, al menos alrededor de 97%, al menos alrededor de 98% o al menos alrededor de 99% de identidad de secuencia con una secuencia como se expone en uno cualquiera de: nucleótidos 61 a 1539 de SEC ID NO: 1; nucleótidos 58 a 1530 de SEC ID NO: 3; nucleótidos 1 a 1479 de SEC ID NO: 5; o nucleótidos 1 a 1473 de SEC ID NO: 7.

En una realización más preferida, tal polinucleótido, o un polipéptido codificado por tal polinucleótido, se puede obtener a partir de un hongo, preferiblemente un hongo filamentoso, tal como del Dikarya, tal como del Basidiomycota, tal como del Agaricomycotina, tal como del Agaricomycetidae, tal como del Agaricales, tal como del Trichomataceae, en particular del género *Marasmius*, por ejemplo *Marasmius scorodonius*. Un polinucleótido adecuado se puede obtener a partir de un hongo que descompone la basura.

Tal polinucleótido puede comprender una secuencia de ácido nucleico que codifica un polipéptido que comprende una secuencia de aminoácidos como se muestra en uno cualquiera de: aminoácidos 21 a 513 de SEC NO: 2; aminoácidos 20 a 510 de SEC ID NO: 4; aminoácidos 1 a 493 de SEC ID NO 6; o aminoácidos 1 a 491 de SEC ID NO: 8, o un equivalente funcional o un fragmento de tal polipéptido.

Tal polinucleótido puede codificar al menos un dominio funcional de un polipéptido que comprende la secuencia expuesta en in aminoácidos 21 a 513 de SEC NO: 2; aminoácidos 20 a 510 de SEC ID NO: 4; aminoácidos 1 a 493 de SEC ID NO 6; o aminoácidos 1 a 491 de SEC ID NO: 8, o al menos un dominio funcional de un equivalente funcional de tal polipéptido.

En una realización preferida, un método de la invención se lleva a cabo usando una enzima peroxidasa codificada por un gen que comprende la secuencia según uno cualquiera de: nucleótidos 61 a 1539 de SEC ID NO: 1; nucleótidos 58 a 1530 de SEC ID NO: 3; nucleótidos 1 a 1479 de SEC ID NO: 5; o nucleótidos 1 a 1473 de SEC ID NO: 7. En otra realización preferida, la invención se lleva a cabo en la que la enzima peroxidasa es codificada por un polinucleótido de manera que la secuencia de aminoácidos comprende uno de: aminoácidos 21 a 513 de SEC NO: 2; aminoácidos 20 a 510 de SEC ID NO: 4; aminoácidos 1 a 493 de SEC ID NO 6; o aminoácidos 1 a 491 de SEC ID NO: 8, o es una variante o un fragmento de cualquiera de esos polipéptidos.

Un polinucleótido que da lugar a un polipéptido adecuado para uso en la invención puede comprender la secuencia codificante que codifica los polipéptidos descritos aquí; se prefiere la secuencia polinucleotídica que comprende uno cualquiera de: nucleótidos 61 a 1539 de SEC ID NO: 1; nucleótidos 58 a 1530 de SEC ID NO: 3; nucleótidos 1 a 1479 de SEC ID NO: 5; o nucleótidos 1 a 1473 de SEC ID NO: 7.

Los polinucleótidos descritos aquí se pueden incorporar en vectores a fin de expresar un polipéptido que se puede usar en un método de la invención.

En tal vector, la secuencia polinucleotídica que codifica una peroxidasa se puede enlazar funcionalmente con secuencias reguladoras adecuadas para la expresión de la secuencia de aminoácidos codificada en una célula hospedante adecuada, tal como una bacteria o un hongo filamentoso, tal como *Marasmius* por ejemplo *M. scorodonius* o *Aspergillus*, por ejemplo *A. niger* o *A. oryzae*.

Un polipéptido adecuado para uso en la invención se puede producir recombinantemente en una célula hospedante que contiene un polinucleótido heterólogo u homólogo como se describe aquí. En tal célula, la expresión de una peroxidasa se puede incrementar significativamente, o se puede incrementar la actividad de la peroxidasa. Típicamente, tal célula es capaz de producir una peroxidasa funcional según la invención, preferiblemente una célula capaz de sobreexpresar la peroxidasa, por ejemplo una cepa de *Aspergillus* que comprende un número mayor de copias de un gen que codifica tal peroxidasa.

El método de la invención requiere el uso de un polipéptido aislado que comprende la secuencia de aminoácidos expuesta en uno cualquiera de: aminoácidos 21 a 513 de SEC NO: 2; aminoácidos 20 a 510 de SEC ID NO: 4; aminoácidos 1 a 493 de SEC ID NO 6; o aminoácidos 1 a 491 de SEC ID NO: 8, o una secuencia de aminoácidos obtenible expresando un polinucleótido que comprende la secuencia expuesta en: nucleótidos 61 a 1539 de SEC ID NO: 1; nucleótidos 58 a 1530 de SEC ID NO: 3; nucleótidos 1 a 1479 de SEC ID NO: 5; o nucleótidos 1 a 1473 de SEC ID NO: 7, en un hospedante apropiado. También, en la presente invención se puede usar un péptido o polipéptido que comprende un equivalente funcional o un fragmento de tal polipéptido.

- Los términos “péptido” y “oligopéptido” se consideran aquí y en lo sucesivo sinónimos (como se reconoce normalmente), y cada término se puede usar de forma intercambiable a medida que lo requiera el contexto para indicar una cadena de al menos dos aminoácidos acoplados mediante enlaces peptídicos. La palabra “polipéptido” se usa aquí para cadenas que contienen más de siete restos de aminoácidos. Todas las fórmulas oligopeptídicas y polipeptídicas o secuencias aquí están escritas de izquierda a derecha y en la dirección desde el término amino al término carboxi. El código de una letra de los aminoácidos usados aquí se conoce normalmente en la técnica y se puede encontrar en Sambrook, et al. (Molecular Cloning: A Laboratory Manual, 2ª ed. Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989).
- Por polipéptido o proteína “aislado” se quiere decir un polipéptido o proteína retirado de su entorno nativo. Por ejemplo, los polipéptidos y proteínas producidos recombinantemente expresados en células hospedantes se consideran aislados para el fin de la invención ya que son polipéptidos nativos o recombinantes que se han purificado sustancialmente mediante cualquier técnica adecuada, tal como, por ejemplo, el método de purificación de una sola etapa descrito en Smith y Johnson, Gene 67:31-40 (1988).
- El polipéptido descrito aquí que tiene actividad de peroxidasa, o un equivalente funcional o un fragmento del mismo, se puede recuperar y purificar de cultivos celulares recombinantes mediante métodos bien conocidos, incluyendo precipitación con sulfato de amonio o etanol, extracción con ácidos, cromatografía de intercambio aniónico o catiónico, cromatografía en fosfocelulosa, cromatografía de interacción hidrófoba, cromatografía de afinidad, cromatografía de hidroxilapatita y cromatografía con lecitina. Lo más preferible, para la purificación se emplea cromatografía de líquidos de altas prestaciones (“HPLC”).
- Los polipéptidos adecuados para uso en la presente invención incluyen productos purificados de forma natural, productos de procedimientos de síntesis química, y productos producidos mediante técnicas recombinantes a partir de un hospedante procarionota o eucariota, incluyendo, por ejemplo, células bacterianas, de levadura, de plantas superiores, de insectos y de mamíferos. Dependiendo del hospedante empleado en un procedimiento de producción recombinante, los polipéptidos de la presente invención pueden estar glucosilados o pueden estar no glucosilados. Además, los polipéptidos también pueden incluir un resto de metionina modificado inicial, en algunos casos como resultado de procesos mediados por el hospedante.
- Las expresiones “equivalente funcional” y “variante funcional” se usan aquí de forma intercambiable. Los equivalentes funcionales de los polinucleótidos que codifican peroxidasa descritos aquí son polinucleótidos aislados que codifican un polipéptido que muestra actividad de peroxidasa, típicamente al menos alrededor de la misma actividad de peroxidasa o mejor que al menos uno de los polipéptidos con actividad de peroxidasa definidos aquí (es decir, aquel que comprende una secuencia expuesta en uno cualquiera de: aminoácidos 21 a 513 de SEC NO: 2; aminoácidos 20 a 510 de SEC ID NO: 4; aminoácidos 1 a 493 de SEC ID NO 6; o aminoácidos 1 a 491 de SEC ID NO: 8).
- Un equivalente funcional de un polipéptido con actividad de peroxidasa es un polipéptido que muestra actividad de peroxidasa, típicamente al menos la misma actividad de peroxidasa o mejor que al menos uno de los polipéptidos con actividad de peroxidasa definidos aquí (es decir, aquel que comprende una secuencia expuesta en uno cualquiera de: aminoácidos 21 a 513 de SEC NO: 2; aminoácidos 20 a 510 de SEC ID NO: 4; aminoácidos 1 a 493 de SEC ID NO 6; o aminoácidos 1 a 491 de SEC ID NO: 8).
- Un equivalente proteico o polipeptídico funcional puede comprender una o más sustituciones, inserciones o supresiones en comparación con un polipéptido que comprende una secuencia expuesta en uno cualquiera de: aminoácidos 21 a 513 de SEC NO: 2; aminoácidos 20 a 510 de SEC ID NO: 4; aminoácidos 1 a 493 de SEC ID NO 6; o aminoácidos 1 a 491 de SEC ID NO: 8. Un equivalente proteico o polipeptídico funcional puede contener sólo sustituciones conservativas de uno o más aminoácidos en comparación con tal polipéptido. Un sustitución, inserción o supresión será típicamente con respecto a un aminoácido no esencial.
- Un aminoácido no esencial es un resto que se puede alterar en una cualquiera de las mencionadas secuencias sin alterar sustancialmente la función biológica del polipéptido. Por ejemplo, se predice que los restos de aminoácidos que están conservados entre las proteínas con actividad de peroxidasa descritas aquí no son particularmente susceptibles a la alteración. Además, los aminoácidos conservados entre las proteínas de peroxidasa descritas aquí y otras peroxidases probablemente no son susceptibles a la alteración.
- La expresión “sustitución conservativa” quiere decir una sustitución en la que el resto de aminoácido es sustituido por un resto de aminoácido que tiene una cadena lateral similar. Estas familias son conocidas en la técnica, e incluyen aminoácidos con cadenas laterales básicas (por ejemplo, lisina, arginina e histidina), cadenas laterales ácidas (por ejemplo ácido aspártico, ácido glutámico), cadenas laterales polares no cargadas (por ejemplo, glicina, asparaginas, glutamina, serina, treonina, tirosina, cisteína), cadenas laterales no polares (por ejemplo, alanina, valina, leucina, isoleucina, prolina, fenilalanina, metionina, triptófano), cadenas laterales ramificadas en beta (por ejemplo, treonina, valina, isoleucina) y cadenas laterales aromáticas (por ejemplo, tirosina, fenilalanina, triptófano, histidina).

Un equivalente funcional de ácido nucleico puede contener típicamente mutaciones en comparación con un polinucleótido que comprende una secuencia como se expone en uno cualquiera de: nucleótidos 61 a 1539 de SEC ID NO: 1; nucleótidos 58 a 1530 de SEC ID NO: 3; nucleótidos 1 a 1479 de SEC ID NO: 5; o nucleótidos 1 a 1473 de SEC ID NO: 7. Típicamente, tales mutaciones serán mutaciones silenciosas o mutaciones que no alteran la función biológica del polipéptido codificado.

En consecuencia, las moléculas de ácido nucleico que codifican un polipéptido de proteína de peroxidasa que comprende una secuencia expuesta en: aminoácidos 21 a 513 de SEC NO: 2; aminoácidos 20 a 510 de SEC ID NO: 4; aminoácidos 1 a 493 de SEC ID NO 6; o aminoácidos 1 a 491 de SEC ID NO: 8, que contienen cambios en restos de aminoácidos que no son esenciales para una actividad biológica particular, se pueden usar para generar una enzima para uso en la invención.

Una variante funcional o equivalente funcional de polipéptido con actividad de peroxidasa diferirá en la secuencia de aminoácidos de una cualquiera de las secuencias específicamente descritas aquí, aunque retendrá al menos una actividad biológica de tal polipéptido.

En una realización, una molécula de ácido nucleico aislada, adecuada para generar un polipéptido para uso en la invención, comprende una secuencia nucleotídica que codifica un polipéptido, en la que el polipéptido comprende una secuencia de aminoácidos que tiene al menos alrededor de 55%, al menos alrededor de 65%, al menos alrededor de 70%, al menos alrededor de 75%, al menos alrededor de 80%, al menos alrededor de 85%, al menos alrededor de 90%, al menos alrededor de 95%, al menos alrededor de 96%, al menos alrededor de 97%, al menos alrededor de 98%, al menos alrededor de 99% o más de identidad de secuencia con la secuencia de aminoácidos mostrada en uno cualquiera de: aminoácidos 21 a 513 de SEC NO: 2; aminoácidos 20 a 510 de SEC ID NO: 4; aminoácidos 1 a 493 de SEC ID NO 6; o aminoácidos 1 a 491 de SEC ID NO: 8. Tal polipéptido puede ser un equivalente funcional o variante funcional adecuado para uso en la invención.

Por ejemplo, en Bowie, J.U. et al., Science 247:1306-1310 (1990), se proporciona una guía con respecto a cómo obtener sustituciones de aminoácidos fenotípicamente silenciosas, en el que los autores indican que hay dos enfoques principales para estudiar la tolerancia de una secuencia de aminoácidos al cambio. El primer método se basa en el proceso de evolución, en el que las mutaciones son aceptadas o rechazadas por selección natural. El segundo enfoque usa ingeniería genética para introducir cambios de aminoácidos en posiciones específicas de un gen clonado, y selecciona o criba para identificar secuencias que mantienen la funcionalidad. Como señalan los autores, estos estudios han revelado que las proteínas son sorprendentemente tolerantes a las sustituciones de aminoácidos. Los autores indican además qué cambios probablemente serán permisivos en cierta posición de la proteína. Por ejemplo, los restos de aminoácidos más enterrados requieren cadenas laterales no polares, mientras que generalmente se conservan algunos rasgos de las cadenas laterales de la superficie. Otras de tales sustituciones fenotípicamente silenciosas están descritas en Bowie et al., más arriba, y las referencias citadas allí.

Una molécula de ácido nucleico aislada que codifica una proteína de peroxidasa como se describe aquí se puede crear introduciendo una o más sustituciones, adiciones o supresiones nucleotídicas en las secuencias codificantes de un polipéptido que comprende un cualquiera de: nucleótidos 61 a 1539 de SEC ID NO: 1; nucleótidos 58 a 1530 de SEC ID NO: 3; nucleótidos 1 a 1479 de SEC ID NO: 5; o nucleótidos 1 a 1473 de SEC ID NO: 7, de manera que se introducen en la proteína codificada una o más sustitución, supresiones o inserciones de aminoácidos. Tales mutaciones se pueden introducir mediante técnicas estándar, tales como mutagénesis dirigida al sitio y mutagénesis mediada por PCR.

La expresión "equivalentes funcionales" también engloba ortólogos de las proteínas de peroxidasas MsP1 o Msp2 de *M. scorodonius* como se describe aquí. Los ortólogos de las proteínas de peroxidasas MsP1 o Msp2 de *M. scorodonius* son proteínas que se pueden aislar de otras cepas o especies, y poseen una actividad biológica similar o idéntica, en particular actividad de peroxidasa. Tales ortólogos se pueden identificar fácilmente puesto que comprenden una secuencia de aminoácidos que es sustancialmente homóloga a uno cualquiera de: aminoácidos 21 a 513 de SEC NO: 2; aminoácidos 20 a 510 de SEC ID NO: 4; aminoácidos 1 a 493 de SEC ID NO 6; o aminoácidos 1 a 491 de SEC ID NO: 8.

Como se define aquí, la expresión "sustancialmente homóloga" se refiere a una primera secuencia de aminoácidos o nucleotídica que contiene un número suficiente o mínimo de aminoácidos o nucleótidos idénticos o equivalentes (por ejemplo, con cadena lateral similar) a una segunda secuencia de aminoácidos o nucleotídica, de tal manera que la primera y la segunda secuencia de aminoácidos o nucleotídica tienen un dominio común. Por ejemplo, las secuencias de aminoácidos o nucleotídicas que contienen un dominio común que tienen al menos alrededor de 55%, preferiblemente al menos alrededor de 65%, al menos alrededor de 70%, más preferiblemente al menos alrededor de 75%, al menos alrededor de 80%, al menos alrededor de 85%, al menos alrededor de 90%, al menos alrededor de 95%, al menos alrededor de 96%, al menos alrededor de 97%, al menos alrededor de 98% o al menos alrededor de 99% de identidad o más se definen aquí como suficientemente idénticas.

También, los ácidos nucleicos que codifican otros miembros de la familia de peroxidasas, que tienen así una secuencia nucleotídica que difiere de las secuencias descritas aquí, pueden ser adecuados para generar una enzima adecuada para uso en la invención. Además, los ácidos nucleicos que codifican las proteínas peroxidasas

descritas aquí de diferentes especies que tienen así una secuencia nucleotídica que difiere de las descritas aquí son también adecuados para uso en la invención.

5 Las moléculas de ácido nucleico que corresponden a variantes (por ejemplo variantes alélicas naturales) y homólogos de los ADN de peroxidasas descritos aquí se pueden aislar basándose en su homología con esos aminoácidos, mediante su uso o un fragmento adecuado de los mismos, como una sonda de hibridación según técnicas de hibridación estándar, preferiblemente en condiciones de hibridación muy restrictivas.

10 Además de las variantes alélicas de origen natural de las secuencias de peroxidasas descritas aquí, la persona experta reconocerá que se pueden introducir cambios mediante mutación en aquellas secuencias nucleotídicas conduciendo de ese modo a cambios en la secuencia de aminoácidos de la proteína peroxidasa sin alterar sustancialmente la función de la proteína.

15 Las proteínas peroxidasas mejoradas se pueden usar en la invención. Las proteínas peroxidasas mejoradas son proteínas en las que al menos una actividad biológica, en particular la actividad de peroxidasa, está mejorada en comparación con un polipéptido que comprende una secuencia como se expone en uno cualquiera de: aminoácidos 21 a 513 de SEC NO: 2; aminoácidos 20 a 510 de SEC ID NO: 4; aminoácidos 1 a 493 de SEC ID NO 6; o aminoácidos 1 a 491 de SEC ID NO: 8. Tales proteínas se pueden obtener introduciendo al azar mutaciones a lo largo de todas o de parte de las secuencias que codifican tal proteína, tal como mediante mutagénesis por saturación, y los mutantes resultantes se pueden expresar recombinantemente y se pueden identificar en busca de su actividad biológica. Por ejemplo, la técnica proporciona ensayos estándar para medir la actividad enzimática de peroxidasas, y de este modo se pueden seleccionar fácilmente proteínas mejoradas.

20 En una realización preferida, la proteína peroxidasa adecuada para uso en la invención tiene una secuencia de aminoácidos según uno cualquiera de: aminoácidos 21 a 513 de SEC NO: 2; aminoácidos 20 a 510 de SEC ID NO: 4; aminoácidos 1 a 493 de SEC ID NO 6; o aminoácidos 1 a 491 de SEC ID NO: 8. En otra realización, el polipéptido con actividad de peroxidasa es sustancialmente homólogo a tal secuencia de aminoácidos y retiene al menos una actividad biológica de tal polipéptido, por ejemplo actividad de peroxidasa, aunque difiere en la secuencia de aminoácidos debido a la variación natural o mutagénesis como se describe anteriormente.

25 En una realización preferida adicional, una proteína peroxidasa adecuada para uso en la invención comprende una secuencia de aminoácidos codificada por un fragmento de ácido nucleico capaz de hibridarse al inverso del complemento de una secuencia de aminoácido que comprende uno cualquiera de: nucleótidos 61 a 1539 de SEC ID NO: 1; nucleótidos 58 a 1530 de SEC ID NO: 3; nucleótidos 1 a 1479 de SEC ID NO: 5; o nucleótidos 1 a 1473 de SEC ID NO: 7, preferiblemente en condiciones de hibridación muy restrictivas.

30 En consecuencia, el polipéptido con actividad de peroxidasa usado en los métodos descritos aquí es una proteína que comprende una secuencia de aminoácidos que tiene al menos alrededor de 55%, al menos alrededor de 65%, al menos alrededor de 70%, al menos alrededor de 75%, al menos alrededor de 80%, al menos alrededor de 85%, al menos alrededor de 90%, al menos alrededor de 95%, al menos alrededor de 96%, al menos alrededor de 97%, al menos alrededor de 98%, al menos alrededor de 99% o más de identidad de secuencia con la secuencia de aminoácidos de uno cualquiera de: aminoácidos 21 a 513 de SEC NO: 2; aminoácidos 20 a 510 de SEC ID NO: 4; aminoácidos 1 a 493 de SEC ID NO 6; o aminoácidos 1 a 491 de SEC ID NO: 8, y que retiene actividad de peroxidasa.

35 Un equivalente funcional de una proteína descrita aquí también se puede identificar, por ejemplo, cribando librerías combinatorias de mutantes, por ejemplo mutantes de truncamiento, de la proteína de la invención para la actividad de peroxidasa. En una realización, se genera una librería abigarrada de variantes mediante mutagénesis combinatoria a nivel del ácido nucleico. Una librería abigarrada de variantes se puede producir, por ejemplo, ligando enzimáticamente una mezcla de oligonucleótidos sintéticos en secuencias génicas de manera que un conjunto degenerado de secuencias proteicas potenciales sea expresable como polipéptidos individuales, o, como alternativa, como un conjunto de proteínas de fusión más grandes (por ejemplo, para la presentación de fagos). Existe una variedad de métodos que se pueden usar para producir librerías de variantes potenciales de los polipéptidos descritos aquí a partir de una secuencia oligonucleotídica degenerada. Los métodos para sintetizar oligonucleótidos degenerados son conocidos en la técnica (véase, por ejemplo, Narang (1983) Tetrahedron 39:3; Itakura et al. (1984) Annu. Rev. Biochem. 53:323; Itakura et al. (1984) Science 198:1056; Ike et al. (1983) Nucleic Acid Res. 11:477).

40 Además, se pueden usar librerías de fragmentos de la secuencia codificante de un polipéptido descrito aquí para generar una población abigarrada de polipéptidos para cribar una selección subsiguiente de variantes. Por ejemplo, una librería de fragmentos de secuencias codificantes se puede generar tratando un fragmento de PCR bicatenario de la secuencia codificante de interés con una nucleasa en condiciones en las que el muescado se produce sólo alrededor de una vez por molécula, desnaturalizando el ADN bicatenario, renaturalizando el ADN para formar ADN bicatenario que puede incluir pares sentido/antisentido de diferentes productos muescados, eliminando porciones monocatenarias de los dúplex reformados mediante tratamiento con nucleasa S1, y ligando la librería de fragmentos resultante en un vector de expresión. Mediante este método, se puede derivar una librería de expresión que codifica fragmentos N-terminales e internos de diversos tamaños de la proteína de interés.

Se conocen en la técnica varias técnicas para cribar productos génicos de librerías combinatorias obtenidos mediante mutaciones puntuales de truncamiento, y para cribar librerías de ADNc para productos génicos que tienen una propiedad seleccionada. Las técnicas más ampliamente usadas, que son susceptibles a un análisis de alta producción, para cribar librerías génicas grandes, incluyen típicamente clonar la librería génica en vectores de expresión replicables, transformar células apropiadas con la librería resultante de vectores, y expresar los genes combinatorios en condiciones en las que la detección de una actividad deseada facilita el aislamiento del vector que codifica el gen cuyo producto se detectó. Se puede usar mutagénesis de conjunto recursiva (REM), una técnica que potencia la frecuencia de mutantes funcionales en las librerías, en combinación con los ensayos de cribado para identificar variantes de una proteína de la invención (Arkin y Yourvan (1992) Proc. Natl. Acad. Sci. USA 89:7811-7815; Delgrave et al. (1993) Protein Engineering 6(3):327-331).

Además de las secuencias que codifican peroxidasas descritas aquí, será manifiesto para la persona experta en la técnica que pueden existir en una población dada polimorfismos de secuencias de ADN, que pueden conducir a cambios en la secuencia de aminoácidos de las proteínas peroxidasas. Tales polimorfismos genéticos pueden existir en células de diferentes poblaciones, o en una población debido a variación alélica natural. Las variantes alélicas también pueden incluir equivalentes funcionales.

Un equivalente funcional de una secuencia que codifica peroxidasa adecuada para uso en la invención se puede obtener usando una sonda marcada que comprende un ácido nucleico aislado que codifica toda o una porción de la secuencia según uno cualquiera de: nucleótidos 61 a 1539 de SEC ID NO: 1; nucleótidos 58 a 1530 de SEC ID NO: 3; nucleótidos 1 a 1479 de SEC ID NO: 5; o nucleótidos 1 a 1473 de SEC ID NO: 7, o una variante de cualquiera de ellos; cribando una librería de fragmentos de ácidos nucleicos con la sonda marcada en condiciones que permiten la hibridación de la sonda a fragmentos de ácidos nucleicos en la librería, formando de ese modo dúplex de ácidos nucleicos, y preparando una secuencia génica de longitud completa a partir de los fragmentos de ácidos nucleicos en cualquier dúplex marcado para obtener un gen relacionado con dicha secuencia codificante.

En una realización, un polinucleótido con actividad de peroxidasa útil para preparar un polipéptido para uso en la invención puede comprender una secuencia que tiene al menos alrededor de 55%, al menos alrededor de 65%, al menos alrededor de 70%, al menos alrededor de 75%, al menos alrededor de 80%, al menos alrededor de 85%, al menos alrededor de 90%, al menos alrededor de 91%, al menos alrededor de 92%, al menos alrededor de 93%, al menos alrededor de 94%, al menos alrededor de 95%, al menos alrededor de 96%, al menos alrededor de 97%, al menos alrededor de 99% o más de identidad de secuencia con uno cualquiera de: nucleótidos 61 a 1539 de SEC ID NO: 1; nucleótidos 58 a 1530 de SEC ID NO: 3; nucleótidos 1 a 1479 de SEC ID NO: 5; o nucleótidos 1 a 1473 de SEC ID NO: 7, o el complemento inverso de cualquiera de ellos.

Un polipéptido con actividad de peroxidasa adecuado para uso en la invención puede comprender una secuencia que tiene al menos alrededor de 55%, al menos alrededor de 65%, al menos alrededor de 70%, al menos alrededor de 75%, al menos alrededor de 80%, al menos alrededor de 85%, al menos alrededor de 90%, al menos alrededor de 91%, al menos alrededor de 92%, al menos alrededor de 93%, al menos alrededor de 94%, al menos alrededor de 95%, al menos alrededor de 96%, al menos alrededor de 97%, al menos alrededor de 98%, al menos alrededor de 99% o más de identidad de secuencia con uno cualquiera de los aminoácidos 21 a 513 de SEC NO: 2; aminoácidos 20 a 510 de SEC ID NO: 4; aminoácidos 1 a 493 de SEC ID NO 6; o aminoácidos 1 a 491 de SEC ID NO: 8.

La presente invención se refiere a métodos en los que se usan polinucleótidos que codifican enzimas peroxidasas, que comprenden una secuencia de aminoácidos según uno cualquiera de: aminoácidos 21 a 513 de SEC NO: 2; aminoácidos 20 a 510 de SEC ID NO: 4; aminoácidos 1 a 493 de SEC ID NO 6; o aminoácidos 1 a 491 de SEC ID NO: 8 o un equivalente funcional o un fragmento de tal polipéptido. Las secuencias que tienen SEC ID NO: 5 y SEC ID NO: 7 (que codifican peroxidasas) son versiones optimizadas de los genes como se aislaron originalmente, en el que tiene lugar una optimización de codones. La optimización de codones se puede usar según métodos conocidos por la persona experta en la técnica, y es especialmente adecuada para la expresión mejorada del gen en una célula hospedante.

También son adecuados para generar enzimas para uso en la invención los polinucleótidos que comprenden una secuencia hibridable en condiciones restrictivas, preferiblemente en condiciones muy restrictivas, al complemento inverso de uno cualquiera de: nucleótidos 61 a 1539 de SEC ID NO: 1; nucleótidos 58 a 1530 de SEC ID NO: 3; nucleótidos 1 a 1479 de SEC ID NO: 5; o nucleótidos 1 a 1473 de SEC ID NO: 7.

Ventajosamente, tales polinucleótidos se pueden obtener a partir de hongos, preferiblemente hongos filamentosos, en particular a partir de *Marasmius scorodoni*.

Como se usa aquí y en lo sucesivo, el término "gen" y la expresión "gen recombinante" se refieren a moléculas de ácidos nucleicos que se pueden aislar de ADN cromosómico, que incluyen un marco de lectura abierto que codifica una proteína, por ejemplo una enzima blanqueante de *M. scorodoni*. Un gen puede incluir secuencias codificantes, secuencias no codificantes, intrones y secuencias reguladoras. Además, un gen se refiere a una molécula de ácido nucleico aislada como se define aquí.

- Una molécula de ácido nucleico útil para generar un polipéptido adecuado para uso en un método de la presente invención, tal como uno descrito aquí, se puede aislar usando técnicas de biología molecular estándar y la información de secuencia proporcionada aquí. Por ejemplo, usando toda o una porción de la secuencia de ácido nucleico que comprende uno o más de: nucleótidos 61 a 1539 de SEC ID NO: 1; nucleótidos 58 a 1530 de SEC ID NO: 3; nucleótidos 1 a 1479 de SEC ID NO: 5; o nucleótidos 1 a 1473 de SEC ID NO: 7 como sonda de hibridación, se pueden aislar moléculas de ácidos nucleicos según la invención usando técnicas de hibridación y de clonación estándar (por ejemplo, como se describe en Sambrook, J., Fritsh, E. F., y Maniatis, T. *Molecular Cloning: A Laboratory Manual*. 2ª ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989).
- Además, una molécula de ácido nucleico que engloba todo o una porción de uno cualquiera de: nucleótidos 61 a 1539 de SEC ID NO: 1; nucleótidos 58 a 1530 de SEC ID NO: 3; nucleótidos 1 a 1479 de SEC ID NO: 5; o nucleótidos 1 a 1473 de SEC ID NO: 7, se puede aislar mediante la reacción en cadena de la polimerasa (PCR) usando cebadores oligonucleotídicos sintéticos diseñados basándose en esa información de secuencia.
- El ácido nucleico adecuado de la invención se puede amplificar usando ADNc, ARNm, o, como alternativa, ADN genómico, como molde, y cebadores oligonucleotídicos apropiados según técnicas de amplificación mediante PCR estándar. El ácido nucleico así amplificado se puede clonar en un vector apropiado y se puede caracterizar mediante análisis de secuencia de ADN.
- Además, los oligonucleótidos que corresponden a o hibridables a secuencias nucleotídicas descritas aquí se pueden preparar mediante técnicas sintéticas estándar, por ejemplo usando un sintetizador de ADN automatizado.
- Una molécula de ácido nucleico útil para generar un polipéptido adecuado para uso en un método de la invención puede comprender una molécula de ácido nucleico que es el complemento inverso de la secuencia nucleotídica descrita aquí, o un equivalente funcional de estas secuencias nucleotídicas.
- Una molécula de ácido nucleico que es complementaria a otra secuencia nucleotídica es aquella que es suficientemente complementaria a la otra secuencia nucleotídica de manera que se puede hibridar a la otra secuencia nucleotídica formando de ese modo un dúplex estable.
- Un aspecto de la descripción se refiere a moléculas de ácidos nucleicos que codifican un polipéptido adecuado para uso en el método de la invención, o un equivalente funcional del mismo, tal como un fragmento o dominio biológicamente activo, así como a moléculas de ácidos nucleicos suficientes para uso como sondas de hibridación para identificar moléculas de ácidos nucleicos que codifican un polipéptido de la invención y fragmentos de tales moléculas de ácidos nucleicos adecuados para uso como cebadores de PCR para la amplificación o mutación de moléculas de ácidos nucleicos.
- Un "polinucleótido aislado" o "ácido nucleico aislado" es un ADN o ARN que no está inmediatamente contiguo con ambas secuencias codificantes con las que está inmediatamente contiguo (una en el extremo 5' y una en el extremo 3') en el genoma de origen natural del organismo a partir del que deriva. De este modo, un ácido nucleico puede incluir algunas o todas las secuencias no codificantes (por ejemplo, promotoras) de 5' que están inmediatamente contiguas a la secuencia codificante. Por lo tanto, la expresión incluye, por ejemplo, un ADN recombinante que se incorpora en un vector, en un plásmido o virus que se replica autónomamente, o en el ADN genómico de una procarionota o eucariota, o que existe como una molécula separada (por ejemplo, un ADNc o un fragmento de ADN genómico producido mediante PCR o tratamiento con endonucleasas de restricción) independiente de otras secuencias. También incluye un ADN recombinante que es parte de un gen híbrido que codifica un polipéptido adicional que está sustancialmente libre de material celular, material vírico, o medio de cultivo (cuando se produce mediante técnicas de ADN recombinante), o precursores químicos u otros compuestos químicos (cuando se sintetiza químicamente). Además, un "fragmento de ácido nucleico aislado" es un fragmento de ácido nucleico que no es de origen natural como fragmento y no se encontraría en el estado natural.
- Como se usa aquí, las expresiones "polinucleótido" o "molécula de ácido nucleico" pretenden incluir moléculas de ADN (por ejemplo, ADNc o ADN genómico) y moléculas de ARN (por ejemplo, ARNm) y análogos del ADN o ARN generados usando análogos nucleotídicos. La molécula de ácido nucleico puede ser monocatenaria o bicatenaria, pero preferiblemente es ADN bicatenario. El ácido nucleico se puede sintetizar usando análogos o derivados oligonucleotídicos (por ejemplo, inosina o nucleótidos de fosforotioato). Tales oligonucleótidos se pueden usar, por ejemplo, para preparar ácidos nucleicos que tienen capacidades de emparejamiento de bases alteradas o una mayor resistencia a nucleasas.
- La información de secuencia como se proporciona aquí no debería interpretarse tan estrechamente como para requerir la inclusión de bases erróneamente identificadas. Las secuencias específicas descritas aquí se pueden usar fácilmente para aislar el gen completo a partir de hongos, en particular *M. scorodoni*, que a su vez se puede someter fácilmente a análisis de secuencia adicionales, identificando de ese modo errores de secuenciación.
- Excepto que se indique de otro modo, todas las secuencias nucleotídicas determinadas secuenciando una molécula de ADN aquí se determinaron usando un secuenciador automatizado de ADN, y todas las secuencias de aminoácidos de polipéptidos codificados por moléculas de ADN determinadas aquí se predijeron mediante

traducción de una secuencia de ADN determinada como anteriormente. Por lo tanto, como se conoce en la técnica para cualquier secuencia de ADN determinada mediante este enfoque automatizado, cualquier secuencia nucleotídica determinada aquí puede contener algunos errores. Las secuencias nucleotídicas determinadas mediante automatización son típicamente al menos alrededor de 90% idénticas, más típicamente al menos alrededor de 95% idénticas, al menos alrededor de 99,9% idénticas a la secuencia nucleotídica real de la molécula de ADN secuenciada. La secuencia real se puede determinar de forma más precisa mediante otros enfoques que incluyen métodos de secuenciación de ADN manuales, bien conocidos en la técnica. Como también se conoce en la técnica, una inserción o supresión individual en una secuencia nucleotídica determinada, en comparación con la secuencia real, provocará un desplazamiento del marco en la traducción de la secuencia nucleotídica, de manera que la secuencia de aminoácidos predicha codificada por una secuencia nucleotídica determinada será completamente diferente de la secuencia de aminoácidos realmente codificada por la molécula de ADN secuenciada, comenzando en el punto de tal inserción o supresión.

La persona experta en la técnica es capaz de identificar tales bases erróneamente identificadas, y sabe cómo corregir tales errores.

Las expresiones "homología" o "porcentaje de identidad" o "identidad de secuencia", y similares, se usan aquí de forma intercambiable. Para los fines de esta invención, se define aquí que, a fin de determinar el grado de identidad de secuencia compartido por dos secuencias de aminoácidos o dos secuencias de ácidos nucleicos, las secuencias se alinean con fines de comparación óptima (por ejemplo, se pueden introducir saltos en la secuencia de una primera secuencia de aminoácidos o de ácido nucleico para el alineamiento óptimo con una segunda secuencia de aminoácidos o de ácido nucleico). Tal alineamiento se puede llevar a cabo a lo largo de las longitudes completas de las secuencias que se comparan. Como alternativa, el alineamiento se puede llevar a cabo a lo largo de una longitud más corta, por ejemplo a lo largo de alrededor de 20, alrededor de 50, alrededor de 100 o más ácidos nucleicos/bases o aminoácidos. Entonces se comparan los restos de aminoácidos o los nucleótidos en las posiciones de aminoácidos o en las posiciones nucleotídicas correspondientes. Cuando una posición en la primera secuencia está ocupada por el mismo resto de aminoácido o el mismo nucleótido que la posición correspondiente en la segunda secuencia, entonces las moléculas son idénticas en esa posición. El porcentaje de identidad entre las dos secuencias es una función del número de posiciones idénticas compartidas por las secuencias (es decir, % de identidad = número de posiciones idénticas/número total de posiciones (es decir, posiciones que se solapan) x 100). Preferiblemente, las dos secuencias tienen la misma longitud. Las dos secuencias se pueden alinear a lo largo de sus longitudes completas.

La persona experta estará al tanto del hecho de que existen varios programas de ordenador diferentes para determinar la homología entre dos secuencias. Por ejemplo, se puede lograr una comparación de secuencias y la determinación del porcentaje de identidad entre las dos secuencias usando un algoritmo matemático.

En una realización preferida, la identidad de secuencia entre dos secuencias de aminoácidos se determina usando el algoritmo de Needleman y Wunsch (J. Mol. Biol. (48):444-453 (1970)) que se ha incorporado en el programa GAP en el paquete de software GCG (disponible en <http://www.accelrys.com/solutions/bioinformatician/>), usando una matriz Blossom 62 o una matriz PAM250, y un peso de salto de 16, 14, 12, 10, 8, 6, ó 4 y un peso de longitud de 1, 2, 3, 4, 5, ó 6. La persona experta apreciará que todos estos parámetros diferentes producirán resultados ligeramente diferentes, pero que el porcentaje global de identidad de las dos secuencias no se ve alterado significativamente cuando se usan algoritmos diferentes.

En otra realización, la identidad de secuencia entre dos secuencias nucleotídicas se determina usando el programa GAP en el paquete de software GCG, usando una matriz NWSgapdna.CMP y un peso de salto de 40, 50, 60, 70, ó 80 y un peso de longitud de 1, 2, 3, 4, 5, ó 6. En otra realización, el porcentaje de identidad de dos secuencias de aminoácidos o nucleotídicas se determina usando el algoritmo de E. Meyers y W. Miller (CABIOS, 4:11-17 (1989)) que se ha incorporado en el programa ALIGN (versión 2.0) (disponible en: <http://vega.igh.cnrs.fr/bin/align-guess.cgi>) usando una tabla de restos de peso PAM120, una penalización de longitud de salto de 12 y una penalización de salto de 4.

Como alternativa, la identidad de secuencia entre dos secuencias de aminoácidos o nucleotídicas se puede determinar usando, por ejemplo, el algoritmo de E. Meyers y W. Miller (CABIOS, 4:11-17 (1989)) que se ha incorporado en el programa ALIGN (versión 2.0) (disponible en el ALIGN Query usando datos de secuencias del servidor Genestream IGH Montpellier Francia <http://vega.igh.cnrs.fr/bin/align-guess.cgi>) usando una tabla de restos de peso PAM120, una penalización de longitud de salto de 12 y una penalización de salto de 4.

Las secuencias de ácidos nucleicos y proteicas de la presente descripción se pueden usar adicionalmente como una "secuencia de consulta" para llevar a cabo una búsqueda frente a bases de datos públicas para, por ejemplo, identificar otros miembros de la familia o secuencias relacionadas. Tales búsquedas se pueden llevar a cabo usando los programas BLASTN, BLASTP y BLASTX (versión 2.0) de Altschul, et al. (1990) J. Mol. Biol. 215:403-10. Las búsquedas de nucleótidos BLAST en las bases de datos de nucleótidos se pueden llevar a cabo con el programa BLASTN para obtener secuencias nucleotídicas homólogas a moléculas de ácidos nucleicos ZFX de la invención. Las búsquedas BLAST con una secuencia nucleotídica traducida en las bases de datos de proteínas se pueden llevar a cabo con el programa BLASTX para obtener secuencias de aminoácidos homólogas al gen ZFX traducido de

la invención. Como alternativa, para comparación de secuencias de proteínas con las bases de datos de proteínas, se puede usar el programa BLASP con la matriz Blosum 62, un umbral esperado = 10, longitud de palabra = 3, costes de existencia de salto de 11 y costes de extensión de salto de 1. Cuando se utilizan programas BLAST, se pueden usar los parámetros por defecto de los programas respectivos (por ejemplo, BLASTX, BLASTP y BLASTN). Véase la página del National Center for Biotechnology Information en <http://www.ncbi.nlm.nih.gov/> para toda la información relevante en búsquedas de homología en bases de datos públicas.

Los algoritmos BLASTP y BLAST N se pueden usar para calcular la identidad de secuencia o alinear secuencias (tal como identificar secuencias equivalentes o correspondientes, por ejemplo en sus ajustes por defecto).

El software para llevar a cabo los análisis BLAST está públicamente disponible a través del National Center for Biotechnology Information (<http://www.ncbi.nlm.nih.gov/>). Este algoritmo implica identificar en primer lugar pares de secuencias de alta puntuación (HSPs) identificando palabras cortas de longitud W en la secuencia de consulta que coincide o satisface alguna puntuación umbral T de valor positivo cuando se alinea con una palabra de la misma longitud en la secuencia de la base de datos. T se denomina como puntuación umbral de palabra vecina. Estos resultados de palabra vecina iniciales actúan como semillas para iniciar búsquedas para encontrar HSPs que los contienen. Los resultados de palabras se extienden en ambas direcciones a lo largo de cada secuencia tanto como se pueda incrementar la puntuación de alineamiento acumulativa. Las extensiones de los resultados de palabra en cada dirección se detienen cuando: la puntuación de alineamiento acumulativa cae en la cantidad X desde su valor máximo logrado; la puntuación acumulativa llega a cero o por debajo, debido a la acumulación de uno o más alineamientos de restos de puntuación negativa; o se alcanza el extremo de cualquier secuencia. Los parámetros del algoritmo BLAST W, T y X determinan la sensibilidad y velocidad del alineamiento. El programa BLASTN de la comparación ADN-ADN usa por defecto una longitud de palabra (W) de 11, una expectativa (E) de 10, y una comparación de ambas hebras. El problema BLASTP para la comparación proteína-proteína usa por defecto una longitud de palabra (W) de 3, la matriz de puntuación BLOSUM62, una penalización de existencia de salto de 11 con una penalización de extensión de salto de 1, y una expectativa (E) de 10.

El algoritmo BLAST lleva a cabo un análisis estadístico de la similitud entre dos secuencias. Una medida de la similitud proporcionada por el algoritmo BLAST es la probabilidad de suma más pequeña (P(N)), que proporciona una indicación de la probabilidad mediante la cual se produciría por casualidad un emparejamiento entre dos secuencias nucleotídicas o de aminoácidos. Por ejemplo, una secuencia se considera similar a otra secuencia si la probabilidad de suma más pequeña en la comparación de la primera secuencia con la segunda secuencia es menor que alrededor de 1, preferiblemente menor que alrededor de 0,1, más preferiblemente menor que alrededor de 0,01, y lo más preferible menor que alrededor de 0,001.

También, se puede usar un motivo peptídico para identificar genes que codifican proteínas que contienen este motivo peptídico. En lugar de un motivo peptídico, se puede usar también una combinación de dos o más motivos peptídicos para identificar genes que codifican proteínas que contienen los motivos peptídicos. Cuando se identifican uno o varios motivos peptídicos que codifican proteasas específicas de prolina, es posible de este modo identificar genes que codifican proteasas específicas de prolina usando uno o una combinación de varios de estos motivos peptídicos. Las proteasas específicas de prolina se usan como ejemplo de cómo se pueden identificar tales genes, pero los métodos descritos son aplicables de forma general. Un motivo peptídico se puede usar para una búsqueda en secuencias de ADN traducidas a partir de un banco de datos de ADN, o secuencias proteicas a partir de un banco de datos de secuencias proteicas usando un programa como PatScan (<http://www.unix.mcs.anl.gov/compbio/PatScan/HTML/>). La secuencia de aminoácidos se ha de introducir en un formato especial que se describe en el sitio web. Otro método que se puede llevar a cabo es usar la secuencia del motivo para una búsqueda en secuencias de ADN traducidas a partir de un banco de datos de ADN o secuencias de proteínas a partir de un banco de datos de secuencias de proteínas usando un programa como <http://myhits.isb-sib.ch/cgi-bin/>. Para este programa, el motivo se introduce en el campo de búsqueda en el formato denominado Prosite, y las bases de datos son investigadas en busca de la presencia del motivo en la secuencia proteica o en la secuencia de ADN traducida. Este método se puede usar para identificar genes fúngicos que codifican proteasas específicas de prolina útiles. Los genes que se identifican usando uno de estos métodos se pueden traducir entonces en una secuencia proteica usando programas conocidos por los expertos en la técnica, y se pueden inspeccionar para determinar la presencia de una secuencia señal en su término amino. Para detectar una secuencia señal, se puede usar un programa como SignalP (<http://www.cbs.dtu.dk/services/SignalP/>). La búsqueda de una secuencia proteica que contenga tanto las secuencias de consenso como una secuencia señal predicha da una gran ventaja para la producción industrial de tal enzima.

Las secuencias, nucleótidos y polipéptidos identificados de esta manera pueden ser adecuados para uso en el método de la invención.

Como se usa aquí, el término "hibridar" pretende describir condiciones para la hibridación y lavado en las que las secuencias nucleotídicas que tienen al menos alrededor de 55%, al menos alrededor de 40%, al menos alrededor de 70%, más preferiblemente al menos alrededor de 80%, incluso más preferiblemente al menos alrededor de 85%, al menos alrededor de 90%, más preferiblemente al menos alrededor de 95%, al menos alrededor de 98%, al menos alrededor de 99% de identidad de secuencia entre sí permanecen hibridadas típicamente entre sí.

Un ejemplo preferido no limitante de tales condiciones de hibridación son la hibridación en 6X cloruro de sodio/citrato de sodio (SSC) a alrededor de 45°C, seguido de uno o más lavados en 1 X SSC, 0,1% de SDS a alrededor de 50°C, preferiblemente a alrededor de 55°C, preferiblemente alrededor de 60°C e incluso más preferiblemente alrededor de 65°C.

- 5 Las condiciones muy restrictivas incluyen, por ejemplo, la hibridación a alrededor de 68°C en 5x SSC/5x disolución de Denhardt/1,0% de SDS y un lavado en 0,2x SSC/0,1% de SDS a alrededor de la temperatura ambiente. Como alternativa, el lavado se puede llevar a cabo a 42°C.

El experto sabrá qué condiciones aplicar para las condiciones de hibridación restrictivas y muy restrictivas. Una guía adicional con respecto a tales condiciones está fácilmente disponible en la técnica, por ejemplo en Sambrook et al., 1989, Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Press, N.Y.; y Ausubel et al. (eds.), 1995, Current Protocols in Molecular Biology, (John Wiley & Sons, N.Y.).

15 Por supuesto, un polinucleótido que se hibrida sólo a una secuencia poly A (tal como el tramo poly(A) terminal de 3' de los ARNm), o a un tramo complementario de restos T (o U), no se incluiría en un polinucleótido de la invención usado para hibridarse específicamente a una porción de un ácido nucleico de la invención, puesto que tal polinucleótido se hibridaría a cualquier molécula de ácido nucleico que contenga un tramo poly(A) o su complemento (por ejemplo, prácticamente cualquier clon de ADNc bicatenario).

En un enfoque típico, las librerías de ADNc construidas a partir de otros organismos, por ejemplo hongos filamentosos, en particular de la especie *Marasmius*, se pueden cribar para obtener polinucleótidos adicionales que codifican polipéptidos que se pueden usar en la invención.

20 Por ejemplo, se pueden cribar cepas de *Marasmius* para polinucleótidos con actividad de peroxidasa homólogos mediante análisis de transferencia Northern. Con la detección de transcritos homólogos a polinucleótidos según la invención, se pueden construir librerías de ADNc a partir de ARN aislado de la cepa apropiada, utilizando técnicas estándar bien conocidas por los expertos en la técnica. Como alternativa, una librería de ADN genómico total se puede cribar usando una sonda hibridable a un polinucleótido con actividad de peroxidasa como se describe aquí.

25 Las secuencias génicas homólogas se pueden aislar, por ejemplo, llevando a cabo PCR usando dos conjuntos de cebadores oligonucleotídicos degenerados, diseñados en base a las secuencias nucleotídicas como se enseñan aquí.

30 El molde para la reacción puede ser ADNc obtenido mediante transcripción inversa de ARNm preparado a partir de cepas que se sabe o que se sospecha que expresan un polinucleótido según la invención. El producto de la PCR se puede subclonar y secuenciar para asegurarse de que las secuencias amplificadas representan las secuencias de una nueva secuencia de ácido nucleico de peroxidasa, o un equivalente funcional de la misma.

35 El fragmento de la PCR se puede usar entonces para aislar el clon de ADNc de longitud completa mediante una variedad de métodos conocidos. Por ejemplo, el fragmento amplificado se puede marcar y usar para cribar una librería de ADNc cosmiótica o bacteriófaga. Como alternativa, el fragmento marcado se puede usar para cribar una librería genómica.

40 La tecnología de la PCR también se puede usar para aislar secuencias de ADNc de longitud completa a partir de otros organismos. Por ejemplo, el ARN se puede aislar, siguiendo procedimientos estándar, a partir de una fuente celular o tisular apropiada. Se puede llevar a cabo una reacción de transcripción inversa en el ARN usando un cebador oligonucleotídico específico para el extremo 5' del fragmento amplificado para el cebado de la síntesis de la primera hebra.

45 Al híbrido resultante ARN/ADN se le pueden añadir entonces "colas" (por ejemplo, con guaninas) usando una reacción de transferasa terminal estándar, el híbrido se puede digerir con ARNasa H, y entonces se puede cebar una síntesis de segunda hebra (por ejemplo, con un cebador poly-C). De este modo, se pueden aislar fácilmente secuencias de ADNc en dirección 5' del fragmento amplificado. Para un repaso de estrategias de clonación útiles, véanse por ejemplo Sambrook et al., más arriba; y Ausubel et al., más arriba.

50 Un vector, preferiblemente un vector de expresión, puede contener un ácido nucleico que codifica una proteína peroxidasa como se define aquí, o un equivalente funcional de la misma. Como se usa aquí, el término "vector" se refiere a una molécula de ácido nucleico capaz de transportar otro ácido nucleico al que se ha enlazado. Un tipo de vector es un "plásmido", que se refiere a un bucle de ADN bicatenario circular en el que se pueden ligar segmentos de ADN adicionales. Otro tipo de vector es un vector vírico, en el que se pueden ligar segmentos de ADN adicionales al genoma vírico. Ciertos vectores son capaces de replicación autónoma en una célula hospedante en la que se introducen (por ejemplo, vectores bacterianos que tienen un origen de replicación bacteriano, y vectores de mamíferos episómicos). Otros vectores (por ejemplo, vectores de mamíferos no episómicos) se integran en el genoma de una célula hospedante con la introducción en la célula hospedante, y de ese modo se replican junto con el genoma del hospedante. Además, ciertos vectores son capaces de dirigir la expresión de genes a los que están enlazados operativamente. Tales vectores se denominan aquí como "vectores de expresión". En general, los vectores de expresión de utilidad en técnicas de ADN recombinante están a menudo en forma de plásmidos. Los

términos “plásmido” y “vector” se pueden usar aquí de forma intercambiable, ya que el plásmido es la forma más habitualmente usada de vector. Sin embargo, el método de la invención pretende incluir tales otras formas de vectores de expresión, tales como vectores víricos (por ejemplo, retrovirus defectuosos en la replicación, adenovirus y virus adenoasociados), que tienen funciones equivalentes.

5 Los vectores de expresión recombinantes descritos aquí comprenden un ácido nucleico en una forma adecuada para la expresión del ácido nucleico en una célula hospedante, lo que significa que el vector de expresión recombinante incluye una o más secuencias reguladoras, seleccionadas en base a las células hospedantes a usar para la expresión, que están enlazadas operativamente a la secuencia de ácido nucleico a expresar. En un vector de expresión recombinante, “enlazada operativamente” quiere decir que la secuencia nucleotídica de interés está  
10 enlazada a la secuencia o secuencias reguladoras de manera que permite la expresión de la secuencia nucleotídica (por ejemplo, en un sistema de transcripción/traducción *in vitro* o en una célula hospedante cuando el vector se introduce en la célula hospedante). La expresión “secuencia reguladora” pretende incluir promotores, potenciadores y otros elementos de control de la expresión (por ejemplo, señal de poliadenilación). Tales secuencias reguladoras se describen, por ejemplo, en Goeddel; Gene Expression Technology: Methods in Enzymology 185, Academic Press, San Diego, CA (1990). Las secuencias reguladoras incluyen aquellas que dirigen la expresión constitutiva de una secuencia nucleotídica en muchos tipos de células hospedantes, y aquellas que dirigen la expresión de la secuencia nucleotídica solamente en una cierta célula hospedante (por ejemplo, secuencias reguladoras específicas de tejidos). Los expertos en la técnica apreciarán que el diseño del vector de expresión puede depender de factores tales como la elección de la célula hospedante a transformar, el nivel de expresión de la proteína deseada, etc.  
20 Tales vectores de expresión se pueden introducir en células hospedantes para producir de ese modo proteínas o péptidos, codificados por ácidos nucleicos como se describen aquí.

Los vectores de expresión recombinantes se pueden diseñar para la expresión de proteínas peroxidasas en células procariontas o eucariotas. Por ejemplo, las proteínas peroxidasas se pueden expresar en células fúngicas, células bacterianas tales como *E. coli*, células de insectos (usando vectores de expresión de baculovirus), células de levadura o células de mamíferos. Las células hospedantes adecuadas se explican adicionalmente en Goeddel, Gene Expression Technology: Methods in Enzymology 185, Academic Press, San Diego, CA (1990). Como alternativa, el vector de expresión recombinante se puede transcribir y traducir *in vitro*, por ejemplo usando secuencias reguladoras del promotor T7 y T7 polimerasa.  
25

Los vectores de expresión que se pueden usar para generar un polipéptido útil en la presente invención incluyen vectores derivados de cromosomas, de episomas y de virus, por ejemplo vectores derivados de plásmidos bacterianos, bacteriófago, episoma de levadura, elementos cromosómicos de levadura, virus tales como baculovirus, papovavirus, virus de la vacuna, adenovirus, virus de la viruela aviar, virus y retrovirus de la pseudorrabia, y vectores derivados de sus combinaciones, tales como aquellos derivados de elementos genéticos plasmídicos y bacteriofágicos, tales como cósmidos y fagómidos.  
30

El inserto de ADN debería enlazarse operativamente a un promotor apropiado, tal como el promotor PL del fago lambda, los promotores lac, trp y tac de *E. coli*, los promotores temprano y tardío de SV40 y los promotores de las LTR retrovíricas, por nombrar unos pocos. Otros promotores adecuados serán conocidos por la persona experta. En una realización específica, se prefieren promotores que son capaces de dirigir un nivel elevado de expresión de peroxidasas en procariontas u hongos filamentosos. Tales promotores son conocidos en la técnica. Los constructos de expresión pueden contener sitios para la iniciación de la transcripción, terminación, y, en la región transcrita, un sitio de unión al ribosoma para la traducción. La porción codificante de los transcritos maduros expresada por los constructos incluirá al comienzo un AUG que inicia la traducción, y un codón de terminación apropiadamente colocado en el extremo del polipéptido a traducir.  
40

El ADN del vector se puede introducir en células procariontas o eucariotas vía técnicas convencionales de transformación o transfección. Como se usa aquí, los términos “transformación” y “transfección” se refieren a una variedad de técnicas reconocidas en la técnica para introducir ácido nucleico extraño (por ejemplo, ADN) en una célula hospedante, incluyendo coprecipitación con fosfato de calcio o cloruro de calcio, transfección mediada por DEAE-dextrano, transducción, infección, lipofección, transfección mediada por lípidos catiónicos, o electroporación. Los métodos adecuados para transformar o transfectar células hospedantes se pueden encontrar en Sambrook, et al. (Molecular Cloning: A Laboratory Manual, 2ª ed. Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989), Davis et al., Basic Methods in Molecular Biology (1986), y otros manuales de laboratorio.  
50

Para la transfección estable de células de mamíferos, se sabe que, dependiendo del vector de expresión y de la técnica de transfección usada, sólo una pequeña fracción de células puede integrar el ADN extraño en su genoma. A fin de identificar y seleccionar estos integrantes, se introduce generalmente en las células hospedantes un gen que codifica un marcador seleccionable (por ejemplo, resistencia a antibióticos) junto con el gen de interés. Los marcadores seleccionables preferidos incluyen aquellos que confieren resistencia a fármacos, tales como G418, higromicina y metatrexato. El ácido nucleico que codifica un marcador seleccionable se puede introducir en una célula hospedante en el mismo vector que aquel que codifica un polipéptido con actividad de peroxidasa, o se puede introducir en un vector distinto. Las células transfectadas de forma estable con el ácido nucleico introducido se  
60

pueden identificar mediante selección con fármacos (por ejemplo, las células que han incorporado el gen del marcador seleccionable sobrevivirán, mientras que las otras células morirán).

La expresión de proteínas en procariotas se lleva a cabo a menudo en *E. coli* con vectores que contienen promotores constitutivos o inducibles que dirigen la expresión de proteínas de fusión o proteínas no de fusión. Los vectores de fusión añaden un número de aminoácidos a una proteína codificada allí, por ejemplo al término amino de la proteína recombinante. Tales vectores de fusión sirven típicamente para tres fines: 1) para incrementar la expresión de proteína recombinante; 2) para incrementar la solubilidad de la proteína recombinante; y 3) para ayudar a la purificación de la proteína recombinante al actuar como ligando en una purificación por afinidad. Además, en vectores de expresión de fusión, se introduce un sitio de escisión proteolítica en la unión del resto de fusión y la proteína recombinante, para permitir la separación de la proteína recombinante del resto de fusión tras la purificación de la proteína de fusión. Tales enzimas, y sus secuencias de reconocimiento relacionadas, incluyen Factor Xa, trombina y enterocinasa.

Como se indica, los vectores de expresión contendrán preferiblemente marcadores seleccionables. Tales marcadores incluirán dihidrofolato reductasa o resistencia a neomicina para el cultivo de células eucariotas, y resistencia a tetraciclina o a ampicilina para el cultivo en *E. coli* y otras bacterias. Los ejemplos representativos de hospedante apropiado incluyen células bacterianas, tales como *E. coli*, *Streptomyces* y *Salmonella typhimurium*; células fúngicas, tales como levadura; células de insecto tales como *Drosophila S2* y *Spodoptera Sf9*; células de animales tales como CHO, COS y melanoma de Bowes; y células vegetales. Los medios de cultivo y condiciones apropiados para las células hospedantes descritas anteriormente son bien conocidos en la técnica.

Entre los vectores preferidos para uso en bacterias están pQE70, pQE60 y pQE-9, disponibles de Qiagen; los vectores pBS, vectores Phagescript, vectores Bluescript, pNH8A, pNH16A, pNH18A, pNH46A, disponibles de Stratagene; y ptrc99a, pKK223-3, pKK233-3, pDR540, pRIT5 disponibles de Pharmacia. Entre los vectores eucariotas preferidos están PWLNEO, pSV2CAT, pOG44, pZT1 y pSG disponibles de Stratagene; y pSVK3, pBPV, pMSG y pSVL disponibles de Pharmacia. Otros vectores adecuados serán fácilmente manifiestos para el experto.

Entre los promotores bacterianos conocidos para uso en la presente invención incluyen los promotores lacI y lacZ de *E. coli*, los promotores T3 y T7, el promotor gpt, los promotores PR, PL lambda y el promotor trp, el promotor de timidina cinasa de HSV, los promotores temprano y tardío de SV40, los promotores de las LTR retroviricas, tales como aquellos del virus del sarcoma de Rous ("RSV"), y los promotores de metalotioneína, tales como el promotor de metalotioneína I de ratón.

La transcripción del ADN que codifica los polipéptidos de la presente invención por eucariotas superiores se puede incrementar insertando en el vector una secuencia potenciadora. Los potenciadores son elementos de ADN que actúan en cis, habitualmente alrededor de 10 a 300 pb que actúan para incrementar la actividad transcripcional de un promotor en un tipo de célula hospedante dado. Los ejemplos de potenciadores incluyen el potenciador de SV40, que está situado en el lado tardío del origen de replicación en los pb 100 a 270, el potenciador del promotor temprano de citomegalovirus, el potenciador de polioma en el lado tardío del origen de replicación, y los potenciadores adenoviricos.

Para la secreción de la proteína traducida en la luz del retículo endoplásmico, en el espacio periplásmico o en el entorno extracelular, se puede incorporar en el polipéptido expresado una señal de secreción apropiada. Las señales pueden ser endógenas al polipéptido, o pueden ser señales heterólogas.

El polipéptido se puede expresar en una forma modificada, tal como una proteína de fusión, y puede incluir no sólo señales de secreción sino también regiones funcionales heterólogas adicionales. De este modo, por ejemplo, se puede añadir al término N del polipéptido una región de aminoácidos adicionales, particularmente aminoácidos cargados, para mejorar la estabilidad y persistencia en la célula hospedante, durante la purificación o durante la manipulación y almacenamiento subsiguientes. También, se pueden añadir restos peptídicos al polipéptido, para facilitar la purificación.

También se describen aquí células, por ejemplo células hospedantes transformadas o células hospedantes recombinantes que contienen un ácido nucleico que codifica un polipéptido adecuado para uso en el método de la invención. Una "célula transformada" o "célula recombinante" es una célula en la que se ha introducido (o en cuyo ancestro se ha introducido), por medio de técnicas de ADN recombinante, un ácido nucleico como se describe aquí. Se incluyen tanto células procariotas como eucariotas, por ejemplo bacterias, hongos, levadura, y similares.

Los ejemplos de células hospedantes adecuadas son *Microcystis*, *Lepista*, por ejemplo *L. irina*, *Cyathus*, por ejemplo *C. pallidus*, *Ganoderma*, por ejemplo *G. applanatum*, *Ischnoderma*, por ejemplo *I. benzoinum*, *Marasmius*, por ejemplo *M. scorodonius*, *Trametes*, por ejemplo *T. suaveolens* de *T. versicolor*, *Cryptococcus*, por ejemplo *C. laurentii*, *Hypomyces*, por ejemplo *H. odoratus* o *Phaffia*, por ejemplo *P. rhodozyma*, *Phanerochaete* por ejemplo *P. chrysosporium*, *Lentinula* por ejemplo *L. edodes*, *Coprinus* por ejemplo *C. cinereus*, *Gloeophyllum* por ejemplo *G. trabeum*, *Ophiostoma* por ejemplo *O. piliferum*, *Aspergillus* por ejemplo *A. niger*, *A. oryzae*, *A. nidulans*, *Thermomyces*, por ejemplo *T. lanuginosa*, *Sporotrichum*, por ejemplo *S. thermophile*, *Aureobasidium* por ejemplo *A.*

*pullulans*, *Amorphotheca*, por ejemplo *A. resiniae*, *Leucosporidium*, por ejemplo *L. scottii*, *Cunninghamella*, por ejemplo *C. elegans*.

5 Se prefieren especialmente las células procedentes de hongos filamentosos, en particular *Aspergillus* - por ejemplo *Aspergillus oryzae* o *Aspergillus niger* - y *Marasmius* - por ejemplo *Marasmius scorodonius* - o células de levaduras tales como *Pichia* - por ejemplo *Pichia Pastoris* - o células de bacterias.

Se puede escoger una célula hospedante que module la expresión de las secuencias insertadas, o que modifique y procese el producto génico de una manera deseada específica. Tales modificaciones (por ejemplo glucosilación) y procesamiento (por ejemplo escisión) de productos proteicos pueden facilitar el funcionamiento óptimo de la proteína.

10 Diversas células hospedantes tienen mecanismos característicos y específicos para el procesamiento y modificación post-traduccionales de proteínas y productos génicos. Se pueden escoger estirpes celulares o sistemas hospedantes apropiados familiares para los expertos en la técnica de biología molecular y/o microbiología para asegurarse la modificación y procesamiento deseados y correctos de la proteína extraña expresada. Para este fin, se pueden usar células hospedantes eucariotas que poseen la maquinaria celular para el procesamiento apropiado del transcrito primario, la glucosilación y fosforilación del producto génico. Tales células hospedantes son bien conocidas en la técnica.

Las células hospedantes también incluyen, pero no se limitan a, estirpes celulares de mamíferos, tales como CHO, VERO, BHK, HeLa, COS, MDCK, 293, 3T3, WI38, y estirpes celulares de plexos coroideos.

20 Si se desea, los polipéptidos se pueden producir mediante una estirpe celular establemente transfectada. Existe para el público un número de vectores adecuados para la transfección estable de células de mamíferos, y los métodos para construir tales estirpes celulares también son públicamente conocidos, por ejemplo en Ausubel et al. (más arriba).

En el método de la invención, un polipéptido se pone en contacto con un material que comprende un hidrato de carbono que no es almidón.

25 El polipéptido se puede añadir como enzima aislada o purificada, una preparación enzimática, o se puede producir *in situ* mediante un microorganismo capaz de producir dicha enzima. La preparación enzimática puede derivar de diversas fuentes, por ejemplo de plantas, animales y microorganismos. Preferiblemente, la preparación enzimática deriva de un microorganismo, puesto que los microorganismos hacen posible obtener la enzima a escala industrial de manera controlada. La preparación enzimática derivada de un microorganismo se puede obtener mediante procedimientos clásicos de fermentación de una cepa microbiana seleccionada, o mediante fermentación de un microorganismo que sobreexpresa la enzima. El microorganismo puede ser una bacteria, un hongo o levadura. Los ejemplos de microorganismos adecuados son *Microcystis*, *Lepista*, por ejemplo *L. irina*, *Cyathus*, por ejemplo *C. pallidus*, *Ganoderma*, por ejemplo *G. applanatum*, *Ischnoderma*, por ejemplo *I. benzoinum*, *Marasmius*, por ejemplo *M. scorodonius*, *Trametes*, por ejemplo *T. suaveoluens* de *T. versicolour*, *Cryptococcus*, por ejemplo *C. laurentii*, *Hypomyces*, por ejemplo *H. odoratus* o *Phaffia*, por ejemplo *P. rhodozyma*, *Phanerochaete* por ejemplo *P. chrysosporium*, *Lentinula* por ejemplo *L. edodes*, *Coprinus* por ejemplo *C. cinereus*, *Gloeophyllum* por ejemplo *G. trabeum*, *Ophiostoma* por ejemplo *O. piliferum*, *Aspergillus* por ejemplo *A. niger*, *A. oryzae*, *A. nidulans*, *Thermomyces*, por ejemplo *T. lanuginosa*, *Sporotrichum*, por ejemplo *S. thermophile*, *Aureobasidium* por ejemplo *A. pullulans*, *Amorphotheca*, por ejemplo *A. resiniae*, *Leucosporidium*, por ejemplo *L. scottii*, *Cunninghamella*, por ejemplo *C. elegans*.

Se pueden llevar a cabo métodos de la invención en los que el polipéptido se añade como una preparación enzimática derivada de o producida *in situ* por una planta, por ejemplo una planta de maíz, una planta de arroz, una planta de caña de azúcar, una planta de trigo o una planta de cebada.

45 Un material lignocelulósico que comprende un hidrato de carbono que no es almidón adecuado para la oxidación mediante un polipéptido es típicamente lignocelulosa. Los principales polisacáridos que comprenden diferentes restos lignocelulósicos, que se pueden considerar como materia prima renovable potencial, son celulosa (glucanos), hemicelulosas (xilanos, heteroxilanos y xiloglucanos). Además, algo de hemicelulosa puede estar presente como glucomanos, por ejemplo en materias primas derivadas de la madera. La hidrólisis enzimática de estos polisacáridos a azúcares solubles, por ejemplo glucosa, xilosa, arabinosa, galactosa, sacarosa, fructosa, manosa, ramnosa, ribosa, ácido D-galacturónico y otras hexosas y pentosas se produce bajo la acción de diferentes enzimas que actúan concertadamente.

Además, las pectinas y otras sustancias pécticas, tales como arabinanos, pueden constituir una proporción considerable de la masa seca de las paredes celulares típicas de tejidos vegetales no leñosos (alrededor de un cuarto a una mitad de la masa seca puede ser pectinas).

55 En el método de la invención, el material lignocelulósico que comprende un hidrato de carbono que no es almidón puede estar en forma de lignocelulosa, es decir, un "sustrato" o "biomasa" o "materia prima" lignocelulósica. Esto engloba cualquier material que contenga celulosa, hemicelulosa, lignina, proteína, e hidratos de carbono, tal como

- almidón y azúcar. "Biomasa" incluye biomasa virgen o biomasa no virgen, tal como biomasa agrícola, compuestos orgánicos comerciales, desechos de construcción y de demolición, residuos sólidos municipales, papel residual y desecho de estercoleros. Las formas habituales de la biomasa incluyen árboles, arbustos y pastos, trigo, paja de trigo, bagazo de caña de azúcar, maíz, cáscaras de maíz, pepita de maíz, incluyendo fibra de las pepitas, productos y subproductos de la molienda de granos tales como maíz (incluyendo molienda húmeda y molienda seca) así como
- 5 residuos sólidos municipales, papel residual y desecho de estercoleros. "Biomasa mezclada" es cualquier mezcla o combinación de biomasa virgen y no virgen, que tiene preferiblemente de alrededor de 5% a alrededor de 95% en peso de biomasa no virgen. "Biomasa agrícola" incluye ramas, arbustos, cañas, maíz y cáscaras de maíz, cosechas energéticas, bosques, frutas, flores, granos, pastos, cosechas herbáceas, hojas, corteza, agujas, leños, raíces, retoños, cultivos leñosos de corta rotación, matorrales, pasto varilla, árboles, vegetales, vides, y maderas duras y blandas (no incluyendo maderas con materiales dañinos). Además, la biomasa agrícola incluye materiales de
- 10 desecho orgánicos generados de procesos agrícolas que incluyen actividades agrícolas y de silvicultura, especialmente incluyendo desecho de madera de silvicultura. La biomasa agrícola puede ser cualquiera de los señalados anteriormente de forma singular o en cualquier combinación de mezcla de los mismos.
- 15 La biomasa rica en almidón, azúcar o proteína, tal como maíz, granos, frutos y vegetales, se consume habitualmente como alimento. Por el contrario, la biomasa rica en celulosa, hemicelulosa y lignina no es fácilmente digerible, y se utiliza principalmente para productos madereros y papeleros, como combustible, y se desechan de forma típica. Generalmente, el sustrato tiene un contenido elevado de lignocelulosa, incluyendo rastrojo de maíz, paja de arroz, heno, bagazo de caña de azúcar, y otra biomasa agrícola, pasto varilla, desechos de silvicultura, virutas de madera de chopo, virutas de madera de pino, serrín, desecho de estercoleros, y similares, incluyendo cualquier combinación del sustrato.
- 20 En un método de la invención, un material lignocelulósico que comprende un hidrato de carbono que no es almidón se pone en contacto con un polipéptido con actividad de peroxidasa como se describe aquí. El material lignocelulósico que comprende un hidrato de carbono que no es almidón se puede poner en contacto entonces con una enzima que degrada, por ejemplo produce azúcares a partir de, el hidrato de carbono que no es almidón comprendido en el material.
- 25 Además, en el método de la invención se pueden usar polipéptidos auxiliares. Tales polipéptidos pueden ser enzimas. Los polipéptidos auxiliares se pueden usar simultáneamente con, antes de o después del uso de un polipéptido con actividad de peroxidasa descrito aquí.
- 30 Un polipéptido auxiliar puede actuar para modificar adicionalmente el material lignocelulósico que comprende un hidrato de carbono que no es almidón, o puede incrementar adicionalmente la susceptibilidad de dicho material lignocelulósico que comprende un hidrato de carbono que no es almidón a la degradación enzimática. La enzima auxiliar puede actuar para incrementar adicionalmente el efecto de un polipéptido con actividad de peroxidasa como se describe aquí.
- 35 Un polipéptido auxiliar puede incrementar la actividad de una enzima que degrada un hidrato de carbono que no es almidón.
- Un polipéptido auxiliar, cuando se pone en contacto con biomasa en una reacción, puede incrementar la actividad de y/o incrementar la susceptibilidad a una enzima que degrada un hidrato de carbono que no es almidón (tal como una celulasa).
- 40 Una enzima auxiliar se puede hacer reaccionar en la misma vasija como un polipéptido con actividad de peroxidasa como se describe aquí y/o como una enzima que degrada un hidrato de carbono que no es almidón (por ejemplo una celulasa).
- Aunque se entiende que muchas clases de enzimas pueden funcionar como enzimas auxiliares, las enzimas auxiliares pueden estar compuestas en particular (pero sin limitarse a) de enzimas de las siguientes clases:
- 45 amilasas, proteasas, lipasas, glucuronidasas u oxidorreductasas.
- Los polipéptidos, por ejemplo enzimas adecuados para uso en un método de la invención pueden estar compuestos de enzimas procedentes de (1) proveedores comerciales; (2) genes clonados que expresan enzimas; (3) caldo complejo (tal como el que resulta del crecimiento de una cepa microbiana en medios, en el que las cepas segregan proteínas y enzimas a los medios); (4) lisados celulares de cepas que se hicieron crecer como en (3); y (5) material vegetal que expresa enzimas capaces de degradar lignocelulosa.
- 50 Se reconoce que se puede utilizar cualquier combinación de enzimas (usadas además de la peroxidasa descrita aquí). Las enzimas se pueden usar solas o en mezclas, que incluyen, pero no se limitan a, al menos una celulasa; al menos una hemicelulasa; al menos una pectinasa. Es decir, se puede llevar a cabo un método de la invención en el que se usa una celulasa, por ejemplo una endocelulasa y/o una exocelulasa y/o una  $\beta$ -glucosidasa.
- 55 Se entiende que, como se describe anteriormente, una mezcla enzimática puede estar compuesta de un miembro de cada una de estas clases de enzimas, varios miembros de una clase de enzimas (tal como dos o más

hemicelulasas), o cualquier combinación de miembros de estas clases de enzimas (tal como una proteasa, una exocelulasa, y una endoxilanasas; o una exoxilanasas, y una lipasa).

5 Aquí, una celulasa es cualquier polipéptido que es capaz de degradar o modificar celulosa. Un polipéptido que es capaz de degradar celulosa es aquel que es capaz de catalizar el proceso de ruptura de celulosa en unidades más pequeñas, ya sea parcialmente, por ejemplo en celodextrinas, o completamente en monómeros de glucosa. Una celulasa puede dar lugar a una población mixta de celodextrinas y monómeros de glucosa cuando se pone en contacto con la celulosa. Tal degradación tendrá lugar típicamente por medio de una reacción de hidrólisis.

10 Aquí, una hemicelulasa es cualquier polipéptido que es capaz de degradar o modificar hemicelulosa. Es decir, una hemicelulasa puede ser capaz de degradar o modificar uno o más de xilano, glucuronoxilano, arabinoxilano, glucomanano y xiloglucano. Un polipéptido que es capaz de degradar una hemicelulosa es aquel que es capaz de catalizar el proceso de ruptura de la hemicelulosa en polisacáridos más pequeños, ya sea parcialmente, por ejemplo en oligosacáridos, o completamente en monómeros de azúcar, por ejemplo monómeros de azúcar de hexosa o pentosa. Una hemicelulasa puede dar lugar a una población mixta de oligosacáridos y monómeros de azúcar cuando se pone en contacto con la hemicelulasa. Tal degradación tendrá lugar típicamente por medio de una  
15 reacción de hidrólisis.

Aquí, una pectinasa es cualquier polipéptido que es capaz de degradar o modificar pectina. Un polipéptido que es capaz de degradar pectina es aquel que es capaz de catalizar el proceso de ruptura de pectina en unidades más pequeñas, ya sea parcialmente, por ejemplo en oligosacáridos, o completamente en monómeros de azúcar. Una pectinasa puede dar lugar a una población mixta de oligosacáridos y monómeros de azúcar cuando se pone en contacto con la pectinasa. Tal degradación tendrá lugar típicamente por medio de una reacción de hidrólisis.  
20

“Celulasa” incluye tanto exohidrolasas como endohidrolasas, que son capaces de reconocer celulosa, o productos que resultan de la ruptura de celulosa, como sustratos. La celulasa incluye mezclas de enzimas que incluyen endoglucanasas, celobiohidrolasas, glucosidasas, o cualquiera de estas enzimas solas o en combinación con otras actividades. Los organismos que producen una actividad que degrada la celulosa producen a menudo una pléyade de enzimas con diferentes especificidades por el sustrato. De este modo, una cepa identificada por digerir celulosa se puede describir como aquella que tiene una celulasa, cuando de hecho varios tipos de enzimas pueden contribuir a la actividad. Por ejemplo, las preparaciones comerciales de “celulasa” son a menudo mezclas de varias enzimas, tales como actividades de endoglucanasa, exoglucanasa, y glucosidasa.  
25

De este modo, “celulasa” incluye aquí mezclas de tales enzimas, e incluye preparaciones comerciales capaces de degradar celulosa, así como sobrenadante de cultivo o extractos celulares que muestran actividad de degradación de la celulosa, o que actúan sobre productos de ruptura de la degradación de celulosa, tales como celotriosa o celobiosa.  
30

“Celobiohidrolasa” o “exoglucanasa” o “exocelulasa” o “1,4-β-D-glucano celobiohidrolasa” o “celulosa 1,4-β-celobiosidasa” o “celobiosidasa” incluye enzimas que hidrolizan enlaces 1,4-β-D-glucosídicos en celulosa y celotetraosa, liberando celobiosa de los extremos reductores o no reductores de las cadenas. Las enzimas en el grupo EC 3.2.1.91 incluyen estas enzimas tales como CBH1 y CBH2.  
35

“β-glucosidasa” o “glucosidasa” o “β-D-glucósido glucohidrolasa” o “celobiasa” EC 3.2.1.21 incluye enzimas que liberan moléculas de glucosa como producto de su acción catalítica. Estas enzimas reconocen oligómeros de glucosa, tales como celobiosa (un dímero de glucosa enlazado mediante enlaces β-1,4) o celotriosa (un trímero de glucosa enlazado mediante enlaces β-1,4) como sustratos. Típicamente, hidrolizan la β-D-glucosa terminal no reductora, con liberación de β-D-glucosa.  
40

“Endoglucanasa” o “1,4-β-D-glucano 4-glucanohidrolasa” o “β-1,4 endocelulasa” o “endocelulasa” o “celulasa” EC 3.2.1.4 incluye enzimas que escinden polímeros de glucosa unidos mediante enlaces β-1,4. Los sustratos sobre los que actúan estas enzimas incluyen celulosa, y sustratos de celulosa modificada, tales como carboximetilcelulosa, RBB-celulosa, y similares. EG1, EG2, EG3, EG4, EG5, EG6 y EG7 caen dentro de esta clase de enzimas.  
45

“Hemicelulasa” o “xilanasas” incluye enzimas tanto exohidrolíticas como endohidrolíticas que son capaces de reconocer e hidrolizar hemicelulosa, o productos que resultan de la ruptura de hemicelulosa, como sustratos. En monocotiledóneas, en las que los heteroxilanos son el constituyente principal de hemicelulosa, se puede usar una combinación de endo-1,4-β-xilanasas (EC 3.2.1.8) y β-D-xilosidasas (EC 3.2.1.37) para romper hemicelulosa en xilosa. Enzimas desramificadoras adicionales son capaces de hidrolizar otros componentes de azúcar (arabinosa, galactosa, manosa) que están situados en puntos de ramificación en la estructura de hemicelulosa. Las enzimas adicionales son capaces de hidrolizar enlaces formados entre azúcares hemicelulósicos (principalmente arabinosa) y lignina. “Endoxilanasas” o “1,4-β-endoxilanasas” o “1,4-β-D-xilano xilanohidrolasa” o (EC 3.2.1.8) incluye enzimas que hidrolizan polímeros de xilosa unidos mediante enlaces β-1,4. Las endoxilanasas se pueden usar para hidrolizar el componente hemicelulósico de la lignocelulosa, así como sustratos de xilano purificados. “Exoxilanasas” o “β-xilosidasas” o “xilano 1,4-β-xilosidasas” o “1,4-β-D-xilano xilohidrolasa” o “xilobiasas” o “exo-1,4-β-xilosidasas (EC 3.2.1.37) incluye enzimas que hidrolizan restos sucesivos de D-xilosa a partir del término no reductor de polímeros  
50  
55

de xilano. "Arabinoxilanasas" o "glucuronarabinoxilano endo-1,4- $\beta$ -xilanasas" o "feraxano endoxilanasas" incluye enzimas que hidrolizan enlaces  $\beta$ -1,4 xilosílicos en algunos sustratos de xilano.

5 "Hemicelulasas", usada en el método de la invención, también incluye enzimas que hidrolizan los enlaces de éster (esterasa, EC 3.1.1) entre unidades de xilosa del polímero de xilano y grupos acetilo (acetil xilano esterasa, EC 3.1.1.6 y EC 3.1.1.72) o entre grupos arabinosilo y restos fenólicos tales como ácido ferúlico (feruloil esterasa, EC 3.1.1.73) y ácido p-cumárico (cumaroil esterasa, EC 3.1.1-).

10 Hemicelulasa también incluye  $\alpha$ -L-arabinofuranosidasa (EC 3.2.1.55) que elimina el sustituyente de arabinosa de la cadena principal de xilano. En consecuencia, se pueden usar en la invención " $\alpha$ -N-arabinofuranosidasa", "arabinosidasa", " $\alpha$ -arabinosidasa", " $\alpha$ -L-arabinosidasa", " $\alpha$ -arabinofuranosidasa", "polisacárido  $\alpha$ -L-arabinofuranosidasa", " $\alpha$ -L-arabinofuranósido hidrolasa", "L-arabinosidasa", " $\alpha$ -L-arabinanasa" o " $\alpha$ -L-arabinofuranósido arabinofuranohidrolasa".

15 "Hemicelulasas" incluye enzimas que hidrolizan enlaces 1,4- $\beta$ -D-manosídicos en mananos, galactomananos y/o glucomananos. En consecuencia, en la invención se pueden usar "manano endo-1,4- $\beta$ -manosidasa", "endo-1,4- $\beta$ -mananasa", "endo- $\beta$ -1,4-manasa", " $\beta$ -mananasa B", " $\beta$ -1, 4-manano 4-mananohidrolasa", "endo- $\beta$ -mananasa", " $\beta$ -D-mananasa" o "1,4- $\beta$ -D-manano mananohidrolasa" (EC 3.1.1.78).

20 "Pectinasa", usada en el método de la invención, incluye cualquier sustancia que puede hidrolizar una pectina o sustancia péctica. Una composición puede comprender cualquier pectinasa, por ejemplo una endo poligalacturonasa, una pectina metil esterasa, una endo-galactanasa, una beta galactosidasa, una pectina acetil esterasa, una endo-pectina liasa, pectato liasa, alfa ramnosidasa, una exo-galacturonasa, una exopoligalacturonato liasa, una ramnogalacturonano hidrolasa, una ramnogalacturonano liasa, una ramnogalacturonano acetil esterasa, una ramnogalacturonano galacturonohidrolasa, una xilogalacturonasa.

25 Aquí, una endo-poligalacturonasa (EC 3.2.1.15) es cualquier polipéptido que sea capaz de catalizar la hidrólisis al azar de enlaces 1,4-  $\alpha$ -D-galactosidurónicos en pectato y otros galacturonanos. Esta enzima también se puede denominar como poligalacturonasa pectina despolimerasa, pectinasa, endopoligalacturonasa, pectolasa, pectina hidrolasa, pectina poligalacturonasa, poli-  $\alpha$ -1,4-galacturónido glucanohidrolasa, endogalacturonasa; endo-D-galacturonasa o poli(1,4-  $\alpha$ -D-galacturónido) glucanohidrolasa.

Aquí, una pectina metil esterasa (EC 3.1.1.11) es cualquier enzima que sea capaz de catalizar la reacción: pectina + n H<sub>2</sub>O = n metanol + pectato. La enzima también puede ser conocida como pectinaesterasa, pectina desmetoxilasa, pectina metoxilasa, pectina metilesterasa, pectasa, pectinoesterasa o pectina pectilhidrolasa.

30 Aquí, una endo-galactanasa (EC 3.2.1.89, EC 3.2.1.90) es cualquier enzima capaz de catalizar la endohidrólisis de enlaces 1,4- $\beta$ -D-galactosídicos en arabinogalactanos. La enzima también puede ser conocida como arabinogalactano endo-1,4- $\beta$ -galactosidasa, endo-1,4- $\beta$ -galactanasa, galactanasa, arabinogalactanasa o arabinogalactano 4- $\beta$ -D-galactanohidrolasa.

35 Aquí, una pectina acetil esterasa se define en este documento como cualquier enzima que tiene una actividad de acetil esterasa que cataliza la desacetilación de los grupos acetilo en los grupos hidroxilo de restos GalUA de pectina.

40 Aquí, una endo-pectina liasa (EC 4.2.2.10) es cualquier enzima capaz de catalizar la escisión eliminativa de éster metílico de (1 $\rightarrow$ 4)-  $\alpha$ -D-galacturonano para dar oligosacáridos con grupos 4-desoxi-6-O-metil-  $\alpha$ -D-galact-4-enuronosilo en sus extremos no reductores. La enzima también puede ser conocida como pectina liasa, pectina *trans*-eliminasa; endo-pectina liasa, polimetilgalacturónico transeliminasa, pectina metiltranseliminasa, pectoliasa, PL, PNL o PMGL o (1 $\rightarrow$ 4)-6-O-metil-  $\alpha$ -D-galacturonano liasa.

45 Aquí, una pectato liasa (EC 4.2.2.2) es cualquier enzima capaz de catalizar la escisión eliminativa de (1 $\rightarrow$ 4)-  $\alpha$ -D-galacturonano para dar oligosacáridos con grupos 4-desoxi-  $\alpha$ -D-galact-4-enuronosilo en sus extremos no reductores. La enzima también puede ser conocida como poligalacturónico transeliminasa, ácido péctico transeliminasa, poligalacturonato liasa, endopectina metiltranseliminasa, pectato transeliminasa, endogalacturonato transeliminasa, ácido péctico liasa, liasa péctica, ácido  $\alpha$ -1,4-D-endopoligalacturónico liasa, PGA liasa, PPasa-N, ácido endo-  $\alpha$ -1,4-poligalacturónico liasa, ácido poligalacturónico liasa, pectina *trans*-eliminasa, ácido poligalacturónico *trans*-eliminasa o (1 $\rightarrow$ 4)-  $\alpha$ -D-galacturonano liasa.

50 Aquí, una alfa ramnosidasa (EC 3.2.1.40) es cualquier polipéptido que sea capaz de catalizar la hidrólisis de restos de  $\alpha$ -L-ramnosa no reductores terminales en  $\alpha$ -L-ramnósidos, o como alternativa en ramnogalacturonano. Esta enzima también puede ser conocida como  $\alpha$ -L-ramnosidasa T,  $\alpha$ -L-ramnosidasa N o  $\alpha$ -L-ramnósido ramnoidrolasa.

Aquí, exo-galacturonasa (EC 3.2.1.82) es cualquier polipéptido capaz de hidrolizar ácido péctico del extremo no reductor, liberando digalacturonato. Esta enzima también puede ser conocida como exo-poli- $\alpha$ -galacturonosidasa, exopoligalacturonosidasa o exopoligalacturanosidasa.

5 Aquí, exo-galacturonasa (EC 3.2.1.67) es cualquier polipéptido capaz de catalizar:  $(1,4\text{-}\alpha\text{-D-galacturónido})_n + \text{H}_2\text{O} = (1,4\text{-}\alpha\text{-D-galacturónido})_{n-1} + \text{D-galacturonato}$ . La enzima también puede ser conocida como galacturano 1,4- $\alpha$ -galacturonidasa, exopoligalacturonasa, poli(galacturonato) hidrolasa, exo-D-galacturonasa, exo-D-galacturonanasa, exopoli-D-galacturonasa o poli(1,4- $\alpha$ -D-galacturónido) galacturonohidrolasa.

10 Aquí, exopoligalacturonato liasa (EC 4.2.2.9) es cualquier polipéptido capaz de catalizar la escisión eliminativa de 4-(4-desoxi- $\alpha$ -D-galact-4-enuronosili)-D-galacturonato a partir del extremo reductor de pectato, es decir, pectina desesterificada. Esta enzima también puede ser conocida como pectato disacárido-liasa, pectato exo-liasa, ácido exopéctico transeliminasa, exopectato liasa, ácido exopoligalacturónico *trans*-eliminasa, PATE, exo-PATE, exo-PGL o (1 $\rightarrow$ 4)- $\alpha$ -D-galacturonano disacárido de extremo reductor liasa.

15 Aquí, ramnogalacturonano hidrolasa es cualquier polipéptido que es capaz de hidrolizar el enlace entre ácido galactosilurónico y ramnopiranosilo de una manera endo en estructuras de ramnogalacturonano que se alternan estrictamente, que consisten en el disacárido [ácido (1,2- $\alpha$ -L-ramnoil-(1,4)- $\alpha$ -galactosilurónico)].

Aquí, ramnogalacturonano liasa es cualquier polipéptido que es capaz de escindir enlaces  $\alpha$ -L-Rap-(1 $\rightarrow$ 4)- $\alpha$ -D-GalpA de una manera endo en ramnogalacturonano mediante eliminación beta.

Aquí, ramnogalacturonano acetil esterasa es cualquier polipéptido que cataliza la desacetilación de la cadena principal de restos de ramnosa y de ácido galacturónico alternos en ramnogalacturonano.

20 Aquí, ramnogalacturonano galacturonohidrolasa es cualquier polipéptido que es capaz de hidrolizar ácido galacturónico a partir del extremo no reductor de estructuras de ramnogalacturonano estrictamente alternas de una manera exo.

25 Aquí, xilogalacturonasa es cualquier polipéptido que actúa sobre xilogalacturonano escindiendo de manera *endo* la cadena principal de ácido galacturónico sustituido con  $\beta$ -xilosa. Esta enzima también puede ser conocida como xilogalacturonano hidrolasa.

Aquí, una  $\alpha$ -L-arabinofuranosidasa (EC 3.2.1.55) es cualquier polipéptido que es capaz de actuar sobre  $\alpha$ -L-arabinofuranósidos,  $\alpha$ -L-arabinanos que contienen enlaces (1,2) y/o (1,3) y/o (1,5), arabinoxilanos y arabinogalactanos. Esta enzima también se puede denominar como  $\alpha$ -N-arabinofuranosidasa, arabinofuranosidasa o arabinosidasa.

30 Aquí, endo-arabinanasa (EC 3.2.1.99) es cualquier polipéptido que es capaz de catalizar endohidrólisis de enlaces 1,5- $\alpha$ -arabinofuranosídicos en 1,5-arabinanos. La enzima también puede ser conocida como endo-arabinasa, arabinano endo-1,5- $\alpha$ -L-arabinosidasa, endo-1,5- $\alpha$ -L-arabinanasa, endo- $\alpha$ -1,5-arabanasa; endo-arabanasa o 1,5- $\alpha$ -L-arabinano 1,5- $\alpha$ -L-arabinanohidrolasa.

35 "Amilasa" o "alfa glucosidasa" incluye enzimas que hidrolizan enlaces 1,4-[alfa]-glucosídicos en oligosacáridos y polisacáridos.

40 "Proteasa" incluye enzimas que hidrolizan enlaces peptídicos (peptidasas), así como enzimas que hidrolizan enlaces entre péptidos y otros restos, tales como azúcares (glicopeptidasas). Muchas proteasas se caracterizan bajo EC 3.4, y son adecuadas para uso en la invención. Algunos tipos específicos de proteasas incluyen cisteína proteasas, incluyendo pepsina, papaína y serina proteasas, incluyendo quimiotripsinas, carboxipeptidasas y metaloendopeptidasas.

"Lipasa" incluye enzimas que hidrolizan lípidos, ácidos grasos, y acilglicéridos, incluyendo fosfoglicéridos, lipoproteínas, diacilglicérols, y similares. En plantas, los lípidos se usan como componentes estructurales para limitar la pérdida de agua y la infección por patógenos. Estos lípidos incluyen ceras derivadas de ácidos grasos, así como cutina y suberina.

45 "Glucuronidasa" incluye enzimas que catalizan la hidrólisis de  $\beta$ -glucuronósido para producir un alcohol y glucuronato. Se han caracterizado muchas glucuronidasas, y pueden ser adecuadas para uso en la invención, por ejemplo  $\beta$ -glucuronidasa (EC 3.2.1.31), hialurono-glucuronidasa (EC 3.2.1.36), glucuronosil-disulfoglucosamina glucuronidasa (3.2.1.56), glicirrizinato  $\beta$ -glucuronidasa (3.2.1.128) o  $\alpha$ -D-glucuronidasa (EC 3.2.1.139).

50 "Oxidorreductasa" es cualquier enzima que cataliza la transferencia de electrones desde una molécula (el agente reductor, también denominado el dador de hidrógenos o de electrones) a otra (el agente oxidante, también denominado el aceptor de hidrógenos o de electrones), por ejemplo una enzima que cae dentro de EC 1. Un ejemplo de tal enzima que se puede usar en un método de la invención es glucosa oxidasa (EC 1.1.3.4), es decir una enzima que cataliza la oxidación de beta-D-glucosa en D-glucono-1,5-lactona.

En tal método descrito anteriormente, se puede usar una combinación de enzimas, ya sea simultánea, separada o secuencialmente, para actuar sobre, por ejemplo, un sustrato lignocelulósico o biomasa vegetal, que sirve como la materia prima, para convertir este sustrato complejo en azúcares simples y oligosacáridos para la producción de etanol u otro producto útil.

- 5 En consecuencia, otro aspecto de la invención incluye métodos que utilizan mezclas de las enzimas como se describen anteriormente, ya sea simultánea, separada o secuencialmente, opcionalmente junto con otras enzimas o tratamientos físicos tales como temperatura y pH para convertir la biomasa vegetal lignocelulósica en azúcares y oligosacáridos.

10 Las combinaciones enzimáticas o tratamientos físicos se pueden administrar concomitantemente, o secuencial o separadamente. Las enzimas se pueden producir exógenamente en microorganismos, levaduras, hongos, bacterias o plantas, después se pueden aislar y se añaden a la materia prima lignocelulósica. Como alternativa, las enzimas se producen, pero no se aíslan, y el caldo de fermentación de la masa celular bruta, o el material vegetal (tal como rastrojo de maíz), y similares, se añaden a la materia prima. Como alternativa, la masa celular bruta o el medio de producción enzimático o el material vegetal se puede tratar para evitar el crecimiento microbiano posterior (por ejemplo, calentando o añadiendo agentes antimicrobianos), y después se puede añadir a la materia prima. Estas mezclas enzimáticas brutas pueden incluir el organismo productor de la enzima. Como alternativa, la enzima se puede producir en una fermentación que usa materia prima (tal como rastrojo de maíz) para proporcionar nutrición a un organismo que produce una enzima o enzimas. De esta manera, las plantas que producen las enzimas pueden servir como la materia prima lignocelulósica y se pueden añadir en la materia prima lignocelulósica.

20 Aunque se han discutido las enzimas auxiliares como mezcla, se reconoce que las enzimas se pueden añadir secuencialmente, en el que la temperatura, el pH y otras condiciones se pueden alterar para incrementar la actividad de cada enzima individual. Como alternativa, se puede determinar para la mezcla enzimática un pH y temperatura óptimos.

25 Un polipéptido se puede usar antes de, al mismo tiempo que, o después de cualquier etapa o etapas de pretratamiento. Como alternativa, o además, un polipéptido se puede usar en combinación con una enzima que degrada un hidrato de carbono que no es almidón (tal como una celulasa). En consecuencia, un polipéptido como se describe aquí se puede usar como parte de un protocolo de pretratamiento, reduciendo así el aporte de compuestos químicos y/o de energía en comparación con un protocolo de pretratamiento convencional. Un polipéptido como se describe aquí se puede usar como parte de un protocolo de sacarificación. El uso del polipéptido puede hacer que se use menos enzima que degrada un hidrato de carbono que no es almidón, o puede permitir que se lleve a cabo un procedimiento que requiera menos tiempo.

30 Se puede llevar a cabo un método de la invención, es decir, se puede hacer reaccionar con un sustrato que comprende un hidrato de carbono que no es almidón, en condiciones suaves que no incluyen tratamiento térmico o ácido extremo, como se utiliza actualmente para la conversión de biomasa usando biorreactores. Un polipéptido con actividad de peroxidasa como se describe aquí se puede usar simultánea, separada o secuencialmente con uno o más de los otros polipéptidos o enzimas descritos aquí.

35 Por ejemplo, las enzimas se pueden incubar a alrededor de 25°C, alrededor de 30°C, alrededor de 35°C, alrededor de 37°C, alrededor de 40°C, alrededor de 45°C, alrededor de 50°C, o alrededor de 55°C. Es decir, se pueden incubar de alrededor de 20°C a alrededor de 70°C, en tampones de fuerza iónica baja a media, y pH neutro. Por "fuerza iónica media" se pretende que el tampón tenga una concentración iónica de alrededor de 200 milimolar (mM) o menos para cualquier componente iónico individual. El pH puede oscilar de alrededor de pH 2,5, alrededor de pH 3,0, alrededor de pH 3,5, alrededor de pH 4,0, alrededor de pH 4,5, alrededor de pH 5, alrededor de pH 5,5, alrededor de pH 6, alrededor de pH 6,5, alrededor de pH 7, alrededor de pH 7,5, alrededor de pH 8,0, a alrededor de pH 8,5. Generalmente, el intervalo de pH será de alrededor de pH 3,0 a alrededor de pH 9.

40 La incubación de combinaciones enzimáticas descritas aquí, es decir, un polipéptido con actividad de peroxidasa y una enzima que degrada un hidrato de carbono que no es almidón, por ejemplo en las condiciones expuestas anteriormente, puede dar como resultado la liberación de azúcar del sustrato, por ejemplo lignocelulosa. La cantidad de azúcar liberada puede ser al menos alrededor de 20%, al menos alrededor de 30%, al menos alrededor de 40%, al menos alrededor de 50%, al menos alrededor de 60%, al menos alrededor de 70%, al menos alrededor de 80%, al menos alrededor de 90%, al menos alrededor de 95% o más de azúcar disponible.

El tratamiento con una enzima que degrada un hidrato de carbono que no es almidón se puede producir desde varios minutos hasta varias horas, tal como de alrededor de 6 horas a alrededor de 120 horas, preferiblemente alrededor de 12 horas a alrededor de 72 horas, más preferiblemente alrededor de 24 a 48 horas.

55 Se puede utilizar una etapa de pretratamiento que implica la incubación con una enzima o una mezcla de enzimas. La etapa de pretratamiento se puede llevar a cabo a muchas temperaturas diferentes, pero se prefiere que el pretratamiento se produzca a la temperatura más adecuada a la mezcla enzimática que se ensaya, o al óptimo enzimático predicho de las enzimas a ensayar. La temperatura del pretratamiento puede oscilar de alrededor de 10°C a alrededor de 80°C, alrededor de 20°C a alrededor de 80°C, alrededor de 30°C a alrededor de 70°C, alrededor

de 40°C a alrededor de 60°C, alrededor de 37°C a alrededor de 50°C, preferiblemente alrededor de 37°C a alrededor de 80°C, más preferiblemente alrededor de 50°C. En ausencia de datos en el óptimo de temperatura, es preferible llevar a cabo las reacciones de pretratamiento a 37°C en primer lugar, después a una temperatura mayor, tal como 50°C. El pH de la mezcla de pretratamiento puede oscilar de alrededor de 2,0 a alrededor de 10,0, pero es preferiblemente alrededor de 3,0 a alrededor de 7,0, más preferiblemente alrededor de 4,0 a alrededor de 6,0, incluso más preferiblemente alrededor de 4,5 a alrededor de 5. Nuevamente, el pH se puede ajustar para maximizar la actividad enzimática, y se puede ajustar con la adición de la enzima. La comparación de los resultados del ensayo a partir de este ensayo permitirá modificar el método para adecuarse mejor a las enzimas que se ensayan.

La reacción de pretratamiento puede producirse desde varios minutos hasta varias horas, tal como de alrededor de 6 horas a alrededor de 120 horas, preferiblemente alrededor de 6 horas a alrededor de 48 horas, más preferiblemente alrededor de 6 a alrededor de 24 horas, lo más preferible durante alrededor de 6 horas.

Si se usa un polipéptido con actividad de peroxidasa en el pretratamiento, esto puede permitir que se usen condiciones suaves descritas anteriormente y evitar el uso de tratamiento térmico o ácido extremo, como se utiliza actualmente para la conversión de biomasa usando biorreactores.

Un método de la invención para producir un azúcar o azúcares es típicamente un procedimiento para convertir un hidrato de carbono complejo, tal como lignocelulosa, en azúcares, preferiblemente azúcares fermentables. Tal procedimiento se puede denominar como "sacarificación". En consecuencia, un método de la invención puede dar como resultado la liberación de uno o más azúcares de hexosa y/o de pentosa, tales como uno o más de glucosa, xilosa, arabinosa, galactosa, ácido D-galacturónico, manosa, ramnosa, sacarosa y fructosa.

Los azúcares fermentables se pueden convertir en productos de fermentación de valor añadido útiles, cuyos ejemplos no limitantes incluyen aminoácidos, vitaminas, productos farmacéuticos, suplementos para piensos de animales, productos químicos de especialidad, materias primas químicas, plásticos, disolventes, combustibles, u otros polímeros orgánicos, ácido láctico, y etanol, incluyendo etanol de combustible.

Los productos de valor añadido específicos que se pueden producir mediante los métodos de la invención incluyen, pero no se limitan a, biocombustibles (incluyendo etanol y butanol y un biogas); ácido láctico; un plástico; un producto químico de especialidad; un ácido orgánico, incluyendo ácido cítrico, ácido succínico, ácido fumárico, ácido itacónico y ácido maleico; ácido 3-hidroxi-propiónico, ácido acrílico; ácido acético; 1,3-propano-diol; etileno, glicerol; un disolvente; un suplemento para piensos de animales; un producto farmacéutico, tal como un antibiótico  $\beta$ -lactámico o una cefalosporina; vitaminas; un aminoácido, tal como lisina, metionina, triptófano, treonina, y ácido aspártico; una enzima industrial, tal como una proteasa, una celulasa, una amilasa, una glucanasa, una lactasa, una lipasa, una liasa, oxidorreductasas, una transferasa o una xilanasas; y una materia prima química.

En consecuencia, la descripción proporciona un método para la preparación de un producto de fermentación, método el cual comprende:

- a. degradar un hidrato de carbono que no es almidón, por ejemplo producir un azúcar o azúcares a partir de un hidrato de carbono que no es almidón, usando un método como se describe aquí; y
  - b. fermentar el material resultante,
- para preparar de ese modo un producto de fermentación.

El producto de fermentación puede ser un aminoácido, una vitamina, un producto farmacéutico, un suplemento para piensos de animales, un producto químico de especialidad, una materia prima química, un plástico, etanol, incluyendo etanol de combustible, entendiéndose el término "etanol" para que incluya alcohol etílico o mezclas de alcohol etílico y agua). Los productos de fermentación más específicos incluyen ácido láctico, ácido 3-hidroxi-propiónico, ácido acrílico, ácido acético, ácido succínico, ácido cítrico, un aminoácido, 1,3-propano-diol, etileno, glicerol, un antibiótico  $\beta$ -lactámico o una cefalosporina.

Un método para la preparación de un producto de fermentación puede comprender opcionalmente la recuperación del producto de fermentación.

Tal procedimiento se puede llevar a cabo en condiciones aerobias o anaerobias. Preferiblemente, el procedimiento se lleva a cabo en condiciones microaerófilas o de oxígeno limitado.

Un proceso de fermentación anaerobia se define aquí como un proceso de fermentación realizado en ausencia de oxígeno o en el que no se consume sustancialmente oxígeno, preferiblemente menos de 5, 2,5 ó 1 mmol/l/h, y en el que moléculas orgánicas sirven tanto como dadores de electrones como aceptores de electrones.

Un proceso de fermentación con oxígeno limitado es un proceso en el que el consumo de oxígeno está limitado por la transferencia de oxígeno desde el gas al líquido. El grado de limitación de oxígeno está determinado por la cantidad y composición del flujo de gas entrante así como de las propiedades de mezclado reales/transferencia másica del equipo de fermentación usado. Preferiblemente, en un proceso en condiciones con oxígeno limitado, la

velocidad de consumo de oxígeno es al menos 5,5, más preferiblemente al menos 6 e incluso más preferiblemente al menos 7 mmoles/l/h.

La descripción también proporciona el uso de un polipéptido que tiene actividad de peroxidasa como se describe aquí para uso en un método para degradar lignina. En tal uso, la lignina puede estar en forma de lignocelulosa.

5 Los siguientes Ejemplos ilustran la invención:

### Ejemplo 1

Degradación de compuesto modelo de lignina por peroxidasas obtenidas a partir de *Marasmius scorodoni*

10 2-Metoxifenol (guayacol), alcohol veratrílico, ácido ferúlico y ácido cumárico sirvieron como compuestos modelo de lignina de bajo peso molecular. Se incubaron 2 mM de cada sustrato con 1,5  $\mu$ l de disolución enzimática y 15  $\mu$ l de H<sub>2</sub>O<sub>2</sub> (20 mM) en tampón de acetato de sodio (50 mM, 1,5 ml de volumen de muestra total, pH 6,0, 27°C). Después de 60 min., se añadieron otros 15  $\mu$ l de H<sub>2</sub>O<sub>2</sub>, y la reacción se detuvo después de 2 h. Se observó un color parduzco con guayacol (véase la Figura 1) y ácido ferúlico. No se produjeron cambios visualmente observables con las muestras de ácido cumárico y de alcohol veratrílico.

15 Para evaluar el efecto de MsP1 sobre una lignina de peso molecular elevado insoluble, se usó lignina organosolv<sup>TM</sup> (polvo marrón, Sigma-Aldrich). Se mezclaron 2,5 mg de lignina organosolv con 20  $\mu$ l de disolución de MsP1 (ya sea activa o inactivada térmicamente) en tampón de acetato de sodio (50 mM, pH 3-6, volumen de muestra total 1,5 ml) y 100  $\mu$ l de disolución de peróxido de hidrógeno (20 mM). Después de 60 min., se añadieron otros 100  $\mu$ l de disolución de peróxido de hidrógeno. El tiempo de reacción total fue 120 min. (incubación a 27°C en un agitador giratorio). El sólido se eliminó mediante centrifugación, y el sobrenadante se sometió a análisis de HPLC. Las  
20 columnas usadas procedieron en Polymer laboratories<sup>TM</sup> PL aquagel-OH MIXED y PL aquagel-OH 20, la detección de UV se llevó a cabo a 195 nm. Puesto que el material de lignina es insoluble en agua, el sobrenadante de las muestras con enzimas inactivadas térmicamente permaneció incoloro. Cuando se usó la enzima activa, se observó un color rojo a parduzco en los sobrenadantes. Los compuestos de peso molecular elevado liberados por el tratamiento enzimático fueron rastreables por medio de cromatografía de exclusión de tamaños (véase la Figura 2).

25 Para producir H<sub>2</sub>O<sub>2</sub> continuamente, se aplicó un sistema de glucosa/glucosa oxidasa. Se incubaron 500 mg de lignina (organosolv), 4 ml de MsP1, 20 mmoles de glucosa, y 10  $\mu$ l (10 U) de glucosa oxidasa en 216 ml de tampón de acetato de sodio (50 mM, pH 6,0) durante 24 h y 140 rpm a temperatura ambiente. Los sobrenadantes se separaron por medio de centrifugación y se analizaron mediante HPLC (detección mediante DAD y dispersión de la luz evaporativa (ELSD)). Se detectó una solubilización significativa del material previamente insoluble (véase la  
30 Figura 3).

Después de llevar a cabo el ensayo, los sólidos se centrifugaron y se secaron. Cuando se usó peroxidasa inactivada por calor, se observó una pérdida de peso seco del pelete de 2,5%. Con peroxidasa activa, la pérdida fue 7,3%. De este modo, la lignina se disuelve parcialmente como resultado de la actividad enzimática.

### Ejemplo 2 (no según la invención)

35 Análisis químico de rastrojo de maíz

La composición química de rastrojo de maíz se caracterizó mediante determinación del nitrógeno de Kjeldahl (proteína), extracción de Soxhlet (lípidos), ensayo de orcinol-ácido sulfúrico (hidratos de carbono), y mediante el método de bromuro de acetilo (lignina) (Tabla 1).

Tabla 1: Composición de rastrojo de maíz

Parámetro	Rastrojo de maíz
Masa seca [%]	98,7
Proteína [% en DM]	12,7
Hidratos de carbono [% en DM]	67,2
Lípidos [% en DM]	0,5
Lignina [% en DM]	20,4
$\Sigma$	100,8

40

**Ejemplo 3 (no según la invención)**Secretoma de *Marasmius scorodoni*

5 Para caracterizar las enzimas extracelulares implicadas en el proceso natural de la degradación de lignina, se hizo crecer *M. scorodoni* en cultivos sumergidos usando rastrojo de maíz como fuente de carbono y de nitrógeno. Los precultivos se hicieron crecer en 100 ml de disolución estándar (SNS), y se usaron cuatro medios diferentes para el cultivo principal:

- Medio A:
 

glucosa	(1,00%)
MgSO <sub>4</sub>	(0,05%)
KH <sub>2</sub> PO <sub>4</sub>	(0,15%)
extracto de levadura	(0,20%)
disolución de oligoelementos	1,0 ml l <sup>-1</sup>
- Medio B: como el medio A, pero sin glucosa
- Medio C: como el medio A, pero sin extracto de levadura
- Medio D: como el medio A, pero con 0,4% de Tween 80 en lugar de glucosa

Los cultivos principales se inocularon con 20 ml de los precultivos. Se extrajeron muestras de los sobrenadantes de los cultivos (dependiendo de la velocidad de crecimiento del micelio fúngico), y el secretoma se analizó por medio de ensayos enzimáticos y mediante técnicas electroforéticas.

## 10 Lacasas y peroxidasas

Las actividades de lacasa y peroxidasa se cuantificaron mediante el ensayo ABTS, con y sin adición de H<sub>2</sub>O<sub>2</sub> respectivamente (véase la Figura 4).

Las mayores actividades de peroxidasa se observaron en los cultivos que se hicieron crecer sin glucosa (medio B y D).

15 Los geles de IEF teñidos para la actividad indicaron la presencia de enzimas de tipo lacasa (pl 4,5) y peroxidasa (pl 3,7) (véase la Figura 5). Puesto que no se observaron bandas de actividad en los cultivos que se hicieron crecer en SNS, estas enzimas son inducibles por el material renovable "rastrojo de maíz".

En comparación con los cultivos que se hicieron crecer en SNS suplementada con lignina (0,1%), se detectaron actividades de peroxidasa elevadas en los cultivos que contienen rastrojo de maíz en medio D (véase la Figura 6).

## 20 Peptidasas

Sólo se detectaron actividades de peptidasa muy bajas en los cultivos suplementados con rastrojo de maíz.

 $\beta$ -glucosidasas

25 4-Nitrofenol- $\beta$ -D-glucósido sirvió como sustrato para cuantificar la actividad de  $\beta$ -glucosidasa en los cultivos sumergidos de *M. scorodoni*. Tras la hidrólisis, la concentración de 4-nitrofenol se cuantificó por medio de espectrofotometría UV-Vis a  $\lambda = 400$  nm (véase la Figura 7).

En el día 16 de cultivo, las actividades todavía crecían en los medios B, C y D.

## Esterasas

La actividad esterolítica se cuantificó usando como un sustrato butirato de nitrofenol (véase la Figura 8).

30 La actividad esterolítica se visualizó directamente en un gel de IEF al teñir la actividad usando fast Blue SB y acetato de 1-naftilo (véase la Figura 9).

**Ejemplo 4**

Liberación de azúcares fermentables a partir de rastrojo de maíz

El sustrato de rastrojo de maíz usado en los experimentos de hidrógenos se pretrató según el siguiente protocolo:

- molienda ( $\varnothing < 1$  mm);

- empapamiento (5 g de rastrojo de maíz + 117 ml de agua, 1 h);
- calentamiento (microondas, 700W, 3 mm);
- lavado (800 ml de agua)
- secado

5 Subsiguientemente, se dispersaron 50 mg del rastrojo de maíz pretratado en 1,5 ml de tampón de acetato de sodio (50 mM, pH 3,5), y se añadieron 6 U de MsP1, y 7,5 U de celulasa (*Trichoderma reesei*, Sigma C8546). La actividad de celulasa se determinó usando una disolución de 2-hidroxietilcelulosa al 1% en tampón de acetato de sodio 0,5 M.

10 En una mezcla de rastrojo de maíz se añadió periódicamente H<sub>2</sub>O<sub>2</sub> (cantidad total de 3,2 μmoles), y en otra mezcla no se añadió H<sub>2</sub>O<sub>2</sub>. Ambas muestras se incubaron a 30°C y se tomaron muestras en tiempo para el análisis del contenido de monómero de glucosa. El contenido de monómero de glucosa se determinó usando el ensayo de azúcar reductor de Nelson Somogyi.

15 En la figura 10, se presenta la liberación de monómero de glucosa como porcentaje de la materia seca de rastrojo de maíz, tanto para MsP1 activa (que incluye adición de H<sub>2</sub>O<sub>2</sub>) como MsP1 inactiva (sin adición de H<sub>2</sub>O<sub>2</sub>). Se muestra claramente que MsP1 activa en la mezcla de incubación da como resultado una liberación adicional de glucosa a partir del rastrojo de maíz por la celulasa de *Trichoderma*.

#### LISTADO DE SECUENCIAS

<110> DSM IP ASSETS B.V.

<120> Método para modificar material de hidrato de carbono que no es almidón

<130> 26485-WO-PCT

20 <160>8

<170> PatentIn version 3.2

<210> 1

<211> 1542

<212> ADN

25 <213> *Marasmius scorodonius*

<220>

<221 > CDS

<222> (1)..(1542)

<400> 1

ES 2 426 954 T3

atg aag ctt ttt tct gcc tct gtt ttt gct gct atc gtc gct agt cac	48
Met Lys Leu Phe Ser Ala Ser Val Phe Ala Ala Ile Val Ala Ser His	
1 5 10 15	
tat gcg tca gcg act gcc cac atc agg gct ccc aat gtg aag cca agg	96
Tyr Ala Ser Ala Thr Ala His Ile Arg Ala Pro Asn Val Lys Pro Arg	
20 25 30	
agg aca agc tca ctt ctg att act ccc cct caa cag cct ccg ctt cca	144
Arg Thr Ser Ser Leu Leu Ile Thr Pro Pro Gln Gln Pro Pro Leu Pro	
35 40 45	
tct gct caa cag gct gca agt gcc tct agc agt gct ggc ttg aat ctc	192
Ser Ala Gln Gln Ala Ala Ser Ala Ser Ser Ser Ala Gly Leu Asn Leu	
50 55 60	
acc gac att cag ggt gat att ctg atc ggc atg aag aag aac aag gaa	240
Thr Asp Ile Gln Gly Asp Ile Leu Ile Gly Met Lys Lys Asn Lys Glu	
65 70 75 80	
cta ttc ttc ttc ttc agt gtc acc gac gca gct act ttc aag gct aag	288
Leu Phe Phe Phe Phe Ser Val Thr Asp Ala Ala Thr Phe Lys Ala Lys	
85 90 95	
ctg gga tcc gac att ctt gga cta atc aca tcc acc gat cag cta ctt	336
Leu Gly Ser Asp Ile Leu Gly Leu Ile Thr Ser Thr Asp Gln Leu Leu	
100 105 110	
gct aat gat act cag cct gtc acg gct gtc aac gtc gct ttc tct agc	384
Ala Asn Asp Thr Gln Pro Val Thr Ala Val Asn Val Ala Phe Ser Ser	
115 120 125	
act ggc ctc aag gca ttg ggt atc aca gat ggt ctg aag gat cct gtc	432
Thr Gly Leu Lys Ala Leu Gly Ile Thr Asp Gly Leu Lys Asp Pro Val	

ES 2 426 954 T3

130	135	140	
ttc gag gcc gga atg ctc agc aac gca gtg agc gac ttg agc gat cca			480
Phe Glu Ala Gly Met Leu Ser Asn Ala Val Ser Asp Leu Ser Asp Pro			
145	150	155	160
ggg acc ggc aat tgg gtc cct ggg ttt gtc ggc acc agt gtt cat ggc			528
Gly Thr Gly Asn Trp Val Pro Gly Phe Val Gly Thr Ser Val His Gly			
	165	170	175
gtt ttc cta ctt gca tcg gac acc att gac aat gta aac acc gag ccg			576
Val Phe Leu Leu Ala Ser Asp Thr Ile Asp Asn Val Asn Thr Glu Pro			
	180	185	190
gcc aac atc caa acc att ttg aat ggc tcg atc acg gag att cat cgt			624
Ala Asn Ile Gln Thr Ile Leu Asn Gly Ser Ile Thr Glu Ile His Arg			
	195	200	205
tta caa ggg gag gct cga ccc ggt gac cag caa ggt cac gaa cac ttt			672
Leu Gln Gly Glu Ala Arg Pro Gly Asp Gln Gln Gly His Glu His Phe			
	210	215	220
gga ttc atg gat gga atc agt aac ccg gcc gtt gat gga ttt aca cct			720
Gly Phe Met Asp Gly Ile Ser Asn Pro Ala Val Asp Gly Phe Thr Pro			
225	230	235	240
cca gcg gaa ata aga cct gga caa gct tta att ccg cct ggt atc atg			768
Pro Ala Glu Ile Arg Pro Gly Gln Ala Leu Ile Pro Pro Gly Ile Met			
	245	250	255
ctt ctc gga gag gca aac gac act ttt cag aat gat cgt cct ccg tgg			816
Leu Leu Gly Glu Ala Asn Asp Thr Phe Gln Asn Asp Arg Pro Pro Trp			
	260	265	270
gcc aaa gat ggt tcc ttc ctt gtc ttc cgt caa atg caa cag cgc gcg			864
Ala Lys Asp Gly Ser Phe Leu Val Phe Arg Gln Met Gln Gln Arg Ala			
	275	280	285
ccc gag ttc aac aag ttc ctg caa gat cac gct ctt aac atg ccg aat			912
Pro Glu Phe Asn Lys Phe Leu Gln Asp His Ala Leu Asn Met Pro Asn			
	290	295	300
atg aca tcc gag caa ggc gct gat ctc ctt ggt gcc agg att gta gga			960
Met Thr Ser Glu Gln Gly Ala Asp Leu Leu Gly Ala Arg Ile Val Gly			
305	310	315	320
cga tgg aaa agt ggt gct cct att gac ctc act ccg ttg gtc gat gac			1008
Arg Trp Lys Ser Gly Ala Pro Ile Asp Leu Thr Pro Leu Val Asp Asp			
	325	330	335
cca gtg ttg gct gct gac aat cag cga aat aac aac ttc gac ttt tct			1056
Pro Val Leu Ala Ala Asp Asn Gln Arg Asn Asn Asn Phe Asp Phe Ser			
	340	345	350
gac gcc acg aat cag aca cgt tgc cct ttc tct gct cat atc cgc aag			1104
Asp Ala Thr Asn Gln Thr Arg Cys Pro Phe Ser Ala His Ile Arg Lys			
	355	360	365

ES 2 426 954 T3

gct aac ccg cgc ggc gat ctt ggg ggt att aat aaa ttc cca aac caa 1152  
Ala Asn Pro Arg Gly Asp Leu Gly Gly Ile Asn Lys Phe Pro Asn Gln  
370 375 380

cac ata atc cga gcg gga att ccg tat gga ccc gaa gtt acc gac gct 1200  
His Ile Ile Arg Ala Gly Ile Pro Tyr Gly Pro Glu Val Thr Asp Ala  
385 390 395 400

gaa aga gcg tca aat agc tct agc act gac cct agt ctg gag cgt ggt 1248  
Glu Arg Ala Ser Asn Ser Ser Ser Thr Asp Pro Ser Leu Glu Arg Gly  
405 410 415

ctg gcg ttt gtg gcc tat cag tct aat atc cag aac gga ttc gta ttc 1296  
Leu Ala Phe Val Ala Tyr Gln Ser Asn Ile Gln Asn Gly Phe Val Phe  
420 425 430

ctt caa aag aat tgg gtt gat aat acg aat ttc ttc cga ccc ggc act 1344  
Leu Gln Lys Asn Trp Val Asp Asn Thr Asn Phe Phe Arg Pro Gly Thr  
435 440 445

ggt gca gat cct ctc atc ggt aca aat tct cgt aac agt ggc acc gat 1392  
Gly Ala Asp Pro Leu Ile Gly Thr Asn Ser Arg Asn Ser Gly Thr Asp  
450 455 460

gcc ccc aac acg cct cgt gtc gtc agc ggc ttg gat cct aat aac gct 1440  
Ala Pro Asn Thr Pro Arg Val Val Ser Gly Leu Asp Pro Asn Asn Ala  
465 470 475 480

acg agc acc atc gaa att gat atc gat ttc gta gtt tct cgt gga gga 1488  
Thr Ser Thr Ile Glu Ile Asp Ile Asp Phe Val Val Ser Arg Gly Gly  
485 490 495

gaa tac ttc ttc tcg ccc tca ctt tct gcg atc agg act gtg ctt tca 1536  
Glu Tyr Phe Phe Ser Pro Ser Leu Ser Ala Ile Arg Thr Val Leu Ser  
500 505 510

gtc tag 1542  
Val

<210> 2

<211> 513

<212> PRT

5 <213> *Marasmius scorodoni*

<400> 2

Met Lys Leu Phe Ser Ala Ser Val Phe Ala Ala Ile Val Ala Ser His  
1 5 10 15

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

ES 2 426 954 T3

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



ES 2 426 954 T3

Glu Tyr Phe Phe Ser Pro Ser Leu Ser Ala Ile Arg Thr Val Leu Ser  
 500 505 510

Val

<210> 3

<211> 1533

<212> ADN

5 <213> *Marasmius scorodonius*

<220>

<221 > CDS

<222> (1)..(1533)

<400> 3

```

atg cgg ctc act tac ctt ccc ttg ttt gcg ggc atc gcc atc cag tct      48
Met Arg Leu Thr Tyr Leu Pro Leu Phe Ala Gly Ile Ala Ile Gln Ser
1           5           10           15

gca tgt gcc ttt cca aac ttc tcc aag tot tcc ata tta aag cct cgt      96
Ala Cys Ala Phe Pro Asn Phe Ser Lys Ser Ser Ile Leu Lys Pro Arg
          20           25           30

agg acg aac tcc cta cta atc aat ccc gat gct cag cca gac ctc ccg     144
Arg Thr Asn Ser Leu Leu Ile Asn Pro Asp Ala Gln Pro Asp Leu Pro
          35           40           45

acc gca aag cag gct tcc act gcg gct gct tct gtg ggt ttg aac ctt     192
Thr Ala Lys Gln Ala Ser Thr Ala Ala Ala Ser Val Gly Leu Asn Leu
          50           55           60

acc gac atc caa gga gac atc tta atc ggt atg aag aag aac aag gaa     240
Thr Asp Ile Gln Gly Asp Ile Leu Ile Gly Met Lys Lys Asn Lys Glu
65           70           75           80

atg ttc ttc ttc ttc agt att gca gat gcc gct gca ttc aaa tcc cac     288
Met Phe Phe Phe Phe Ser Ile Ala Asp Ala Ala Phe Lys Ser His
          85           90           95

ttg ggc tcc gct att ctc cct ctt atc acc tgg acg caa cag ctg ctt     336
Leu Gly Ser Ala Ile Leu Pro Leu Ile Thr Ser Thr Gln Gln Leu Leu
          100          105          110

gct gtt gcc agc cag cct acc acc gct gtc aac ctt gcc ttc tcc caa     384
Ala Val Ala Ser Gln Pro Thr Thr Ala Val Asn Leu Ala Phe Ser Gln
          115          120          125

acc gga ttg aat gcg ctg ggg ctt gct gct caa ggc ttg ggg gac tcc     432
Thr Gly Leu Asn Ala Leu Gly Leu Ala Ala Gln Gly Leu Gly Asp Ser
          130          135          140
    
```

ES 2 426 954 T3

ctc ttt gcc agt ggt cag ttc agc ggt gca cag agt ctc ggc gac ccg 480  
 Leu Phe Ala Ser Gly Gln Phe Ser Gly Ala Gln Ser Leu Gly Asp Pro  
 145 150 155 160

gga acc tcg aac tgg gtc caa gcg ttc gct ggt aca ggc ata cat ggt 528  
 Gly Thr Ser Asn Trp Val Gln Ala Phe Ala Gly Thr Gly Ile His Gly  
 165 170 175

gtc ttc ctt ctc gca tcg gat acc gtc gac aac gtg aat gcc gag ctg 576  
 Val Phe Leu Leu Ala Ser Asp Thr Val Asp Asn Val Asn Ala Glu Leu  
 180 185 190

tcg caa atc cag tcc atc ttg gga acc tcc atc act gag gcg tac cgc 624  
 Ser Gln Ile Gln Ser Ile Leu Gly Thr Ser Ile Thr Glu Ala Tyr Arg  
 195 200 205

ctt cag ggt gag gct cgt ccc ggc gat cag caa ggc cac gaa cat ttc 672  
 Leu Gln Gly Glu Ala Arg Pro Gly Asp Gln Gln Gly His Glu His Phe  
 210 215 220

gga ttc atg gac gga atc agc aac cct gcc atc gac gga ttc tcc act 720  
 Gly Phe Met Asp Gly Ile Ser Asn Pro Ala Ile Asp Gly Phe Ser Thr  
 225 230 235 240

gcg ctg cct ggt caa gcc gtt ctc agt ccc gga ctt ttc ctg cta gga 768  
 Ala Leu Pro Gly Gln Ala Val Leu Ser Pro Gly Leu Phe Leu Leu Gly  
 245 250 255

gag gac ggc gat ggt tcc tcg tct tcg cgt ccg tct tgg gca aag gac 816  
 Glu Asp Gly Asp Gly Ser Ser Ser Ser Arg Pro Ser Trp Ala Lys Asp  
 260 265 270

ggc tct ttc ctt gct ttc cgc cag ctt caa cag cgt gtc cca gag ttc 864  
 Gly Ser Phe Leu Ala Phe Arg Gln Leu Gln Gln Arg Val Pro Glu Phe  
 275 280 285

aac aag ttc ctc gct gac aac gct gcg cta aca cag ggt aac gct gat 912  
 Asn Lys Phe Leu Ala Asp Asn Ala Ala Leu Thr Gln Gly Asn Ala Asp  
 290 295 300

ctt ctc ggt gcc cga atg atg gga cgg tgg aaa tct ggt gct ccg gtc 960  
 Leu Leu Gly Ala Arg Met Met Gly Arg Trp Lys Ser Gly Ala Pro Val  
 305 310 315 320

gac ctt gcc ccc acc gcg gac gat gtt gat ctc gcc aat gac ccc caa 1008  
 Asp Leu Ala Pro Thr Ala Asp Asp Val Asp Leu Ala Asn Asp Pro Gln  
 325 330 335

agg aac aac aac ttc aac ttt acc cac gcc ggt ttc acc gag acc act 1056  
 Arg Asn Asn Asn Phe Asn Phe Thr His Ala Gly Phe Thr Glu Thr Thr  
 340 345 350

gac gaa act cac tgc ccc ttc tcc gct cac atc cgt aag acg aac cct 1104  
 Asp Glu Thr His Cys Pro Phe Ser Ala His Ile Arg Lys Thr Asn Pro  
 355 360 365

cga tcg gac ttc aac ccc cag aac acc aac aac cac atc atc cgt gct 1152

ES 2 426 954 T3

Arg Ser Asp Phe Asn Pro Gln Asn Thr Asn Asn His Ile Ile Arg Ala  
 370 375 380

ggt att cct tac gga cct gaa gtc act gac gct gaa gcc tca tcc aac 1200  
 Gly Ile Pro Tyr Gly Pro Glu Val Thr Asp Ala Glu Ala Ser Ser Asn  
 385 390 395 400

acg tcc agc acg gac gct agc ctc gag cgt ggc ttg gct ttc gtt gca 1248  
 Thr Ser Ser Thr Asp Ala Ser Leu Glu Arg Gly Leu Ala Phe Val Ala  
 405 410 415

tac caa tcg aac att ggc aac ggc ttt gca ttc att caa cag aat tgg 1296  
 Tyr Gln Ser Asn Ile Gly Asn Gly Phe Ala Phe Ile Gln Gln Asn Trp  
 420 425 430

ggt gac aac gca aac ttc ttc ttc gga aaa acc acc ccg cct ggt att 1344  
 Val Asp Asn Ala Asn Phe Phe Phe Gly Lys Thr Thr Pro Pro Gly Ile  
 435 440 445

gac ccc atc atc ggc tcg aat gct gca cag aac aac ttc gct ccc aac 1392  
 Asp Pro Ile Ile Gly Ser Asn Ala Ala Gln Asn Asn Phe Ala Pro Asn  
 450 455 460

tct cca cgt cct gtg tcc gga ctt gac ccc aca gat tcg acc acg atc 1440  
 Ser Pro Arg Pro Val Ser Gly Leu Asp Pro Thr Asp Ser Thr Thr Ile  
 465 470 475 480

gtc acc tta aac acc gac ttc gtt gtt tct cgt gga gga gag tac ttc 1488  
 Val Thr Leu Asn Thr Asp Phe Val Val Ser Arg Gly Gly Glu Tyr Phe  
 485 490 495

ttc tcc ccc tcg ctc tct gcg atc cag aac acg ctt tct gtt tga 1533  
 Phe Ser Pro Ser Leu Ser Ala Ile Gln Asn Thr Leu Ser Val  
 500 505 510

<210> 4

<211> 510

<212> PRT

5 <213> *Marasmius scorodonius*

<400> 4

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

Ala Cys Ala Phe Pro Asn Phe Ser Lys Ser Ser Ile Leu Lys Pro Arg  
 20 25 30

Arg Thr Asn Ser Leu Leu Ile Asn Pro Asp Ala Gln Pro Asp Leu Pro  
 35 40 45

Thr Ala Lys Gln Ala Ser Thr Ala Ala Ala Ser Val Gly Leu Asn Leu  
 50 55 60

ES 2 426 954 T3

Thr Asp Ile Gln Gly Asp Ile Leu Ile Gly Met Lys Lys Asn Lys Glu  
 65 70 75 80

Met Phe Phe Phe Phe Ser Ile Ala Asp Ala Ala Ala Phe Lys Ser His  
 85 90 95

Leu Gly Ser Ala Ile Leu Pro Leu Ile Thr Ser Thr Gln Gln Leu Leu  
 100 105 110

Ala Val Ala Ser Gln Pro Thr Thr Ala Val Asn Leu Ala Phe Ser Gln  
 115 120 125

Thr Gly Leu Asn Ala Leu Gly Leu Ala Ala Gln Gly Leu Gly Asp Ser  
 130 135 140

Leu Phe Ala Ser Gly Gln Phe Ser Gly Ala Gln Ser Leu Gly Asp Pro  
 145 150 155 160

Gly Thr Ser Asn Trp Val Gln Ala Phe Ala Gly Thr Gly Ile His Gly  
 165 170 175

Val Phe Leu Leu Ala Ser Asp Thr Val Asp Asn Val Asn Ala Glu Leu  
 180 185 190

Ser Gln Ile Gln Ser Ile Leu Gly Thr Ser Ile Thr Glu Ala Tyr Arg  
 195 200 205

Leu Gln Gly Glu Ala Arg Pro Gly Asp Gln Gln Gly His Glu His Phe  
 210 215 220

Gly Phe Met Asp Gly Ile Ser Asn Pro Ala Ile Asp Gly Phe Ser Thr  
 225 230 235 240

Ala Leu Pro Gly Gln Ala Val Leu Ser Pro Gly Leu Phe Leu Leu Gly  
 245 250 255

Glu Asp Gly Asp Gly Ser Ser Ser Ser Arg Pro Ser Trp Ala Lys Asp  
 260 265 270

Gly Ser Phe Leu Ala Phe Arg Gln Leu Gln Gln Arg Val Pro Glu Phe  
 275 280 285

ES 2 426 954 T3

Asn Lys Phe Leu Ala Asp Asn Ala Ala Leu Thr Gln Gly Asn Ala Asp  
 290 295 300

Leu Leu Gly Ala Arg Met Met Gly Arg Trp Lys Ser Gly Ala Pro Val  
 305 310 315 320

Asp Leu Ala Pro Thr Ala Asp Asp Val Asp Leu Ala Asn Asp Pro Gln  
 325 330 335

Arg Asn Asn Asn Phe Asn Phe Thr His Ala Gly Phe Thr Glu Thr Thr  
 340 345 350

Asp Glu Thr His Cys Pro Phe Ser Ala His Ile Arg Lys Thr Asn Pro  
 355 360 365

Arg Ser Asp Phe Asn Pro Gln Asn Thr Asn Asn His Ile Ile Arg Ala  
 370 375 380

Gly Ile Pro Tyr Gly Pro Glu Val Thr Asp Ala Glu Ala Ser Ser Asn  
 385 390 395 400

Thr Ser Ser Thr Asp Ala Ser Leu Glu Arg Gly Leu Ala Phe Val Ala  
 405 410 415

Tyr Gln Ser Asn Ile Gly Asn Gly Phe Ala Phe Ile Gln Gln Asn Trp  
 420 425 430

Val Asp Asn Ala Asn Phe Phe Phe Gly Lys Thr Thr Pro Pro Gly Ile  
 435 440 445

Asp Pro Ile Ile Gly Ser Asn Ala Ala Gln Asn Asn Phe Ala Pro Asn  
 450 455 460

Ser Pro Arg Pro Val Ser Gly Leu Asp Pro Thr Asp Ser Thr Thr Ile  
 465 470 475 480

Val Thr Leu Asn Thr Asp Phe Val Val Ser Arg Gly Gly Glu Tyr Phe  
 485 490 495

Phe Ser Pro Ser Leu Ser Ala Ile Gln Asn Thr Leu Ser Val  
 500 505 510

<210> 5

<211> 1482

<212> ADN

ES 2 426 954 T3

<213> *Marasmius scorodoni*

<220>

<221 > CDS

<222> (1)..(1482)

5 <400> 5

act gcc cac atc cgc gct ccc aac gtg aag ccc cgc cgc acc aac tcc	48
Thr Ala His Ile Arg Ala Pro Asn Val Lys Pro Arg Arg Thr Asn Ser	
1 5 10 15	
ctt ctg atc act ccc cct cag cag cct ccc ctt ccc tct gct cag cag	96
Leu Leu Ile Thr Pro Pro Gln Gln Pro Pro Leu Pro Ser Ala Gln Gln	
20 25 30	
gct gcc tcc gcc tcc agc tcc gct ggc ctc aac ctt acc gac att cag	144
Ala Ala Ser Ala Ser Ser Ser Ala Gly Leu Asn Leu Thr Asp Ile Gln	
35 40 45	
ggg gac att ctg atc ggc atg aag aag aac aag gaa ctc ttc ttc ttc	192
Gly Asp Ile Leu Ile Gly Met Lys Lys Asn Lys Glu Leu Phe Phe Phe	
50 55 60	
ttc tcc gtc acc gac gcc gct act ttc aag gct aag ctc gga tcc gac	240
Phe Ser Val Thr Asp Ala Ala Thr Phe Lys Ala Lys Leu Gly Ser Asp	
65 70 75 80	
att ctc ggt ctg atc acc tcc acc gat cag ctg ctc gct aac gac act	288
Ile Leu Gly Leu Ile Thr Ser Thr Asp Gln Leu Leu Ala Asn Asp Thr	
85 90 95	
cag cct gtc acc gct gtc aac gtc gct ttc tot agc act ggc ctc aag	336
Gln Pro Val Thr Ala Val Asn Val Ala Phe Ser Ser Thr Gly Leu Lys	
100 105 110	
gcc ttg ggt atc acc gac gat ctg aag gat ccc gtc ttc gag gcc ggt	384
Ala Leu Gly Ile Thr Asp Asp Leu Lys Asp Pro Val Phe Glu Ala Gly	
115 120 125	
atg ctc agc aac gcc gtg agc gac ttg agc gat ccc ggt acc ggc aac	432
Met Leu Ser Asn Ala Val Ser Asp Leu Ser Asp Pro Gly Thr Gly Asn	
130 135 140	
tgg gtc cct ggc ttc gtc ggc acc tcc gtt cac ggc gtt ttc ctg ctc	480
Trp Val Pro Gly Phe Val Gly Thr Ser Val His Gly Val Phe Leu Leu	
145 150 155 160	
gcc tcg gac acc atc gac aac gtc aac acc gag ctg gcc aac atc cag	528
Ala Ser Asp Thr Ile Asp Asn Val Asn Thr Glu Leu Ala Asn Ile Gln	
165 170 175	
acc att ttg aac ggc tcg atc acc gag att cac cgt ctg cag ggt gag	576
Thr Ile Leu Asn Gly Ser Ile Thr Glu Ile His Arg Leu Gln Gly Glu	

ES 2 426 954 T3

180					185					190						
gct	cgt	ccc	ggt	gac	cag	cag	ggt	cac	gag	cac	ttc	gga	ttc	atg	gac	624
Ala	Arg	Pro	Gly	Asp	Gln	Gln	Gly	His	Glu	His	Phe	Gly	Phe	Met	Asp	
		195					200					205				
ggt	atc	tcc	aac	ccc	gcc	gtt	gac	ggc	ttc	acc	cct	ccc	gcg	gaa	atc	672
Gly	Ile	Ser	Asn	Pro	Ala	Val	Asp	Gly	Phe	Thr	Pro	Pro	Ala	Glu	Ile	
	210					215					220					
cgt	cct	gga	cag	gct	ctc	att	ccc	ccc	ggt	atc	atg	ctt	ctc	ggt	gag	720
Arg	Pro	Gly	Gln	Ala	Leu	Ile	Pro	Pro	Gly	Ile	Met	Leu	Leu	Gly	Glu	
225					230					235					240	
gcc	aac	gac	act	ttc	cag	aac	gat	cgt	cct	ccc	tgg	gcc	aag	gac	ggt	768
Ala	Asn	Asp	Thr	Phe	Gln	Asn	Asp	Arg	Pro	Pro	Trp	Ala	Lys	Asp	Gly	
				245					250					255		
tcc	ttc	ctt	gtc	ttc	cgt	cag	atg	cag	cag	cgc	gcg	ccc	gag	ttc	aac	816
Ser	Phe	Leu	Val	Phe	Arg	Gln	Met	Gln	Gln	Arg	Ala	Pro	Glu	Phe	Asn	
			260					265					270			
aag	ttc	ctg	cag	gat	cac	gcc	ctc	aac	atg	ccc	aac	atg	acc	tcc	gag	864
Lys	Phe	Leu	Gln	Asp	His	Ala	Leu	Asn	Met	Pro	Asn	Met	Thr	Ser	Glu	
		275					280					285				
cag	ggc	gct	gat	ctc	ctt	ggt	gcc	cgc	att	gtt	gga	cgc	tgg	aag	tcc	912
Gln	Gly	Ala	Asp	Leu	Leu	Gly	Ala	Arg	Ile	Val	Gly	Arg	Trp	Lys	Ser	
	290					295					300					
ggt	gct	cct	atc	gac	ctc	act	ccc	ttg	gtc	gat	gac	ccc	gtg	ttg	gcc	960
Gly	Ala	Pro	Ile	Asp	Leu	Thr	Pro	Leu	Val	Asp	Asp	Pro	Val	Leu	Ala	
305					310					315				320		
gcc	gac	aac	cag	cgc	aac	aac	aac	ttc	gac	ttc	tct	gac	gcc	acc	aac	1008
Ala	Asp	Asn	Gln	Arg	Asn	Asn	Asn	Phe	Asp	Phe	Ser	Asp	Ala	Thr	Asn	
				325					330					335		
cag	acc	cgt	tgc	cct	ttc	tct	gct	cac	atc	cgc	aag	gct	aac	ccc	cgc	1056
Gln	Thr	Arg	Cys	Pro	Phe	Ser	Ala	His	Ile	Arg	Lys	Ala	Asn	Pro	Arg	
			340					345					350			
ggc	gat	ctc	ggt	ggt	atc	aac	aag	ttc	ccc	aac	cag	cac	atc	atc	cgc	1104
Gly	Asp	Leu	Gly	Gly	Ile	Asn	Lys	Phe	Pro	Asn	Gln	His	Ile	Ile	Arg	
		355					360					365				
gcc	ggt	att	ccc	tac	ggt	ccc	gaa	gtt	acc	gac	gct	gag	aag	gcg	tcc	1152
Ala	Gly	Ile	Pro	Tyr	Gly	Pro	Glu	Val	Thr	Asp	Ala	Glu	Lys	Ala	Ser	
	370					375					380					
aac	agc	tct	agc	act	gac	cct	tcg	ctg	gag	cgt	ggt	ctg	gcg	ttc	gtg	1200
Asn	Ser	Ser	Ser	Thr	Asp	Pro	Ser	Leu	Glu	Arg	Gly	Leu	Ala	Phe	Val	
385					390					395				400		
gcc	tac	cag	tct	aac	atc	cag	aac	gga	ttc	gtg	ttc	ctt	cag	aag	aac	1248
Ala	Tyr	Gln	Ser	Asn	Ile	Gln	Asn	Gly	Phe	Val	Phe	Leu	Gln	Lys	Asn	
				405					410					415		

ES 2 426 954 T3

tgg gtt gac aac acc aac ttc ttc cgt ccc ggc act ggt gtc gat cct 1296  
 Trp Val Asp Asn Thr Asn Phe Phe Arg Pro Gly Thr Gly Val Asp Pro  
                   420                                  425                                  430

ctc atc ggt acc aac tct cgt aac tcc ggc acc gat gcc ccc aac acc 1344  
 Leu Ile Gly Thr Asn Ser Arg Asn Ser Gly Thr Asp Ala Pro Asn Thr  
                   435                                  440                                  445

cct cgt gtc gtc agc ggc ttg gat cct aac aac gct acc agc acc atc 1392  
 Pro Arg Val Val Ser Gly Leu Asp Pro Asn Asn Ala Thr Ser Thr Ile  
                   450                                  455                                  460

gag atc gat atc gac ttc gtc gtt tct cgt gga ggc gaa tac ttc ttc 1440  
 Glu Ile Asp Ile Asp Phe Val Val Ser Arg Gly Gly Glu Tyr Phe Phe  
 465                                  470                                  475                                  480

tcg ccc tcg ctc tct gcc atc cgc act gtg ctc tcc gtc taa 1482  
 Ser Pro Ser Leu Ser Ala Ile Arg Thr Val Leu Ser Val  
                                   485                                  490

<210> 6

<211> 493

<212> PRT

5 <213> *Marasmius scorodonius*

<400> 6

Thr Ala His Ile Arg Ala Pro Asn Val Lys Pro Arg Arg Thr Asn Ser  
 1                  5                                  10                                  15

Leu Leu Ile Thr Pro Pro Gln Gln Pro Pro Leu Pro Ser Ala Gln Gln  
                   20                                  25                                  30

Ala Ala Ser Ala Ser Ser Ser Ala Gly Leu Asn Leu Thr Asp Ile Gln  
                   35                                  40                                  45

Gly Asp Ile Leu Ile Gly Met Lys Lys Asn Lys Glu Leu Phe Phe Phe  
                   50                                  55                                  60

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

Ile Leu Gly Leu Ile Thr Ser Thr Asp Gln Leu Leu Ala Asn Asp Thr  
                   85                                  90                                  95

Gln Pro Val Thr Ala Val Asn Val Ala Phe Ser Ser Thr Gly Leu Lys  
                   100                                  105                                  110

ES 2 426 954 T3

Ala Leu Gly Ile Thr Asp Asp Leu Lys Asp Pro Val Phe Glu Ala Gly  
 115 120 125

Met Leu Ser Asn Ala Val Ser Asp Leu Ser Asp Pro Gly Thr Gly Asn  
 130 135 140

Trp Val Pro Gly Phe Val Gly Thr Ser Val His Gly Val Phe Leu Leu  
 145 150 155 160

Ala Ser Asp Thr Ile Asp Asn Val Asn Thr Glu Leu Ala Asn Ile Gln  
 165 170 175

Thr Ile Leu Asn Gly Ser Ile Thr Glu Ile His Arg Leu Gln Gly Glu  
 180 185 190

Ala Arg Pro Gly Asp Gln Gln Gly His Glu His Phe Gly Phe Met Asp  
 195 200 205

Gly Ile Ser Asn Pro Ala Val Asp Gly Phe Thr Pro Pro Ala Glu Ile  
 210 215 220

Arg Pro Gly Gln Ala Leu Ile Pro Pro Gly Ile Met Leu Leu Gly Glu  
 225 230 235 240

Ala Asn Asp Thr Phe Gln Asn Asp Arg Pro Pro Trp Ala Lys Asp Gly  
 245 250 255

Ser Phe Leu Val Phe Arg Gln Met Gln Gln Arg Ala Pro Glu Phe Asn  
 260 265 270

Lys Phe Leu Gln Asp His Ala Leu Asn Met Pro Asn Met Thr Ser Glu  
 275 280 285

Gln Gly Ala Asp Leu Leu Gly Ala Arg Ile Val Gly Arg Trp Lys Ser  
 290 295 300

Gly Ala Pro Ile Asp Leu Thr Pro Leu Val Asp Asp Pro Val Leu Ala  
 305 310 315 320

Ala Asp Asn Gln Arg Asn Asn Asn Phe Asp Phe Ser Asp Ala Thr Asn  
 325 330 335

Gln Thr Arg Cys Pro Phe Ser Ala His Ile Arg Lys Ala Asn Pro Arg



ES 2 426 954 T3

cag gct tct act gcc gcc gcc tcc gtt gga ctc aac ttg acc gac att Gln Ala Ser Thr Ala Ala Ala Ser Val Gly Leu Asn Leu Thr Asp Ile 35 40 45	144
cag ggc gat att ctc atc ggt atg aag aag aac aag gaa atg ttc ttc Gln Gly Asp Ile Leu Ile Gly Met Lys Lys Asn Lys Glu Met Phe Phe 50 55 60	192
ttc ttc tcc att gcc gat gct gcc gct ttc aag agc cac ctc ggt tcg Phe Phe Ser Ile Ala Asp Ala Ala Ala Phe Lys Ser His Leu Gly Ser 65 70 75 80	240
gcc att ctc ccc ctt atc acc agc act cag cag ctt ctg gct gtt gcc Ala Ile Leu Pro Leu Ile Thr Ser Thr Gln Gln Leu Leu Ala Val Ala 85 90 95	288
tcc cag cct acc act gcg gtt aac ttg gcc ttc tcc cag acc ggc ctc Ser Gln Pro Thr Thr Ala Val Asn Leu Ala Phe Ser Gln Thr Gly Leu 100 105 110	336
aac gcg ctg ggc ctc gct gcc cag ggt ctt gga gac tcc ctg ttc gct Asn Ala Leu Gly Leu Ala Ala Gln Gly Leu Gly Asp Ser Leu Phe Ala 115 120 125	384
tct ggt cag ttc tct ggc gcc cag tcc ctg gga gat cct ggt act tcg Ser Gly Gln Phe Ser Gly Ala Gln Ser Leu Gly Asp Pro Gly Thr Ser 130 135 140	432
aac tgg gtc cag gcc ttc gct ggc acc ggt atc cac ggt gtc ttc ctc Asn Trp Val Gln Ala Phe Ala Gly Thr Gly Ile His Gly Val Phe Leu 145 150 155 160	480
ttg gct tcc gat acc gtt gac aac gtg aac gcc gag ctc agc cag atc Leu Ala Ser Asp Thr Val Asp Asn Val Asn Ala Glu Leu Ser Gln Ile 165 170 175	528
cag tcc atc ctc ggc acc tcc att act gag gct tac cgc ctc cag gga Gln Ser Ile Leu Gly Thr Ser Ile Thr Glu Ala Tyr Arg Leu Gln Gly 180 185 190	576
gag gcg cgt ccc ggt gac cag cag ggt cac gag cac ttc ggc ttc atg Glu Ala Arg Pro Gly Asp Gln Gln Gly His Glu His Phe Gly Phe Met 195 200 205	624
gac ggc atc tcg aac cct gcc atc gac ggt ttc agc acc gct ctg ccc Asp Gly Ile Ser Asn Pro Ala Ile Asp Gly Phe Ser Thr Ala Leu Pro 210 215 220	672
ggc cag gct gtg ctt tcc ccc ggc ctg ttc ctg ttg ggt gaa gac ggt Gly Gln Ala Val Leu Ser Pro Gly Leu Phe Leu Leu Gly Glu Asp Gly 225 230 235 240	720
gat ggt agc tcg tcc tct cgc ccc agc tgg gct aag gac gga tcc ttc Asp Gly Ser Ser Ser Ser Arg Pro Ser Trp Ala Lys Asp Gly Ser Phe 245 250 255	768

ES 2 426 954 T3

ctc gcc ttc cgt cag ctc cag cag cgt gtt ccc gag ttc aac aag ttc	816
Leu Ala Phe Arg Gln Leu Gln Gln Arg Val Pro Glu Phe Asn Lys Phe	
260 265 270	
ctc gcc gac aac gct gcc ctg acc cag ggt aac gcc gat ctt ctc gga	864
Leu Ala Asp Asn Ala Ala Leu Thr Gln Gly Asn Ala Asp Leu Leu Gly	
275 280 285	
gcc cgc atg atg ggc cgc tgg aag tct ggt gcc ccc gtc gac ctg gct	912
Ala Arg Met Met Gly Arg Trp Lys Ser Gly Ala Pro Val Asp Leu Ala	
290 295 300	
cct acc gcg gac gat gtg gac ttg gcc aac gac cct cag cgc aac aac	960
Pro Thr Ala Asp Asp Val Asp Leu Ala Asn Asp Pro Gln Arg Asn Asn	
305 310 315 320	
aac ttc aac ttc acc cac gcc ggc ttc acc gag acc acc gac gag act	1008
Asn Phe Asn Phe Thr His Ala Gly Phe Thr Glu Thr Thr Asp Glu Thr	
325 330 335	
cac tgc ccc ttc tcc gcc cac atc cgc aag acc aac ccc cgt tcc gat	1056
His Cys Pro Phe Ser Ala His Ile Arg Lys Thr Asn Pro Arg Ser Asp	
340 345 350	
ttc aac cct cag aac act aac aac cac atc atc cgt gct ggt atc cct	1104
Phe Asn Pro Gln Asn Thr Asn Asn His Ile Ile Arg Ala Gly Ile Pro	
355 360 365	
tac ggt ccc gaa gtc acc gac gct gag gcc tct tcg aac acc tcc tcc	1152
Tyr Gly Pro Glu Val Thr Asp Ala Glu Ala Ser Ser Asn Thr Ser Ser	
370 375 380	
acc gac gct tcc ctc gag cgc ggc ctt gcc ttc gtc gcg tac cag agc	1200
Thr Asp Ala Ser Leu Glu Arg Gly Leu Ala Phe Val Ala Tyr Gln Ser	
385 390 395 400	
aac att ggt aac ggc ttc gct ttc ctg cag cag gcc tgg gtt gat aac	1248
Asn Ile Gly Asn Gly Phe Ala Phe Leu Gln Gln Ala Trp Val Asp Asn	
405 410 415	
gcc aac ttc ttc ttc ggc aag acc act ccc ccc gga gtg gac ccc atc	1296
Ala Asn Phe Phe Phe Gly Lys Thr Thr Pro Pro Gly Val Asp Pro Ile	
420 425 430	
att ggt tct gtg gct gcc cag aac aac ttc gct cct aac ggt ccc cgt	1344
Ile Gly Ser Val Ala Ala Gln Asn Asn Phe Ala Pro Asn Gly Pro Arg	
435 440 445	
ccc gtt agc ggt ctc gac cct acc gat tcc act aag atc gtc acc atc	1392
Pro Val Ser Gly Leu Asp Pro Thr Asp Ser Thr Lys Ile Val Thr Ile	
450 455 460	
aac acc gac ttc gtg tcc tcg cgc ggc ggt gag tac ttc ttc agc ccc	1440
Asn Thr Asp Phe Val Ser Ser Arg Gly Gly Glu Tyr Phe Phe Ser Pro	
465 470 475 480	
tct ctg tcc gcc atc cag aac acc ctc tcc gtc taa	1476

ES 2 426 954 T3

Ser Leu Ser Ala Ile Gln Asn Thr Leu Ser Val  
 485 490

<210> 8

<211> 491

<212> PRT

5 <213> *Marasmius scorodoni*

<400> 8

Phe Pro Asn Phe Ser Lys Ser Ser Ile Leu Lys Pro Arg Arg Thr Asn  
 1 5 10 15

Ser Leu Leu Ile Asn Pro Asp Ala Gln Pro Asp Leu Pro Thr Ala Lys  
 20 25 30

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

Gln Gly Asp Ile Leu Ile Gly Met Lys Lys Asn Lys Glu Met Phe Phe  
 50 55 60

Phe Phe Ser Ile Ala Asp Ala Ala Ala Phe Lys Ser His Leu Gly Ser  
 65 70 75 80

Ala Ile Leu Pro Leu Ile Thr Ser Thr Gln Gln Leu Leu Ala Val Ala  
 85 90 95

Ser Gln Pro Thr Thr Ala Val Asn Leu Ala Phe Ser Gln Thr Gly Leu  
 100 105 110

Asn Ala Leu Gly Leu Ala Ala Gln Gly Leu Gly Asp Ser Leu Phe Ala  
 115 120 125

Ser Gly Gln Phe Ser Gly Ala Gln Ser Leu Gly Asp Pro Gly Thr Ser  
 130 135 140

Asn Trp Val Gln Ala Phe Ala Gly Thr Gly Ile His Gly Val Phe Leu  
 145 150 155 160

Leu Ala Ser Asp Thr Val Asp Asn Val Asn Ala Glu Leu Ser Gln Ile  
 165 170 175

Gln Ser Ile Leu Gly Thr Ser Ile Thr Glu Ala Tyr Arg Leu Gln Gly  
 180 185 190

ES 2 426 954 T3

Glu Ala Arg Pro Gly Asp Gln Gln Gly His Glu His Phe Gly Phe Met  
 195 200 205

Asp Gly Ile Ser Asn Pro Ala Ile Asp Gly Phe Ser Thr Ala Leu Pro  
 210 215 220

Gly Gln Ala Val Leu Ser Pro Gly Leu Phe Leu Leu Gly Glu Asp Gly  
 225 230 235 240

Asp Gly Ser Ser Ser Ser Arg Pro Ser Trp Ala Lys Asp Gly Ser Phe  
 245 250 255

Leu Ala Phe Arg Gln Leu Gln Gln Arg Val Pro Glu Phe Asn Lys Phe  
 260 265 270

Leu Ala Asp Asn Ala Ala Leu Thr Gln Gly Asn Ala Asp Leu Leu Gly  
 275 280 285

Ala Arg Met Met Gly Arg Trp Lys Ser Gly Ala Pro Val Asp Leu Ala  
 290 295 300

Pro Thr Ala Asp Asp Val Asp Leu Ala Asn Asp Pro Gln Arg Asn Asn  
 305 310 315 320

Asn Phe Asn Phe Thr His Ala Gly Phe Thr Glu Thr Thr Asp Glu Thr  
 325 330 335

His Cys Pro Phe Ser Ala His Ile Arg Lys Thr Asn Pro Arg Ser Asp  
 340 345 350

Phe Asn Pro Gln Asn Thr Asn Asn His Ile Ile Arg Ala Gly Ile Pro  
 355 360 365

Tyr Gly Pro Glu Val Thr Asp Ala Glu Ala Ser Ser Asn Thr Ser Ser  
 370 375 380

Thr Asp Ala Ser Leu Glu Arg Gly Leu Ala Phe Val Ala Tyr Gln Ser  
 385 390 395 400

Asn Ile Gly Asn Gly Phe Ala Phe Leu Gln Gln Ala Trp Val Asp Asn  
 405 410 415

ES 2 426 954 T3

Ala Asn Phe Phe Phe Gly Lys Thr Thr Pro Pro Gly Val Asp Pro Ile  
420 425 430

Ile Gly Ser Val Ala Ala Gln Asn Asn Phe Ala Pro Asn Gly Pro Arg  
435 440 445

Pro Val Ser Gly Leu Asp Pro Thr Asp Ser Thr Lys Ile Val Thr Ile  
450 455 460

Asn Thr Asp Phe Val Ser Ser Arg Gly Gly Glu Tyr Phe Phe Ser Pro  
465 470 475 480

Ser Leu Ser Ala Ile Gln Asn Thr Leu Ser Val  
485 490

## REIVINDICACIONES

1. Un método para la oxidación de un material lignocelulósico, método el cual comprende poner en contacto dicho material lignocelulósico con un polipéptido que tiene actividad de peroxidasa y que es:
- 5 a. un polipéptido que comprende una secuencia de aminoácidos que tiene al menos 55% de homología con una secuencia de aminoácidos expuesta en uno cualquiera de: aminoácidos 21 a 513 de SEC NO: 2; aminoácidos 20 a 510 de SEC ID NO: 4; aminoácidos 1 a 493 de SEC ID NO 6; o aminoácidos 1 a 491 de SEC ID NO: 8, o
- 10 b. un polipéptido codificado por un polinucleótido que comprende una secuencia nucleotídica que tiene al menos 55% de homología con la secuencia nucleotídica expuesta en uno cualquiera de: nucleótidos 61 a 1539 de SEC ID NO: 1; nucleótidos 58 a 1530 de SEC ID NO: 3; nucleótidos 1 a 1479 de SEC ID NO: 5; o nucleótidos 1 a 1473 de SEC ID NO: 7.
2. Un método según la reivindicación 1, en el que la oxidación aumenta la susceptibilidad del material lignocelulósico a la degradación enzimática.
3. Un método para la producción de un azúcar o azúcares a partir de un material lignocelulósico, método el cual comprende las siguientes etapas:
- 15 (i) pretratamiento del material lignocelulósico;
- (ii) oxidación de dicho material lignocelulósico usando un método según la reivindicación 1 ó 2; y
- (iii) poner en contacto el material lignocelulósico así modificado con una o más celulasas y/o una o más hemicelulasas y/o una o más pectinasas.
- 20 4. Un método según la reivindicación 3, en el que el polipéptido que tiene actividad de peroxidasa se pone en contacto con el material que comprende un hidrato de carbono que no es almidón durante, antes de, o después de la etapa de pretratamiento.
5. Un método según una cualquiera de las reivindicaciones anteriores, en el que el polipéptido que tiene actividad de peroxidasa se pone en contacto con el material lignocelulósico en combinación con una enzima auxiliar, enzima auxiliar la cual modifica además dicho material que comprende un hidrato de carbono que no es almidón o incrementa la susceptibilidad de dicho material que comprende un hidrato de carbono que no es almidón a la degradación enzimática, en el que la enzima auxiliar es una enzima de las clases: amilasas, proteasas, lipasas, glucuronidasas u oxidorreductasas.
- 25 6. Un método según una cualquiera de las reivindicaciones anteriores, en el que el polipéptido que tiene actividad de peroxidasa es codificado por un polinucleótido obtenible de *Marasmius scorodoni*.
- 30 7. Un método según una cualquiera de las reivindicaciones anteriores, en el que el material lignocelulósico es una preparación de huertos de árboles frutales, matorral, desecho de la industria maderera, desecho maderero urbano, desecho municipal, desecho de explotación de la madera, maderas de pequeño calibre, cultivos leñosos de corta rotación, desecho industrial, paja de trigo, paja de avena, paja de arroz, paja de cebada, paja de centeno, paja de cáñamo, vainas de soja, vainas de arroz, paja de arroz, alimento a base de gluten de maíz, vainas de avena, caña de azúcar, rastrojo de maíz, tallos de maíz, mazorcas de maíz, cáscaras de maíz, hierba de pradera, hierba grama, panizo, pulpa de remolacha azucarera, pulpa de frutos cítricos, cáscaras de semillas, desechos celulósicos de animales, recortes de césped, algodón, algas, árboles, arbustos, hierbas, trigo, paja de trigo, bagazo de caña de azúcar, maíz, cáscaras de maíz, mazorcas de maíz, pepita de maíz, fibras procedentes de pepitas, productos y subproductos de la molienda húmeda y seca de granos, residuos sólidos municipales, papel residual, desecho de estercoleros, material herbáceo, residuos agrícolas, residuos de silvicultura, residuos sólidos municipales, papel residual, pasta papelera, residuos de la molienda del papel, ramas, arbustos, cañas, maíz, cáscaras de maíz, un cultivo energético, bosque, una fruta, una flor, un grano, una hierba, un cultivo herbáceo, una hoja, corteza, una aguja, un tronco, una raíz, un retoño, un arbusto, pasto varilla, un árbol, un vegetal, piel de fruta, una vid, pulpa de caña de azúcar, salvado de trigo, vainas de avena, madera dura o blanda, material de desecho orgánico generado a partir de un proceso industrial o desecho maderero de silvicultura, o una combinación de dos cualquiera o más de los mismos.
- 35 40 45 8. Un método para la preparación de un producto de fermentación, método el cual comprende:
- a. tratar un material lignocelulósico según un método de una cualquiera de las reivindicaciones 1-7, y
- 50 b. fermentar el material resultante,
- para preparar de ese modo un producto de fermentación.
9. Un método según la reivindicación 8, en el que el producto de fermentación es etanol, butanol; ácido láctico, ácido

## ES 2 426 954 T3

3-hidroxi-propiónico, ácido acrílico, ácido acético, ácido succínico, ácido cítrico, ácido málico, ácido fumárico, ácido itacónico, un aminoácido, 1,3-propano-diol, etileno, glicerol, un antibiótico  $\beta$ -lactámico o una cefalosporina.

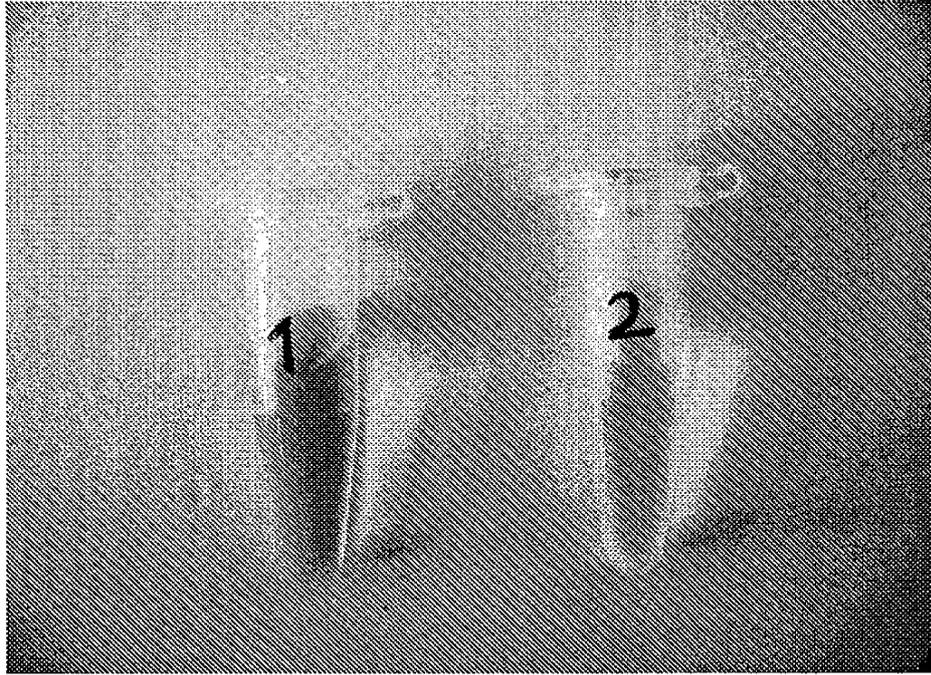


Fig. 1

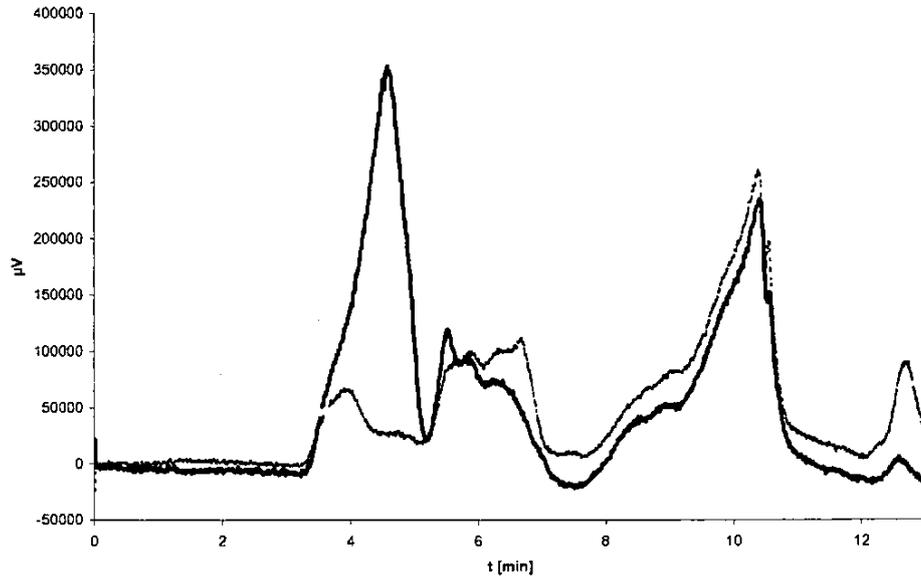


Fig. 2

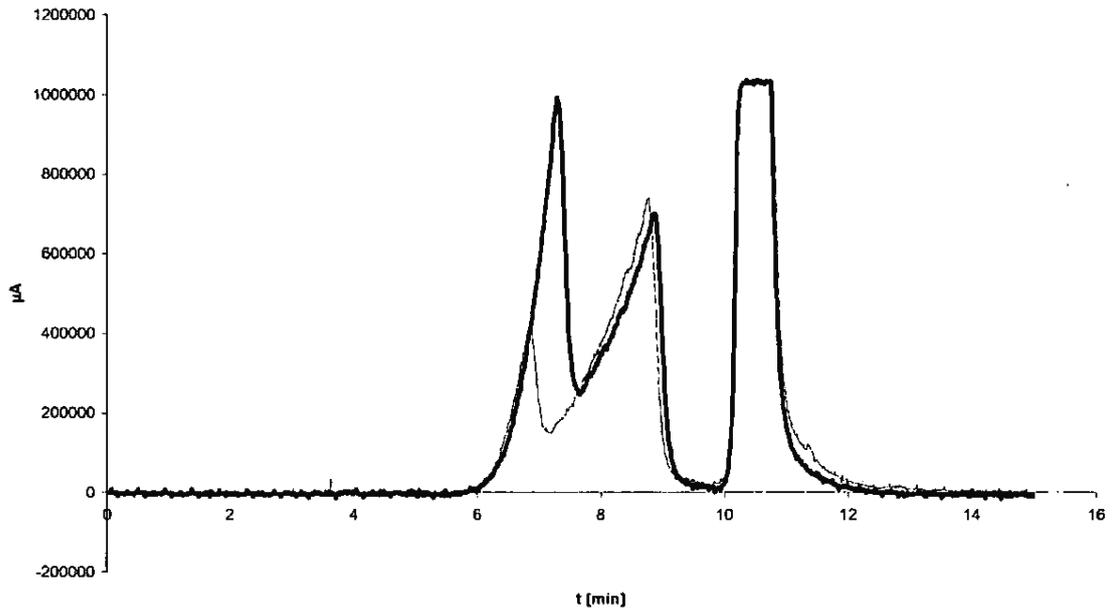


Fig. 3

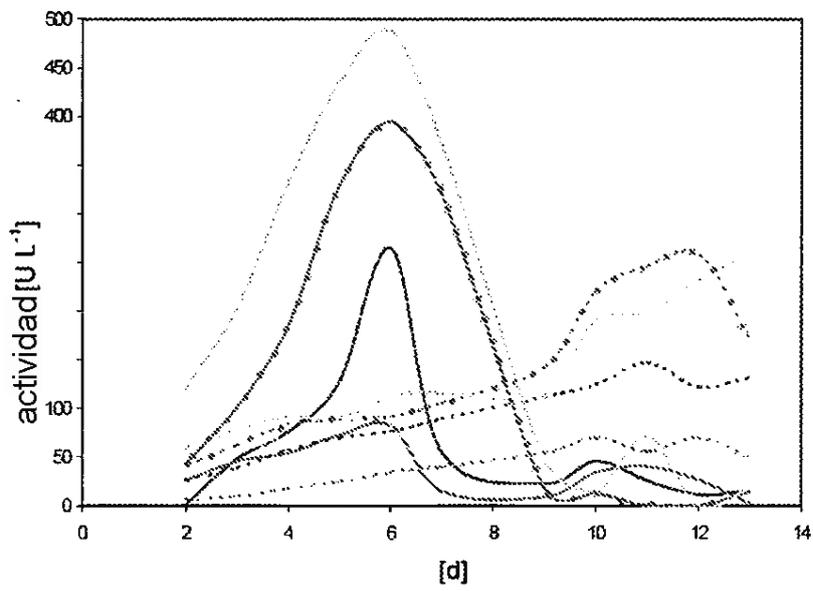


Fig. 4

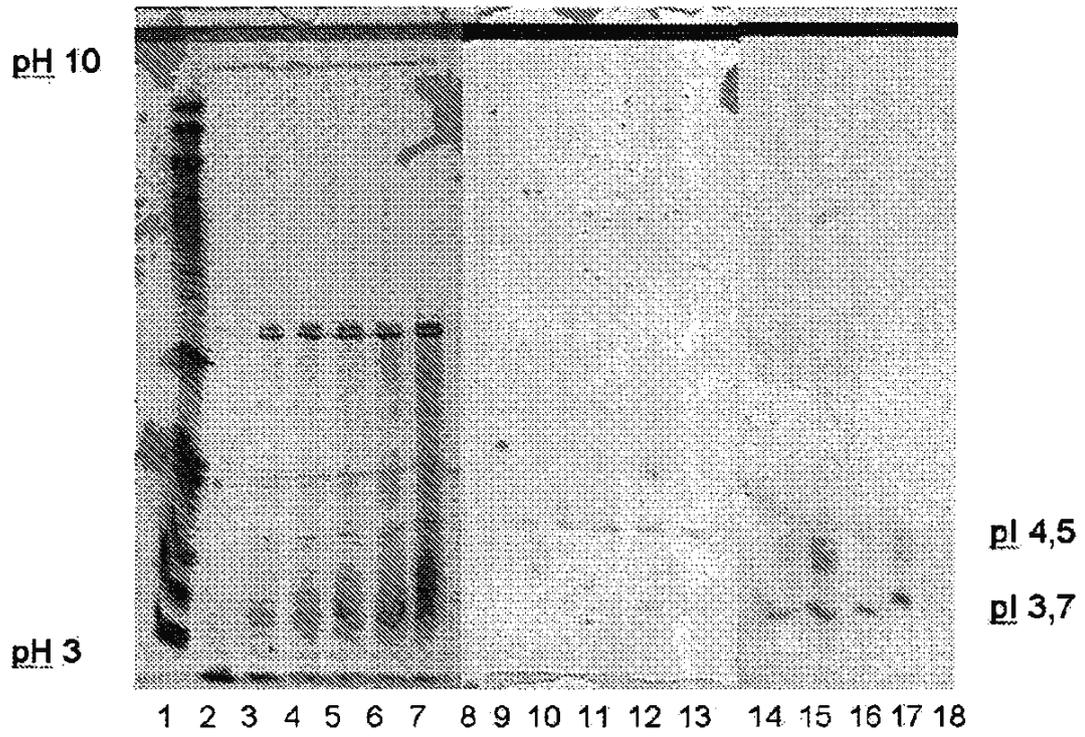


Fig. 5

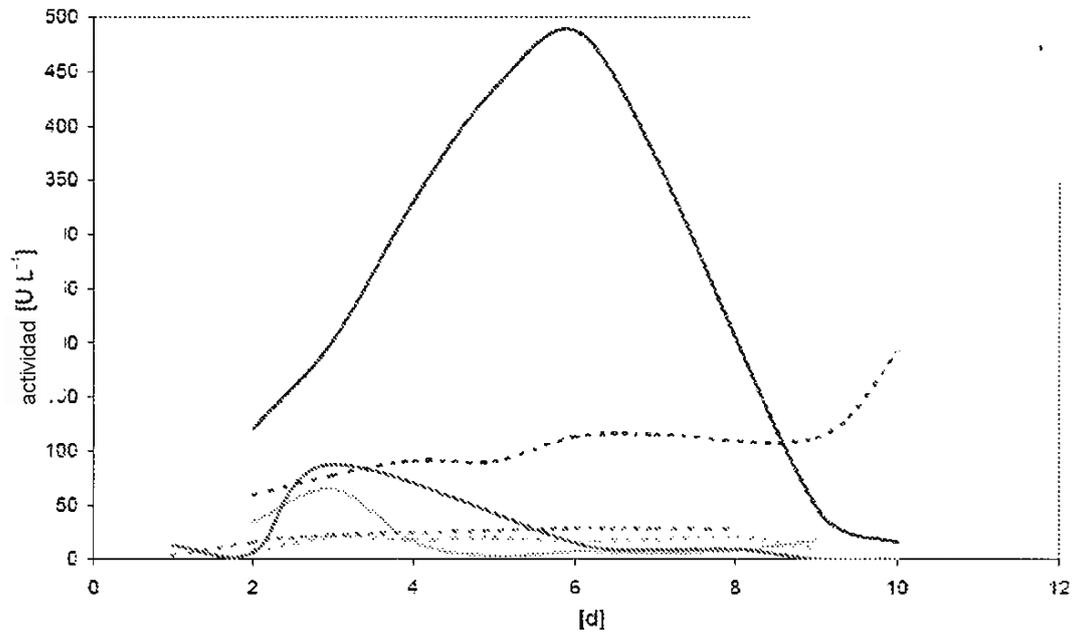


Fig. 6

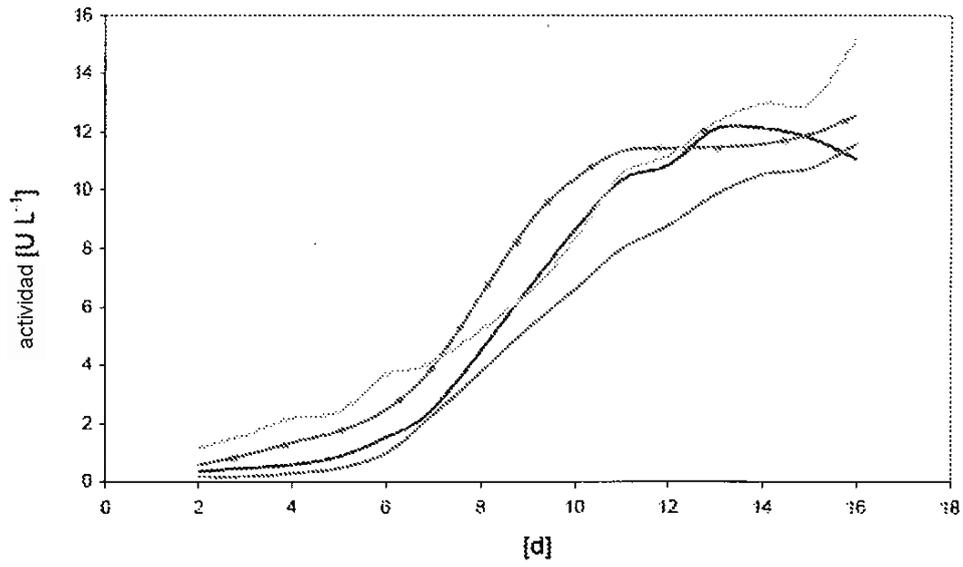


Fig. 7

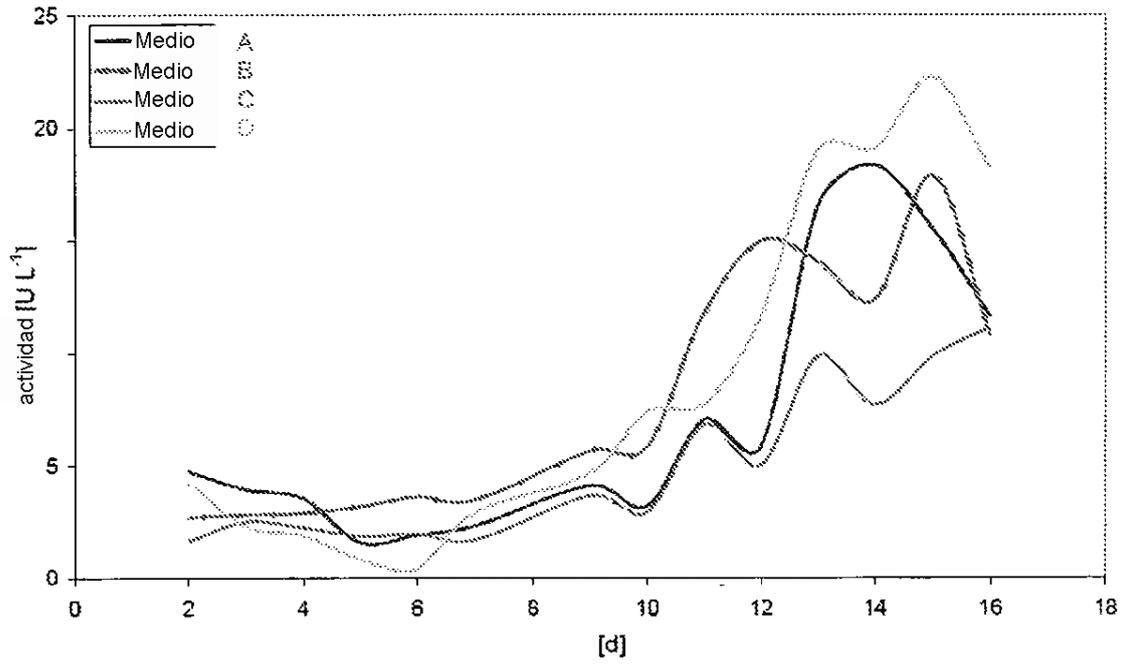


Fig. 8

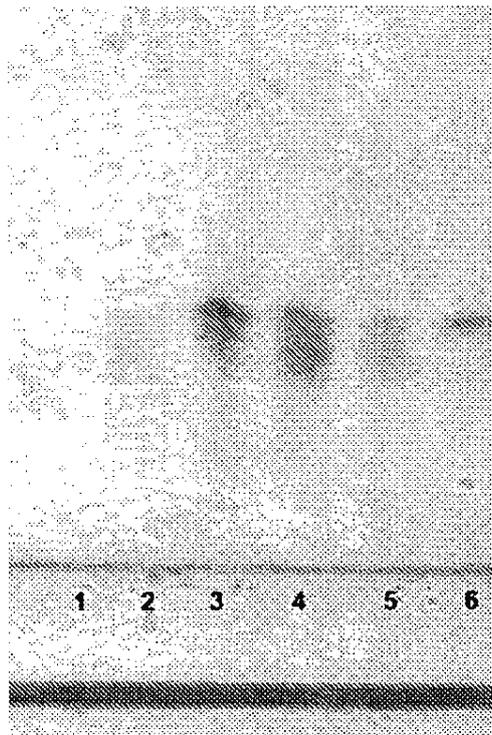


Fig. 9

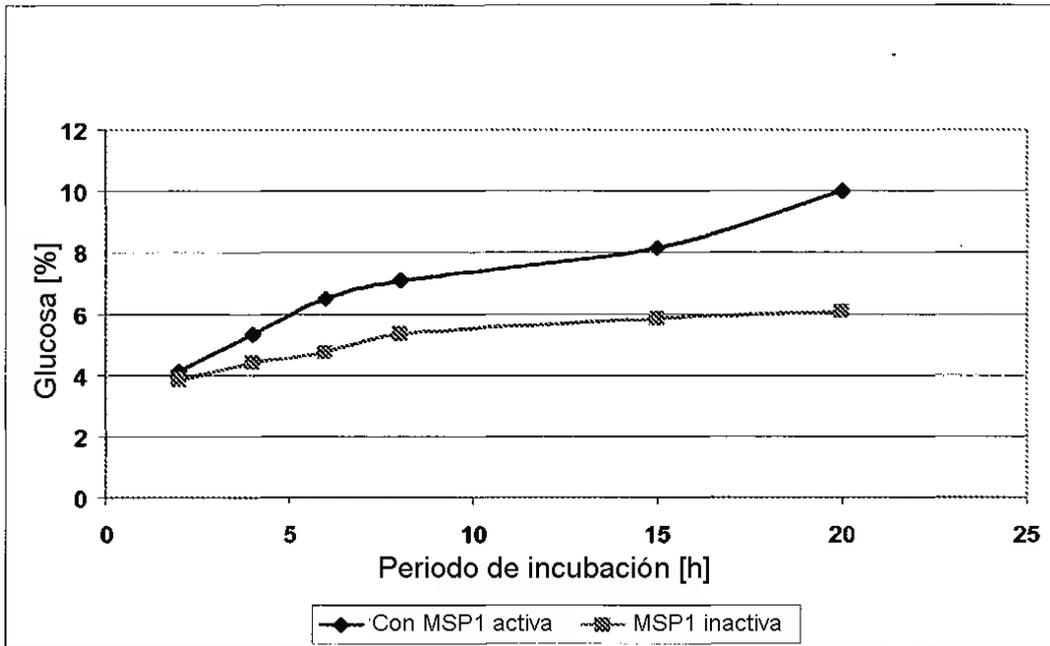


Fig . 10