



República Federativa do Brasil
Ministério da Indústria, Comércio Exterior
e Serviços
Instituto Nacional da Propriedade Industrial

(11) PI 1012923-5 B1

(22) Data do Depósito: 27/05/2010

(45) Data de Concessão: 16/10/2018



(54) Título: CÉLULA HOSPEDEIRA MICROBIANA TRANSGÊNICA, MÉTODOS PARA PRODUZIR UM POLIPEPTÍDEO, PARA PRODUZIR UM MUTANTE DE UMA CÉLULA MÃE, PARA DEGRADAR OU CONVERTER UM MATERIAL CELULÓSICO, PARA PRODUZIR UM PRODUTO DE FERMENTAÇÃO, PARA FERMENTAR UM MATERIAL CELULÓSICO, CONSTRUTO DE ÁCIDO NUCLEICO, E, VETOR DE EXPRESSÃO

(51) Int.Cl.: C12N 9/14

(30) Prioridade Unionista: 02/06/2009 US 61/183,490

(73) Titular(es): NOVOZYMES, INC.. NOVOZYMES A/S

(72) Inventor(es): YE LIU; LAN TANG; PAUL HARRIS; WENPING WU

(85) Data do Início da Fase Nacional: 02/12/2011

“CÉLULA HOSPEDEIRA MICROBIANA TRANSGÊNICA, MÉTODOS PARA PRODUZIR UM POLIPEPTÍDEO, PARA PRODUZIR UM MUTANTE DE UMA CÉLULA MÃE, PARA DEGRADAR OU CONVERTER UM MATERIAL CELULÓSICO, PARA PRODUZIR UM
5 PRODUTO DE FERMENTAÇÃO, PARA FERMENTAR UM MATERIAL CELULÓSICO, CONSTRUTO DE ÁCIDO NUCLEICO, E, VETOR DE EXPRESSÃO”

DECLARAÇÃO DOS DIREITOS PARA AS INVENÇÕES REALIZADAS EM PESQUISA E DESENVOLVIMENTO PATROCINADAS PELO
10 GOVERNO FEDERAL

Esta invenção foi realizada com o auxílio do Governo no acordo de cooperação DE-FC36-08GO18080, concedida pelo Departamento de Energia. O Governo apresenta certos direitos nesta invenção.

REFERÊNCIA A UMA LISTAGEM DE SEQUÊNCIA

15 Este pedido contém uma listagem de sequência na forma legível em computador, que é aqui incorporada pela referência.

REFERÊNCIA A UM DEPÓSITO DE MATERIAL BIOLÓGICO

Este pedido contém uma referência a um depósito de material biológico, cujo depósito é aqui incorporado pela referência.

20 FUNDAMENTOS DA INVENÇÃO

CAMPO DA INVENÇÃO

A presente invenção diz respeito a polipeptídeos com atividade celobiohidrolase e polinucleotídeos que codificam os polipeptídeos. A invenção também diz respeito a construtos de ácido nucleico, vetores e células
25 hospedeiras que compreendem os polinucleotídeos, bem como métodos de

produzir e usar os polipeptídeos.

DESCRIÇÃO DA TÉCNICA RELACIONADA

A celulose é um polímero do açúcar simples glicose por ligação do tipo beta-1,4. Muitos microrganismos produzem enzimas que
5 hidrolisam beta-glicanas ligadas. Estas enzimas incluem endoglicanases, celobiohidrolases e beta-glicosidases. As endoglicanases digerem o polímero de celulose em locais aleatórios, abrindo-o para ataque por celobiohidrolases. As celobiohidrolases liberam sequencialmente moléculas de celobiose a partir das extremidades do polímero de celulose. A celobiose é um dímero ligado a
10 1,4-beta de glicose solúvel em água. As beta-glicosidases hidrolisam celobiose em glicose.

A conversão de matérias primas lignocelulíticas em etanol apresenta as vantagens da disponibilidade imediata de grandes quantidades de matéria prima, a vantagem de evitar queimar ou aterrar os materiais, e a
15 limpeza do etanol combustível. Madeira, resíduos agrícolas, culturas herbáceas e resíduos sólidos municipais são considerados matérias primas para a produção de etanol. Estes materiais consistem principalmente em celulose, hemicelulose, e lignina. Uma vez que a celulose é convertida em glicose, a glicose é facilmente fermentada por levedura em etanol.

20 Existe uma necessidade na técnica de melhorar as composições de proteína celulolítica por meio da suplementação com enzimas adicionais para aumentar a eficiência e fornecer soluções de enzima econômicas para a degradação de lignocelulose.

A presente invenção fornece polipeptídeos com atividade
25 celobiohidrolase e polinucleotídeos que codificam os polipeptídeos.

SUMÁRIO DA INVENÇÃO

A presente invenção diz respeito a polipeptídeos isolados com atividade celobiohidrolase selecionados do grupo que consiste em:

(a) um polipeptídeo com pelo menos 90 % de identidade de

sequência com o polipeptídeo maduro de SEQ ID NO: 2 ou SEQ ID NO: 4;

(b) um polipeptídeo codificado por um polinucleotídeo que hibridiza em condições de severidade elevada (i) na sequência que codifica o polipeptídeo maduro de SEQ ID NO: 1 ou SEQ ID NO: 3, (ii) na sequência de DNAC contido no polipeptídeo maduro de SEQ ID NO: 1 ou SEQ ID NO: 3, ou (iii) na fita complementar de comprimento total de (i) ou (ii);

(c) um polipeptídeo codificado por um polinucleotídeo com pelo menos 90 % de identidade de sequência com a sequência que codifica o polipeptídeo maduro de SEQ ID NO: 1 ou SEQ ID NO: 3;

(d) um variante que compreende uma substituição, deleção, e/ou inserção de um ou mais (vários) aminoácidos do polipeptídeo maduro de SEQ ID NO: 2 ou SEQ ID NO: 4; e

(e) um fragmento do polipeptídeo de (a), (b), (c), ou (d) que apresenta atividade celobiohidrolase.

A presente invenção também diz respeito a polinucleotídeos isolados que codificam os polipeptídeos da presente invenção; construtos de ácido nucleico, vetores de expressão recombinantes e células hospedeiras recombinantes que compreendem os polinucleotídeos; e métodos de produzir os polipeptídeos.

A presente invenção também diz respeito a métodos para degradar ou converter um material celulósico compreendendo: tratar o material celulósico com uma composição de enzima na presença de um polipeptídeo com atividade celobiohidrolase da presente invenção.

A presente invenção também diz respeito a métodos para produzir um produto de fermentação que compreende:

(a) sacarificar um material celulósico com uma composição de enzima na presença de um polipeptídeo com atividade celobiohidrolase da presente invenção;

(b) fermentar o material celulósico sacarificado com um ou

mais microrganismos fermentadores para sintetizar o produto de fermentação;
e

(c) recuperar o produto de fermentação a partir da fermentação.

5 A presente invenção também diz respeito a métodos de fermentar um material celulósico compreendendo: fermentar o material celulósico com um ou mais microrganismos fermentadores, em que o material celulósico é sacarificado com uma composição de enzima na presença de um polipeptídeo com atividade celobiohidrolase da presente invenção.

10 A presente invenção também diz respeito a um polinucleotídeo que codifica um peptídeo sinal que compreende ou que consiste em aminoácidos 1 a 17 de SEQ ID NO: 2 ou aminoácidos 1 a 17 de SEQ ID NO:4, que é operavelmente ligado a um gene que codifica uma proteína; construtos de ácido nucleico, expressão vetores e células hospedeiras
15 recombinantes que compreendem os polinucleotídeos; e métodos de produzir uma proteína.

DESCRIÇÃO RESUMIDA DAS FIGURAS

A figura 1 mostra a sequência de DNA genômico e a sequência de aminoácidos deduzida de um gene de celobiohidrolase da
20 família 6 de *Thielavia hyrcaniae* NN045097 (SEQ ID NOs: 1 e 2, respectivamente).

A figura 2 mostra a sequência de DNA genômico e a sequência de aminoácidos deduzida de um gene de celobiohidrolase da família 6 de *Thielavia hyrcaniae* NN045178 (SEQ ID NOs: 3 e 4,
25 respectivamente).

A figura 3 mostra um mapa de restrição de pCBHII45178.

A figura 4 mostra um mapa de restrição de pGEM-T-CBHII45178.

A figura 5 mostra um mapa de restrição de pCBHII45097.

A figura 6 mostra um mapa de restrição de pGEM-T-CBHII45097.

A figura 7 mostra os perfis de atividade de pH da celobiohidrolase de *Thielavia hyrcaniae* NN045097 e NN045178 a 50 °C.

5 As figuras 8A e 8B mostram a estabilidade da celobiohidrolase de *Thielavia hyrcaniae* NN045097 e NN045178 como uma função do pH a 50 °C.

A figura 9 mostra os perfis de atividade da temperatura da celobiohidrolase de *Thielavia hyrcaniae* NN045097 e NN045178 em pH 5.

10 As figuras 10A e 10B mostram a estabilidade das celobiohidrolases de *Thielavia hyrcaniae* NN045097 e NN045178 como uma função da temperatura em pH 5.

Definições

Celobiohidrolase: O termo “celobiohidrolase” significa uma
15 1,4-beta-D-glicano celobiohidrolase (E.C. 3.2.1.91), que catalisa a hidrólise de ligações 1,4-beta-D-glicosídicas em celulose, celo-oligossacarídeos, ou qualquer polímero contendo glicose ligada a beta-1,4, que libera celobiose a partir das extremidades reduzidas ou não reduzidas da cadeia (Teeri, 1997, Crystalline cellulose degradation: New insight into the function of
20 cellobiohydrolases, *Trends in Biotechnology* 15: 160-167; Teeri *et al.*, 1998, *Trichoderma reesei* cellobiohydrolases: why so efficient on crystalline cellulose?, *Biochem. Soc. Trans.* 26: 173-178). Com propósitos da presente invenção, a atividade celobiohidrolase é determinada de acordo com os procedimentos descritos Lever *et al.*, 1972, *Anal. Biochem.* 47: 273-279; van
25 Tilbeurgh *et al.*, 1982, *FEBS Letters*, 149: 152-156; van Tilbeurgh e Claeysens, 1985, *FEBS Letters*, 187: 283-288; e Tomme *et al.*, 1988, *Eur. J. Biochem.* 170: 575-581. O método de Lever *et al.* pode ser empregado para avaliar a hidrólise de celulose na palha de milho, enquanto os métodos de van Tilbeurgh *et al.* e Tomme *et al.* podem ser usados para determinar a atividade

celobiohidrolase em um derivado de dissacarídeo fluorescente, 4-metilumbeliferil- β -D-lactosídeo. Além disso, PASC pode ser usado como um substrato da maneira aqui descrita.

Os polipeptídeos da presente invenção apresentam pelo menos 20 %, por exemplo pelo menos 40 %, pelo menos 50 %, pelo menos 60 %, pelo menos 70 %, pelo menos 80 %, pelo menos 90 %, pelo menos 95 % e pelo menos 100 % da atividade celobiohidrolase do polipeptídeo maduro de SEQ ID NO: 2 ou SEQ ID NO: 4.

Família 6, GH6 ou CEL6: Os termos “família 6”, “GH6”, ou “CEL6” são aqui definidos como um polipeptídeo que está na família 6 de glicosídeo hidrolase, de acordo com Henrissat B., 1991, A classification of glycosyl hydrolases based on amino-acid sequence similarities, *Biochem. J.* 280: 309-316, e Henrissat e Bairoch, 1996, Updating the sequence-based classification of glycosyl hydrolases, *Biochem. J.* 316: 695-696.

Enzima celulolítica ou celulase: O termo “enzima celulolítica” ou “celulase” significa uma ou mais (várias) enzimas que hidrolisam um material celulolítico. Tais enzimas incluem endoglucanase(s), celobiohidrolase(s), beta-glicosidase(s) ou combinações destas. As duas abordagens básicas para medir atividade celulolítica incluem: (1) medir a atividade celulolítica total, e (2) medir as atividades celulolíticas individuais (endoglicanases, celobiohidrolases, e beta-glicosidases) da maneira revisada em Zhang *et al.*, Outlook for cellulase improvement: Screening and selection strategies, 2006, *Biotechnology Advances* 24: 452-481. A atividade celulolítica total é medida em geral usando substratos insolúveis, incluindo filtro de papel Whatman Nº1, celulose microcristalina, celulose bacteriana, celulose de alga, algodão, lignocelulose pré-tratada, etc. O ensaio de atividade celulolítica total mais comum é o ensaio de filtro de papel usando filtro de papel Whatman Nº1 como o substrato. O ensaio foi estabelecido pelo International Union of Pure and Applied Chemistry (IUPAC) (Ghose, 1987,

Measurement of cellulase activities, *Pure Appl. Chem.* 59: 257-68).

Com propósitos da presente invenção, a atividade da enzima celulolítica é determinada medindo o aumento na hidrólise de um material celulolítico pela(s) enzima(s) celulolítica(s) nas condições a seguir: 1-20 mg
5 de proteína da enzima celulolítica/g de celulose em PCS por 3-7 dias a 50 °C comparada a uma hidrólise controle sem a adição de proteína da enzima celulolítica. As condições típicas são reações de 1 mL, PCS lavado ou não lavado, 5 % de sólidos insolúveis, acetato de sódio 50 mM em pH 5, MnSO₄ 1 mM, 50-65 °C, 72 horas, análise de açúcar por coluna AMINEX® HPX-87H
10 (Bio-Rad Laboratories, Inc., Hercules, CA, USA).

Endoglicanase: O termo “endoglicanase” significa uma endo-1,4-(1,3;1,4)-beta-D-glicano 4-glicanoidrolase (E.C. 3.2.1.4), que catalisa a endo-hidrólise de ligações 1,4-beta-D-glicosídicas em celulose, derivados de celulose (tais como carboximetil celulose e hidroxietil celulose), liquenina,
15 ligações beta-1,4 em beta-1,3 glicanos mistos, tais como beta-D-glicanos ou xiloglicanos de cereal e outro material vegetal contendo componentes celulolíticos. A atividade de endoglicanase pode ser determinada medindo a redução na viscosidade do substrato ou o aumento na redução das extremidades determinado por um ensaio de redução de açúcar (Zhang *et al.*,
20 2006, *Biotechnology Advances* 24: 452-481). Com propósitos da presente invenção, a atividade de endoglicanase é determinada usando carboximetil celulose (CMC) como substrato de acordo com o procedimento de Ghose, 1987, *Pure e Appl. Chem.* 59: 257-268, em pH 5, 40 °C.

Beta-glicosidase: O termo “beta-glicosidase” significa uma
25 atividade de beta-D-glicosídeo glicoidrolase (E.C. 3.2.1.21), que catalisa a hidrólise de resíduos de beta-D-glicose com a liberação de beta-D-glicose. A celobiase é sinônimo de beta-glicosidade. Com propósitos da presente invenção, a atividade de beta-glicosidase é determinada a 25 °C usando 4-nitrofenil-beta-D-glicopiranosídeo 1mM como substrato em citrato de sódio

50 mM em pH 4,8. Uma unidade da atividade de beta-glicosidase é definida como 1,0 μ mol de ânion *p*-nitrofenolato produzido por minuto a 25 °C, pH 4,8.

Melhor atividade celulolítica: O termo “melhor atividade celulolítica” significa uma atividade biológica catalisada por um polipeptídeo GH61 que melhora a hidrólise de um material celulolítico por enzima com atividade celulolítica. Com propósitos da presente invenção, melhor atividade celulolítica é determinada medindo o aumento na redução de açúcares ou o aumento do total de celobiose e glicose a partir da hidrólise de um material celulolítico por enzima celulolítica nas condições a seguir: 1-50 mg de proteína total/g de celulose em PCS, em que proteína total é compreendida de 50-99,5 % p/p de proteína as enzima celulolítica e 0.5-50 % p/p de proteína de um polipeptídeo GH61 com melhor atividade celulolítica por 1-7 dias a 50 °C, comparada a uma hidrólise controle com igual carregamento total protéico sem melhor atividade celulolítica (1-50 mg de proteína celulolítica/g de celulose em PCS). Em um aspecto preferido, uma mistura de CELLUCLAST® 1,5L (Novozymes A/S, Bagsværd, Dinamarca) na presença de 2-3 % de peso protéico total de beta-glicosidase de *Aspergillus oryzae* (produzida de maneira recombinante em *Aspergillus oryzae*, de acordo com WO 02/095014) ou 3 % de peso protéico total de beta-glicosidase de *Aspergillus fumigatus* (produzida de maneira recombinante em *Aspergillus oryzae*, da maneira descrita em WO 2002/095014) de carregamento de proteína celulase é usada como a fonte da atividade celulolítica.

Os polipeptídeos GH61 com melhor atividade celulolítica melhoram a hidrólise de um material celulolítico catalisado por enzima com atividade celulolítica, reduzindo a quantidade de enzima celulolítica exigida para atingir o mesmo grau de hidrólise, preferivelmente pelo menos 1,01 vez, mais preferivelmente pelo menos 1,05 vez, mais preferivelmente pelo menos 1,10 vez, mais preferivelmente pelo menos 1,25 vez, mais preferivelmente

pelo menos 1,5 vez, mais preferivelmente pelo menos 2 vezes, mais preferivelmente pelo menos 3 vezes, mais preferivelmente pelo menos 4 vezes, mais preferivelmente pelo menos 5 vezes, ainda mais preferivelmente pelo menos 10 vezes, e acima de tudo preferivelmente pelo menos 20 vezes.

- 5 **Família do glicosídeo hidrolase 61:** O termo “família do glicosídeo hidrolase 61” ou “família GH61” ou “GH61” significa um polipeptídeo que está na família 61 de glicosídeo hidrolase de acordo com Henrissat B., 1991, A classification of glycosyl hydrolases based on amino-acid sequence similarities, *Biochem. J.* 280: 309-316, e Henrissat B., e
10 Bairoch A., 1996, Updating the sequence-based classification of glycosyl hydrolases, *Biochem. J.* 316: 695-696.

- Enzima hemicelulolítica ou hemicelulase:** O termo “enzima hemicelulolítica” ou “hemicelulase” significa um ou mais (várias) enzimas que hidrolizam um material hemicelulósico. Ver, por exemplo, Shallom, D. e
15 Shoham, Y. Microbial hemicellulases. *Current Opinion In Microbiology*, 2003, 6(3): 219-228). As hemicelulases são componentes chave na degradação da biomassa de planta. Exemplos de hemicelulases incluem, mas sem limitação, uma acetilmanana esterase, uma acetixilano esterase, uma arabinanase, uma arabinofuranosidase, uma ácido coumárico esterase, uma
20 feruloil esterase, uma galactosidase, uma glicuronidase, uma glicuronoil esterase, uma mananase, uma manosidase, uma xilanase e uma xilosidase. O substratos destas enzimas, as hemiceluloses, são um grupo heterogêneo de polissacarídeos ramificados e lineares que são ligados por meio de ligações de hidrogênio às microfibrilas de celulose na parede celular vegetal, reticulando-
25 as em uma rede grande. As hemiceluloses também são covalentemente anexadas à lignina, formando junto com a celulose uma estrutura altamente complexa. A estrutura variável e organização de hemiceluloses exige a ação combinada de muitas enzimas para sua completa degradação. Os módulos catalíticos de hemicelulases são tanto glicosídeo hidrolases (GHs), que

hidrolisam ligações glicosídicas, quanto carboidrato esterases (CEs), que hidrolisam ligações éster de grupos laterais de acetato ou ácido ferúlico. Estes módulos catalíticos, com base na homologia de sua sequência primária, podem ser determinados nas famílias GH e CE marcadas por números.

5 Algumas famílias, com dobramentos similares completos, podem ser agrupadas adicionalmente em clãs, marcadas alfabeticamente (por exemplo, GH-A). Uma classificação mais informativa e atualizada destas e outras enzimas ativas em carboidrato está disponível na base de dados Carbohydrate-Active Enzymes (CAZy). As atividades hemicelulolíticas da enzima podem

10 ser medidas de acordo com Ghose e Bisaria, 1987, *Pure & Appl. Chem.* 59: 1739-1752.

Atividade que degrada xilano ou atividade xilanolítica: O termo “atividade que degrada xilano” ou “atividade xilanolítica” significa uma atividade biológica que hidrolisa material contendo xilano. As duas

15 abordagens básicas para medir atividade xilanolítica incluem: (1) medir a atividade xilanolítica total, e (2) medir as atividades xilanolíticas individuais (endoxilanases, beta-xilosidases, arabinofuranosidases, alfa-glicuronidases, acetil-xilano esterases, feruloil esterases, e alfa-glicuronil esterases). Recentes progressos nos ensaios de enzimas xilanolíticas foram sumarizados em por

20 exemplo, várias publicações, incluindo Biely e Puchard, Recent progress in the assays of xylanolytic enzymes, 2006, *Journal of the Science of Food and Agriculture* 86(11): 1636-1647; Spanikova e Biely, 2006, Glicuronoyl esterase - Novel carbohydrate esterase produced by *Schizophyllum commune*, *FEBS Letters* 580(19): 4597-4601; Herrmann, Vrsanska, Jurickova, Hirsch,

25 Biely, e Kubicek, 1997, The beta-D-xylosidase of *Trichoderma reesei* is a multifunctional beta-D-xilan xylohydrolase, *Biochemical Journal* 321: 375-381.

A atividade total que degrada xilano pode ser medida determinando a redução de açúcares formada a partir de vários tipos de xilano

incluindo, por exemplo, xilanos de espelta de aveia, madeira de faia e madeira de lariço, ou por determinação fotométrica de fragmentos de xilano corado liberados de vários xilanos covalentemente corados. O ensaio da atividade xilanolítica total mais comum baseia-se na produção de açúcares reduzidos de 4-O-metil glicuronoxilano polimérico, da maneira descrita em Bailey, Biely, Poutanen, 1992, Interlaboratory testing of methods for assay of xylanase activity, *Journal of Biotechnology* 23(3): 257-270.

Com propósitos da presente invenção, a atividade que degrada xilano é determinada medindo o aumento na hidrólise de xilano de vidoeiro (Sigma Chemical Co., Inc., St. Louis, MO, USA) por enzima(s) que degrada(m) xilano nas condições típicas a seguir: reações de 1 mL, 5 mg/mL de substrato (sólidos totais), 5 mg de proteína xilanolítica/g de substrato, acetato de sódio 50 mM em pH 5, 50 °C, 24 horas, análise de açúcar usando o ensaio de hidrazida do ácido *p*-hidroxibenzóico (PHBAH) da maneira descrita por Lever, 1972, A new reaction for colorimetric determination of carbohydrates, *Anal. Biochem* 47: 273-279.

Xilanase: O termo “xilanase” significa uma 1,4-beta-D-xilan-xiloidrolase (E.C. 3.2.1.8) que catalisa a endo-hidrólise de ligações 1,4-beta-D-xilosídicas em xilanos. Com propósitos da presente invenção, a atividade de xilanase é determinada com AZCL-arabinoxilano 0,2 % como substrato em Triton X-100 0,01% e tampão de fosfato de sódio 200 mM, pH 6 a 37 °C. Uma unidade da atividade de xilanase é definida como 1,0 µmol de azurina produzida por minuto a 37 °C, pH 6, a partir de AZCL-arabinoxilano 0,2% como substrato em tampão de fosfato de sódio 200 mM, pH 6.

Beta-xilosidase: O termo “beta-xilosidase” significa uma beta-D-xilosídeo xiloidrolase (E.C. 3.2.1.37) que catalisa a exo-hidrólise de beta →(4)-xilo-oligosacarídeos curtos para remover sucessivos resíduos de D-xilose a partir das terminações não reduzidas. Com propósitos da presente invenção, uma unidade de beta-xilosidase é definida como 1,0 µmol de ânion

p-nitrofenolato produzido por minuto a 40 °C, pH 5, a partir de *p*-nitrofenil-beta-D-xilosídeo 1 mM como substrato em citrato de sódio 100 mM contendo TWEEN® 20 0,01 %.

Acetil-xilano esterase: O termo “acetil-xilano esterase”
5 significa uma carboxilesterase (EC 3.1.1.72) que catalisa a hidrólise de grupos acetila a partir de xilano polimérico, xilose acetilada, glicose acetilada, acetato de alfa-naftila e acetato de *p*-nitrofenila. Com propósitos da presente invenção, a atividade de acetil-xilano esterase é determinada usando *p*-nitrofenilacetato 0,5 mM como substrato em acetato de sódio 50 mM, pH 5,0,
10 contendo TWEEN™ 20 0,01 %. Uma unidade de acetil-xilano esterase é definida como a quantidade de enzima capaz de liberar 1 µmol de ânion *p*-nitrofenolato por minuto em pH 5, 25 °C.

Feruloil esterase: O termo “feruloil esterase” significa uma 4-hidróxi-3-metoxicinamoil-açúcar hidrolase (EC 3.1.1.73), que catalisa a
15 hidrólise do grupo 4-hidróxi-3-metoxicinamoil (feruloil) a partir um açúcar esterificado, que é em geral arabinose em substratos "naturais", para produzir ferulato (4-hidróxi-3-metoxicinamato). A feruloil esterase também é conhecida como ésteres de ácido ferúlico, hidroxicinamoil esterase, FAE-III, cinamoil éster hidrolase, FAEA, cinnAE, FAE-I, ou FAE-II. Com propósitos
20 da presente invenção, a atividade da feruloil esterase é determinada usando *p*-nitrofenilferulato 0,5 mM como substrato em acetato de sódio 50 mM, pH 5,0. Uma unidade de feruloil esterase é igual a quantidade de enzima capaz de liberar 1 µmol de *p*-nitrofenolato anion por minuto em pH 5, 25 °C.

Alfa-glicuronidase: O termo “alfa-glicuronidase” significa
25 uma alfa-D-glucosiduronato glicuronoidrolase (EC 3.2.1.139) que catalisa a hidrólise de um alfa-D-glicuronosídeo em D-glicuronato e um álcool. Com propósitos da presente invenção, a atividade de alfa-glicuronidase é determinada de acordo com de Vries, 1998, *J. Bacteriol.* 180: 243-249. Uma unidade de alfa-glicuronidase é igual a quantidade de enzima capaz de liberar

1 μ mol de ácido glucurônico ou 4-O-metilglucurônico por minuto em pH 5, 40 °C.

Alfa-L-arabinofuranosidase: O termo “alfa-L-arabinofuranosidase” significa uma alfa-L-arabinofuranosídeo
5 arabinofuranoidrolase (EC 3.2.1.55) que catalisa a hidrólise de resíduos de alfa-L-arabinofuranosídeo não reduzidos terminais em alfa-L-arabinosídeos. A enzima age em alfa-L-arabinofuranosídeos, alfa-L-arabinanos contendo ligações (1,3) e/ou (1,5), arabinoxilanos e arabinogalactanos. A alfa-L-arabinofuranosidase é também conhecida como arabinosidase, alfa-
10 arabinosidase, alfa-L-arabinosidase, alfa-arabinofuranosidase, polissacarídeo de alfa-L-arabinofuranosidase, alfa-L-arabinofuranosídeo hidrolase, L-arabinosidase, ou alfa-L-arabinanase. Com propósitos da presente invenção, a atividade de alfa-L-arabinofuranosidase é determinada usando 5 mg de arabinoxilano de trigo de viscosidade média (Megazyme International Ireland,
15 Ltd., Bray, Co. Wicklow, Irlanda) por mL de acetato de sódio 100 mM, pH 5, em um volume total de 200 μ L por 30 minutos a 40 °C, seguido por análise de arabinose por cromatografia de coluna AMINEX® HPX-87H (Bio-Rad Laboratories, Inc., Hercules, CA, USA).

Material celulolítico: O material celulolítico pode ser
20 qualquer material contendo celulose. O polissacarídeo predominante na parede celular primária da biomassa é a celulose, o segundo mais abundante é a hemicelulose e o terceiro é a pectina. A parede celular secundária, produzida após a célula parar de crescer, também contém polissacarídeos e é reforçada por lignina polimérica covalentemente reticulada em hemicelulose. A celulose
25 é um homopolímero de anidrocetobiose e, assim, um beta-(1-4)-D-glicano linear, enquanto as hemiceluloses incluem uma variedade de compostos, tais como xilanos, xiloglicanos, arabinoxilanos e mananas em estruturas ramificadas complexas com um espectro de substituintes. Embora em geral polimórfica, a celulose é encontrada no tecido vegetal principalmente como

uma matriz cristalina insolúvel de cadeias paralelas de glicano. As hemiceluloses em geral ligam hidrogênio à celulose, bem como a outras hemiceluloses, o que ajuda a estabilizar a matriz da parede celular.

5 A celulose é encontrada em geral, por exemplo, nos caules, folhas, sépalas, cascas e sabugos de plantas ou folhas, ramos e madeira de árvores. O material celulolítico pode ser, mas sem limitação, material herbáceo, resíduo agrícola, resíduo florestal, resíduo sólido municipal, resíduo de papel, e resíduo de polpa e papel moído (ver, por exemplo, Wiselogle *et al.*, 1995, em *Handbook on Bioethanol* (Charles E. Wyman, editor), pp.105-10
118, Taylor & Francis, Washington D.C.; Wyman, 1994, *Bioresource Technology* 50: 3-16; Lynd, 1990, *Applied Biochemistry and Biotechnology* 24/25: 695-719; Mosier *et al.*, 1999, Recent Progress in Bioconversion of Lignocellulosics, em *Advances in Biochemical Engineering/Biotechnology*, T. Scheper, managing editor, Volume 65, pp.23-40, Springer-Verlag, Nova
15 Iorque). Entende-se aqui que a celulose pode estar na forma de lignocelulose, um material de parede celular de planta contendo lignina, celulose e hemicelulose em uma matriz mista. Em um aspecto preferido, o material celulolítico é lignocelulose, que compreende celulose, hemicelulose e lignina.

Em um aspecto, o material celulolítico é material herbáceo.
20 Em um outro aspecto, o material celulolítico é resíduo agrícola. Em um outro aspecto, o material celulolítico é resíduo florestal. Em um outro aspecto, o material celulolítico é resíduo sólido municipal. Em um outro aspecto, o material celulolítico é resíduo de papel. Em um outro aspecto, o material celulolítico é resíduo de polpa e papel moído.

25 Em um outro aspecto, o material celulolítico é palha de milho. Em um outro aspecto, o material celulolítico é fibra de milho. Em um outro aspecto, o material celulolítico é espiga de milho. Em um outro aspecto, o material celulolítico é casca de laranja. Em um outro aspecto, o material celulolítico é palha de arroz. Em um outro aspecto, o material celulolítico é

palha de trigo. Em um outro aspecto, o material celulolítico é painço amarelo. Em um outro aspecto, o material celulolítico é *Miscanthus*. Em um outro aspecto, o material celulolítico é bagaço.

Em um outro aspecto, o material celulolítico é celulose microcristalina. Em um outro aspecto, o material celulolítico é celulose bacteriana. Em um outro aspecto, o material celulolítico é celulose de alga. Em um outro aspecto, o material celulolítico é línter de algodão. Em um outro aspecto, o material celulolítico é celulose amorfa tratada com ácido fosfórico. Em um outro aspecto, o material celulolítico é papel de filtro.

O material celulolítico pode ser usado como é, ou pode ser submetido a pré-tratamento, usando métodos convencionais conhecidos na técnica, da maneira descrita aqui. Em um aspecto preferido, o material celulolítico é pré-tratado.

Palha de milho pré-tratada: O termo “PCS” ou “palha de milho pré-tratada” significa um material celulolítico derivado de palha de milho por tratamento com calor e ácido sulfúrico diluído.

Isolado ou purificado: O termo “isolado” ou “purificado” significa um polipeptídeo ou polinucleotídeo que é removido de pelo menos um componente com o qual é naturalmente associado. Por exemplo, um polipeptídeo pode ser pelo menos 1 % puro, por exemplo, pelo menos 5 % puro, pelo menos 10 % puro, pelo menos 20 % puro, pelo menos 40 % puro, pelo menos 60 % puro, pelo menos 80 % puro, pelo menos 90 % puro ou pelo menos 95 % puro, da maneira determinada por SDS-PAGE e um polinucleotídeo pode ser pelo menos 1 % puro, por exemplo, pelo menos 5 % puro, pelo menos 10 % puro, pelo menos 20 % puro, pelo menos 40 % puro, pelo menos 60 % puro, pelo menos 80 % puro, pelo menos 90 % puro, ou pelo menos 95 % puro, da maneira determinada por eletroforese em agarose.

Polipeptídeo maduro: O termo “polipeptídeo maduro” significa um polipeptídeo em sua forma final após a tradução e quaisquer

modificações pós-translacionais, tais como processamento N-terminal, truncação C-terminal, glicosilação, fosforilação, etc. Em um aspecto, o polipeptídeo maduro são os aminoácidos 18 a 493 de SEQ ID NO: 2 baseados no programa SignalP (Nielsen *et al.*, 1997, *Protein Engineering* 10: 1-6) que

5 prevê que aminoácidos 1 a 17 de SEQ ID NO: 2 são um peptídeo sinal. Em um outro aspecto, o polipeptídeo maduro são os aminoácidos 18 to 487 de SEQ ID NO: 4 com base no programa SignalP (Nielsen *et al.*, 1997, *supra*) que prevê que os aminoácidos 1 a 17 de SEQ ID NO: 4 são um peptídeo sinal. É conhecido na técnica que uma célula hospedeira pode produzir uma mistura

10 de dois dos polipeptídeos maduros mais diferentes (isto é, com um aminoácido C-terminal e/ou N-terminal diferente) expressa pelo mesmo polinucleotídeo.

Sequência que codifica polipeptídeo maduro: O termo “sequência que codifica polipeptídeo maduro” significa um polinucleotídeo

15 que codifica um polipeptídeo maduro com atividade celobiohidrolase. Em um aspecto, a sequência que codifica o polipeptídeo maduro são os nucleotídeos 52 a 1595 de SEQ ID NO: 1 baseados no programa SignalP (Nielsen *et al.*, 1997, *supra*) que prevê que os nucleotídeos 1 a 51 de SEQ ID NO: 1 codificam um peptídeo sinal. Em um outro aspecto, a sequência que codifica

20 o polipeptídeo maduro é a sequência de DNAC contida em nucleotídeos 52 a 1581 de SEQ ID NO: 3 com base no programa SignalP (Nielsen *et al.*, 1997, *supra*) que prevê nucleotídeos 1 a 51 de SEQ ID NO: 3 que codificam um peptídeo sinal. Em um outro aspecto, a sequência que codifica o polipeptídeo maduro é a sequência de DNAC contida em nucleotídeos 52 a 1595 de SEQ

25 ID NO: 1. Em um outro aspecto, a sequência que codifica o polipeptídeo maduro é a sequência de DNAC contida nos nucleotídeos 52 a 1581 de SEQ ID NO: 3.

Identidade de sequência: A relação entre duas sequências de aminoácidos ou entre duas sequências de nucleotídeos é descrita pelo

parâmetro “identidade de sequência”.

Com propósitos da presente invenção, o grau de identidade de sequência entre as duas sequências de aminoácidos é determinado usando o algoritmo de Needleman-Wunsch (Needleman e Wunsch, 1970, *J. Mol. Biol.* 48: 443-453) da maneira implementada no programa Needle do pacote EMBOSS (EMBOSS: The European Molecular Biology Open Software Suite, Rice *et al.*, 2000, *Trends Genet.* 16: 276-277), preferivelmente a versão 3.0.0 ou superior. Os parâmetros opcionais usados são penalidade de abertura de intervalo de 10, penalidade de extensão de intervalo de 0,5, e a matriz de substituição EBLOSUM62 (EMBOSS versão de BLOSUM62). O rendimento da “maior identidade” marcado em Needle (obtido usando a opção não resumida) é usado como a identidade percentual e é calculado da maneira a seguir:

(Resíduos idênticos x 100)/(Tamanho do alinhamento – Número total de intervalos em alinhamento)

Com propósitos da presente invenção, o grau de identidade de sequência entre duas sequências de desoxirribonucleotídeos é determinado usando o algoritmo de Needleman-Wunsch (Needleman e Wunsch, 1970, *supra*), da maneira implementada no programa Needle do pacote EMBOSS (EMBOSS: The European Molecular Biology Open Software Suite, Rice *et al.*, 2000, *supra*), preferivelmente a versão 3.0.0 ou superior. Os parâmetros opcionais usados são penalidade de abertura de intervalo de 10, penalidade de extensão de intervalo de 0,5, e a matriz de substituição EDNAFULL (EMBOSS versão de NCBI NUC4.4). O rendimento da “maior identidade” marcada por Needle (obtido usando a opção não resumida) é usado como a identidade percentual e é calculado da maneira a seguir:

(Desoxirribinucleotídeos idênticos x 100)/(Tamanho do alinhamento – Número total de intervalos em alinhamento)

Fragmento: O termo “fragmento” significa um polipeptídeo com um ou mais (vários) aminoácidos deletados da terminação amino e/ou carboxil de um polipeptídeo maduro; em que o fragmento apresenta atividade celobiohidrolase. Em um aspecto, um fragmento contém pelo menos 400
5 resíduos de aminoácido, mais preferivelmente pelo menos 425 resíduos de aminoácido, e acima de tudo preferivelmente pelo menos 450 resíduos de aminoácido, do polipeptídeo maduro de SEQ ID NO: 2. Em um outro aspecto, um fragmento contém pelo menos 400 resíduos de aminoácido, mais preferivelmente pelo menos 425 resíduos de aminoácido, e acima de tudo
10 preferivelmente pelo menos 450 resíduos de aminoácido, do polipeptídeo maduro de SEQ ID NO: 4.

Subsequência: O termo “subsequência” significa um polinucleotídeo com um ou mais (vários) nucleotídeos deletados da extremidade 5' e/ou 3' de uma sequência que codifica polipeptídeo maduro;
15 em que a subsequência codifica um fragmento com atividade celobiohidrolase. Em um aspecto, uma subsequência contém pelo menos 1200 nucleotídeos, mais preferivelmente pelo menos 1275 nucleotídeos e acima de tudo preferivelmente pelo menos 1350 nucleotídeos da sequência que codifica polipeptídeo maduro de SEQ ID NO: 1. Em um outro aspecto, uma
20 subsequência contém pelo menos 1200 nucleotídeos, mais preferivelmente pelo menos 1275 nucleotídeos, e acima de tudo preferivelmente pelo menos 1350 nucleotídeos da sequência codificante de polipeptídeo maduro de SEQ ID NO: 3.

Variante alélico: O termo “variante alélico” significa
25 qualquer de duas ou mais formas alternativas de um gene que ocupa o mesmo locus cromossômico. A variação alélica aumenta naturalmente por meio de mutação, e pode resultar em polimorfismo nas populações. As mutações de gene podem ser silenciosas (nenhuma mudança no polipeptídeo codificado) ou podem codificar polipeptídeos com sequências alteradas de aminoácidos.

Um variante alélico de um polipeptídeo é um polipeptídeo codificado por um variante alélico de um gene.

Sequência codificante: O termo “sequência codificante” significa um polinucleotídeo, que especifica diretamente a sequência de aminoácidos de um polipeptídeo. As extremidades da sequência codificante são em geral determinadas por um quadro aberto de leitura, que em geral começa com o códon inicial ATG ou códons iniciais alternativos, tais como GTG e TTG, e termina com um códon de parada, tais como TAA, TAG e TGA. A sequência codificante pode ser um DNA, DNAC, sintética, ou polinucleotídeo recombinante.

DNAC: O termo "DNAC" significa uma molécula de DNA que pode ser preparada por transcrição reversa de uma molécula RNAm madura, unida obtida de uma célula eucariótica. O DNAC precisa de sequências intron que podem estar presentes no DNA genômico correspondente. O transcrito de RNA inicial e primário é um precursor para o RNAm que é processado por meio de uma série de etapas, incluindo união, antes de aparecer como RNAm unido maduro.

Construto de ácido nucleico: O termo "construto de ácido nucleico" significa uma molécula de ácido nucleico, tanto de fita simples quanto de fita dupla, que é isolada de um gene de ocorrência natural ou é modificada para conter segmentos de ácidos nucleicos de uma maneira que não pode existir de outra forma na natureza ou que é sintética. O termo construto de ácido nucleico é sinônimo do termo “cassete de expressão” quando o construto de ácido nucleico contém as sequências controle exigidas para a expressão de uma sequência codificante da presente invenção.

Sequências controle: O termo “sequências controle” significa todos os componentes necessários para a expressão de um polinucleotídeo que codifica um polipeptídeo da presente invenção. Cada sequência controle pode ser natural ou estrangeira ao polinucleotídeo que codifica o polipeptídeo ou

natural ou estrangeira uma a outra. Tais sequências controle incluem, mas sem limitação, uma sequência líder, sequência de poliadenilação, sequência de pro-peptídeo, promotor, sequência de peptídeo sinal e finalizador de transcrição. No mínimo, as sequências controle incluem um promotor e sinais
5 de parada transcrpcionais e translacionais. As sequências controle podem ser fornecidas com ligadores com o propósito de introduzir sítios de restrição específicos que facilitam a ligação das sequências controle com a região codificante do polinucleotídeo que codifica um polipeptídeo.

Operavelmente ligado: O termo “operavelmente ligado”
10 significa uma configuração na qual uma sequência controle é colocada em uma posição adequada com relação à sequência codificante de um polinucleotídeo, de maneira tal que a sequência controle direcione a expressão da sequência codificante.

Expressão: O termo “expressão” inclui qualquer etapa
15 envolvida na produção do polipeptídeo incluindo, mas sem limitação, transcrição, modificação pós-transcricional, tradução, modificação pós-translacionais e secreção.

Vetor de expressão: O termo “vetor de expressão” significa
uma molécula de DNA linear ou circular que compreende um polinucleotídeo
20 que codifica um polipeptídeo, e é operavelmente ligada a nucleotídeos adicionais que fornecem sua expressão.

Célula hospedeira: O termo “célula hospedeira” significa
qualquer tipo de célula que é susceptível à transformação, transfecção,
transdução, e similares a um construto de ácido nucleico ou vetor de
25 expressão que compreende um polinucleotídeo da presente invenção. O termo “célula hospedeira” inclui qualquer progênie de uma célula mãe que não é idêntica à célula mãe em virtude das mutações que ocorrem durante a replicação.

Variante: O termo “variante” significa um polipeptídeo com

atividade celobiohidrolase que compreende uma alteração, isto é, uma substituição, inserção, e/ou deleção de um ou mais (vários) resíduos de aminoácido em uma ou mais (várias) posições. Uma substituição significa uma alteração de um aminoácido que ocupa uma posição com um aminoácido diferente; uma deleção significa remoção de um aminoácido que ocupa uma posição; e uma inserção significa adicionar um ou mais (vários) aminoácidos, por exemplo, 1-5 aminoácidos, adjacentes a um aminoácido que ocupa uma posição.

DESCRIÇÃO DETALHADA DA INVENÇÃO

10 **Polipectídeos Com Atividade de celobiohidrolase**

A presente invenção diz respeito a polipeptídeos isolados com atividade celobiohidrolase selecionados do grupo que consiste em:

- (a) um polipeptídeo com pelo menos 90 % de identidade de sequência com o polipeptídeo maduro de SEQ ID NO: 2 ou SEQ ID NO: 4;
- 15 (b) um polipeptídeo codificado por um polinucleotídeo que hibridiza em condições de severidade elevada (i) na sequência que codifica o polipeptídeo maduro de SEQ ID NO: 1 ou SEQ ID NO: 3, (ii) na sequência de DNAc contida na sequência que codifica o polipeptídeo maduro de SEQ ID NO: 1 ou SEQ ID NO: 3, ou (iii) na fita complementar de comprimento total
- 20 de (i) ou (ii);
- (c) um polipeptídeo codificado por um polinucleotídeo com pelo menos 90 % de identidade de sequência com a sequência que codifica o polipeptídeo maduro de SEQ ID NO: 1 ou SEQ ID NO: 3 ou a sequência de DNAc deste;
- 25 (d) um variante que compreende uma substituição, deleção, e/ou inserção de um ou mais (vários) aminoácidos do polipeptídeo maduro de SEQ ID NO: 2 ou SEQ ID NO: 4; e
- (e) um fragmento de um polipeptídeo de (a), (b), (c) ou (d) que apresenta atividade celobiohidrolase.

A presente invenção diz respeito a polipeptídeos isolados com uma identidade de sequência com o polipeptídeo maduro de SEQ ID NO: 2 ou SEQ ID NO: 4 de pelo menos 90 %, de pelo menos 91 %, de pelo menos 92 %, de pelo menos 93 %, de pelo menos 94 %, de pelo menos 95 %, pelo menos 96 %, pelo menos 97 %, pelo menos 98 %, pelo menos 99 %, ou 100 %, que apresentam atividade celobiohidrolase. Em um aspecto, os polipeptídeos diferem em não mais que dez aminoácidos, por exemplo, em five aminoácidos, em quatro aminoácidos, em três aminoácidos, em dois aminoácidos, e em um aminoácido do polipeptídeo maduro de SEQ ID NO: 2 ou SEQ ID NO: 4.

Um polipeptídeo da presente invenção compreende preferivelmente ou consiste na sequência de aminoácidos de SEQ ID NO: 2 ou um variante alélico deste; ou é um fragmento deste com atividade celobiohidrolase. Em um outro aspecto, o polipeptídeo compreende ou consiste no polipeptídeo maduro de SEQ ID NO: 2. Em um outro aspecto preferido, o polipeptídeo compreende ou consiste em aminoácidos 18 a 493 de SEQ ID NO: 2.

Um polipeptídeo da presente invenção também compreende ou consiste em preferivelmente a sequência de aminoácidos de SEQ ID NO: 4 ou um variante alélico desta; ou é um fragmento desta com atividade celobiohidrolase. Em um outro aspecto, o polipeptídeo compreende ou consiste em SEQ ID NO: 4. Em um outro aspecto, o polipeptídeo compreende ou consiste no polipeptídeo maduro de SEQ ID NO: 4. Em um outro aspecto preferido, o polipeptídeo compreende ou consiste em aminoácidos 18 a 487 de SEQ ID NO: 4.

A presente invenção também diz respeito a polipeptídeos isolados com atividade celobiohidrolase que são codificados por polinucleotídeos que hidrolisam em condições de severidade elevada ou em condições de severidade muito elevada com (i) a sequência que codifica o

polipeptídeo maduro de SEQ ID NO: 1 ou SEQ ID NO: 3, (ii) a sequência de DNAC contida na sequência que codifica o polipeptídeo maduro de SEQ ID NO: 1 ou SEQ ID NO: 3, ou (iii) a fita complementar de comprimento total de (i) ou (ii) (J. Sambrook, E.F. Fritsch, e T. Maniatis, 1989, *Molecular Cloning, A Laboratory Manual*, 2ª edição, Cold Spring Harbor, Nova Iorque).

O polinucleotídeo de SEQ ID NO: 1 ou SEQ ID NO: 3 ou uma subsequência deste, bem como a sequência de aminoácidos de SEQ ID NO: 2 ou SEQ ID NO: 4 ou um fragmento desta, pode ser usado para determinar sondas de ácido nucleico para identificar e clonar DNA que codifica polipeptídeos com atividade celobiohidrolase a partir de cepas de diferentes gêneros ou espécies, de acordo com métodos bem conhecidos na técnica. Em particular, tais sondas podem ser usadas para hibridização com o DNA genômico ou DNAC do gênero ou espécie de interesse, seguindo procedimentos padrões de Southern, a fim de identificar e isolar o gene correspondente neste. Tais sondas pode ser consideravelmente menores que a sequência completa, mas podem ter pelo menos 14, por exemplo, pelo menos 25, pelo menos 35, ou pelo menos 70 nucleotídeos em tamanho. Preferivelmente, a sonda de ácido nucleico tem pelo menos 100 nucleotídeos em tamanho, por exemplo, pelo menos 200 nucleotídeos, pelo menos 300 nucleotídeos, pelo menos 400 nucleotídeos, pelo menos 500 nucleotídeos, pelo menos 600 nucleotídeos, pelo menos 700 nucleotídeos, pelo menos 800 nucleotídeos, ou pelo menos 900 nucleotídeos em tamanho. Tanto as sondas de DNA quanto de RNA podem ser usadas. As sondas são tipicamente marcadas para detectar o gene correspondente (por exemplo, com ^{32}P , ^3H , ^{35}S , biotina ou avidina). Tais sondas são incluídas na presente invenção.

Uma biblioteca de DNA genômico ou DNAC preparada a partir de tais outras cepas pode ser selecionada para o DNA que hibridiza com as sondas descritas anteriormente e codifica um polipeptídeo com atividade celobiohidrolase. O DNA genômico ou outro, a partir de tais outras cepas,

pode ser separado por eletroforese em gel de agarose ou poliacrilamida, ou outras técnicas de separação. O DNA das bibliotecas ou o DNA separado pode ser transferido e imobilizado em nitrocelulose ou outro material carreador adequado. A fim de identificar um clone ou DNA que é homólogo à
5 SEQ ID NO: 1 ou SEQ ID NO: 3 ou uma subsequência desta, o material carreador é preferivelmente usado em um Southern blot.

Com propósitos da presente invenção, hibridização indica que o polinucleotídeo hibridiza em uma sonda de ácido nucleico marcada que corresponde à sequência que codifica o polipeptídeo maduro de SEQ ID NO:
10 1 ou SEQ ID NO: 3; a sequência de DNAC contida na sequência que codifica o polipeptídeo maduro de SEQ ID NO: 1 ou SEQ ID NO: 3; sua fita complementar de tamanho total; ou uma subsequência desta; em condições de severidade muito baixa a muito elevada. As moléculas nas quais a sonda de ácido nucleico hibridiza nestas condições podem ser detectadas usando, por
15 exemplo, filme de raios-X.

Em um aspecto, a sonda de ácido nucleico é a sequência que codifica o polipeptídeo maduro de SEQ ID NO: 1 ou SEQ ID NO: 3 ou a sequência de DNAC deste. Em um outro aspecto, a sonda de ácido nucleico é um polinucleotídeo que codifica o polipeptídeo de SEQ ID NO: 2 ou SEQ ID
20 NO: 4, ou um fragmento deste. Em um outro aspecto preferido, a sonda de ácido nucleico é SEQ ID NO: 1 ou SEQ ID NO: 3, ou a sequência de DNAC desta. Em um outro aspecto, a sonda de ácido nucleico é o polinucleotídeo contido no plasmídeo pGEM-T-CBHII45097 que está contido em *E. coli* DSM 22598, em que o polinucleotídeo codifica um polipeptídeo com
25 atividade celobiohidrolase. Em um outro aspecto, a sonda de ácido nucleico é o polipeptídeo maduro que codifica a região contida no plasmídeo pGEM-T-CBHII45097 que está contido em *E. coli* DSM 22598. Em um outro aspecto, a sonda de ácido nucleico é o polinucleotídeo contínuo no plasmídeo pGEM-T-CBHII45178, que está contido em *E. coli* DSM 22599, em que o

polinucleotídeo codifica um polipeptídeo com atividade celobiohidrolase. Em um outro aspecto, a sonda de ácido nucleico é a região codificante do polipeptídeo maduro contida no plasmídeo pGEM-T-CBHIII45178, que está contido em *E. coli* DSM 22599.

5 Para sondas longas de pelo menos 100 nucleotídeos em tamanho, as condições de severidade muito baixa a muito elevada são definidas como pré-hibridização e hibridização a 42 °C em SSPE 5X, SDS 0,3 %, 200 microgramas/mL de DNA cisalhado e desnaturado de espermatozoides de salmão, e tanto formamida a 25 % para severidades muito baixas e baixas, 10 formamida a 35 % para severidades médias a médias-altas, quanto formamida a 50 % para severidades elevadas e muito elevadas, seguindo procedimentos padrões de Southern blotting por 12 a 24 horas, de maneira ideal. O material carreador é finalmente lavado três vezes cada por 15 minutos usando SSC 2X, SDS 0,2 % a 45 °C (severidade muito baixa), a 50 °C (severidade baixa), a 55 15 °C (severidade média), a 60 °C (severidade média-alta), a 65 °C (severidade alta) e a 70 °C (severidade muito elevada).

 Para sondas curtas de cerca de 15 nucleotídeos a cerca de 70 nucleotídeos em tamanho, as condições de severidade são definidas como pré-hibridização e hibridização em cerca de 5 °C a cerca de 10 °C abaixo da T_m calculada usando o cálculo de acordo com Bolton e McCarthy (1962, *Proc. Natl. Acad. Sci. USA* 48:1390) em NaCl 0,9 M, Tris-HCl 0,09 M, pH 7,6, EDTA 6 mM, NP-40 0,5 %, solução de Denhardt 1X, pirofosfato de sódio 1 mM, fosfato monobásico de sódio 1 mM, ATP 0,1 mM, e 0,2 mg de RNA de levedura por mL, seguindo os procedimentos padrões de Southern blotting 20 por 12 a 24 horas de maneira ideal. O material carreador é finalmente lavado uma vez em SSC 6X e SDS 0,1 % por 15 minutos e duas vezes cada por 15 minutos usando SSC 6X a 5 °C a 10 °C abaixo da T_m calculada. 25

 A presente invenção também diz respeito a polipeptídeos isolados com atividade celobiohidrolase codificada por polinucleotídeos com

uma identidade de sequência com a sequência que codifica o polipeptídeo maduro de SEQ ID NO: 1 ou SEQ ID NO: 3, ou a sequência de DNAC deste de pelo menos 90 %, pelo menos 91 %, pelo menos 92 %, pelo menos 93 %, pelo menos 94 %, pelo menos 96 %, pelo menos 97 %, pelo menos 98 %, pelo menos 99 %, ou 100 %.

A presente invenção também diz respeito a variantes que compreendem uma substituição, deleção, e/ou inserção de um ou mais (ou vários) aminoácidos do polipeptídeo maduro de SEQ ID NO: 2 ou SEQ ID NO: 4, ou uma sequência homóloga deste. Preferivelmente, as alterações de aminoácido são de uma natureza inferior, isto é, substituições ou inserções conservativas de aminoácido que não afetam significativamente o dobramento e/ou atividade da proteína; pequenas deleções, tipicamente de um a cerca de 30 aminoácidos; pequenas extensões amino ou carboxil-terminais, tal como um resíduo de metionina amino-terminal; um pequeno peptídeo ligador de até cerca de 20-25 resíduos; ou uma pequena extensão que facilita a purificação mudando a carga líquida ou uma outra função, tal como uma região de polihistidina, um epítipo antigênico ou um domínio de ligação.

Exemplos de substituições conservativas estão no grupo de aminoácidos básicos (arginina, lisina e histidina), aminoácidos acídicos (ácido glutâmico e ácido aspártico), aminoácidos polares (glutamina e asparagina), aminoácidos hidrofóbicos (leucina, isoleucina e valina), aminoácidos aromáticos (fenilalanina, triptofano e tirosina) e aminoácidos pequenos (glicina, alanina, serina, treonina e metionina). As substituições de aminoácido que não alteram, em geral, a atividade específica são conhecidas na técnica e são descritas, por exemplo, por H. Neurath e R.L. Hill, 1979, em *The Proteins*, Academic Press, Nova Iorque. As mudanças que ocorrem mais comumente são Ala/Ser, Val/Ile, Asp/Glu, Thr/Ser, Ala/Gly, Ala/Thr, Ser/Asn, Ala/Val, Ser/Gly, Tyr/Phe, Ala/Pro, Lys/Arg, Asp/Asn, Leu/Ile, Leu/Val, Ala/Glu, e Asp/Gly.

Alternativamente, as mudanças de aminoácidos são de uma natureza tal que as propriedades físico-químicas dos polipeptídeos são alteradas. Por exemplo, as mudanças de aminoácido podem melhorar a estabilidade térmica do polipeptídeo, alteram a especificidade do substrato, mudam o pH ideal e similares.

Os aminoácidos essenciais em um polipeptídeo pai podem ser identificados de acordo com procedimentos conhecidos na técnica, tal como mutagênese sítio direcionada ou mutagênese de varredura de alanina (Cunningham e Wells, 1989, *Science* 244: 1081-1085). Na última técnica, as mutações de alanina simples são introduzidas e cada resíduo na molécula, e as moléculas mutantes resultantes são testadas com relação à atividade celobiohidrolase para identificar resíduos de aminoácido que são importantes para a atividade da molécula. Ver também, Hilton *et al.*, 1996, *J. Biol. Chem.* 271: 4699-4708. O sítio ativo da enzima ou outra interação biológica também pode ser determinada por análise física da estrutura, da maneira determinada por tais técnica como ressonância magnética nuclear, cristalografia, difração de elétron, ou marcação de foto afinidade, junto com mutação de aminoácidos de sítio de contato suposto. Ver, por exemplo, de Vos *et al.*, 1992, *Science* 255: 306-312; Smith *et al.*, 1992, *J. Mol. Biol.* 224: 899-904; Wlodaver *et al.*, 1992, *FEBS Lett.* 309: 59-64. As identidades dos aminoácidos essenciais também podem ser inferidas a partir da análise de identidades com polipeptídeos que estão relacionados com o polipeptídeo pai.

As substituições, deleções, e/ou inserções de aminoácido simples ou únicas podem ser realizadas e testadas usando métodos conhecidos de mutagênese, recombinação, e/ou embaralhamento, seguido por um procedimento de seleção relevante, tais como aqueles revelados por Reidhaar-Olson e Sauer, 1988, *Science* 241: 53-57; Bowie e Sauer, 1989, *Proc. Natl. Acad. Sci. USA* 86: 2152-2156; WO 95/17413; ou WO 95/22625. Outros métodos que podem ser usados incluem PCR propensa ao erro, exibição de

fago (por exemplo, Lowman *et al.*, 1991, *Biochemistry* 30: 10832-10837; patente U.S. 5.223.409; WO 92/06204), e mutagênese região-direcionada (Derbyshire *et al.*, 1986, *Gene* 46: 145; Ner *et al.*, 1988, *DNA* 7: 127).

Os métodos de mutagênese/embaralhamento podem ser
5 combinados com métodos de seleção automatizados de alto rendimento para detectar a atividade dos polipeptídeos clonados e mutagenizados expressos por células hospedeiras (Ness *et al.*, 1999, *Nature Biotechnology* 17: 893-896). As moléculas de DNA mutagenizado que codificam polipeptídeos
10 ativos podem ser recuperadas a partir das células hospedeiras e rapidamente sequenciadas usando métodos padrões na técnica. Estes métodos permitem a rápida determinação da importância de resíduos individuais de aminoácido em um polipeptídeo.

O número total de substituições, deleções e/ou inserções de aminoácido do polipeptídeo maduro de SEQ ID NO: 2 ou SEQ ID NO: 4 não
15 é mais que 10, por exemplo, 1, 2, 3, 4, 5, 6, 7, 8 ou 9.

O polipeptídeo pode ser polipeptídeo híbrido, no qual uma porção de um polipeptídeo é fundida na terminação N ou na terminação C de uma porção de um outro polipeptídeo.

O polipeptídeo pode ser um polipeptídeo fundido, ou
20 polipeptídeo de fusão clivável, em que um outro polipeptídeo é fundido na terminação N ou na terminação C do polipeptídeo da presente invenção. Um polipeptídeo fundido é produzido fundindo um polinucleotídeo que codifica um outro polipeptídeo a um polinucleotídeo da presente invenção. As técnicas para produzir polipeptídeos de fusão são conhecidas na tecnologia, e incluem
25 ligar as sequências codificantes que codificam os polipeptídeos, de maneira tal que eles estejam em alinhamento e que a expressão do polipeptídeo fundido esteja no controle do(s) mesmo(s) promotor(s) e finalizador(s). As proteínas de fusão também podem ser construídas usando tecnologia de inteína na qual as fusões são criadas pós-translacionalmente (Cooper *et al.*,

1993, *EMBO J.* 12: 2575-2583; Dawson *et al.*, 1994, *Science* 266: 776-779).

Um polipeptídeo de fusão podem compreender adicionalmente um sítio de clivagem entre os dois polipeptídeos. Mediante a secreção da proteína de fusão, o sítio é clivado liberando os dois polipeptídeos. Exemplos de sítios de clivagem incluem, mas sem limitação, o sítios revelados em Martin *et al.*, 2003, *J. Ind. Microbiol. Biotechnol.* 3: 568-576; Svetina *et al.*, 2000, *J. Biotechnol.* 76: 245-251; Rasmussen-Wilson *et al.*, 1997, *Appl. Environ. Microbiol.* 63: 3488-3493; Ward *et al.*, 1995, *Biotechnology* 13: 498-503; e Contreras *et al.*, 1991, *Biotechnology* 9: 378-381; Eaton *et al.*, 1986, *Biochemistry* 25: 505-512; Collins-Racie *et al.*, 1995, *Biotechnology* 13: 982-987; Carter *et al.*, 1989, *Proteins: Structure, Function, e Genetics* 6: 240-248; e Stevens, 2003, *Drug Discovery World* 4: 35-48.

Fontes de polipeptídeos com atividade celobiohidrolase

Um polipeptídeo com atividade celobiohidrolase da presente invenção pode ser obtido de microrganismos de qualquer gênero. Com propósitos da presente invenção, o termo “obtido de”, da maneira aqui usada junto com uma dada fonte, pode significar que o polipeptídeo codificado por um polinucleotídeo é produzido pela fonte ou por uma cepa na qual o polinucleotídeo da fonte foi inserido. Em um aspecto, o polipeptídeo obtido de uma dada fonte é secretado extracelularmente.

O polipeptídeo pode ser um polipeptídeo bacteriano. Por exemplo, o polipeptídeo pode ser um polipeptídeo de bactéria Gram-positiva, tal como um polipeptídeo com atividade celobiohidrolase de *Bacillus*, *Clostridium*, *Enterococcus*, *Geobacillus*, *Lactobacillus*, *Lactococcus*, *Oceanobacillus*, *Staphylococcus*, *Streptococcus*, ou *Streptomyces*, ou um polipeptídeo de bactéria Gram-negativa, tal como um polipeptídeo de *Campylobacter*, *E. coli*, *Flavobacterium*, *Fusobacterium*, *Helicobacter*, *Ilyobacter*, *Neisseria*, *Pseudomonas*, *Salmonella*, ou *Ureaplasma*.

Em um aspecto, o polipeptídeo é um polipeptídeo de *Bacillus*

alkalophilus, *Bacillus amyloliquefaciens*, *Bacillus brevis*, *Bacillus circulans*, *Bacillus clausii*, *Bacillus coagulans*, *Bacillus firmus*, *Bacillus lautus*, *Bacillus lentus*, *Bacillus licheniformis*, *Bacillus megaterium*, *Bacillus pumilus*, *Bacillus stearothermophilus*, *Bacillus subtilis*, ou *Bacillus thuringiensis*.

- 5 Em um outro aspecto, o polipeptídeo é um polipeptídeo de *Streptococcus equisimilis*, *Streptococcus pyogenes*, *Streptococcus uberis*, ou *Streptococcus equi* subsp. *Zooepidemicus*.

- Em um outro aspecto, o polipeptídeo é um polipeptídeo de *Streptomyces achromogenes*, *Streptomyces avermitilis*, *Streptomyces coelicolor*, *Streptomyces griseus*, ou *Streptomyces lividans*.
- 10

- O polipeptídeo também pode ser um polipeptídeo fúngico. Por exemplo, o polipeptídeo pode ser um polipeptídeo de levedura, tal como um polipeptídeo de *Candida*, *Kluyveromyces*, *Pichia*, *Saccharomyces*, *Schizosaccharomyces*, ou *Yarrowia*; ou um polipeptídeo de fungo
- 15 filamentoso, tal como um polipeptídeo de *Acremonium*, *Agaricus*, *Alternaria*, *Aspergillus*, *Aureobasidium*, *Botryosphaeria*, *Ceriporiopsis*, *Chaetomidium*, *Chrysosporium*, *Claviceps*, *Cochliobolus*, *Coprinopsis*, *Coptotermes*, *Corynascus*, *Cryphonectria*, *Cryptococcus*, *Diplodia*, *Exidia*, *Filibasidium*, *Fusarium*, *Gibberella*, *Holomastigotoides*, *Humicola*, *Irpex*, *Lentinula*,
- 20 *Leptosphaeria*, *Magnaporthe*, *Melanocarpus*, *Meripilus*, *Mucor*, *Myceliophthora*, *Neocallimastix*, *Neurospora*, *Paecilomyces*, *Penicillium*, *Phanerochaete*, *Piromyces*, *Poitrasia*, *Pseudoplectania*, *Pseudotrichonympha*, *Rhizomucor*, *Schizophyllum*, *Scytalidium*, *Talaromyces*, *Thermoascus*, *Thielavia*, *Tolypocladium*, *Trichoderma*, *Trichophaea*,
- 25 *Verticillium*, *Volvariella*, ou *Xylaria*.

 Em um outro aspecto, o polipeptídeo é um polipeptídeo de *Saccharomyces carlsbergensis*, *Saccharomyces cerevisiae*, *Saccharomyces diastaticus*, *Saccharomyces douglasii*, *Saccharomyces kluyveri*, *Saccharomyces norbensis*, ou *Saccharomyces oviformis*.

Em um outro aspecto, o polipeptídeo é um polipeptídeo de *Acremonium cellulolyticus*, *Aspergillus aculeatus*, *Aspergillus awamori*, *Aspergillus foetidus*, *Aspergillus fumigatus*, *Aspergillus japonicus*, *Aspergillus nidulans*, *Aspergillus niger*, *Aspergillus oryzae*, *Chrysosporium*
 5 *inops*, *Chrysosporium keratinophilum*, *Chrysosporium lucknowense*, *Chrysosporium merdarium*, *Chrysosporium pannicola*, *Chrysosporium queenslandicum*, *Chrysosporium tropicum*, *Chrysosporium zonatum*, *Fusarium bactridioides*, *Fusarium cerealis*, *Fusarium crookwellense*, *Fusarium culmorum*, *Fusarium graminearum*, *Fusarium graminum*,
 10 *Fusarium heterosporum*, *Fusarium negundi*, *Fusarium oxysporum*, *Fusarium reticulatum*, *Fusarium roseum*, *Fusarium sambucinum*, *Fusarium sarcochroum*, *Fusarium sporotrichioides*, *Fusarium sulphureum*, *Fusarium torulosum*, *Fusarium trichothecioides*, *Fusarium venenatum*, *Humicola grisea*, *Humicola insolens*, *Humicola lanuginosa*, *Irpex lacteus*, *Mucor*
 15 *miehei*, *Myceliophthora thermophila*, *Neurospora crassa*, *Penicillium funiculosum*, *Penicillium purpurogenum*, *Phanerochaete chrysosporium*, *Thielavia achromatica*, *Thielavia albomyces*, *Thielavia albopilosa*, *Thielavia australeinsis*, *Thielavia fimeti*, *Thielavia microspora*, *Thielavia ovispora*, *Thielavia peruviana*, *Thielavia setosa*, *Thielavia spededonium*, *Thielavia*
 20 *subthermophila*, *Thielavia terrestris*, *Trichoderma harzianum*, *Trichoderma koningii*, *Trichoderma longibrachiatum*, *Trichoderma reesei*, ou *Trichoderma viride*

Em um outro aspecto, o polipeptídeo é um polipeptídeo de *Thielavia hyrcaniae* com atividade celobiohidrolase, por exemplo, um
 25 polipeptídeo obtido de *Thielavia hyrcaniae* NN045097 ou *Thielavia hyrcaniae* NN045178 (CGMCC 0864).

Será entendido que, para as espécies anteriormente mencionadas, a invenção inclui tanto os estágios perfeitos quanto os imperfeitos, e outros equivalentes taxonômicos, por exemplo, anamorfos,

sem levar em consideração o nome da espécie pelo qual são conhecidas. Os versados na técnica reconhecerão facilmente a identidade dos equivalentes adequados.

As cepas destas espécies são facilmente acessíveis ao público em inúmeras coleções de cultura, tais como o American Type Culture Collection (ATCC), Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH (DSMZ), Centraalbureau Voor Schimmelcultures (CBS), e Agricultural Research Service Patent Culture Collection, Northern Regional Research Center (NRRL).

O polipeptídeo pode ser identificado e obtido de outras fontes, incluindo microrganismos isolados da natureza (por exemplo, soli, compostos, água, etc.), usando as sondas anteriormente mencionadas. As técnicas para isolar microrganismos de habitats naturais são bem conhecidas na tecnologia. O polinucleotídeo que codifica o polipeptídeo pode então ser obtido selecionando similarmente uma biblioteca de DNA genômico ou DNAC de um outro microrganismo ou amostra mista de DNA. Uma vez que um polinucleotídeo que codifica um polipeptídeo foi detectado com a(s) sonda(s), o polinucleotídeo pode ser isolado ou clonado utilizando técnicas que são bem conhecidas pelos versados na técnica (ver, por exemplo, Sambrook *et al.*, 1989, *supra*).

Polinucleotídeos

A presente invenção também diz respeito a polinucleotídeos isolados que codificam um polipeptídeo da presente invenção.

As técnicas usadas para isolar ou clonar um polinucleotídeo que codifica um polipeptídeo são conhecidas na tecnologia e incluem o isolamento do DNA genômico, preparação do DNAC, ou uma combinação destes. A clonagem dos polinucleotídeos a partir de tal DNA genômico pode ser realizada, por exemplo, usando a bem conhecida reação em cadeia da polimerase (PCR) ou seleção de anticorpo de bibliotecas de expressão para

detectar fragmentos de DNA clonado com características estruturas compartilhadas. Ver, por exemplo, Innis *et al.*, 1990, *PCR: A Guide to Methods and Application*, Academic Press, Nova Iorque. Outros procedimentos de amplificação de ácido nucleico, tais como reação de ligação da ligase (LCR), transcrição ativada de ligação (LAT) e amplificação a base de nucleotídeos (NASBA) podem ser usados. Os polinucleotídeos podem ser clonados a partir de uma cepa de *Thielavia*, ou um outro organismo relacionado e assim, por exemplo, pode ser um alélico ou espécie variante do polipeptídeo que codifica a região do polinucleotídeo.

10 A presente invenção também diz respeito a polinucleotídeos isolados que compreendem ou que consistem em polinucleotídeos com um grau de identidade de sequência com a sequência que codifica o polipeptídeo maduro de SEQ ID NO: 1 ou SEQ ID NO: 3, ou a sequência de DNAC deste de pelo menos 90 %, de pelo menos 91 %, de pelo menos 92 %, de pelo menos 93 %, de pelo menos 94 %, de pelo menos 95 %, pelo menos 96 %, 15 pelo menos 97 %, pelo menos 98 %, pelo menos 99 %, ou 100 %, que codifica um polipeptídeo com atividade celobiohidrolase.

 A modificação de um polinucleotídeo que codifica um polipeptídeo da presente invenção pode ser necessária para a síntese de polipeptídeos substancialmente similar ao polipeptídeo. O termo 20 “substancialmente similar” ao polipeptídeo refere-se a formas de ocorrência não natural do polipeptídeo. Estes polipeptídeos podem diferir, de alguma maneira modificada por engenharia, do polipeptídeo isolado de sua fonte natural, por exemplo, variantes que diferem em atividade específica, termoestabilidade, pH ideal ou similares. O variante pode ser construído com 25 base no polinucleotídeo apresentado como a sequência que codifica o polipeptídeo maduro de SEQ ID NO: 1 ou SEQ ID NO: 3, ou na sequência de DNAC deste, por exemplo, uma subsequência deste, e/ou por introdução de substituições de nucleotídeo que não resultam em uma mudança na sequência

de aminoácidos do polipeptídeo, mas que correspondem ao uso do códon do organismo hospedeiro pretendido para a produção da enzima, ou por introdução de substituições de nucleotídeo que podem resultar em uma sequência de aminoácidos diferente. Para uma descrição geral de substituição de nucleotídeo, ver, por exemplo, Ford *et al.*, 1991, *Protein Expression and Purification* 2: 95-107.

A presente invenção também diz respeito a polinucleotídeos isolados que codificam polipeptídeos da presente invenção, que hibridizam em condições de severidade elevada ou condições de severidade muito elevada (i) na sequência que codifica o polipeptídeo maduro de SEQ ID NO: 1 ou SEQ ID NO: 3, (ii) na sequência de DNAc contida na sequência que codifica o polipeptídeo maduro de SEQ ID NO: 1 ou SEQ ID NO: 3, ou (iii) na fita complementar de comprimento total de (i) ou (ii); ou variantes alélicas e subsequências destes (Sambrook *et al.*, 1989, *supra*), da maneira aqui definida.

Em um aspecto, o polinucleotídeo compreende ou consiste em SEQ ID NO: 1, a sequência que codifica polipeptídeo maduro de SEQ ID NO: 1 ou a sequência contida no plasmídeo pGEM-T-CBHII45097 que está contido em *E. coli* DSM 22598, ou uma subsequência de SEQ ID NO: 1 que codifica um fragmento de SEQ ID NO: 2 com atividade celobiohidrolase, tal como o polinucleotídeo de nucleotídeos 52 a 1595 de SEQ ID NO: 1. Em um outro aspecto, o polinucleotídeo compreende ou consiste em SEQ ID NO: 3, a sequência que codifica polipeptídeo maduro de SEQ ID NO: 3, ou a sequência contida no plasmídeo pGEM-T-CBHII45178, que está contido em *E. coli* DSM 22599, ou uma subsequência de SEQ ID NO: 3 que codifica um fragmento de SEQ ID NO: 4 com atividade celobiohidrolase, tal como o polinucleotídeo de nucleotídeos 52 a 1581 de SEQ ID NO: 3.

Construtos de ácido nucleico

A presente invenção também diz respeito a construtos de ácido

nucleico que compreende um polinucleotídeo da presente invenção, operavelmente ligado a um ou mais (vários) sequências controle, que direciona a expressão da sequência codificante em uma célula hospedeira adequada em condições compatíveis com as sequências controle.

5 Um polinucleotídeo pode ser manipulado em uma variedade de maneiras para fornecer a expressão do polipeptídeo. A manipulação do polinucleotídeo antes de sua inserção em um vetor pode ser desejável ou necessária dependendo do vetor de expressão. As técnicas para modificar polinucleotídeos que utilizam métodos de DNA recombinante são bem
10 conhecidas na tecnologia.

A sequência controle pode ser uma sequência promotora, um polinucleotídeo que é reconhecido por uma célula hospedeira para a expressão de um polinucleotídeo que codifica um polipeptídeo da presente invenção. A sequência promotora contém sequências controle transcricionais que medeiam
15 a expressão do polipeptídeo. O promotor pode ser qualquer polinucleotídeo que mostra a atividade transcricional na célula hospedeira de escolha, incluindo promotores mutantes, truncados e híbridos, e pode ser obtido de genes que codificam polipeptídeos extracelulares ou intracelulares tanto homólogos quanto heterólogos à célula hospedeira.

20 Exemplos de promotores adequados para direcionar a transcrição dos construtos de ácido nucleico da presente invenção em uma célula hospedeira bacteriana são os promotores obtidos do gene alfa-amilase de *Bacillus amyloliquefaciens* (*amyQ*), gene alfa-amilase de *Bacillus licheniformis* (*amyL*), gene penicilinase de *Bacillus licheniformis* (*penP*),
25 gene amilase maltogênica de *Bacillus stearothermophilus* (*amyM*), gene levansucrase de *Bacillus subtilis* (*sacB*), genes *xylA* e *xylB* de *Bacillus subtilis*, operon *lac* de *E. coli*, gene agarase de *Streptomyces coelicolor* (*dagA*) e gene beta-lactamase de procarioto (Villa-Kamaroff *et al.*, 1978, *Proc. Natl. Acad. Sci. USA* 75: 3727-3731), bem como o promotor *tac* (DeBoer *et al.*,

1983, *Proc. Natl. Acad. Sci. USA* 80: 21-25). Os promotores adicionais são descritos em "Useful proteins from recombinant bacteria" em Gilbert *et al.*, 1980, *Scientific American*, 242: 74-94; e em Sambrook *et al.*, 1989, *supra*.

Exemplos de promotores adequados para direcionar a transcrição do construtos de ácido nucleico da presente invenção em células hospedeiras de fungo filamentoso são promotores obtidos dos genes para acetamidase de *Aspergillus nidulans*, alfa-amilase neutra de *Aspergillus niger*, alfa-amilase ácida estável de *Aspergillus niger*, glucoamilase de *Aspergillus niger* ou *Aspergillus awamori* (*glaA*), TAKA amilase de *Aspergillus oryzae*, protease alcalina de *Aspergillus oryzae*, triose fosfato isomerase de *Aspergillus oryzae*, protease do tipo tripsina de *Fusarium oxysporum* (WO 96/00787), amiloglicosidase de *Fusarium venenatum* (WO 00/56900), Daria de *Fusarium venenatum* (WO 00/56900), Quinn de *Fusarium venenatum* (WO 00/56900), lipase de *Rhizomucor miehei*, proteinase aspártica de *Rhizomucor miehei*, beta-glicosidase de *Trichoderma reesei*, celobiohidrolase I de *Trichoderma reesei*, celobiohidrolase II de *Trichoderma reesei*, endoglicanase I de *Trichoderma reesei*, endoglicanase II de *Trichoderma reesei*, endoglicanase III de *Trichoderma reesei*, endoglicanase IV de *Trichoderma reesei*, endoglicanase V de *Trichoderma reesei*, xilanase I de *Trichoderma reesei*, xilanase II de *Trichoderma reesei*, beta-xilosidase de *Trichoderma reesei*, bem como o promotor NA2-tpi (um promotor modificado a partir de um gene que codifica uma alfa-amilase neutra em *Aspergilli*, no qual o líder não traduzido é substituído por um líder não traduzido de um gene que codifica triose fosfato isomerase em *Aspergilli*; exemplos não limitantes incluem promotores modificados a partir do gene que codifica alfa-amilase neutra em *Aspergillus niger*, em que o líder não traduzido é substituído por um líder não traduzido a partir do gene que codifica triose fosfato isomerase em *Aspergillus nidulans* ou *Aspergillus oryzae*); e promotores mutantes, truncados e híbridos deste.

Em uma levedura hospedeira, os promotores usados são obtidos dos genes para enolase de *Saccharomyces cerevisiae* (ENO-1), galactokinase de *Saccharomyces cerevisiae* (GAL1), álcool desidrogenase/gliceraldeído-3-fosfato desidrogenase de *Saccharomyces cerevisiae* (ADH1, ADH2/GAP), triose fosfato isomerase de *Saccharomyces cerevisiae* (TPI), metalotioneína de *Saccharomyces cerevisiae* (CUP1) e 3-fosfoglicerato quinase de *Saccharomyces cerevisiae*. Outros promotores usados para leveduras como células hospedeiras são descritos por Romanos *et al.*, 1992, *Yeast* 8: 423-488.

10 A sequência controle também pode ser uma sequência finalizadora de transcrição adequada, que é reconhecida por uma célula hospedeira para finalizar a transcrição. A sequência finalizadora é operavelmente ligado à terminação 3' do polinucleotídeo que codifica o polipeptídeo. Qualquer finalizador que é funcional na célula hospedeira de
15 escolha pode ser usado na presente invenção.

Os finalizadores preferidos para células hospedeiras de fungo filamentoso são obtidos dos genes para antranilato sintase de *Aspergillus nidulans*, glucoamilase de *Aspergillus niger*, alfa-glicosidase de *Aspergillus niger*, TAKA amilase de *Aspergillus oryzae* e protease do tipo tripsina de
20 *Fusarium oxysporum*.

Os finalizadores preferidos para levedura como células hospedeiras são obtidos dos genes para enolase de *Saccharomyces cerevisiae*, citocromo C de *Saccharomyces cerevisiae* (CYC1) e gliceraldeído-3-fosfato desidrogenase de *Saccharomyces cerevisiae*. Outros finalizadores usados para
25 levedura como células hospedeiras são descritos por Romanos *et al.*, 1992, *supra*.

A sequência controle também pode ser uma sequência líder adequada, quando transcrita é uma região não traduzida de um RNAm que é importante para a tradução pela célula hospedeira. A sequência líder é

operavelmente ligada à terminação 5' do polinucleotídeo que codifica o polipeptídeo. Qualquer sequência líder que é funcional na célula hospedeira de escolha pode ser usada.

Os líderes preferidos para as células hospedeiras de fungo filamentoso são obtidos dos genes para TAKA amilase de *Aspergillus oryzae* e triose fosfato isomerase de *Aspergillus nidulans*.

Os líderes adequados para levedura como células hospedeiras são obtidos dos genes para enolase de *Saccharomyces cerevisiae* (ENO-1), 3-fosfoglicerato quinase de *Saccharomyces cerevisiae*, fator alfa de *Saccharomyces cerevisiae* e álcool desidrogenase/gliceraldeído-3-fosfato desidrogenase de *Saccharomyces cerevisiae* (ADH2/GAP).

A sequência controle também pode ser uma sequência de poliadenilação, uma sequência operavelmente ligada à terminação 3' do polinucleotídeo e, quando transcrita, é reconhecida pela célula hospedeira como um sinal para adicionar resíduos de poliadenosina ao RNAm transcrito. Qualquer sequência de poliadenilação que é funcional na célula hospedeira de escolha pode ser usada.

As sequências de poliadenilação preferidas para as células hospedeiras de fungo filamentoso são obtidas dos genes para TAKA amilase de *Aspergillus oryzae*, glucoamilase de *Aspergillus niger*, antranilato sintase de *Aspergillus nidulans*, protease do tipo tripsina de *Fusarium oxysporum* e alfa-glicosidase de *Aspergillus niger*.

As sequências de poliadenilação usadas para levedura como células hospedeiras são descritas por Guo e Sherman, 1995, *Mol. Cellular Biol.* 15: 5983-5990.

A sequência controle também pode ser uma região que codifica um peptídeo sinal, que codifica um peptídeo sinal ligado à terminação N de um polipeptídeo, e direciona o polipeptídeo na via de secreção da célula. A extremidade 5' da sequência codificante do

polinucleotídeo pode conter intrinsecamente uma sequência de peptídeo sinal codificante naturalmente ligada no quadro aberto de leitura, com o segmento da sequência codificante que codifica o polipeptídeo. Alternativamente, a extremidade 5' da sequência codificante pode conter uma sequência de peptídeo sinal codificante que é estrangeira à sequência codificante. A sequência estrangeira de peptídeo sinal codificante pode ser exigida onde a sequência codificante não contém naturalmente uma sequência de peptídeo sinal codificante. Alternativamente, a sequência estrangeira de peptídeo sinal codificante pode simplesmente substituir a sequência natural de peptídeo sinal codificante a fim de melhorar a secreção do polipeptídeo. Entretanto, qualquer sequência de peptídeo sinal codificante que direciona o polipeptídeo expresso na via secretória de uma célula hospedeira de escolha pode ser usada.

As sequências de peptídeos sinais codificantes eficientes para células hospedeiras bacterianas são as sequências de peptídeos sinais codificantes obtidas dos genes para NCIB 11837 amilase maltogênica de *Bacillus*, subtilisina de *Bacillus licheniformis*, beta-lactamase de *Bacillus licheniformis*, alfa-amilase de *Bacillus stearothermophilus*, proteases neutras de *Bacillus stearothermophilus* (*nprT*, *nprS*, *nprM*) e *prsA* de *Bacillus subtilis*. Os peptídeos sinais adicionais são descritos por Simonen e Palva, 1993, *Microbiological Reviews* 57: 109-137.

As sequências de peptídeos sinais codificantes eficientes para células hospedeiras de fungo filamentoso são as sequências de peptídeos sinais codificantes obtidas dos genes para amilase neutra de *Aspergillus niger*, glucoamilase de *Aspergillus niger*, TAKA amilase de *Aspergillus oryzae*, celulase de *Humicola insolens*, endoglicanase V de *Humicola insolens*, lipase de *Humicola lanuginosa* e proteinase aspártica de *Rhizomucor miehei*.

Os peptídeos sinais usados para levedura como células hospedeiras são obtidos dos genes para fator alfa de *Saccharomyces cerevisiae* e invertase de *Saccharomyces cerevisiae*. Outras sequências de

peptídeos sinais codificantes usadas são descritas por Romanos *et al.*, 1992, *supra*.

A sequência controle também pode ser uma sequência de pro-peptídeo codificante que codifica um pro-peptídeo posicionado na terminação N de um polipeptídeo. O polipeptídeo resultante é conhecido como uma pro-enzima ou pro-polipeptídeo (ou um zimogênio em alguns casos). Um pro-polipeptídeo é em geral inativo e pode ser convertido em um polipeptídeo ativo por clivagem catalítica ou autocatalítica do pro-peptídeo a partir do pro-polipeptídeo. A sequência de pro-peptídeo codificante pode ser obtida dos genes para protease alcalina de *Bacillus subtilis* (*aprE*), protease neutra de *Bacillus subtilis* (*nprT*), lacase de *Myceliophthora thermophila* (WO 95/33836), proteinase aspártica de *Rhizomucor miehei* e fator alfa de *Saccharomyces cerevisiae*.

Onde tanto o peptídeo sinal quanto a sequência de pro-peptídeos estão presentes na terminação N de um polipeptídeo, a sequência de pro-peptídeo está posicionada próxima à terminação N de um polipeptídeo, e a sequência de peptídeo sinal está posicionada próxima à terminação N da sequência de pro-peptídeo.

Também pode ser desejável adicionar sequências regulatórias que permitem a regulação da expressão do polipeptídeo com relação ao crescimento da célula hospedeira. Exemplos de sistemas regulatórios são aqueles que fazem com que a expressão do gene seja ligada e desligada em resposta a um estímulo químico ou físico, incluindo a presença de um composto regulatório. Os sistemas regulatórios em sistemas procariotos incluem os sistemas operadores *lac*, *tac* e *trp*. Em levedura, o sistema ADH2 ou sistema GAL1 podem ser usados. Em fungos filamentosos, o promotor de glucoamilase de *Aspergillus niger*, promotor de TAKA alfa-amilase de *Aspergillus oryzae* e promotor de glucoamilase de *Aspergillus oryzae* podem ser usados. Outros exemplos de sequências regulatórias são aquelas que

permitem a amplificação de gene. Em sistemas eucarióticos, estas sequências regulatórias incluem o gene diidrofolato redutase que é amplificado na presença de metotrexato, e o genes de metalotioneína que são amplificados com metais pesados. Nestes casos, o polinucleotídeo que codifica o polipeptídeo pode ser operavelmente ligado com a sequência regulatória.

Vetores de expressão

A presente invenção também diz respeito a vetores de expressão recombinantes que compreendem um polinucleotídeo da presente invenção, um promotor, e sinais de parada transcrpcionais e translacionais. Os vários nucleotídeos e sequências controle podem ser reunidos para produzir um vetor de expressão recombinante que pode incluir um ou mais (vários) sítios de restrição convenientes para permitir a inserção ou substituição do polinucleotídeo que codifica o polipeptídeo em tais sítios. Alternativamente, o polinucleotídeo pode ser expresso inserindo o polinucleotídeo, ou um construto de ácido nucleico que compreende a sequência, em um vetor apropriado para a expressão. Na criação do vetor de expressão, a sequência codificante está localizada no vetor de maneira tal que a sequência codificante seja operavelmente ligada com as sequências controle adequadas para a expressão.

O vetor de expressão recombinante pode ser qualquer vetor (por exemplo, um plasmídeo ou vírus) que pode ser convenientemente submetido aos procedimentos de DNA recombinante e pode resultar na expressão do polinucleotídeo. A escolha do vetor dependerá tipicamente da compatibilidade do vetor com a célula hospedeira na qual o vetor será introduzido. O vetor pode ser um plasmídeo linear ou circular fechado.

O vetor pode ser um vetor que se replica autonomamente, isto é, um vetor que existe como uma entidade extracromossômica, cuja a replicação é independente da replicação cromossômica, por exemplo, um plasmídeo, um elemento extracromossômico, um minicromossomo, ou um

cromossomo artificial. O vetor pode conter qualquer meio para garantir a auto-replicação. Alternativamente, o vetor pode ser um que, quando introduzido na célula hospedeira é integrado no genoma e replicado junto com o(s) cromossomo(s) nos(s) qual(s) foi integrado. Além do mais, um vetor
 5 simples, ou plasmídeo, ou dois ou mais vetores ou plasmídeos que juntos contêm o DNA total a ser introduzido no genoma da célula hospedeira, ou um transposon, podem ser usados.

O vetor contém preferivelmente um ou mais (vários) marcadores selecionáveis que permitem a fácil seleção de células
 10 transformadas, transitadas, transduzidas ou similares. Um marcador selecionável é um produto do gene que fornece resistência biocida ou viral, resistência a metais pesados, prototrofia a auxotófos e similares.

Exemplos de marcadores selecionáveis bacterianos são os genes *dal* de *Bacillus subtilis* ou *Bacillus licheniformis*, ou marcadores que
 15 conferem resistência a antibióticos, tal como resistência à ampicilina, cloranfenicol, canamicina ou tetraciclina. Os marcadores adequados para levedura como células hospedeiras são ADE2, HIS3, LEU2, LYS2, MET3, TRP1, e URA3. Os marcadores selecionáveis para uso em células hospedeiras de fungo filamentoso incluem, mas sem limitação, *amdS* (acetamidase), *argB*
 20 (ornitina carbamotransferase), *bar* (fosfinotricina acetiltransferase), *hph* (higromicina fosfotransferase), *niaD* (nitrato redutase), *pyrG* (orotidina-5'-fosfato decarboxilase), *sC* (sulfato adeniltransferase), e *trpC* (antranilato sintase), bem como equivalentes destes. São preferidos para uso em uma célula de *Aspergillus* os genes *amdS* e *pyrG* de *Aspergillus nidulans* ou
 25 *Aspergillus oryzae* e o gene *bar* de *Streptomyces hygroscopicus*.

O vetor contém preferivelmente um(s) elemento(s) que permite(s) a integração do vetor no genoma da célula hospedeira ou a replicação autônoma do vetor na célula independente do genoma.

Para a integração no genoma da célula hospedeira, o vetor

pode depender da sequência de polinucleotídeo que codifica o polipeptídeo, ou qualquer outro elemento do vetor para a integração no genoma por recombinação homóloga ou não homóloga. Alternativamente, o vetor pode conter polinucleotídeos adicionais para direcionar a integração por recombinação homóloga no genoma da célula hospedeira em um(s) local(s) exato(s) no(s) cromossomo(s). Para aumentar a probabilidade de integração em um local preciso, os elementos integracionais podem conter um número suficiente de ácidos nucleicos, tais como 100 a 10.000 pares de bases, 400 a 10.000 pares de bases e 800 a 10.000 pares de bases, que apresenta um alto grau de identidade de sequência com a sequência alvo correspondente para melhorar a probabilidade de recombinação homóloga. Os elementos integracionais podem ser qualquer sequência que é homóloga à sequência alvo no genoma da célula hospedeira. Além do mais, os elementos integracionais podem ser polinucleotídeos não codificantes ou codificantes. Por outro lado, o vetor pode ser integrado no genoma da célula hospedeira por recombinação não homóloga.

Para replicação autônoma, o vetor pode compreender adicionalmente uma origem de replicação que possibilita o vetor a replicar autonomamente na célula hospedeira em questão. A origem de replicação pode ser qualquer plasmídeo replicador que medeia a replicação autônoma que funciona em uma célula. O termo “origem de replicação” ou “plasmídeo replicador” significa um polinucleotídeo que possibilita que um plasmídeo ou vetor se replique *in vivo*.

Exemplos de origens de replicação bacterianas são as origens de replicação de plasmídeos pBR322, pUC19, pACYC177, e pACYC184 que permitem a replicação em *E. coli*, e pUB110, pE194, pTA1060, e pAM β 1 que permitem a replicação em *Bacillus*.

Exemplos de origens de replicação para uso em uma levedura como célula hospedeira são as origem de replicação de 2 micron, ARS1,

ARS4, a combinação de ARS1 e CEN3, e a combinação de ARS4 e CEN6.

Exemplos de origens de replicação usadas em uma célula de filamentosos são AMA1 e ANS1 (Gems *et al.*, 1991, *Gene* 98: 61-67; Cullen *et al.*, 1987, *Nucleic Acids Res.* 15: 9163-9175; WO 00/24883). O isolamento do gene AMA1 e a construção de plasmídeos ou vetores que compreendem o gene podem ser realizados de acordo com os métodos revelados em WO 00/24883.

Mais de uma cópia de um polinucleotídeo da presente invenção pode ser inserida em uma célula hospedeira para aumentar a produção de um polipeptídeo. Um aumento no número de cópias do polinucleotídeo pode ser obtido integrando pelo menos uma cópia adicional da sequência no genoma da célula hospedeira, ou incluindo um gene marcador selecionável amplificável com o polinucleotídeo, onde as células contendo cópias amplificadas do gene marcador selecionável e, por meio disso, cópias adicionais do polinucleotídeo, podem ser selecionadas cultivando as células na presença do agente selecionável adequado.

Os procedimentos usados para ligar os elementos descritos anteriormente para construir os vetores de expressão recombinantes da presente invenção são bem conhecidos pelos versados na técnica (ver, por exemplo, Sambrook *et al.*, 1989, *supra*).

Células hospedeiras

A presente invenção também diz respeito a células hospedeiras recombinantes, que compreendem um polinucleotídeo da presente invenção operavelmente ligado a uma ou mais (várias) sequências controle que direcionam a produção de um polipeptídeo da presente invenção. Um construto ou vetor que compreende um polinucleotídeo é introduzido em uma célula hospedeira, de maneira tal que o construto ou vetor seja mantido como um integrante cromossomal ou como um vetor auto replicante extra-cromossomal da maneira descrita anteriormente. O termo "célula hospedeira"

inclui qualquer progênie de uma célula mãe que não é idêntica à célula mãe em virtude das mutações que ocorrem durante a replicação. A escolha de uma célula hospedeira dependerá muito do gene que codifica o polipeptídeo e sua fonte.

5 A célula hospedeira pode ser qualquer célula usada na produção recombinante de um polipeptídeo da presente invenção, por exemplo, um procarioto ou um eucarioto.

 A célula hospedeira procariota pode ser qualquer bactéria Gram-positiva ou Gram-negativa. As bactérias Gram-positivas incluem, mas
10 sem limitação, *Bacillus*, *Clostridium*, *Enterococcus*, *Geobacillus*, *Lactobacillus*, *Lactococcus*, *Oceanobacillus*, *Staphylococcus*, *Streptococcus*, e *Streptomyces*. As bactérias Gram-negativas incluem, mas sem limitação, *Campylobacter*, *E. coli*, *Flavobacterium*, *Fusobacterium*, *Helicobacter*, *Ilyobacter*, *Neisseria*, *Pseudomonas*, *Salmonella*, e *Ureaplasma*.

15 A célula hospedeira bacteriana pode ser qualquer célula de *Bacillus* incluindo, mas sem limitação, células de *Bacillus alkalophilus*, *Bacillus amyloliquefaciens*, *Bacillus brevis*, *Bacillus circulans*, *Bacillus clausii*, *Bacillus coagulans*, *Bacillus firmus*, *Bacillus lautus*, *Bacillus lentus*, *Bacillus licheniformis*, *Bacillus megaterium*, *Bacillus pumilus*, *Bacillus*
20 *stearothermophilus*, *Bacillus subtilis*, e *Bacillus thuringiensis*.

 A célula hospedeira bacteriana também pode ser qualquer célula de *Streptococcus* incluindo, mas sem limitação, células de *Streptococcus equisimilis*, *Streptococcus pyogenes*, *Streptococcus uberis*, e *Streptococcus equi* subsp. *Zooepidemicus*.

25 A célula hospedeira bacteriana também pode ser qualquer célula de *Streptomyces* incluindo, mas sem limitação, células de *Streptomyces achromogenes*, *Streptomyces avermitilis*, *Streptomyces coelicolor*, *Streptomyces griseus*, e *Streptomyces lividans*.

 A introdução de DNA em uma célula de *Bacillus*, por

exemplo, pode ser realizada por transformação de protoplasto (ver, por exemplo, Chang e Cohen, 1979, *Mol. Gen. Genet.* 168: 111-115), usando células competentes (ver, por exemplo, Young e Spizizen, 1961, *J. Bacteriol.* 81: 823-829, ou Dubnau e Davidoff-Abelson, 1971, *J. Mol. Biol.* 56: 209-221), por eletroporação (ver, por exemplo, Shigekawa e Dower, 1988, *Biotechniques* 6: 742-751), ou por conjugação (ver, por exemplo, Koehler e Thorne, 1987, *J. Bacteriol.* 169: 5271-5278). A introdução de DNA em uma célula de *E. coli*, por exemplo, pode ser realizada por transformação de protoplasto (ver, por exemplo, Hanahan, 1983, *J. Mol. Biol.* 166: 557-580) ou eletroporação (ver, por exemplo, Dower *et al.*, 1988, *Nucleic Acids Res.* 16: 6127-6145). A introdução de DNA em uma célula de *Streptomyces*, por exemplo, pode ser realizada por transformação de protoplasto e eletroporação (ver, por exemplo, Gong *et al.*, 2004, *Folia Microbiol. (Praha)* 49: 399-405), por conjugação (ver, por exemplo, Mazodier *et al.*, 1989, *J. Bacteriol.* 171: 3583-3585), ou por transdução (ver, por exemplo, Burke *et al.*, 2001, *Proc. Natl. Acad. Sci. USA* 98: 6289-6294). A introdução de DNA em uma célula de *Pseudomonas*, por exemplo, pode ser realizada por eletroporação (ver, por exemplo, Choi *et al.*, 2006, *J. Microbiol. Methods* 64: 391-397) ou por conjugação (ver, por exemplo, Pinedo e Smets, 2005, *Appl. Environ. Microbiol.* 71: 51-57). A introdução de DNA em uma célula de *Streptococcus*, por exemplo, pode ser realizada competência natural (ver, por exemplo, Perry e Kuramitsu, 1981, *Infect. Immun.* 32: 1295-1297), por transformação de protoplasto (ver, por exemplo, Catt e Jollick, 1991, *Microbios* 68: 189-207, por eletroporação (ver, por exemplo, Buckley *et al.*, 1999, *Appl. Environ. Microbiol.* 65: 3800-3804) ou por conjugação (ver, por exemplo, Clewell, 1981, *Microbiol. Rev.* 45: 409-436). Entretanto, qualquer método conhecido na técnica para introduzir DNA em uma célula hospedeira pode ser usado.

A célula hospedeira também pode ser um eucarioto, tal como

uma célula de mamífero, inseto, planta ou fungo.

A célula hospedeira pode ser uma célula fúngica. “Fungos”, da maneira aqui usada, inclui os filos *Ascomycota*, *Basidiomycota*, *Chytridiomycota* e *Zygomycota* (da maneira definida por Hawksworth *et al.*,
5 *Em, Ainsworth and Bisby's Dictionary of the Fungi*, 8ª edição, 1995, CAB International, University Press, Cambridge, UK), bem como o *Oomycota* (da maneira citada em Hawksworth *et al.*, 1995, *supra*, page 171) e todos os fungos mitospóricos (Hawksworth *et al.*, 1995, *supra*).

A célula hospedeira fúngica pode ser uma célula de levedura.
10 “Levedura”, da maneira aqui usada, inclui leveduras ascosporogêneas (*Endomycetales*), leveduras basidiosporogêneas e leveduras que pertencem aos fungos imperfeitos (*Blastomycetes*). Uma vez que a classificação de levedura pode mudar no futuro, com propósitos desta invenção, as leveduras podem ser definidas da maneira descrita em *Biology and Activities of Yeast*
15 (Skinner, F.A., Passmore, S.M., e Davenport, R.R., eds, *Soc. App. Bacteriol. Symposium Series No. 9*, 1980).

A levedura como célula hospedeira pode ser uma célula de *Candida*, *Hansenula*, *Kluyveromyces*, *Pichia*, *Saccharomyces*, *Schizosaccharomyces*, ou *Yarrowia*, tal como uma célula de *Kluyveromyces*
20 *lactis*, *Saccharomyces carlsbergensis*, *Saccharomyces cerevisiae*, *Saccharomyces diastaticus*, *Saccharomyces douglasii*, *Saccharomyces kluyveri*, *Saccharomyces norbensis*, *Saccharomyces oviformis*, ou *Yarrowia lipolytica*.

A célula hospedeira fúngica pode ser uma célula de fungo
25 filamentoso. “Fungos filamentosos” incluem todas as formas filamentosas da subdivisão *Eumycota* e *Oomycota* (da maneira definida por Hawksworth *et al.*, 1995, *supra*). Os fungos filamentosos são caracterizados em geral por uma parede de micélio composta de quitina, celulose, glicana, quitosana, manana e outros polissacarídeos complexos. O crescimento vegetativo é por

alongamento de hifas e o catabolismo do carbono é obrigatoriamente aeróbico. Ao contrário, o crescimento vegetativo de leveduras, tal como *Saccharomyces cerevisiae*, é por brotamento de um talo unicelular e o catabolismo do carbono pode ser fermentativo.

- 5 As células hospedeiras de fungo filamentoso podem ser uma célula de *Acremonium*, *Aspergillus*, *Aureobasidium*, *Bjerkandera*, *Ceriporiopsis*, *Chrysosporium*, *Coprinus*, *Coriolus*, *Cryptococcus*, *Filibasidium*, *Fusarium*, *Humicola*, *Magnaporthe*, *Mucor*, *Myceliophthora*, *Neocallimastix*, *Neurospora*, *Paecilomyces*, *Penicillium*, *Phanerochaete*,
10 *Phlebia*, *Piromyces*, *Pleurotus*, *Schizophyllum*, *Talaromyces*, *Thermoascus*, *Thielavia*, *Tolypocladium*, *Trametes*, ou *Trichoderma*.

- Por exemplo, as células hospedeiras de fungo filamentoso podem ser uma célula de *Aspergillus awamori*, *Aspergillus foetidus*, *Aspergillus fumigatus*, *Aspergillus japonicus*, *Aspergillus nidulans*,
15 *Aspergillus niger*, *Aspergillus oryzae*, *Bjerkandera adusta*, *Ceriporiopsis aneirina*, *Ceriporiopsis caregiea*, *Ceriporiopsis gilvescens*, *Ceriporiopsis pannocinta*, *Ceriporiopsis rivulosa*, *Ceriporiopsis subrufa*, *Ceriporiopsis subvermispora*, *Chrysosporium inops*, *Chrysosporium keratinophilum*, *Chrysosporium lucknowense*, *Chrysosporium merdarium*, *Chrysosporium*
20 *pannicola*, *Chrysosporium queenslandicum*, *Chrysosporium tropicum*, *Chrysosporium zonatum*, *Coprinus cinereus*, *Coriolus hirsutus*, *Fusarium bactridioides*, *Fusarium cerealis*, *Fusarium crookwellense*, *Fusarium culmorum*, *Fusarium graminearum*, *Fusarium graminum*, *Fusarium heterosporum*, *Fusarium negundi*, *Fusarium oxysporum*, *Fusarium*
25 *reticulatum*, *Fusarium roseum*, *Fusarium sambucinum*, *Fusarium sarcochroum*, *Fusarium sporotrichioides*, *Fusarium sulphureum*, *Fusarium torulosum*, *Fusarium trichothecioides*, *Fusarium venenatum*, *Humicola insolens*, *Humicola lanuginosa*, *Mucor miehei*, *Myceliophthora thermophila*, *Neurospora crassa*, *Penicillium purpurogenum*, *Phanerochaete*

chrysosporium, *Phlebia radiata*, *Pleurotus eryngii*, *Thielavia terrestris*, *Trametes villosa*, *Trametes versicolor*, *Trichoderma harzianum*, *Trichoderma koningii*, *Trichoderma longibrachiatum*, *Trichoderma reesei*, ou *Trichoderma viride*.

5 As células fúngicas podem ser transformadas por um processo que envolve formação de protoplasto, transformação dos protoplastos e regeneração da parede celular de uma maneira conhecida *por se*. Os procedimentos adequados para a transformação de células hospedeiras de *Aspergillus* e *Trichoderma* são descritos em EP 238023 e Yelton *et al.*, 1984, 10 *Proc. Natl. Acad. Sci. USA* 81: 1470-1474. Os métodos adequados para transformar as espécies de *Fusarium* são descritos por Malardier *et al.*, 1989, *Gene* 78: 147-156, e WO 96/00787. A levedura pode ser transformada usando os procedimentos descritos por Becker e Guarente, Em Abelson, J.N. e Simon, M.I., editors, *Guide to Yeast Genetics and Molecular Biology*, 15 *Methods in Enzymology*, Volume 194, pp 182-187, Academic Press, Inc., Nova Iorque; Ito *et al.*, 1983, *J. Bacteriol.* 153: 163; e Hinnen *et al.*, 1978, *Proc. Natl. Acad. Sci. USA* 75: 1920.

Métodos de produção

20 A presente invenção também diz respeito a métodos de produzir um polipeptídeo da presente invenção, que compreende: (a) cultivar uma célula, que em sua forma tipo selvagem produz o polipeptídeo, em condições que conduzem para a produção do polipeptídeo; e (b) recuperar o polipeptídeo. Em um aspecto preferido, a célula é do gênero *Thielavia*. Em um aspecto mais preferido, a célula é *Thielavia hyrcaniae*. Em um aspecto 25 acima de tudo preferido, a célula é *Thielavia hyrcaniae* NN045097. Em um outro aspecto mais preferido, a célula é *Thielavia hyrcaniae* NN045178 (CGMCC 0864).

 A presente invenção também diz respeito a métodos de produzir um polipeptídeo da presente invenção, que compreende: (a) cultivar

uma célula hospedeira recombinante da presente invenção em condições que conduzem para a produção do polipeptídeo; e (b) recuperar o polipeptídeo.

As células hospedeiras são cultivadas em um meio nutriente adequado para a produção do polipeptídeo usando métodos bem conhecidos na técnica. Por exemplo, a célula pode ser cultivada por cultivo em frasco de agitação, e fermentação em pequena escala ou grande escala (incluindo fermentações contínuas, em lotes, em lote alimentado ou em estado sólido) em fermentadores de laboratório ou industriais, realizada em um meio e condições adequadas que permitem que o polipeptídeo seja expresso e/ou isolado. O cultivo ocorre em um meio nutriente adequado que compreende fontes de carbono e nitrogênio e sais inorgânicos, usando os procedimentos conhecidos na técnica. Os meios adequados são disponíveis de fornecedores comerciais, ou podem ser preparados de acordo com as composições publicadas (por exemplo, em catálogos do American Type Culture Collection). Se o polipeptídeo for secretado no meio nutriente, o polipeptídeo pode ser recuperado diretamente do meio. Se o polipeptídeo não for secretado, ele pode ser recuperado de lisados celulares.

O polipeptídeo pode ser detectado usando métodos conhecidos na técnica que são específicos para os polipeptídeos. Estes métodos de detecção podem incluir o uso de anticorpos específicos, a formação de um produto de enzima, ou o desaparecimento de um substrato de enzima. Por exemplo, um ensaio de enzima pode ser usado para determinar a atividade do polipeptídeo.

O polipeptídeo pode ser recuperado usando métodos conhecidos na técnica. Por exemplo, o polipeptídeo pode ser recuperado do meio nutriente por procedimentos convencionais incluindo, mas sem limitação, centrifugação, filtração, extração, secagem por aspersão, evaporação ou precipitação.

O polipeptídeo pode ser purificado por uma variedade de

procedimentos conhecidos na técnica incluindo, mas sem limitação, cromatografia (por exemplo, de troca iônica, de afinidade, hidrofóbica, focagem isoeletrica e por exclusão de tamanho), procedimentos eletroforéticos (por exemplo, focagem isoeletrica preparativa), solubilidade diferencial (por exemplo, precipitação de sulfato de amônio), SDS-PAGE, ou extração (ver, por exemplo, *Protein Purification*, J.-C. Janson e Lars Ryden, editores, VCH Publishers, Nova Iorque, 1989) para obter polipeptídeos substancialmente puros.

Em um aspecto alternativo, o polipeptídeo não é recuperado, mas certamente uma célula hospedeira da presente invenção que expressa o polipeptídeo é usada como uma fonte do polipeptídeo.

Plantas

A presente invenção também diz respeito a plantas isoladas, por exemplo, uma planta transgênica, parte da planta ou célula de planta, que compreende um polinucleotídeo isolado da presente invenção de maneira a expressar e produzir o polipeptídeo em quantidades recuperáveis. O polipeptídeo pode ser recuperado da planta ou parte da planta. Alternativamente, a planta ou parte da planta contendo o polipeptídeo pode ser usada como tal para melhorar a qualidade de um alimento ou ração, por exemplo, melhorar o valor nutricional, palatabilidade e propriedades reológicas, ou para destruir um fator anti-nutritivo.

A planta transgênica pode ser dicotiledônea (um dicotilédone) ou monocotiledônea (um monocotilédone). Exemplos de plantas monocotiledôneas são gramíneas, tais como grama do prado (grama azul, *Poa*), gramíneas forrageiras tais como *Festuca*, *Lolium*, gramíneas temperadas, tais como *Agrostis*, e cereais, por exemplo, trigo, aveias, centeio, cevada, arroz, sorgo e milho (grão de milho).

Exemplos de plantas dicotiledôneas são tabaco, legumes, tais como tremoços, batata, beterraba, ervilha, feijão e soja, e plantas crucíferas

(família *Brassicaceae*), tais como couve flor, canola e o organismo modelo muito relacionado *Arabidopsis thaliana*.

Exemplos de parte das plantas são caule, calo, folhas, raiz, frutas, sementes e tubérculos, bem como os tecidos individuais que compreendem estas partes, por exemplo, epiderme, mesófilo, parênquima, tecidos vasculares, meristemas. Os compartimentos celulares de planta específicos, tais como cloroplastos, apoplastos, mitocôndria, vacúolos, peroxissomos e citoplasma também são considerados como uma parte da planta. Além do mais, qualquer célula de planta, seja qual for a origem do tecido, é considerada como uma parte da planta. Da mesma maneira, partes da planta tais como tecidos específicos e células isoladas para facilitar a utilização da invenção são também consideradas partes da planta, por exemplo, embriões, endospermas, aleurona e revestimentos de semente.

São também incluídas no escopo da presente invenção são a progênie de tais plantas, partes da planta e células da planta.

A planta transgênica ou célula de planta que expressa um polipeptídeo pode ser construída de acordo com métodos conhecidos na tecnologia. Em resumo, a planta ou célula de planta é construída incorporando um ou mais (vários) construtos de expressão que codificam um polipeptídeo no genoma da planta hospedeira ou genoma do cloroplasto e que propagam a planta modificada resultante ou célula de planta em uma planta transgênica ou célula de planta.

O construto de expressão é convenientemente um construto de ácido nucleico, que compreende um polinucleotídeo que codifica um polipeptídeo operavelmente ligado a sequências regulatórias adequadas exigidas para a expressão do polinucleotídeo na planta ou parte da planta de escolha. Além do mais, o construto de expressão pode compreender um marcador selecionável usado para identificar células hospedeiras, nas quais o construto de expressão foi integrado e as sequências de DNA necessárias para

a introdução do construto na planta em questão (a última depende do método de introdução de DNA a ser usado).

A escolha das sequências regulatórias, tais como sequências promotoras e finalizadoras e, opcionalmente, sequências sinal ou de trânsito, é determinada, por exemplo, com base em onde, quando e como deseja-se expressar o polipeptídeo. Por exemplo, a expressão do gene que codifica um polipeptídeo pode ser constitutiva ou indutível, ou pode ser para o desenvolvimento, específica de estágio ou tecido, e o produto do gene pode ser alvejado em um tecido específico ou parte da planta tal como sementes ou folhas. As sequências regulatórias são, por exemplo, descritas por Tague *et al.*, 1988, *Plant Physiology* 86: 506.

Para a expressão constitutiva, o 35S-CaMV, a ubiquitina 1 do milho, e o promotor de actina 1 de arroz podem ser usados (Franck *et al.*, 1980, *Cell* 21: 285-294; Christensen *et al.*, 1992, *Plant Mol. Biol.* 18: 675-689; Zhang *et al.*, 1991, *Plant Cell* 3: 1155-1165). Os promotores específicos de órgão podem ser, por exemplo, um promotor de tecidos de armazenamento tais como sementes, tubérculos de batata e frutas (Edwards e Coruzzi, 1990, *Ann. Rev. Genet.* 24: 275-303), ou de tecidos de armazenamento metabólico tais como meristemas (Ito *et al.*, 1994, *Plant Mol. Biol.* 24: 863-878), um promotor específico de semente tal como o promotor de glutelina, prolamina, globulina, ou albumina de arroz (Wu *et al.*, 1998, *Plant Cell Physiol.* 39: 885-889), um promotor de *Vicia faba* da legumina B4 e o gene de proteína de semente desconhecido de *Vicia faba* (Conrad *et al.*, 1998, *J. Plant Physiol.* 152: 708-711), um promotor de uma proteína de corpo oleoso de semente Chen *et al.*, 1998, *Plant Cell Physiol.* 39: 935-941), o promotor de proteína de armazenamento *napA* de *Brassica napus*, ou qualquer outro promotor específico de semente conhecido na técnica, por exemplo, da maneira descrita em WO 91/14772. Além do mais, o promotor pode ser um promotor específico de folha tal como o promotor *rbcs* de arroz ou tomate (Kyoizuka *et*

al., 1993, *Plant Physiol.* 102: 991-1.000), o promotor do gene adenina metiltransferase do vírus clorela (Mitra e Higgins, 1994, *Plant Mol. Biol.* 26: 85-93), o promotor do gene *aldP* de arroz (Kagaya *et al.*, 1995, *Mol. Gen. Genet.* 248: 668-674), ou um promotor indutível de ferida tal como o
 5 promotor *pin2* da batata (Xu *et al.*, 1993, *Plant Mol. Biol.* 22: 573-588). Da mesma forma, o promotor pode ser induzido por tratamentos abióticos, tais como temperatura, aridez, ou alterações na salinidade, ou induzido por substâncias aplicadas exogenamente que ativam o promotor, por exemplo, etanol, estrogênio, hormônios vegetais, tais como etileno, ácido abscísico e
 10 ácido giberélico e metais pesados.

Um elemento melhorador de promotor também pode ser usado para atingir maior expressão de um polipeptídeo na planta. Por exemplo, o elemento melhorador do promotor pode ser um intron que é colocado entre o promotor e o polinucleotídeo que codifica um polipeptídeo. Por exemplo, Xu
 15 *et al.*, 1993, *supra*, revelam o uso do primeiro intron do gene actina 1 de arroz para melhorar a expressão.

O gene marcador selecionável e quaisquer outras partes do construto de expressão podem ser escolhidos daqueles disponíveis na técnica.

O construto de ácido nucleico é incorporado no genoma da
 20 planta de acordo com técnicas convencionais conhecidas na técnica, incluindo transformação mediada por *Agrobacterium*, transformação mediada por vírus, microinjeção, bombardeamento de partícula, transformação biolística e eletroporação (Gasser *et al.*, 1990, *Science* 244: 1293; Potrykus, 1990, *Bio/Technology* 8: 535; Shimamoto *et al.*, 1989, *Nature* 338: 274).

25 Atualmente, a transferência de genes mediada por *Agrobacterium tumefaciens* é o método de escolha para gerar dicotiledôneas transgênicas (para uma revisão, ver Hooykas e Schilperoort, 1992, *Plant Mol. Biol.* 19: 15-38) e também pode ser usado para transformar monocotiledôneas, embora outros métodos de transformação sejam frequentemente usados por

estas plantas. Atualmente, o método de escolha para gerar monocotiledôneas transgênicas é o bombardeamento de partículas (partículas microscópicas de ouro ou tungstênio revestidas com o DNA transformante) de calos embrionários ou embriões em desenvolvimento (Christou, 1992, *Plant J.* 2: 275-281; Shimamoto, 1994, *Curr. Opin. Biotechnol.* 5: 158-162; Vasil *et al.*, 1992, *Bio/Technology* 10: 667-674). Um método alternativo para a transformação de monocotiledôneas baseia-se na transformação de protoplasto, da maneira descrita por Omirulleh *et al.*, 1993, *Plant Mol. Biol.* 21: 415-428. Os métodos de transformação adicionais para uso de acordo com a presente revelação incluem aqueles descritos nas patentes U.S. 6.395.966 e 7.151.204 (ambas as quais são aqui incorporadas pela referência na sua íntegra).

Após a transformação, os transformantes com o construto de expressões incorporados são selecionados e regenerados nas plantas completas, de acordo com métodos bem conhecidos na técnica. Em geral, o procedimento da transformação é designado para a eliminação seletiva de genes de seleção tanto durante a regeneração quanto após as gerações usando, por exemplo, co-transformação com dois construtos de DNA-T separados ou excisão sítio específica do gene de seleção por uma recombinase específica.

Além de direcionar a transformação de um genótipo particular de planta com um construto preparado de acordo com a presente invenção, as plantas transgênicas podem ser preparadas cruzando uma planta com o construto em uma segunda planta que precisa do construto. Por exemplo, um construto que codifica um polipeptídeo pode ser introduzido em uma variedade particular de planta por cruzamento, sem a necessidade de transformar diretamente uma planta daquela dada variedade. Portanto, a presente invenção inclui não apenas uma planta regenerada diretamente das células que foram transformadas de acordo com a presente invenção, mas também a progênie de tais plantas. Da maneira aqui usada, progênie pode se

referir à descendência de qualquer geração de uma planta mãe preparada de acordo com a presente invenção. Tal progênie pode incluir um construto de DNA preparado de acordo com a presente invenção, ou uma porção de um construto de DNA preparado de acordo com a presente invenção. Os
5 cruzamentos resultam na introdução de um transgene em uma linhagem de planta polinizando de maneira cruzada uma linhagem inicial com uma linhagem de planta doadora. Os exemplos não limitantes de tais etapas são articulados adicionalmente na patente U.S. 7.151.204.

As plantas podem ser geradas por meio de um processo de
10 conversão de retrocruzamento. Por exemplo, as plantas incluem plantas referidas como um genótipo, linhagem, inato, híbrido convertido por retrocruzamento.

Os marcadores genéticos podem ser usados para auxiliar na introgressão de um ou mais transgenes da invenção a partir de um antecedente
15 genético em um outro. A seleção auxiliada por marcador oferece vantagens com relação à reprodução convencional, na qual ele pode ser usado para evitar erros causados por variações fenotípicas. Adicionalmente, os marcadores genéticos podem fornecer dados com relação ao grau relativo de germoplasma elite na progênie individual de um cruzamento particular. Por
20 exemplo, quando uma planta com uma característica desejada, que de outra forma apresenta um antecedente genético agronomicamente não desejável, é cruzada com uma mãe elite, os marcadores genéticos podem ser usados para selecionar a progênie que possui não apenas a característica de interesse, mas também apresentam uma proporção relativamente grande do germoplasma
25 desejado. Desta maneira, o número de gerações exigidas para introgressar uma ou mais características em um antecedente genético particular é minimizado.

A presente invenção também diz respeito a métodos de produzir um polipeptídeo da presente invenção que compreende: (a) cultivar

uma planta transgênica ou uma célula de planta, que compreende um polinucleotídeo que codifica o polipeptídeo, em condições que conduzem para a produção do polipeptídeo; e (b) recuperar o polipeptídeo.

Remoção ou redução da atividade da celobiohidrolase

5 A presente invenção também diz respeito a métodos de produzir um mutante de uma célula mãe, que compreendem interromper ou deletar um polinucleotídeo, ou uma porção deste, codificar um polipeptídeo da presente invenção, que resulta na célula mutante que produz menos do polipeptídeo do que a célula mãe quando cultivada nas mesmas condições.

10 A célula mutante pode ser construída reduzindo ou eliminando a expressão do polinucleotídeo usando métodos bem conhecidos na técnica, por exemplo, inserções, interrupções, substituições ou deleções. Em um aspecto preferido, o polinucleotídeo é inativado. O polinucleotídeo a ser modificado ou inativado pode ser, por exemplo, a região codificante, ou uma
15 parte desta, essencial para a atividade, ou um elemento regulatório exigido para a expressão da região codificante. Um exemplo de uma sequência regulatória ou controle como esta pode ser uma sequência promotora ou uma parte funcional desta, isto é, uma parte que é suficiente para afetar a expressão do polinucleotídeo. Outras sequências controle para possível modificação
20 incluem, mas sem limitação, um líder, sequência de poliadenilação, sequência de pro-peptídeo, sequência de peptídeo sinal, finalizador de transcrição e ativador transcricional.

 A modificação ou inativação do polinucleotídeo pode ser realizada submetendo a célula mãe à mutagênese e selecionando células
25 mutantes nas quais a expressão do polinucleotídeo foi reduzida ou eliminada. A mutagênese, que pode ser específica ou aleatória, pode ser realizada, por exemplo, pelo uso de um agente mutagênico físico ou químico adequado, pelo uso de um oligonucleotídeo adequado, ou submetendo a sequência de DNA à mutagênese gerada por PCR. Além do mais, a mutagênese pode ser realizada

pelo uso de qualquer combinação destes agentes mutagênicos.

Exemplos de um agente mutagênico físico ou químico adequado para o presente propósito incluem irradiação ultravioleta (UV), hidroxilamina, N-metil-N'-nitro-N-nitrosoguanidina (MNNG), O-metil
5 hidroxilamina, ácido nitroso, etil metano sulfonato (EMS), bissulfito de sódio, ácido fórmico e análogos de nucleotídeo.

Quando tais agentes são usados, a mutagênese é realizada tipicamente incubando a célula mãe a ser mutagenizada na presença do agente mutagênico de escolha em condições adequadas, e triando e/ou selecionando
10 células mutantes que exibem expressão reduzida ou nenhuma expressão do gene.

A modificação ou inativação do polinucleotídeo pode ser realizada por introdução, substituição, ou remoção de um ou mais (vários) nucleotídeos no gene ou um elemento regulatório exigido para a transcrição
15 ou tradução deste. Por exemplo, os nucleotídeos podem ser inseridos ou removidos de maneira a resultar na introdução de um códon de parada, na remoção do códon inicial, ou em uma mudança no quadro aberto de leitura. Tal modificação ou inativação pode ser realizada por mutagênese sítio direcionada ou mutagênese gerada por PCR de acordo com métodos
20 conhecidos na técnica. Embora a princípio a modificação possa ser realizada *in vivo*, isto é, diretamente na célula que expressa o polinucleotídeo a ser modificado, é preferível que a modificação seja realizada *in vitro* da maneira exemplificada a seguir.

Um exemplo de uma maneira conveniente para eliminar ou
25 reduzir a expressão de um polinucleotídeo baseia-se nas técnicas de substituição de gene, deleção de gene ou rompimento de gene. Por exemplo, no método de rompimento de gene, uma sequência de ácidos nucleicos que corresponde ao polinucleotídeo endógeno é mutagenizada *in vitro* para produzir uma sequência de ácidos nucleicos defeituosa que é então

transformada na célula mãe para produzir um gene defeituoso. Por recombinação homóloga, a sequência de ácidos nucleicos defeituosa substitui o polinucleotídeo endógeno. Pode ser desejável que o polinucleotídeo defeituoso também codifique um marcador que pode ser usado para a seleção de transformantes em que o polinucleotídeo foi modificado ou destruído. Em um aspecto particularmente preferido, o polinucleotídeo é interrompido com um marcador selecionável tais como aqueles aqui descritos.

A presente invenção também diz respeito a métodos de inibir a expressão de um polipeptídeo com atividade celobiohidrolase em uma célula, que compreende administrar à célula ou expressar na célula uma molécula de RNA de fita dupla (RNAds), em que o RNAds compreende uma subsequência de um polinucleotídeo da presente invenção. Em um aspecto preferido, o RNAds tem cerca de 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25 ou mais nucleotídeos duplex em tamanho.

O RNAds é preferivelmente um RNA de interferência pequeno (RNAsi) ou um RNA micro (RNAmi). Em um aspecto preferido, o RNAds é RNA de interferência pequeno (RNAssi) para inibir a transcrição. Em um outro aspecto preferido, o RNAds é RNA micro (RNAsmi) para inibir a tradução.

A presente invenção também diz respeito a tais moléculas de RNA de fita dupla (RNAds) que compreendem uma porção da sequência que codifica o polipeptídeo maduro de SEQ ID NO: 1 ou SEQ ID NO: 3 para inibir expressão do polipeptídeo em uma célula. Embora a presente invenção não seja limitada por nenhum mecanismo de ação particular, o RNAds pode entrar em uma célula e causar a degradação de um RNA de fita simples (RNAss) de sequências similares ou idênticas, incluindo RNAm endógenos. Quando uma célula é exposta ao RNAds, o RNAm do gene homólogo é seletivamente degradado por um processo denominado RNA de interferência (RNAi).

Os RNAsds da presente invenção podem ser usados no silenciamento de gene. Em um aspecto, a invenção fornece métodos para degradar seletivamente o RNA usando um RNAids da presente invenção. O processo pode ser praticado *in vitro*, *ex vivo* ou *in vivo*. Em um aspecto, as moléculas de RNAds podem ser usadas para gerar uma mutação de perda-de-função em uma célula, um órgão ou um animal. Os métodos de preparar e usar moléculas de RNAds para degradar seletivamente o RNA são bem conhecidos na técnica; ver, por exemplo, patentes U.S. 6.489.127, 6.506.559, 6.511.824 e 6.515.109.

10 A presente invenção diz respeito adicionalmente a uma célula mutante de uma célula mãe que compreende uma interrupção ou deleção de um polinucleotídeo que codifica o polipeptídeo, ou uma sequência controle deste, ou um gene silenciado que codifica o polipeptídeo, que resulta na célula mutante processando menos do polipeptídeo ou nenhum polipeptídeo, comparada à célula mãe.

As células mutantes deficientes em polipeptídeo são particularmente usadas nas células hospedeiras para a expressão de polipeptídeos naturais e heterólogos. Portanto, a presente invenção diz respeito adicionalmente a métodos de produzir um polipeptídeo heterólogo ou natural que compreende: (a) cultivar a célula mutante em condições que conduzem para a produção do polipeptídeo; e (b) recuperar o polipeptídeo. O termo "polipeptídeos heterólogos" significa polipeptídeos que não são naturais da célula hospedeira, por exemplo, um variante de uma proteína natural. A célula hospedeira pode compreender mais de uma cópia de um polinucleotídeo que codifica o polipeptídeo heterólogo ou natural.

25 Os métodos usados para o cultivo e purificação do produto de interesse podem ser realizados por métodos conhecidos na técnica.

Os métodos da presente invenção para produzir um produto essencialmente livre de celobiohidrolase são de particular interesse na

produção de polipeptídeos de eucariotos, em particular, proteínas fúngicas tais como enzimas. As células deficientes em celobiohidrolase também podem ser usadas para expressar proteínas heterólogas de interesse farmacêutico, tais como hormônios, fatores de crescimento, receptores e similares. O termo "polipeptídeos de eucarioto" inclui não apenas polipeptídeos naturais, mas também aqueles polipeptídeos, por exemplo, enzimas, que foram modificados por substituições, deleções ou adições de aminoácido, ou outras tais modificações para melhorar a atividade, termoestabilidade, tolerância ao pH e similares.

Em um aspecto adicional, a presente invenção diz respeito a um produto de proteína essencialmente livre de atividade celobiohidrolase que é sintetizado por um método da presente invenção.

Composições

A presente invenção também diz respeito a composições que compreendem um polipeptídeo da presente invenção. Preferivelmente, as composições são enriquecidas em um polipeptídeo como este. O termo "enriquecido" indica que a atividade celobiohidrolase da composição aumentou, por exemplo, com um fator de enriquecimento de pelo menos 1,1.

A composição pode compreender um polipeptídeo da presente invenção como o principal componente enzimático, por exemplo, uma composição de mono-componente. Alternativamente, a composição pode compreender múltiplas atividades enzimáticas, tais como uma aminopeptidase, amilase, carboidrase, carboxipeptidase, catalase, celulase, quitinase, cutinase, ciclodextrina glicosiltransferase, desoxirribonuclease, esterase, alfa-galactosidase, beta-galactosidase, glucoamilase, alfa-glicosidase, beta-glicosidase, haloperoxidase, invertase, lacase, lipase, manosidase, oxidase, enzima pectinolítica, peptidoglutaminase, peroxidase, fitase, polifenoloxidase, enzima proteolítica, ribonuclease, transglutaminase ou xilanase. A(s) enzima(s) adicional(s) pode(m) ser produzida(s), por

exemplo, por um microrganismo que pertence ao gênero *Aspergillus*, preferivelmente *Aspergillus aculeatus*, *Aspergillus awamori*, *Aspergillus fumigatus*, *Aspergillus foetidus*, *Aspergillus japonicus*, *Aspergillus nidulans*, *Aspergillus niger*, ou *Aspergillus oryzae*; *Fusarium*, preferivelmente

5 *Fusarium bactridioides*, *Fusarium cerealis*, *Fusarium crookwellense*, *Fusarium culmorum*, *Fusarium graminearum*, *Fusarium graminum*, *Fusarium heterosporum*, *Fusarium negundi*, *Fusarium oxysporum*, *Fusarium reticulatum*, *Fusarium roseum*, *Fusarium sambucinum*, *Fusarium sarcochroum*, *Fusarium sulphureum*, *Fusarium toruloseum*, *Fusarium*

10 *trichothecioides*, ou *Fusarium venenatum*; *Humicola*, preferivelmente *Humicola insolens* ou *Humicola lanuginosa*; ou *Trichoderma*, preferivelmente *Trichoderma harzianum*, *Trichoderma koningii*, *Trichoderma longibrachiatum*, *Trichoderma reesei*, ou *Trichoderma viride*.

As composições de polipeptídeo podem ser preparadas de

15 acordo com métodos conhecidos na técnica e podem estar na forma de uma composição líquida ou seca. Por exemplo, a composição de polipeptídeo pode estar na forma de um granulado ou um microgranulado. O polipeptídeo pode ser estabilizado de acordo com métodos conhecidos na técnica.

São fornecidos a seguir exemplos de usos preferidos das

20 composições de polipeptídeo da invenção. A dosagem da composição de polipeptídeo da invenção e outras condições nas quais a composição é usada podem ser determinadas com base nos métodos conhecidos na técnica.

Usos

A presente invenção é também diz respeito aos seguintes

25 métodos para usar os polipeptídeos com atividade celobiohidrolase, ou composições destes.

A presente invenção também diz respeito a métodos para degradar ou converter um material celulósico que compreendem: tratar o material celulósico com um composição de enzima na presença de um

polipeptídeo com atividade celobiohidrolase da presente invenção. Em um aspecto preferido, o método compreende recuperar adicionalmente o material celulósico degradado ou convertido.

5 A presente invenção também diz respeito a métodos de produzir um produto de fermentação que compreendem: (a) sacarificar um material celulósico com uma composição de enzima, na presença de um polipeptídeo com atividade celobiohidrolase da presente invenção; (b) fermentar o material celulósico sacarificado com um ou mais (vários) microrganismos fermentadores para produzir o produto de fermentação; e (c)
10 recuperar o produto de fermentação a partir da fermentação.

A presente invenção também diz respeito a métodos de fermentar um material celulósico que compreendem: fermentar o material celulósico com um ou mais (vários) microrganismos fermentadores, em que o material celulósico é sacarificado com uma composição de enzima na
15 presença de um polipeptídeo com atividade celobiohidrolase da presente invenção. Em um aspecto preferido, a fermentação do material celulósico produz um produto de fermentação. Em um outro aspecto preferido, o método compreende recuperar adicionalmente o produto de fermentação a partir da fermentação.

20 Os métodos da presente invenção podem ser usados para sacarificar um material celulósico em açúcares fermentáveis e converter os açúcares fermentáveis em muitas substâncias utilizáveis, por exemplo, combustível, etanol potável e/ou produtos de fermentação (por exemplo, ácidos, álcoois, cetonas, gases e similares). A produção de um produto de
25 fermentação desejado a partir de material celulósico envolve tipicamente pré-tratamento, hidrólise enzimática (sacarificação) e fermentação.

O processamento de material celulósico ou contendo xilano, de acordo com a presente invenção, pode ser realizado usando processos convencionais na técnica. Além do mais, os métodos da presente invenção

podem ser implementados usando qualquer aparelho de processamento de biomassa convencional configurado para operar de acordo com a invenção.

A hidrólise (sacarificação) e fermentação, separadas ou simultâneas, incluem, mas sem limitação, hidrólise e fermentação separadas (SHF); sacarificação e fermentação simultânea (SSF); sacarificação e co-fermentação simultânea (SSCF); hidrólise e fermentação híbridas (HHF); hidrólise e co-fermentação separadas (SHCF); hidrólise e co-fermentação híbridas (HHCF) e conversão microbiana direta(DMC). SHF usa etapas de processo separadas para hidrolisar de maneira enzimática primeiro o material
5 celulósico em açúcares fermentáveis, por exemplo, açúcares do tipo glicose, celobiose, celotriose e pentose, e então fermentar os açúcares fermentáveis em etanol. Em SSF, a hidrólise enzimática de material celulósico e a fermentação de açúcares em etanol são combinadas em uma etapa (Philippidis, G. P., 1996, Cellulose bioconversion technology, em *Handbook on Bioethanol: Production and Utilization*, Wyman, C. E., ed., Taylor & Francis,
10 Washington, DC, 179-212). SSCF envolve a co-fermentação de múltiplos açúcares (Sheehan, J., e Himmel, M., 1999, Enzymes, energy and the environment: A strategic perspective on the U.S. Department of Energy's research and development activities for bioethanol, *Biotechnol. Prog.* 15: 817-
15 827). HHF envolve uma etapa separada de hidrólise e, além disso, um etapa simultânea de sacarificação e hidrólise que pode ser realizada no mesmo reator. As etapas em um processo de HHF podem ser realizadas em diferentes temperaturas, isto é, sacarificação enzimática em alta temperatura seguida por SSF em uma temperatura menor que a cepa da fermentação possa tolerar.
20 DMC combina todos os três processos (produção, hidrólise e fermentação enzimática) em uma ou mais (várias) etapas, onde o mesmo organismo é usado para produzir as enzimas para a conversão do material celulósico em açúcares fermentáveis e para converter os açúcares fermentáveis em um produto final (Lynd, L. R., Weimer, P. J., van Zyl, W. H., e Pretorius, I. S.,
25

2002, Microbial cellulose utilization: Fundamentals and biotechnology, *Microbiol. Mol. Biol. Reviews* 66: 506-577). Entende-se aqui que qualquer método conhecido na técnica que compreende pré-tratamento, hidrólise enzimática (sacarificação), fermentação, ou uma combinação destes, pode ser
5 usado na prática dos métodos da presente invenção.

Um aparelho convencional pode incluir um reator agitado com lote alimentado, um reator agitado em lotes, um reator agitado de fluxo contínuo com ultrafiltração e/ou um reator de coluna de fluxo em pistão contínuo (Fernanda de Castilhos Corazza, Flávio Faria de Moraes, Gisella
10 Maria Zanin e Ivo Neitzel, 2003, Optimal control in fed-batch reactor for the cellobiose hydrolysis, *Acta Scientiarum. Technology* 25: 33-38; Gusakov, A. V., e Sinitsyn, A. P., 1985, Kinetics of the enzymatic hydrolysis of cellulose: 1. A mathematical model for a batch reactor process, *Enz. Microb. Technol.* 7: 346-352), um reator de atrito (Ryu, S. K., e Lee, J. M., 1983, Bioconversion
15 of waste cellulose using a attrition bioreactor, *Biotechnol. Bioeng.* 25: 53-65), ou um reator com agitação intensiva induzida por um campo eletromagnético (Gusakov, A. V., Sinitsyn, A. P., Davydkin, I. Y., Davydkin, V. Y., Protas, O. V., 1996, Enhancement of enzymatic cellulose hydrolysis using a novel type of bioreactor with intensive stirring induced by electromagnetic field, *Appl.*
20 *Biochem. Biotechnol.* 56: 141-153). Os tipos de reatores adicionais incluem: reatores do tipo leito fluidizado, de manta com fluxo ascendente, imobilizado e extrudor para hidrólise e/ou fermentação.

Pré-tratamento. Na prática dos métodos da presente invenção, qualquer processo de pré-tratamento conhecido na técnica pode ser usado para
25 romper os componentes das paredes celulares das plantas de material celulósico e/ou contendo xilano (Chandra *et al.*, 2007, Substrate pretreatment: The key to effective enzymatic hydrolysis enzimática of lignocellulosics? *Adv. Biochem. Engin./Biotechnol.* 108: 67-93; Galbe e Zacchi, 2007, Pretreatment of lignomaterial celulósicos for efficient bioethanol production,

Adv. Biochem. Engin. / Biotechnol. 108: 41-65; Hendriks e Zeeman, 2009, Pretreatments to enhance the digestibility of lignocellulosic biomass, *Bioresource Technol.* 100: 10-18; Mosier *et al.*, 2005, Features of promising technologies for pretreatment of lignocellulosic biomass, *Bioresource Technol.* 96: 673-686; Taherzadeh e Karimi, 2008, Pretreatment of lignocellulosic wastes to improve ethanol and biogas production: A review, *Int. J. de Mol. Sci.* 9: 1621-1651; Yang e Wyman, 2008, Pretreatment: The key to unlocking low-cost cellulosic ethanol, *Biofuels Bioproducts and Biorefining-Biofpr.* 2: 26-40).

10 O material celulósico também pode ser submetido a redução de tamanho de partícula, pré-enxarcamento, umidificação, lavagem, ou condicionamento antes do pré-tratamento usando os métodos conhecidos na técnica.

Os pré-tratamentos convencionais incluem, mas sem limitação, pré-tratamento com vapor (com ou sem explosão), pré-tratamento com ácido diluído, pré-tratamento com água quente, pré-tratamento alcalino, pré-tratamento com cal, oxidação úmida, explosão úmida, explosão com fibra de amônia, pré-tratamento com organosolvente e pré-tratamento biológico. Os pré-tratamentos adicionais incluem tratamentos com percolação com amônia, ultrassom, eletroporação, micro-ondas, CO₂ supercrítico, H₂O supercrítica, ozônio e irradiação gama.

O material celulósico pode ser pré-tratado antes da hidrólise e/ou fermentação. O pré-tratamento é preferivelmente realizado antes da hidrólise. Alternativamente, o pré-tratamento pode ser realizado simultaneamente com hidrólise enzimática para liberar açúcares fermentáveis, tais como glicose, xilose e/ou celobiose. Em muitos casos a etapa de pré-tratamento resulta em alguma conversão de biomassa em açúcares fermentáveis (mesmo na ausência de enzimas).

Pré-tratamento com vapor. No pré-tratamento com vapor, o

material celulósico é aquecido para romper os componentes das paredes celulares das plantas, incluindo lignina, hemicelulose e celulose, para tornar a celulose e outras frações, por exemplo, hemicelulose, acessíveis às enzimas. O material celulósico passa por um vaso de reação onde o vapor é injetado para aumentar a temperatura para a temperatura e a pressão exigidas e é mantido ali pelo tempo de reação desejado. O pré-tratamento com vapor é preferivelmente realizado a 140-230 °C, mais preferivelmente 160-200 °C, e acima de tudo preferivelmente 170-190 °C, onde a faixa de temperatura ideal depende de qualquer adição de um catalisador químico. O tempo de permanência para o pré-tratamento com vapor é preferivelmente 1-15 minutos, mais preferivelmente 3-12 minutos, e acima de tudo preferivelmente 4-10 minutos, onde o tempo de permanência ideal depende da faixa de temperatura e de qualquer adição de um catalisador químico. O pré-tratamento com vapor permite carregamentos sólidos relativamente altos, de maneira tal que o material celulósico é úmido em geral apenas durante o pré-tratamento. O pré-tratamento com vapor é frequentemente combinado com uma descarga explosiva do material após o pré-tratamento, que é conhecida como explosão de vapor, isto é, queima rápida em pressão atmosférica e fluxo turbulento do material para aumentar a área de superfície acessível por fragmentação (Duff e Murray, 1996, *Bioresource Technology* 855: 1-33; Galbe e Zacchi, 2002, *Appl. Microbiol. Biotechnol.* 59: 618-628; pedido de patente U.S. 20020164730). Durante o pré-tratamento com vapor, os grupos de acetil hemicelulose são clivados e o ácido resultante autocatalisa a hidrólise parcial da hemicelulose em monossacarídeos e oligossacarídeos. A lignina é removida em apenas uma extensão limitada.

Um catalisador, tal como H₂SO₄ ou SO₂ (tipicamente 0,3 a 3 % p/p), é em geral adicionado ao pré-tratamento com vapor que diminui o tempo e a temperatura, aumenta a recuperação e melhora a hidrólise enzimática (Ballesteros *et al.*, 2006, *Appl. Biochem. Biotechnol.* 129-132:

496-508; Varga *et al.*, 2004, *Appl. Biochem. Biotechnol.* 113-116: 509-523; Sassner *et al.*, 2006, *Enzyme Microb. Technol.* 39: 756-762).

Pré-tratamento químico: O termo “tratamento químico” refere-se a qualquer pré-tratamento químico que promove a separação e/ou liberação de celulose, hemicelulose e/ou lignina. Exemplos de processo de pré-tratamento químico adequados incluem, por exemplo, pré-tratamento com ácido diluído, pré-tratamento com cal, oxidação úmida, explosão de fibra com amônia/congelamento (AFEX), percolação com amônia (APR) e pré-tratamentos com organosolv.

No pré-tratamento com ácido diluído, o material celulósico é misturado com o ácido diluído, tipicamente H_2SO_4 , e água para formar uma lama, aquecido por vapor na temperatura desejada, e após um tempo de permanência é queimado em pressão atmosférica. O pré-tratamento com ácido diluído pode ser realizado com inúmeros projetos de reator, por exemplo, reatores de fluxo em pistão, reatores contra corrente, ou reatores de leito agitado em contra corrente contínua (Duff e Murray, 1996, *supra*; Schell *et al.*, 2004, *Bioresource Technol.* 91: 179-188; Lee *et al.*, 1999, *Adv. Biochem. Eng. Biotechnol.* 65: 93-115).

Vários métodos de pré-tratamento em condições alcalinas também podem ser usados. Estes pré-tratamentos alcalinos incluem, mas sem limitação, pré-tratamento com cal, oxidação úmida, percolação com amônia (APR), e explosão de fibra com amônia/congelamento (AFEX).

O pré-tratamento com cal é realizado com carbonato de cálcio, hidróxido de sódio, ou amônia em baixas temperaturas de 85-150 °C e tempo de permanência de 1 hora a vários dias (Wyman *et al.*, 2005, *Bioresource Technol.* 96: 1959-1966; Mosier *et al.*, 2005, *Bioresource Technol.* 96: 673-686). WO 2006/110891, WO 2006/11899, WO 2006/11900, e WO 2006/110901 revelam métodos de pré-tratamento que usa amônia.

A oxidação úmida é um pré-tratamento térmico realizado

tipicamente a 180-200 °C por 5-15 minutos com a adição de um agente oxidante, tal como peróxido de hidrogênio ou super-pressão de oxigênio (Schmidt e Thomsen, 1998, *Bioresource Technol.* 64: 139-151; Palonen *et al.*, 2004, *Appl. Biochem. Biotechnol.* 117: 1-17; Varga *et al.*, 2004, *Biotechnol. Bioeng.* 88: 567-574; Martin *et al.*, 2006, *J. Chem. Technol. Biotechnol.* 81: 1669-1677). O pré-tratamento é realizado preferivelmente em 1-40 % de material seco, mais preferivelmente 2-30 % de material seco, e acima de tudo preferivelmente 5-20 % de material seco, e frequentemente o pH inicial aumenta pela adição de álcali, tal como carbonato de sódio.

Uma modificação do método de pré-tratamento com oxidação úmida, conhecido como explosão úmida (combinação de explosão de oxidação úmida e vapor), pode manusear material seco até 30 %. Na explosão úmida, o agente oxidante é introduzido durante o pré-tratamento após um certo tempo de permanência. O pré-tratamento é então finalizado queimando em pressão atmosférica (WO 2006/032282).

A explosão de fibra com amônia (AFEX) envolve tratar o material celulósico com amônia líquida ou gasosa em temperaturas moderadas, tal como 90-100 °C, e alta pressão tal como 17-20 bar por 5-10 minutos, onde o conteúdo do material seco pode ser tão alto quanto 60 % (Gollapalli *et al.*, 2002, *Appl. Biochem. Biotechnol.* 98: 23-35; Chundawat *et al.*, 2007, *Biotechnol. Bioeng.* 96: 219-231; Alizadeh *et al.*, 2005, *Appl. Biochem. Biotechnol.* 121: 1133-1141; Teymouri *et al.*, 2005, *Bioresource Technol.* 96: 2014-2018). O pré-tratamento com AFEX resulta na despolimerização de celulose e hidrólise parcial de hemicelulose. Os complexos de lignina-carboidrato são clivados.

O pré-tratamento com organosolv delignifica o material celulósico por extração usando etanol aquoso (40-60 % etanol) a 160-200 °C por 30-60 minutos (Pan *et al.*, 2005, *Biotechnol. Bioeng.* 90: 473-481; Pan *et al.*, 2006, *Biotechnol. Bioeng.* 94: 851-861; Kurabi *et al.*, 2005, *Appl. Biochem. Biotechnol.* 121: 219-230). O ácido sulfúrico é em geral adicionado como um

catalisador. Em pré-tratamento com organosolv, a maioria da hemicelulose é removida.

Outros exemplos de métodos de pré-tratamento adequados são descritos por Schell *et al.*, 2003, *Appl. Biochem. e Biotechnol.* Vol. 105-108, p. 69-85, e Mosier *et al.*, 2005, *Bioresource Technology* 96: 673-686, e pedido publicado U.S. 2002/0164730.

Em um aspecto, o pré-tratamento químico é realizado preferivelmente como um tratamento ácido e, mais preferivelmente, como um tratamento com ácido diluído contínuo e/ou fraco. O ácido é tipicamente ácido sulfúrico, mas outros ácidos também podem ser usados, tais como ácido acético, ácido cítrico, ácido nítrico, ácido fosfórico, ácido tartárico, ácido succínico, cloreto de hidrogênio ou misturas destes. O tratamento com ácido fraco é conduzido na faixa de pH de preferivelmente 1-5, mais preferivelmente 1-4, e acima de tudo preferivelmente 1-3. Em um aspecto, a concentração do ácido está na faixa de preferivelmente 0,01 a 20 % em peso de ácido, mais preferivelmente 0,05 a 10 % em peso de ácido, ainda mais preferivelmente 0,1 a 5 % em peso de ácido, e acima de tudo preferivelmente 0,2 a 2,0 % em peso de ácido. O ácido é colocado em contato com material celulósico e mantido a uma temperatura na faixa de preferivelmente 160-220 °C, e mais preferivelmente 165-195 °C, por períodos que variam de segundos a minutos, por exemplo, 1 segundo a 60 minutos.

Em um outro aspecto, pré-tratamento é realizado como uma etapa de explosão de fibra (etapa de pré-tratamento com AFEX).

Em um outro aspecto, o pré-tratamento ocorre em uma lama aquosa. Em aspectos preferidos, o material celulósico está presente durante o pré-tratamento em quantidades preferivelmente entre 10-80 % em peso, mais preferivelmente entre 20-70 % em peso, e acima de tudo preferivelmente entre 30-60 % em peso, tal como em torno de 50 % em peso. O material celulósico pré-tratado pode ser não lavado ou lavado usando qualquer método

conhecido na técnica, por exemplo, lavado com água.

Pré-tratamento mecânico: O termo “pré-tratamento mecânico” refere-se a vários tipos de trituração ou moagem (por exemplo, moagem a seco, moagem úmida, ou moagem com bola vibratória).

5 Pré-tratamento físico: O termo “pré-tratamento físico” refere-se a qualquer pré-tratamento que promove a separação e/ou liberação de celulose, hemicelulose, e/ou lignina do material celulósico ou contendo xilano. Por exemplo, o pré-tratamento físico pode envolver irradiação (por exemplo, irradiação por micro-ondas), explosão vaporosa/vapor, hidrotermólise e
10 combinações destes.

 O pré-tratamento físico pode envolver alta pressão e/ou alta temperatura (explosão por vapor). Em um aspecto, alta pressão significa pressão na faixa de preferivelmente cerca de 300 a cerca de 600 psi, mais preferivelmente cerca de 350 a cerca de 550 psi, e acima de tudo
15 preferivelmente cerca de 400 a cerca de 500 psi, tal como em torno de 450 psi. Em um outro aspecto, alta temperatura significa temperaturas na faixa de cerca de 100 a cerca de 300 °C, preferivelmente cerca de 140 a cerca de 235 °C. Em um aspecto preferido, o pré-tratamento mecânico é realizado em um processo em lotes, sistema hidrolisador de arma a vapor que usa alta pressão e alta
20 temperatura da maneira definida anteriormente, por exemplo, um hidrolisador Sunds disponível pela Sunds Defibrator AB, Suécia.

 Pré-tratamento químico e físico combinado: O material celulósico pode ser pré-tratado tanto física quanto quimicamente. Por exemplo, a etapa de pré-tratamento pode envolver o tratamento com ácido fraco ou
25 diluído e tratamento com alta temperatura e/ou pressão. Os pré-tratamentos físicos e químicos podem ser realizados sequencialmente ou simultaneamente, da maneira desejada. Um pré-tratamento mecânico também pode ser incluído.

 Dessa maneira, em um aspecto preferido, o material celulósico é submetido aos pré-tratamento mecânico, químico ou físico, ou qualquer

combinação destes, para promover a separação e/ou liberação de celulose, hemicelulose e/ou lignina.

Pré-tratamento biológico: O termo “pré-tratamento biológico” refere-se a qualquer pré-tratamento biológico que promove a separação e/ou liberação de celulose, hemicelulose e/ou lignina do material celulósico ou contendo xilano. As técnicas de pré-tratamento biológico podem envolver aplicar microrganismos que solubilizam lignina (ver, por exemplo, Hsu, T.-A., 1996, Pretreatments of biomass, em *Handbook on Bioethanol: Production and Utilization*, Wyman, C. E., ed., Taylor & Francis, Washington, DC, 179-212; Ghosh e Singh, 1993, Physicochemical and biological treatments for enzymatic/microbial conversion of cellulosic biomass, *Adv. Appl. Microbiol.* 39: 295-333; McMillan, J. D., 1994, Pretreating lignocellulosic biomass: a review, em *Enzymatic Conversion of Biomass for Fuels Production*, Himmel, M. E., Baker, J. O., e Overend, R. P., eds., ACS Symposium Series 566, American Chemical Society, Washington, DC, capítulo 15; Gong, C. S., Cao, N. J., Du, J., e Tsao, G. T., 1999, Ethanol production from renewable resources, em *Advances in Biochemical Engineering/Biotechnology*, Scheper, T., ed., Springer-Verlag Berlin Heidelberg, Alemanha, 65: 207-241; Olsson e Hahn-Hagerdal, 1996, Fermentation of lignocellulosic hydrolysates for ethanol production, *Enz. Microb. Tech.* 18: 312-331; e Vallander e Eriksson, 1990, Production of ethanol from lignomaterial celulósicos: State of the art, *Adv. Biochem. Eng./Biotechnol.* 42: 63-95).

Sacarificação. Na etapa da hidrólise, também conhecida como sacarificação, o material celulósico ou contendo xilano, por exemplo, pré-tratado, é hidrolisado para quebrar a celulose e, alternativamente também, hemicelulose em açúcares fermentáveis, tais como glicose, celobiose, xilose, xilulose, arabinose, manose, galactose e/ou oligossacarídeos solúveis. A hidrólise é realizada enzimaticamente por uma composição de enzimas na presença de um polipeptídeo com atividade celobiohidrolase da presente

invenção. A composição pode compreender adicionalmente uma ou mais (várias) enzimas hemicelulolíticas ou que degradam xilano. Os componentes da enzima e proteína das composições podem ser adicionados sequencialmente.

5 A hidrólise enzimática é realizada preferivelmente em um ambiente aquoso adequado em condições que podem ser facilmente determinadas pelos versados na técnica. Em um aspecto preferido, a hidrólise é realizada em condições adequadas para a atividade da(s) enzima(s), isto é, ideal para a(s) enzima(s). A hidrólise pode ser realizada como um processo em lote alimentado ou contínuo, onde o material celulósico pré-tratado (substrato) é
10 alimentado gradualmente, por exemplo, por uma enzima contendo a solução de hidrólise.

A sacarificação é realizada em geral em reatores de tanque agitado ou fermentadores em pH, temperatura e condições de mistura controlados. O tempo do processo, temperatura e condições de pH adequados
15 podem ser facilmente determinados pelos versados na técnica. Por exemplo, a sacarificação pode durar até 200 horas, mas é tipicamente realizada de maneira preferível em cerca de 12 a cerca de 96 horas, mais preferivelmente cerca de 16 a cerca de 72 horas, e acima de tudo preferivelmente cerca de 24 a cerca de 48 horas. A temperatura está na faixa de preferivelmente cerca de 25 °C a cerca
20 de 70 °C, mais preferivelmente cerca de 30 °C a cerca de 65 °C, e mais preferivelmente cerca de 40 °C to 60 °C, em particular cerca de 50 °C. O pH está na faixa de preferivelmente cerca de 3 a cerca de 8, mais preferivelmente cerca de 3,5 a cerca de 7, e acima de tudo preferivelmente cerca de 4 a cerca de 6, em particular cerca de pH 5. O teor dos sólidos secos está na faixa de
25 preferivelmente cerca de 5 a cerca de 50 % em peso, mais preferivelmente cerca de 10 a cerca de 40 % em peso, e acima de tudo preferivelmente cerca de 20 a cerca de 30 % em peso.

Nos métodos da presente invenção, a composição de enzima pode compreender qualquer proteína que é usada na degradação ou conversão

do material celulósico.

Em um aspecto, a composição de enzima compreende uma ou mais (várias) enzimas/proteínas celulolíticas. Em um outro aspecto, a composição de enzima compreende ou compreende adicionalmente uma ou mais (várias) enzimas hemicelulolíticas. Em um outro aspecto, a composição de enzima compreende uma ou mais (várias) enzimas celulolíticas e uma ou mais (várias) enzimas hemicelulolíticas. Em um outro aspecto, a composição de enzima compreende uma ou mais (várias) enzimas selecionadas do grupo de enzimas celulolíticas e enzimas hemicelulolíticas.

Em um outro aspecto, a composição de enzima compreende uma ou mais (várias) enzimas/proteínas celulolíticas selecionadas do grupo que consiste em uma endoglucanase, uma celobiohidrolase, uma beta-glucosidase e um polipetídeo com melhor atividade celulolítica. Em um outro aspecto, a composição de enzima compreende ou compreende adicionalmente uma ou mais (várias) proteínas selecionadas do grupo que consiste em uma hemicelulase, uma expansina, uma esterase, uma enzima ligninolítica, uma pectinase, uma peroxidase, uma protease e uma swolenina. A hemicelulase é preferivelmente uma ou mais (várias) enzimas selecionadas do grupo que consiste em uma acetilmanana esterase, uma acetixilano esterase, uma arabinanase, uma arabinofuranosidase, um ácido coumárico esterase, uma feruloil esterase, uma galactosidase, uma glucuronidase, uma glicuronoil esterase, uma mananase, uma manosidase, uma xilanase e uma xilosidase.

Em um outro aspecto, a composição de enzima compreende uma endoglucanase. Em um outro aspecto, a composição de enzima compreende uma celobiohidrolase. Em um outro aspecto, a composição de enzima compreende uma beta-glicosidase. Em um outro aspecto, a composição de enzima compreende um polipetídeo com melhor atividade celulolítica. Em um outro aspecto, a composição de enzima compreende uma acetilmanana esterase. Em um outro aspecto, a composição de enzima

compreende uma acetilxilano esterase. Em um outro aspecto, a composição de enzima compreende uma arabinanase (por exemplo, alfa-L-arabinanase). Em um outro aspecto, a composição de enzima compreende uma arabinofuranosidase (por exemplo, alfa-L-arabinofuranosidase). Em um outro aspecto, a composição de enzima compreende uma ácido coumárico esterase. Em um outro aspecto, a composição de enzima compreende uma feruloil esterase. Em um outro aspecto, a composição de enzima compreende uma galactosidase (por exemplo, alfa-galactosidase e/ou beta-galactosidase). Em um outro aspecto, a composição de enzima compreende uma glicuronidase (por exemplo, alfa-D-glicuronidase). Em um outro aspecto, a composição de enzima compreende uma glicuronoil esterase. Em um outro aspecto, a composição de enzima compreende uma mananase. Em um outro aspecto, a composição de enzima compreende uma manosidase (por exemplo, beta-manosidase). Em um outro aspecto, a composição de enzima compreende uma xilanase. Em um aspecto preferido, a xilanase é uma xilanase da família 10. Em um outro aspecto, a composição de enzima compreende uma xilosidase.

Em um outro aspecto, a composição de enzima compreende uma expansina. Em um outro aspecto, a composição de enzima compreende uma esterase. Em um outro aspecto, a composição de enzima compreende uma enzima ligninolítica. Em um aspecto preferido, a enzima ligninolítica é uma lacase. Em um outro aspecto preferido, a enzima ligninolítica é uma manganês peroxidase. Em um outro aspecto preferido, a enzima ligninolítica é uma lignina peroxidase. Em um outro aspecto preferido, a enzima ligninolítica é uma enzima que produz H_2O_2 . Em um outro aspecto, a composição de enzima compreende uma pectinase. Em um outro aspecto, a composição de enzima compreende uma peroxidase. Em um outro aspecto, a composição de enzima compreende uma protease. Em um outro aspecto, a composição de enzima compreende uma swollenina.

Nos métodos da presente invenção, a(s) enzima(s)/proteína(s) pode(m) ser adicionada(s) antes ou durante a fermentação, por exemplo, durante a sacarificação ou durante ou após a propagação do(s) microrganismo(s) fermentador(s).

5 Um ou mais (vários) componentes da composição de enzima podem ser proteínas tipo selvagem, proteínas recombinantes, ou uma combinação de proteínas tipo selvagem e proteínas recombinantes. Por exemplo, um ou mais (vários) componentes podem ser proteínas naturais de uma célula, que é usada as uma célula hospedeira para expressar
10 recombinantemente um ou mais (vários) outros componentes da composição de enzima. Um ou mais (vários) componentes da composição de enzima podem ser produzidos como monocomponentes, que são então combinados para formar a composição de enzima. A composição de enzima pode ser uma combinação de preparações protéicas multicomponente e monocomponente.

15 As enzimas usadas nos métodos da presente invenção pode estar em qualquer forma adequada para uso nos processos aqui descritos tais como, por exemplo, um caldo bruto de fermentação com ou sem células removidas, um lisado celular com ou sem restos celulares, uma preparação de enzima semi-purificada ou purificada, ou uma célula hospedeira como uma
20 fonte da enzimas. A composição de enzima pode ser um pó seco ou granulado, um granulado que não está em pó, um líquido, um líquido estabilizado, ou uma enzima protegida estabilizada. As preparações de enzima líquida, por exemplo, podem ser estabilizadas adicionado estabilizadores tal como um açúcar, um álcool de açúcar ou um outro poliol, e/ou ácido lático ou
25 um outro ácido orgânico, de acordo com os processos estabelecidos.

As quantidades ideais das enzimas e polipeptídeos com atividade celobiohidrolase dependem de vários fatores incluindo, mas sem limitação, a mistura de enzima celulolíticas do componente, o substrato celulósico, a concentração do substrato celulósico, o(s) pré-tratamento(s) do

substrato celulósico, temperatura, tempo, pH, e inclusão de organismo fermentador (por exemplo, levedura para sacarificação e fermentação simultânea).

Em um aspecto preferido, uma quantidade eficiente de
5 proteína celulolítica para material celulolítico está em cerca de 0,5 a cerca de 50 mg, preferivelmente em cerca de 0,5 a cerca de 40 mg, mais preferivelmente cerca de 0,5 a cerca de 25 mg, mais preferivelmente cerca de 0,75 a cerca de 20 mg, mais preferivelmente cerca de 0,75 a cerca de 15 mg, ainda mais preferivelmente cerca de 0,5 a cerca de 10 mg, e acima de tudo
10 preferivelmente cerca de 2,5 a cerca de 10 mg por g de material celulósico.

Em um outro aspecto preferido, uma quantidade eficiente de polipeptídeo(s) com atividade celobiohidrolase no material celulósico é cerca de 0,01 a cerca de 50,0 mg, preferivelmente cerca de 0,01 a cerca de 40 mg, mais preferivelmente cerca de 0,01 a cerca de 30 mg, mais preferivelmente
15 cerca de 0,01 a cerca de 20 mg, mais preferivelmente cerca de 0,01 a cerca de 10 mg, mais preferivelmente cerca de 0,01 a cerca de 5 mg, mais preferivelmente cerca de 0,025 a cerca de 1,5 mg, mais preferivelmente cerca de 0,05 a cerca de 1,25 mg, mais preferivelmente cerca de 0,075 a cerca de 1,25 mg, mais preferivelmente cerca de 0,1 a cerca de 1,25 mg, ainda mais
20 preferivelmente cerca de 0,15 a cerca de 1,25 mg, e acima de tudo preferivelmente cerca de 0,25 a cerca de 1,0 mg por g de material celulósico.

Em um outro aspecto preferido, uma quantidade eficiente de polipeptídeo(s) com atividade celobiohidrolase na proteína celulolítica é cerca de 0,005 a cerca de 1,0 g, preferivelmente em cerca de 0,01 a cerca de 1,0 g, mais preferivelmente em cerca de 0,15 a cerca de 0,75 g, mais preferivelmente
25 preferivelmente cerca de 0,15 a cerca de 0,5 g, mais preferivelmente cerca de 0,1 a cerca de 0,5 g, ainda mais preferivelmente cerca de 0,1 a cerca de 0,5 g, e acima de tudo preferivelmente cerca de 0,05 a cerca de 0,2 g por g de proteína celulolítica.

As enzimas podem ser derivadas ou obtidas de qualquer origem adequada, incluindo, origem bacteriana, fúngica, levedura, vegetal ou de mamífero. O termo “obtida” significa aqui que a enzima pode ter sido isolada de um organismo que produz naturalmente a enzima como uma natural. O termo “obtido” também significa aqui que a enzima pode ter sido produzida recombinantemente em um organismo hospedeiro que emprega métodos aqui descritos, em que a enzima produzida de maneira recombinante é tanto natural ou estrangeira ao organismo hospedeiro quanto apresenta uma sequência modificada de aminoácidos, por exemplo, com um ou mais (vários) aminoácidos que são deletados, inseridos e/ou substituídos, isto é, uma enzima produzida de maneira recombinante que é um mutante e/ou um fragmento de uma sequência natural de aminoácidos ou uma enzima produzida por processos de embaralhamento de ácido nucleico conhecidos na técnica. São incluídos no significado de uma enzima natural os variantes naturais, e no significado de uma enzima estrangeira, os variantes obtidos de maneira recombinante, tal como por mutagênese sítio direcionada ou embaralhamento.

O polipeptídeo com atividade de enzima pode ser um polipeptídeo bacteriano. Por exemplo, o polipeptídeo pode ser um polipeptídeo de bactéria Gram-positiva, tal como um polipeptídeo com atividade de enzima de *Bacillus*, *Streptococcus*, *Streptomyces*, *Staphylococcus*, *Enterococcus*, *Lactobacillus*, *Lactococcus*, *Clostridium*, *Geobacillus*, ou *Oceanobacillus*, ou um polipeptídeo de bactéria Gram-negativa, tal como um polipeptídeo com atividade de enzima de *E. coli*, *Pseudomonas*, *Salmonella*, *Campylobacter*, *Helicobacter*, *Flavobacterium*, *Fusobacterium*, *Ilyobacter*, *Neisseria*, ou *Ureaplasma*.

Em um aspecto preferido, o polipeptídeo é um polipeptídeo com atividade de enzima de *Bacillus alkalophilus*, *Bacillus amyloliquefaciens*, *Bacillus brevis*, *Bacillus circulans*, *Bacillus clausii*,

Bacillus coagulans, *Bacillus firmus*, *Bacillus lautus*, *Bacillus lentus*, *Bacillus licheniformis*, *Bacillus megaterium*, *Bacillus pumilus*, *Bacillus stearothermophilus*, *Bacillus subtilis*, ou *Bacillus thuringiensis*.

Em um outro aspecto preferido, o polipeptídeo é um polipeptídeo com atividade de enzima de *Streptococcus equisimilis*,
5 *Streptococcus pyogenes*, *Streptococcus uberis*, ou *Streptococcus equi* subsp. *Zooepidemicus*.

Em um outro aspecto preferido, o polipeptídeo é um polipeptídeo com atividade de enzima de *Streptomyces achromogenes*,
10 *Streptomyces avermitilis*, *Streptomyces coelicolor*, *Streptomyces griseus*, ou *Streptomyces lividans*.

O polipeptídeo com atividade de enzima também pode ser um polipeptídeo de fungo, e mais preferivelmente um polipeptídeo de levedura, tal como um polipeptídeo com atividade de enzima de *Candida*,
15 *Kluyveromyces*, *Pichia*, *Saccharomyces*, *Schizosaccharomyces*, ou *Yarrowia*; ou mais preferivelmente um polipeptídeo de fungo filamentoso, tal como um polipeptídeo com atividade de enzima de *Acremonium*, *Agaricus*, *Alternaria*, *Aspergillus*, *Aureobasidium*, *Botryosphaeria*, *Ceriporiopsis*, *Chaetomidium*, *Chrysosporium*, *Claviceps*, *Cochliobolus*, *Coprinopsis*, *Coptotermes*,
20 *Corynascus*, *Cryphonectria*, *Cryptococcus*, *Diplodia*, *Exidia*, *Filibasidium*, *Fusarium*, *Gibberella*, *Holomastigotoides*, *Humicola*, *Irpex*, *Lentinula*, *Leptosphaeria*, *Magnaporthe*, *Melanocarpus*, *Meripilus*, *Mucor*, *Myceliophthora*, *Neocallimastix*, *Neurospora*, *Paecilomyces*, *Penicillium*, *Phanerochaete*, *Piromyces*, *Poitrasia*, *Pseudoplectania*,
25 *Pseudotrichonympha*, *Rhizomucor*, *Schizophyllum*, *Scytalidium*, *Talaromyces*, *Thermoascus*, *Thielavia*, *Tolypocladium*, *Trichoderma*, *Trichophaea*, *Verticillium*, *Volvariella*, ou *Xylaria*.

Em um aspecto preferido, o polipeptídeo é um polipeptídeo com atividade de enzima de *Saccharomyces carlsbergensis*, *Saccharomyces*

cerevisiae, *Saccharomyces diastaticus*, *Saccharomyces douglasii*, *Saccharomyces kluyveri*, *Saccharomyces norbensis*, ou *Saccharomyces oviformis*.

Em um outro aspecto preferido, o polipeptídeo é um polipeptídeo com atividade de enzima de *Acremonium cellulolyticus*, *Aspergillus aculeatus*, *Aspergillus awamori*, *Aspergillus fumigatus*, *Aspergillus foetidus*, *Aspergillus japonicus*, *Aspergillus nidulans*, *Aspergillus niger*, *Aspergillus oryzae*, *Chrysosporium keratinophilum*, *Chrysosporium lucknowense*, *Chrysosporium tropicum*, *Chrysosporium merdarium*, *Chrysosporium inops*, *Chrysosporium pannicola*, *Chrysosporium queenslandicum*, *Chrysosporium zonatum*, *Fusarium bactridioides*, *Fusarium cerealis*, *Fusarium crookwellense*, *Fusarium culmorum*, *Fusarium graminearum*, *Fusarium graminum*, *Fusarium heterosporum*, *Fusarium negundi*, *Fusarium oxysporum*, *Fusarium reticulatum*, *Fusarium roseum*, *Fusarium sambucinum*, *Fusarium sarcochroum*, *Fusarium sporotrichioides*, *Fusarium sulphureum*, *Fusarium torulosum*, *Fusarium trichothecioides*, *Fusarium venenatum*, *Humicola grisea*, *Humicola insolens*, *Humicola lanuginosa*, *Irpex lacteus*, *Mucor miehei*, *Myceliophthora thermophila*, *Neurospora crassa*, *Penicillium funiculosum*, *Penicillium purpurogenum*, *Phanerochaete chrysosporium*, *Thielavia achromatica*, *Thielavia albomyces*, *Thielavia albopilosa*, *Thielavia australeinsis*, *Thielavia fimeti*, *Thielavia microspora*, *Thielavia ovispora*, *Thielavia peruviana*, *Thielavia spededonium*, *Thielavia setosa*, *Thielavia subthermophila*, *Thielavia terrestris*, *Trichoderma harzianum*, *Trichoderma koningii*, *Trichoderma longibrachiatum*, *Trichoderma reesei*, *Trichoderma viride*, ou *Trichophaea saccata*.

Mutantes dos polipeptídeos com atividade de enzima quimicamente modificados ou modificados por engenharia protéica também podem ser usados.

Um ou mais (vários) componentes da composição de enzima

pode ser um componente recombinante, isto é, produzido por clonagem de uma sequência de DNA que codifica o componente único e a subsequente célula transformada com a sequência de DNA e expressa em um hospedeiro (ver, por exemplo, WO 91/17243 e WO 91/17244). O hospedeiro é preferivelmente um hospedeiro heterólogo (a enzima é estrangeira ao hospedeiro), mas o hospedeiro também pode, em certas condições, ser um hospedeiro homólogo (a enzima é natural do hospedeiro). As proteínas celulolíticas monocomponentes também podem ser preparadas purificando uma proteína como esta a partir de um caldo de fermentação.

Exemplos de preparações de proteína celulolítica comerciais adequadas para uso na presente invenção incluem, por exemplo, CELLIC™ Ctec (Novozymes A/S), CELLUCLAST™ (Novozymes A/S), NOVOZYM™ 188 (Novozymes A/S), CELLUZYME™ (Novozymes A/S), CEREFLO™ (Novozymes A/S), e ULTRAFLO™ (Novozymes A/S), ACCELERASE™ (Genencor Int.), LAMINEX™ (Genencor Int.), SPEZYME™ CP (Genencor Int.), ROHAMENT™ 7069 W (Röhm GmbH), FIBREZYME® LDI (Dyadic International, Inc.), FIBREZYME® LBR (Dyadic International, Inc.), ou VISCOSTAR® 150L (Dyadic International, Inc.). As enzimas celulase são adicionadas em quantidades eficiente de cerca de 0,001 a cerca de 5,0 % em peso de sólidos, mais preferivelmente de cerca de 0,025 a cerca de 4,0 % em peso de sólidos, e acima de tudo preferivelmente de cerca de 0,005 a cerca de 2,0 % em peso de sólidos. As enzimas celulase são adicionadas em quantidades eficientes de cerca de 0,001 a cerca de 5,0 % em peso de sólidos, mais preferivelmente de cerca de 0,025 a cerca de 4,0 % em peso de sólidos, e mais preferivelmente de cerca de 0,005 a cerca de 2,0 % em peso de sólidos.

Exemplos de endoglicanases bacterianas que podem ser usadas nos métodos da presente invenção, incluem, mas sem limitação, uma endoglicanase de *Acidothermus cellulolyticus* (WO 91/05039; WO 93/15186; patente U.S. 5.275.944; WO 96/02551; patente U.S. 5.536.655, WO

00/70031, WO 05/093050); endoglicanase III de *Thermobifida fusca* (WO 05/093050) e endoglicanase V de *Thermobifida fusca* (WO 05/093050).

Exemplos de endoglicanases fúngicas que podem ser usadas nos métodos da presente invenção, incluem, mas sem limitação, uma

5 endoglicanase I de *Trichoderma reesei* (Penttilä *et al.*, 1986, *Gene* 45: 253-263; GENBANK™ número de acesso M15665); endoglicanase II de *Trichoderma reesei* (Saloheimo, *et al.*, 1988, *Gene* 63:11-22; GENBANK™ número de acesso M19373); endoglicanase III de *Trichoderma reesei* (Okada

10 *et al.*, 1988, *Appl. Environ. Microbiol.* 64: 555-563; GENBANK™ número de acesso AB003694) e endoglicanase V de *Trichoderma reesei* (Saloheimo *et al.*, 1994, *Molecular Microbiology* 13: 219-228; GENBANK™ número de acesso Z33381); endoglicanase de *Aspergillus aculeatus* (Ooi *et al.*, 1990, *Nucleic Acids Research* 18: 5884); endoglicanase de *Aspergillus kawachii* (Sakamoto *et al.*, 1995, *Current Genetics* 27: 435-439); endoglicanase de

15 *Erwinia carotovora* (Saarilahti *et al.*, 1990, *Gene* 90: 9-14); endoglicanase de *Fusarium oxysporum* (GENBANK™ número de acesso L29381); endoglicanase de *Humicola grisea* var. *thermoidea* (GENBANK™ número de acesso AB003107); endoglicanase de *Melanocarpus albomyces* (GENBANK™ número de acesso MAL515703); endoglicanase de

20 *Neurospora crassa* (GENBANK™ número de acesso XM_324477); endoglicanase V de *Humicola insolens*; endoglicanase de CBS 117.65 *Myceliophthora thermophila*; endoglicanase de CBS 495.95 basidiomiceto; endoglicanase de CBS 494.95 basidiomiceto; endoglicanase de NRRL 8126 CEL6B *Thielavia terrestris*; endoglicanase de NRRL 8126 CEL6C *Thielavia*

25 *terrestris*; endoglicanase de NRRL 8126 CEL7C *Thielavia terrestris*; endoglicanase de NRRL 8126 CEL7E *Thielavia terrestris*; endoglicanase de NRRL 8126 CEL7F *Thielavia terrestris*; endoglicanase de ATCC 62373 CEL7A *Cladorrhinum foecundissimum*; e endoglicanase de *Trichoderma reesei* cepa número VTT-D-80133 (GENBANK™ número de acesso

M15665).

Exemplos de celobiohidrolases usadas nos métodos da presente invenção incluem, mas sem limitação, celobiohidrolase I de *Trichoderma reesei*; celobiohidrolase II de *Trichoderma reesei*;
 5 celobiohidrolase I de *Humicola insolens*, celobiohidrolase II de *Myceliophthora thermophila*, celobiohidrolase II de *Thielavia terrestris* (CEL6A), celobiohidrolase I de *Chaetomium thermophilum* e celobiohidrolase II de *Chaetomium thermophilum*.

Exemplos de beta-glicosidases usadas nos métodos da presente
 10 invenção incluem, mas sem limitação, beta-glicosidase de *Aspergillus oryzae*; beta-glicosidase de *Aspergillus fumigatus*; beta-glicosidase de *Penicillium brasilianum* IBT 20888; beta-glicosidase de *Aspergillus niger* e beta-glicosidase de *Aspergillus aculeatus*.

O polipeptídeo de *Aspergillus oryzae* com atividade de beta-
 15 glicosidase pode ser obtido de acordo com WO 2002/095014. O polipeptídeo de *Aspergillus fumigatus* com atividade de beta-glicosidase pode ser obtido de acordo com WO 2005/047499. O polipeptídeo de *Penicillium brasilianum* com atividade de beta-glicosidase pode ser obtido de acordo com WO 2007/019442. O polipeptídeo de *Aspergillus niger* com atividade de beta-
 20 glicosidase pode ser obtido de acordo com Dan *et al.*, 2000, *J. Biol. Chem.* 275: 4973-4980. O polipeptídeo de *Aspergillus aculeatus* com atividade de beta-glicosidase pode ser obtido de acordo com Kawaguchi *et al.*, 1996, *Gene* 173: 287-288.

A beta-glicosidase pode ser uma proteína de fusão. Em um
 25 aspecto, a beta-glicosidase é a proteína de fusão BG variante de beta-glicosidase de *Aspergillus oryzae* ou a proteína de fusão beta-glicosidase de *Aspergillus oryzae* obtida de acordo com WO 2008/057637.

Outras endoglicanases, celobiohidrolases e beta-glicosidases são reveladas em várias famílias de glicosil hidrolase usando a classificação

de acordo com Henrissat B., 1991, A classification of glycosyl hydrolases based on amino-acido sequence similarities, *Biochem. J.* 280: 309-316, e Henrissat B., e Bairoch A., 1996, Updating the sequence-based classification of glycosyl hydrolases, *Biochem. J.* 316: 695-696.

5 Outras enzima celulolíticas que podem ser usadas na presente invenção são descritas em EP 495,257, EP 531,315, EP 531,372, WO 89/09259, WO 94/07998, WO 95/24471, WO 96/11262, WO 96/29397, WO 96/034108, WO 97/14804, WO 98/08940, WO 98/012307, WO 98/13465, WO 98/015619, WO 98/015633, WO 98/028411, WO 99/06574, WO 10 99/10481, WO 99/025846, WO 99/025847, WO 99/031255, WO 2000/009707, WO 2002/050245, WO 2002/0076792, WO 2002/101078, WO 2003/027306, WO 2003/052054, WO 2003/052055, WO 2003/052056, WO 2003/052057, WO 2003/052118, WO 2004/016760, WO 2004/043980, WO 2004/048592, WO 2005/001065, WO 2005/028636, WO 2005/093050, WO 15 2005/093073, WO 2006/074005, WO 2006/117432, WO 2007/071818, WO 2007/071820, WO 2008/008070, WO 2008/008793, patente U.S. 4.435.307, patente U.S. 5.457.046, patente U.S. 5.648.263, patente U.S. 5.686.593, patente U.S. 5.691.178, patente U.S. 5.763.254 e patente U.S. 5.776.757.

20 Nos métodos da presente invenção, qualquer polipeptídeo GH61 com melhor atividade celulolítica pode ser usado.

Em um primeiro aspecto, o polipeptídeo com melhor atividade celulolítica compreende os motivos a seguir:

[ILMV]-P-X(4,5)-G-X-Y-[ILMV]-X-R-X-[EQ]-X(4)-[HNQ] e [FW]-[TF]-K-[AIV],

25 em que X é qualquer aminoácido, X(4,5) é qualquer aminoácido nas posições contíguas 4 ou 5, e X(4) é qualquer aminoácido nas posições contíguas 4.

O polipeptídeo que compreende os motivos anteriormente observados podem compreender adicionalmente:

H-X(1,2)-G-P-X(3)-[YW]-[AILMV],
 [EQ]-X-Y-X(2)-C-X-[EHQN]-[FILV]-X-[ILV], ou
 H-X(1,2)-G-P-X(3)-[YW]-[AILMV] e [EQ]-X-Y-X(2)-C-X-
 [EHQN]-[FILV]-X-[ILV],

5 em que X é qualquer aminoácido, X(1,2) é qualquer aminoácido na posição 1 ou posições contíguas 2, X(3) é qualquer aminoácido nas posições contíguas 3, e X(2) é qualquer aminoácido nas posições contíguas 2. Nos motivos anteriores, a abreviação do aminoácido de única letra IUPAC aceita é empregada.

10 Em um aspecto preferido, o polipeptídeo com melhor atividade celulolítica compreende adicionalmente H-X(1,2)-G-P-X(3)-[YW]-[AILMV]. Em um outro aspecto preferido, o polipeptídeo isolado com melhor atividade celulolítica compreende adicionalmente [EQ]-X-Y-X(2)-C-X-[EHQN]-[FILV]-X-[ILV]. Em um outro aspecto preferido, o polipeptídeo com melhor
 15 atividade celulolítica compreende adicionalmente H-X(1,2)-G-P-X(3)-[YW]-[AILMV] e [EQ]-X-Y-X(2)-C-X-[EHQN]-[FILV]-X-[ILV].

Em um segundo aspecto, o polipeptídeo com melhor atividade celulolítica compreende o motivo a seguir:

[ILMV]-P-x(4,5)-G-x-Y-[ILMV]-x-R-x-[EQ]-x(3)-UM-[HNQ],

20 em que x é qualquer aminoácido, x(4,5) é qualquer aminoácido nas posições contíguas 4 ou 5, e x(3) é qualquer aminoácido nas posições contíguas 3. No motivo anterior, a abreviação do aminoácido de única letra IUPAC aceita é empregada.

Exemplos de polipeptídeos com melhor atividade celulolítica
 25 usados nos métodos da presente invenção incluem, mas sem limitação, polipeptídeos com melhor atividade celulolítica de *Thielavia terrestris* (WO 2005/074647); polipeptídeos com melhor atividade celulolítica de *Thermoascus aurantiacus* (WO 2005/074656); polipeptídeos com melhor atividade celulolítica de *Trichoderma reesei* (WO 2007/089290); e

polipeptídeos com melhor atividade celulolítica de *Myceliophthora thermophila* (WO 2009/085935; WO 2009/085859; WO 2009/085864; WO 2009/085868).

Exemplos de preparações de enzima hemicelulolítica comerciais que (por exemplo, que degradam xilano) adequadas para uso na presente invenção incluem, por exemplo, SHEARZYME™ (Novozymes A/S), CELLIC™ Htec (Novozymes A/S), VISCOZYME® (Novozymes A/S), ULTRAFLO® (Novozymes A/S), PULPZYME® HC (Novozymes A/S), MULTIFECT® Xilanase (Genencor), ECOPULP® TX-200A (AB Enzimas), HSP 6000 Xilanase (DSM), DEPOL™ 333P (Biocatalysts Limit, Wales, UK), DEPOL™ 740L (Biocatalysts Limit, Wales, UK), e DEPOL™ 762P (Biocatalysts Limit, Wales, UK).

Exemplos de xilanases usadas nos métodos da presente invenção incluem, mas sem limitação, xilanase de *Aspergillus aculeatus* (GeneSeqP:AAR63790; WO 94/21785), xilanases de *Aspergillus fumigatus* (WO 2006/078256) e xilanases de *Thielavia terrestris* NRRL 8126 (WO 2009/079210).

Exemplos de beta-xilosidases usadas nos métodos da presente invenção incluem, mas sem limitação, beta-xilosidase de *Trichoderma reesei* (UniProtKB/TrEMBL número de acesso Q92458), *Talaromyces emersonii* (SwissProt número de acesso Q8X212), e *Neurospora crassa* (SwissProt número de acesso Q7SOW4).

Exemplos de acetil-xilano esterases usadas nos métodos da presente invenção incluem, mas sem limitação, acetil-xilano esterase de *Hypocrea jecorina* (WO 2005/001036), acetil-xilano esterase de *Neurospora crassa* (UniProt número de acesso q7s259), acetil-xilano esterase de *Thielavia terrestris* NRRL 8126 (WO 2009/042846), acetil-xilano esterase de *Chaetomium globosum* (Uniprot número de acesso Q2GWX4), acetil-xilano esterase de *Chaetomium gracile* (GeneSeqP número de acesso AAB82124),

acetil-xilano esterase de *Phaeosphaeria nodorum* (Uniprot número de acesso Q0UHI1), e acetil-xilano esterase de *Humicola insolens* DSM 1800 (WO 2009/073709).

Exemplos de esterases de ácido ferílico usadas nos métodos da presente invenção incluem, mas sem limitação, feruloil esterase de *Humicola insolens* DSM 1800 (WO 2009/076122), feruloil esterase de *Neurospora crassa* (UniProt número de acesso Q9HGR3) e feruloil esterase de *Neosartorya fischeri* (UniProt Número de acesso A1D9T4).

Exemplos de arabinofuranosidases usadas nos métodos da presente invenção incluem, mas sem limitação, arabinofuranosidase de *Humicola insolens* DSM 1800 (WO 2009/073383) e arabinofuranosidase de *Aspergillus niger* (GeneSeqP número de acesso AAR94170).

Exemplos de alfa-glicuronidases usadas nos métodos da presente invenção incluem, mas sem limitação, alfa-glicuronidase de *Aspergillus clavatus* (UniProt número de acesso alcc12), alfa-glicuronidase de *Trichoderma reesei* (Uniprot número de acesso Q99024), alfa-glicuronidase de *Talaromyces emersonii* (UniProt número de acesso Q8X211), alfa-glicuronidase de *Aspergillus niger* (Uniprot número de acesso Q96WX9), alfa-glicuronidase de *Aspergillus terreus* (SwissProt número de acesso Q0CJP9) e alfa-glicuronidase de *Aspergillus fumigatus* (SwissProt número de acesso Q4WW45).

As enzimas e proteínas usadas nos métodos da presente invenção podem ser produzidas por fermentação das cepas microbianas observadas anteriormente em um meio nutriente contendo fontes adequadas de carbono e nitrogênio e sais inorgânicos, usando os procedimentos conhecidos na técnica (ver, por exemplo, Bennett, J.W. e LaSure, L. (eds.), *More Gene Manipulations in Fungi*, Academic Press, CA, 1991). Os meios adequados são disponíveis de fornecedores comerciais ou podem ser preparados de acordo com composições publicadas (por exemplo, em

catálogos do American Type Culture Collection). As faixas de temperatura e outras condições adequadas para o crescimento e produção de enzima são conhecidos na técnica (ver, por exemplo, Bailey, J.E., e Ollis, D.F., *Biochemical Engineering Fundamentals*, McGraw-Hill Book Company, NY, 1986).

A fermentação pode ser qualquer método de cultivo de uma célula que resulta na expressão ou isolamento de uma enzima. A fermentação, portanto, pode ser entendida como aquela que compreende cultivo em frasco de agitação, ou fermentação em pequena ou grande escala (incluindo fermentações contínuas, em lotes, em lote alimentado ou de estado sólido) em laboratório ou fermentadores industriais, realizada em um meio adequado e em condições que permitem que a enzima seja expressa ou isolada. As enzimas resultantes produzidas pelos métodos descritos anteriormente podem ser recuperadas do meio de fermentação e purificadas por procedimentos convencionais.

Fermentação. Os açúcares fermentáveis obtidos do material celulósico hidrolisado podem ser fermentados por um ou mais (vários) microrganismos fermentadores capazes de fermentar o açúcares direta ou indiretamente em um produto de fermentação desejado. “Fermentação” ou “processo de fermentação” refere-se a qualquer processo de fermentação ou qualquer processo que compreende uma etapa de fermentação. Processos de fermentação também incluem processos de fermentação usados na indústria de consumo de álcool (por exemplo, cerveja e vinho), indústria de panificação (por exemplo, produtos de panificação fermentados), indústria de couro e indústria do tabaco. As condições de fermentação dependem do produto de fermentação desejado e do organismo fermentador, e podem ser facilmente determinadas pelos versados na técnica.

Na etapa de fermentação, os açúcares, liberados do material celulósico como um resultado das etapas de pré-tratamento e hidrólise

enzimática e, são fermentados em um produto, por exemplo, etanol, por um organismo fermentador, tal como levedura. A hidrólise (sacarificação) e fermentação podem ser separadas ou simultâneas, da maneira descrita aqui.

5 Qualquer material celulósico hidrolisado adequado pode ser usado na etapa de fermentação na prática da presente invenção. O material é em geral selecionado com base no produto de fermentação desejado, isto é, a substância a ser obtida da fermentação, e no processo empregado, da maneira bem conhecida na técnica.

10 Entende-se aqui que o termo “meio de fermentação” refere-se a um meio antes do(s) microrganismo(s) fermentador(s) ser(m) adicionado(s), tal como um meio que resulta de um processo de sacarificação, bem como um meio usado em um processo de sacarificação e fermentação simultânea (SSF).

“Microrganismo fermentador” refere-se a qualquer microrganismo, incluindo organismos bacterianos e fúngicos, adequados para
15 uso em um processo de fermentação desejado para produzir um produto de fermentação. O organismo fermentador pode ser organismos fermentadores de C_6 e/ou C_5 , ou uma combinação destes. Tanto os organismos fermentadores C_6 quanto C_5 são bem conhecidos na técnica. Os microrganismos fermentadores adequados são capazes de fermentar, isto é, converteres, açúcares, tais como
20 glicose, xilose, xilulose, arabinose, maltose, manose, galactose, ou oligossacarídeos, direta ou indiretamente no produto de fermentação desejado.

Exemplos de organismos fermentadores bacterianos e fúngicos que produzem etanol são descritos por Lin *et al.*, 2006, *Appl. Microbiol. Biotechnol.* 69: 627-642.

25 Exemplos de microrganismos fermentadores que podem fermentar açúcares C_6 incluem organismos bacterianos e fúngicos, tal como levedura. As leveduras preferidas incluem cepas de *Saccharomyces* spp., preferivelmente *Saccharomyces cerevisiae*.

Exemplos de organismos fermentadores que podem fermentar

açúcares C₅ incluem organismos bacterianos e fúngicos, tal como levedura. As leveduras que fermentam C₅ preferidas incluem cepas de *Pichia*, preferivelmente *Pichia stipitis*, tal como *Pichia stipitis* CBS 5773; cepas de *Candida*, preferivelmente *Candida boidinii*, *Candida brassicae*, *Candida sheatae*, *Candida diddensii*, *Candida pseudotropicalis*, ou *Candida utilis*.

Outros organismos fermentadores incluem cepas de *Zymomonas*, tal como *Zymomonas mobilis*; *Hansenula*, tal como *Hansenula anomala*; *Kluyveromyces*, tal como *K. fragilis*; *Schizosaccharomyces*, tal como *S. pombe*; e *E. coli*, especialmente cepas de *E. coli* que foram geneticamente modificadas para melhorar o rendimento de etanol.

Em um aspecto preferido, a levedura é uma *Saccharomyces* spp. Em um aspecto mais preferido, a levedura é *Saccharomyces cerevisiae*. Em um outro aspecto mais preferido, a levedura é *Saccharomyces distaticus*. Em um outro aspecto mais preferido, a levedura é *Saccharomyces uvarum*. Em um outro aspecto preferido, a levedura é uma *Kluyveromyces*. Em um outro aspecto mais preferido, a levedura é *Kluyveromyces marxianus*. Em um outro aspecto mais preferido, a levedura é *Kluyveromyces fragilis*. Em um outro aspecto preferido, a levedura é uma *Candida*. Em um outro aspecto mais preferido, a levedura é *Candida boidinii*. Em um outro aspecto mais preferido, a levedura é *Candida brassicae*. Em um outro aspecto mais preferido, a levedura é *Candida diddensii*. Em um outro aspecto mais preferido, a levedura é *Candida pseudotropicalis*. Em um outro aspecto mais preferido, a levedura é *Candida utilis*. Em um outro aspecto preferido, a levedura é uma *Clavispora*. Em um outro aspecto mais preferido, a levedura é *Clavispora lusitaniae*. Em um outro aspecto mais preferido, a levedura é *Clavispora opuntiae*. Em um outro aspecto preferido, a levedura é uma *Pachysolen*. Em um outro aspecto mais preferido, a levedura é *Pachysolen tannophilus*. Em um outro aspecto preferido, a levedura é uma *Pichia*. Em um outro aspecto mais preferido, a levedura é uma *Pichia stipitis*. Em um outro

aspecto preferido, a levedura é uma *Bretannomyces*. Em um outro aspecto mais preferido, a levedura é *Bretannomyces clausenii* (Philippidis, G. P., 1996, Cellulose bioconversion technology, em *Handbook on Bioethanol: Production and Utilization*, Wyman, C. E., ed., Taylor & Francis, Washington, DC, 179-212).

As bactérias que podem fermentar de maneira eficiente a hexose e pentose em etanol incluem, por exemplo, *Zymomonas mobilis* e *Clostridium thermocellum* (Philippidis, 1996, *supra*).

Em um aspecto preferido, a bactéria é uma *Zymomonas*. Em um aspecto mais preferido, a bactéria é *Zymomonas mobilis*. Em um outro aspecto preferido, a bactéria é um *Clostridium*. Em um outro aspecto mais preferido, a bactéria é *Clostridium thermocellum*.

As leveduras comercialmente disponíveis para a produção de etanol incluem, por exemplo, levedura ETANOL RED™ (disponível pela Fermentis/Lesaffre, USA), FALI™ (disponível pela Fleischmann's Yeast, USA), levedura fresca SUPERSTART™ e THERMOSACC™ (disponível pela Ethanol Technology, WI, USA), BIOFERM™ AFT e XR (disponível pela NABC - North American Bioproducts Corporation, GA, USA), GERT STRAND™ (disponível pela Gert Strand AB, Suécia), e FERMIOL™ (disponível pela DSM Specialties).

Em um aspecto preferido, o microrganismo fermentador foi geneticamente modificado para fornecer a capacidade de fermentar açúcares pentose, tais como como microrganismos que utilizam xilose, que utilizam arabinose e que co-utilizam xilose e arabinose.

A clonagem de genes heterólogos em vários microrganismos fermentadores levou à construção de organismos capazes de converter hexoses e pentoses em etanol (co-fermentação) (Chen e Ho, 1993, Cloning and improving the expression of *Pichia stipitis* xylose reductase gene in *Saccharomyces cerevisiae*, *Appl. Biochem. Biotechnol.* 39-40: 135-147; Ho et

al., 1998, Genetically engineered *Saccharomyces* yeast capable of effectively
 cofermenting glucose and xylose, *Appl. Environ. Microbiol.* 64: 1852-1859;
 Kotter e Ciriacy, 1993, Xylose fermentation by *Saccharomyces cerevisiae*,
Appl. Microbiol. Biotechnol. 38: 776-783; Walfridsson *et al.*, 1995, Xylose-
 5 metabolizing *Saccharomyces cerevisiae* strains overexpressing the TKL1 and
 TAL1 genes encoding the pentose fosfato pathway enzymes transketolase and
 transaldolase, *Appl. Environ. Microbiol.* 61: 4184-4190; Kuyper *et al.*, 2004,
 Minimal metabolic engineering of *Saccharomyces cerevisiae* for efficient
 anaerobic xylose fermentation: a proof of principle, *FEMS Yeast Research* 4:
 10 655-664; Beall *et al.*, 1991, Parametric studies of ethanol production from
 xylose and other sugars by recombinant *Escherichia coli*, *Biotech. Bioeng.* 38:
 296-303; Ingram *et al.*, 1998, Metabolic engineering of bacteria for ethanol
 production, *Biotechnol. Bioeng.* 58: 204-214; Zhang *et al.*, 1995, Metabolic
 engineering of a pentose metabolism pathway in ethanologenic *Zymomonas*
 15 *mobilis*, *Science* 267: 240-243; Deanda *et al.*, 1996, Development of an
 arabinose-fermenting *Zymomonas mobilis* strain by metabolic pathway
 engineering, *Appl. Environ. Microbiol.* 62: 4465-4470; WO 2003/062430,
 xilose isomerase).

Em um aspecto preferido, o microrganismo fermentador
 20 geneticamente modificado é *Saccharomyces cerevisiae*. Em um outro aspecto
 preferido, o microrganismo fermentador geneticamente modificado é
Zymomonas mobilis. Em um outro aspecto preferido, o microrganismo
 fermentador geneticamente modificado é *Escherichia coli*. Em um outro
 aspecto preferido, o microrganismo fermentador geneticamente modificado é
 25 *Klebsiella oxytoca*. Em um outro aspecto preferido, o microrganismo
 fermentador geneticamente modificado é *Kluyveromyces* sp.

Sabe-se bem na técnica que os organismos descritos
 anteriormente também podem ser usados para produzir outras substâncias, da
 maneira descrita aqui.

O microrganismo fermentador é tipicamente adicionado à lignocelulose degradada ou hidrolisado e a fermentação é realizada por cerca de 8 a cerca de 96 horas, tal como cerca de 24 a cerca de 60 horas. A temperatura está tipicamente entre cerca de 26 °C a cerca de 60 °C, em particular cerca de 32 °C ou 50 °C, e em cerca de pH 3 a cerca de pH 8, tal como em torno de pH 4-5, 6, ou 7.

Em um aspecto preferido, a levedura e/ou um outro microrganismo é aplicado ao material celulósico degradado e a fermentação é realizada por cerca de 12 a cerca de 96 horas, tal como tipicamente 24-60 horas. Em um aspecto preferido, o temperatura está preferivelmente entre cerca de 20 °C a cerca de 60 °C, mais preferivelmente cerca de 25 °C a cerca de 50 °C, e acima de tudo preferivelmente cerca de 32 °C a cerca de 50 °C, em particular cerca de 32 °C ou 50 °C, e o pH está em geral em cerca de pH 3 a cerca de pH 7, preferivelmente em torno de pH 4-7. Entretanto, alguns organismos fermentadores, por exemplo, bactérias, apresentam maior temperatura ideal de fermentação. A levedura ou um outro microrganismo que preferivelmente aplicado em quantidades de aproximadamente 10^5 a 10^{12} , de maneira preferível de aproximadamente 10^7 a 10^{10} , de maneira essencial aproximadamente contagens de 2×10^8 células viáveis por mL de caldo de fermentação. Diretriz adicional com relação ao uso de levedura para a fermentação pode ser encontrada em, por exemplo, "O Alcohol Textbook" (Editors K. Jacques, T.P. Lyons e D.R. Kelsall, Nottingham University Press, Reino Unido 1999), que é aqui incorporado pela referência.

Para a produção de etanol, após a fermentação, a lama fermentada é destilada para extrair o etanol. O etanol obtido de acordo com os métodos da invenção pode ser usado, por exemplo, como combustível etanol, etanol para bebidas, isto é, espíritos neutros potáveis ou etanol industrial.

Um estimulador de fermentação pode ser usado em combinação com qualquer dos processos aqui descritos para melhorar

adicionalmente o processo de fermentação e, em particular, o desempenho do microrganismo fermentador, tal como, melhoria da taxa e rendimento de etanol. Um “estimulador de fermentação” refere-se aos estimuladores para crescimento dos microrganismos fermentadores, em particular, leveduras. Os

5 estimuladores de fermentação preferíveis para crescimento incluem vitaminas e minerais. Exemplos de vitaminas incluem multivitaminas, biotina, pantotenato, ácido nicotínico, meso-inositol, tiamina, piridoxina, ácido para-aminobenzóico, ácido fólico, riboflavina e vitaminas A, B, C, D, e E. Ver, por exemplo, Alfenore *et al.*, Improving ethanol production and viability of

10 *Saccharomyces cerevisiae* by a vitamin feeding strategy during fed-batch process, Springer-Verlag (2002), que é aqui incorporado pela referência. Exemplos de minerais incluem minerais e sais minerais que podem suprir nutrientes que compreendem P, K, Mg, S, Ca, Fe, Zn, Mn, e Cu.

Produtos de fermentação: Um produto de fermentação pode

15 ser qualquer substância derivada da fermentação. O produto de fermentação pode ser, sem limitação, um álcool (por exemplo, arabinitol, butanol, etanol, glicerol, metanol, 1,3-propanediol, sorbitol e xilitol); um ácido orgânico (por exemplo, ácido acético, ácido acetônico, ácido adípico, ácido ascórbico, ácido cítrico, ácido 2,5-diceto-D-glucônico, ácido fórmico, ácido fumárico, ácido

20 glucárico, ácido glucônico, ácido glucurônico, ácido glutárico, ácido 3-hidroxipropiônico, ácido itacônico, ácido láctico, ácido málico, ácido malônico, ácido oxálico, ácido oxaloacético, ácido propiônico, ácido succínico e ácido xilônico); uma cetona (por exemplo, acetona); um aminoácido (por exemplo, ácido aspártico, ácido glutâmico, glicina, lisina,

25 serina e treonina) e um gás (por exemplo, metano, hidrogênio (H₂), dióxido de carbono (CO₂) e monóxido de carbono (CO)). O produto de fermentação também pode ser proteína como um produto de alto valor.

Em um aspecto preferido, o produto de fermentação é um álcool. Será entendido que o termo “álcool” inclui uma substância que contém

uma ou mais frações hidroxila. Em um aspecto mais preferido, o álcool é arabinitol. Em um outro aspecto mais preferido, o álcool é butanol. Em um outro aspecto mais preferido, o álcool é etanol. Em um outro aspecto mais preferido, o álcool é glicerol. Em um outro aspecto mais preferido, o álcool é metanol. Em um outro aspecto mais preferido, o álcool é 1,3-propanediol. Em um outro aspecto mais preferido, o álcool é sorbitol. Em um outro aspecto mais preferido, o álcool é xilitol. Ver, por exemplo, Gong, C. S., Cao, N. J., Du, J., e Tsao, G. T., 1999, Ethanol production from renewable resources, em *Advances em Biochemical Engineering/Biotechnology*, Scheper, T., ed., Springer-Verlag Berlin Heidelberg, Alemanha, 65: 207-241; Silveira, M. M., e Jonas, R., 2002, The biotechnological production of sorbitol, *Appl. Microbiol. Biotechnol.* 59: 400-408; Nigam, P., e Singh, D., 1995, Processs for fermentative production of xilitol – a sugar substitute, *Process Biochemistry* 30 (2): 117-124; Ezeji, T. C., Qureshi, N. e Blaschek, H. P., 2003, Production of acetone, butanol e ethanol by *Clostridium beijerinckii* BA101 e *in situ* recovery by gas stripping, *World Journal of Microbiology and Biotechnology* 19 (6): 595-603.

Em um outro aspecto preferido, o produto de fermentação é um ácido orgânico. Em um outro aspecto mais preferido, o ácido orgânico é ácido acético. Em um outro aspecto mais preferido, o ácido orgânico é ácido acetônico. Em um outro aspecto mais preferido, o ácido orgânico é ácido adípico. Em um outro aspecto mais preferido, o ácido orgânico é ácido ascórbico. Em um outro aspecto mais preferido, o ácido orgânico é ácido cítrico. Em um outro aspecto mais preferido, o ácido orgânico é ácido 2,5-diceto-D-glucônico. Em um outro aspecto mais preferido, o ácido orgânico é ácido fórmico. Em um outro aspecto mais preferido, o ácido orgânico é ácido fumárico. Em um outro aspecto mais preferido, o ácido orgânico é ácido glucárico. Em um outro aspecto mais preferido, o ácido orgânico é ácido glucônico. Em um outro aspecto mais preferido, o ácido orgânico é ácido

glucurônico. Em um outro aspecto mais preferido, o ácido orgânico é ácido glutárico. Em um outro aspecto preferido, o ácido orgânico é ácido 3-hidroxipropiônico. Em um outro aspecto mais preferido, o ácido orgânico é ácido itacônico. Em um outro aspecto mais preferido, o ácido orgânico é ácido láctico. Em um outro aspecto mais preferido, o ácido orgânico é ácido málico. Em um outro aspecto mais preferido, o ácido orgânico é ácido malônico. Em um outro aspecto mais preferido, o ácido orgânico é ácido oxálico. Em um outro aspecto mais preferido, o ácido orgânico é ácido propiônico. Em um outro aspecto mais preferido, o ácido orgânico é ácido sucínico. Em um outro aspecto mais preferido, o ácido orgânico é ácido xilônico. Ver, por exemplo, Chen, R., e Lee, Y. Y., 1997, Membrane-mediated extractive fermentation for lactic acid production from cellulosic biomass, *Appl. Biochem. Biotechnol.* 63-65: 435-448.

Em um outro aspecto preferido, o produto de fermentação é uma cetona. Será entendido que o termo “cetona” inclui uma substância que contém uma ou mais frações de cetona. Em um outro aspecto mais preferido, a cetona é acetona. Ver, por exemplo, Qureshi e Blaschek, 2003, *supra*.

Em um outro aspecto preferido, o produto de fermentação é um aminoácido. Em um outro aspecto mais preferido, o ácido orgânico é ácido aspártico. Em um outro aspecto mais preferido, o aminoácido é ácido glutâmico. Em um outro aspecto mais preferido, o aminoácido é glicina. Em um outro aspecto mais preferido, o aminoácido é lisina. Em um outro aspecto mais preferido, o aminoácido é serina. Em um outro aspecto mais preferido, o aminoácido é treonina. Ver, por exemplo, Richard, A., e Margaritis, A., 2004, Empirical modeling of batch fermentation kinetics for poly(glutamic acid) production and other microbial biopolymers, *Biotechnology e Bioengineering* 87 (4): 501-515.

Em um outro aspecto preferido, o produto de fermentação é um gás. Em um outro aspecto mais preferido, o gás é metano. Em um outro

aspecto mais preferido, o gás é H_2 . Em um outro aspecto mais preferido, o gás é CO_2 . Em um outro aspecto mais preferido, o gás é CO. Ver, por exemplo, Kataoka, N., A. Miya, e K. Kiriya, 1997, Studies on hydrogen production by continuous culture system of hydrogen-producing anaerobic bacteria, *Water Science and Technology* 36 (6-7): 41-47; e Gunaseelan V.N. em *Biomass and Bioenergy*, Vol. 13 (1-2), pp. 83-114, 1997, Anaerobic digestion of biomass for methane production: a review.

Recuperação. O(s) produto(s) de fermentação pode(m) ser opcionalmente recuperado(s) do meio de fermentação usando qualquer método conhecido na técnica incluindo, mas sem limitação, cromatografia, procedimentos eletroforéticos, solubilidade diferencial, destilação ou extração. Por exemplo, o álcool é separado do material celulósico fermentado e purificado por métodos convencionais de destilação. O etanol com uma pureza de até cerca de 96 vol. % pode ser obtido, o qual pode ser usado, por exemplo, como combustível etanol, etanol para bebidas, isto é, espíritos neutros potáveis ou etanol industrial.

Peptídeo sinal

A presente invenção também diz respeito a um polinucleotídeo isolado que codifica um peptídeo sinal compreendendo ou consistindo em aminoácidos 1 a 17 de SEQ ID NO: 2 ou aminoácidos 1 a 17 de SEQ ID NO: 4. O polinucleotídeo pode compreender adicionalmente um gene que codifica uma proteína, que é operavelmente ligado ao peptídeo sinal. A proteína é preferivelmente estrangeira ao peptídeo sinal.

A presente invenção também diz respeito o construtos de ácido nucleico, vetores de expressão e células hospedeiras recombinantes que compreendem tais polinucleotídeos.

A presente invenção também diz respeito a métodos de produzir uma proteína, que compreendem: (a) cultivar uma célula hospedeira recombinante que compreende um polinucleotídeo como este; e (b) recuperar

a proteína.

A proteína pode ser natural ou heteróloga a um célula hospedeira. O termo “proteína” não significa aqui que refere-se a um tamanho específico do produto codificado e, portanto, inclui peptídeos, oligopeptídeos e polipeptídeos. O termo “proteína” também inclui dois ou mais polipeptídeos combinados para formar o produto codificado. As proteínas também incluem polipeptídeos híbridos e polipeptídeos fundidos.

Preferivelmente, a proteína é um hormônio ou variante deste, enzima, receptor ou porção deste, anticorpo ou porção deste, ou repórter. Por exemplo, a proteína pode ser uma oxidoredutase, transferase, hidrolase, liase, isomerase, ou ligase tais como uma aminopeptidase, amilase, carboidrase, carboxipeptidase, catalase, celulase, quitinase, cutinase, ciclodextrina glicosiltransferase, desoxirribonuclease, esterase, alfa-galactosidase, beta-galactosidase, glucoamilase, alfa-glicosidase, beta-glicosidase, invertase, lacase, another lipase, manosidase, mutanase, oxidase, enzima pectinolítica, peroxidase, fitase, polifenoloxidase, enzima proteolítica, ribonuclease, transglutaminase ou xilanase.

O gene pode ser obtido de qualquer fonte de procarioto, eucarioto ou outra.

A presente invenção é descrita adicionalmente pelos exemplos a seguir que não devem ser interpretados como limitantes do escopo da invenção.

Exemplos

Os produtos químicos usados como tampões e substratos foram produtos comerciais de pelo menos grau de reagente.

Meios

As placas de ágar YG foram compostas de 5,0 g de extrato de levedura em pó da Difco, 10,0 g de glicose, 20,0 g de ágar e água deionizada para 1 litro

As placas de PDA foram compostas de 39 g ágar batata dextrose e água deionizada para 1 litro.

as placas de meio mínimo foram compostas de 6 g de NaNO_3 , 0,52 g de KCl, 1,52 g de KH_2PO_4 , 1 mL de solução de elementos traço COVE, 20 g de ágar nobre, 20 mL de glicose 50 %, 2,5 mL de $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$, 20 mL de uma solução de biotina 0,02 %, e água deionizada para 1 litro.

A solução de elementos traço COVE foi composta de 0,04 g de $\text{Na}_2\text{B}_4\text{O}_7 \cdot 10\text{H}_2\text{O}$, 0,4 g de $\text{CuSO}_4 \cdot 5\text{H}_2\text{O}$, 1,2 g de $\text{FeSO}_4 \cdot 7\text{H}_2\text{O}$, 0,7 g de $\text{MnSO}_4 \cdot \text{H}_2\text{O}$, 0,8 g de $\text{Na}_2\text{MoO}_4 \cdot 2\text{H}_2\text{O}$, 10 g de $\text{ZnSO}_4 \cdot 7\text{H}_2\text{O}$, e água deionizada para 1 litro.

O meio YPM foi composto de extrato de levedura 2%, peptona a 2% e maltose 1%.

Exemplo 1: Isolamento de cepas de *Thielavia hyrcania* NN045178 e NN045097

As cepas de *Thielavia hyrcania* NN045178 e NN045097 foram isoladas de amostras de solo coletadas em Yunnan, China, em 9 de março de 1999. As cepas foram isoladas aplicando uma solução de suspensão do solo nas placas de ágar YG, seguido por incubação a 45 °C por 3-5 dias e purificação adicional de colônias fúngicas. As cepas foram identificadas tanto por características morfológicas quanto por seqüenciamento de DNA das regiões espaçadoras transcritas internas (ITS). *Thielavia hyrcania* NN045178 foi depositada no China General Microbiological Culture Collectio Center e o número de depósito foi determinado CGMCC 0864.

Exemplo 2: Extração de DNA genômico de *Thielavia hyrcania* NN045178 e NN045097

Ambas as cepas de *Thielavia hyrcania*, NN045178 e NN045097, foram crescidas separadamente em placas de PDA a 45 °C por 3 dias. Os micélios foram cada qual coletados diretamente das placas de ágar em almofariz esterilizado, e congelados em nitrogênio líquido. Os micélios

congelados foram moídos, por almofariz e pilão, em um pó fino e o DNA genômico foi isolado usando um estojo DNEASY® Plant Mini (QIAGEN Inc., Valencia, CA, USA).

Exemplo 3: Isolamento de genes de celobiohidrolase de tamanho total da família 6 das cepas de *Thielavia hyrcania* NN045178 e NN045097

Uma sequência parcial de um gene de celobiohidrolase da família 6 foi identificado a partir de *Thielavia hyrcania* NN045178 por seleção de PCR usando as sondas mostradas a seguir, que são derivadas das regiões conservadas dos genes de celobiohidrolase da família 6.

Oligonucleotídeo iniciador CBHIIDF1:

t(gt)cc(tc)ga(tc)cg(tc)ga(tc)tg(tc)gc (SEQ ID NO: 5)

Oligonucleotídeo iniciador CBHIIDR2:

tc(ag)ccacc(gt)ggctt(gt)um(tc)cca (SEQ ID NO: 6)

Um gene de celobiohidrolase de tamanho total da família 6 foi isolado de *Thielavia hyrcania* NN045178 usando um estojo Genome Walking (Takara Bio Inc., Seta 3-4-1, Otsu, Shiga, 520-2193, Japão), de acordo com as instruções do fabricante. Resumidamente, o DNA genômico total de *Thielavia hyrcania* NN045178 foi usado como molde para amplificações por PCR da extremidade 5' e da extremidade 3' do gene. Os dois oligonucleotídeos iniciadores gene específicos mostrados a seguir foram usados para cada clonagem, um para a PCR primária e um para a PCR secundária. Os oligonucleotídeos iniciadores foram determinados com base na sequência do gene parcial celobiohidrolase da família 6 de *Thielavia hyrcania* NN045178.

Oligonucleotídeo iniciador CBHII45178DPf1:

5'-AACTACGACGAGAAGCACTACGTCGAG-3' (SEQ ID NO: 7)

Oligonucleotídeo iniciador CBHII45178DPf3:

5'-AGCTCCTGGATGCCTTTGTCTGGA-3' (SEQ ID NO: 8)

Oligonucleotídeo iniciador CBHII45178DPas1:

5'-AGGCGTTGTAGTTGGCGACGTTG-3' (SEQ ID NO: 9)

Oligonucleotídeo iniciador CBHII45178DPas3:

5'-CAATCGAGAACTCGCCGTTGGA-3' (SEQ ID NO: 10)

As amplificações primárias foram compostas de 2 µL de DNA genômico como molde, 2,5 mM de cada um de dATP, dTTP, dGTP e dCTP, 100 pmol de AP2 (Takara Bio Inc., Seta 3-4-1, Otsu, Shiga, 520-2193, Japão), e 10 pmol de oligonucleotídeo iniciador CBHII45178DPas1 para a clonagem da extremidade 5' ou 10 pmol de oligonucleotídeo iniciador CBHII45178DPf1 para clonagem da extremidade 3', 5 µL de tampão de PCR II 10X LA (Takara Bio Inc., Seta 3-4-1, Otsu, Shiga, 520-2193, Japão), e 2,5 unidades de TakaRa LA Taq DNA polimerase (Takara Bio Inc., Seta 3-4-1, Otsu, Shiga, 520-2193, Japão) em um volume final de 50 µL. As amplificações foram realizadas usando um termociclador Peltier (Bio-Rad Laboratories, Inc. Hercules, CA, USA) programado para pré-desnaturação a 94 °C por 1 minuto e 98 °C por 1 minuto; cinco ciclos, cada um em uma temperatura de desnaturação de 94 °C por 30 segundos; anelamento a 60 °C por 1 minuto, e alongamento a 72 °C por 2 minutos; 1 ciclo de desnaturação a 94 °C por 30 segundos; anelamento a 25 °C por 3 minutos e alongamento a 72 °C por 2 minutos; quinze repetições de 2 ciclos a 94 °C por 30 segundos; 62 °C por 1 minuto; e 72 °C por 2 minutos; seguido por 1 ciclo a 94 °C por 30 segundos; 44 °C por 1 minuto; e 72 °C por 2 minutos; e uma extensão final a 72 °C por 10 minutos. O bloco aquecido passou então pra um ciclo de absorção a 4 °C.

As amplificações secundárias foram compostas de 2 µL de produto de PCR primária diluído 20X como molde, 2,5 mM de cada um de dATP, dTTP, dGTP, e dCTP, 100 pmol de AP2, e 10 pmol de oligonucleotídeo iniciador CBHII45178DPas3 para a clonagem da extremidade 5' ou 10 pmol de oligonucleotídeo iniciador CBHII45178DPf3 para clonagem da extremidade 3', 5 µL de tampão de PCR II LA 10X, e 2,5 unidades de TakaRa LA Taq DNA polimerase em um volume final de 50 µL.

As amplificações foram realizadas usando um termociclador Peltier programado para quinze repetições de 2 ciclos de 94 °C por 30 segundos; 62 °C por 1 minuto; 72 °C por 2 minutos; seguido por 1 ciclo de 94 °C por 30 segundos; 44 °C por 1 minuto; 72 °C por 2 minutos; e uma extensão final a 72 °C por 10 minutos. O bloco aquecido passou então para um ciclo de absorção a 4 °C.

Os produtos de reação foram isolados por eletroforese em gel de agarose a 1,0 % usando tampão Tris-borato 90 mM e EDTA 1 mM (TBE), onde uma banda de produto de 600 bp da reação de clonagem da extremidade 3' e uma banda do produto de 1,1 kb da reação de clonagem da extremidade 5' foram excisadas dos géis, purificadas usando illustra GFX PCR DNA e um estojo de purificação de banda de gel (GE Healthcare, Buckinghamshire, UK), de acordo com as instruções do fabricante, e confirmados por sequenciamento de DNA usando um analisador de DNA 3730XL (Applied Biosystems Inc, Foster City, CA, USA).

Como a sequência parcial de um outro gene de celobiohidrolase da família 6 de *Thielavia hyrcania* NN045097 mostrou identidade elevada (97%) com o gene de celobiohidrolase de *Thielavia hyrcania* NN045178, o gene de tamanho total do gene de celobiohidrolase da família 6 de *Thielavia hyrcania* NN045097 foi clonado usando o mesmo par de oligonucleotídeo iniciador para a clonagem do gene de celobiohidrolase de *Thielavia hyrcania* NN045178.

Exemplo 4: Caracterização das sequências genômicas de *Thielavia hyrcania* NN045097 e NN045178 que codificam as celobiohidrolases da família 6

A sequência de nucleotídeos do gene de celobiohidrolase de tamanho total da família 6 de *Thielavia hyrcania* NN045097 apresenta 1598 bp incluindo o códon de parada. A sequência de proteína deduzida da proteína

madura mostrou similaridade com a exoglucanase GH6A de *Humicola insolens* (uniprot:Q9C1S9).

A sequência de nucleotídeos (SEQ ID NO: 1) e a sequência deduzida de aminoácidos (SEQ ID NO: 2) do gene de celobiohidrolase de tamanho total de família 6 de *Thielavia hyrcanica* NN045097 são mostradas na figura 1. O fragmento genômico codifica um polipeptídeo de 493 aminoácidos, interrompido por 2 introns de 58 bp cada. O teor % de G+C do gene e a sequência codificante madura são cada qual 65,7 %. Usando o programa de software SignalP (Nielsen *et al.*, 1997, *supra*), um peptídeo sinal de 17 resíduos foi previsto. A proteína madura prevista contém 476 aminoácidos com uma massa molecular prevista de 50,0 kDa.

Um alinhamento global aos pares comparativo de sequências de aminoácidos foi determinado usando o algoritmo de Needleman-Wunsch (Needleman e Wunsch, 1970, *supra*), implementado no programa Needle de EMBOSS, com penalidade de abertura de intervalo de 10, penalidade de extensão de intervalo de 0,5, e a matriz EBLOSUM62. O alinhamento mostrou que a sequência deduzida de aminoácidos do gene de celobiohidrolase da família 6 de *Thielavia hyrcanica* NN045097, que codifica o polipeptídeo maduro, compartilha 84,1 % e 83,7 % de identidade (excluindo os intervalos) com as sequências deduzidas de aminoácidos de duas proteínas da enzima lignocelulósica (WO 2008/095033; GeneSeqP números de acesso ASR94104 e ASR94110, respectivamente).

A sequência de nucleotídeos do gene de celobiohidrolase de tamanho total da família 6 de *Thielavia hyrcanica* apresenta 1584 bp, incluindo o códon de parada. A sequência de proteína deduzida da proteína madura mostrou similaridade com a exoglucanase GH6A de *Humicola insolens* (uniprot:Q9C1S9).

A sequência de nucleotídeos (SEQ ID NO: 3) e a sequência deduzida de aminoácidos (SEQ ID NO: 4) do gene de celobiohidrolase de

tamanho total da família 6 de *Thielavia hyrcania* são mostradas na figura 2. O fragmento genômico codifica um polipeptídeo de 487 aminoácidos, interrompido por 2 introns de 59 e 61 bp. O teor % de G+C do gene e a sequência codificante madura são cada qual 65,4%. Usando o programa do software SignalP (Nielsen *et al.*, 1997, *supra*), um peptídeo sinal de 17
5 resíduos foi previsto. A proteína madura prevista contém 470 aminoácidos com uma massa molecular prevista de 49,5 kDa.

Um alinhamento global aos pares comparativo de sequências de aminoácidos foi determinado usando o algoritmo de Needleman-Wunsch
10 (Needleman e Wunsch, 1970, *supra*), implementado no programa Needle de EMBOSS, com penalidade de abertura de intervalo de 10, penalidade de extensão de intervalo de 0,5, e a matriz EBLOSUM62. O alinhamento mostrou que a sequência deduzida de aminoácidos do gene de celobiohidrolase da família 6 de *Thielavia hyrcania* NN045178, que codifica
15 o polipeptídeo maduro, compartilha 85,3 % e 83,6 % de identidade (excluindo os intervalos) com as sequências deduzidas de aminoácidos de duas proteínas da enzima lignocelulósica (WO 2008/095033; GeneSeqP números de acesso ASR94104 e ASR94110, respectivamente).

Exemplo 5: Clonagem do gene de celobiohidrolase da família 6 de
20 ***Thielavia hyrcania* NN045178 e construção de um vetor de expressão de**
Aspergillus oryzae

Dois oligonucleotídeos iniciadores sintéticos mostrados a seguir foram determinados para amplificar por PCR o gene de celobiohidrolase de *Thielavia hyrcania* NN045178 a partir do DNA genômico
25 preparado no exemplo 2. Um estojo de clonagem IN-FUSION® CF Dry-down (Clontech Laboratories, Inc., Mountain View, CA, USA) foi usado para clonar o fragmento diretamente no vetor de expressão pPFJO355, sem a necessidade de digestão por restrição e ligação. O vetor de expressão pPFJO355 contém o promotor TAKA alfa-amilase de *Aspergillus oryzae* e o

finalizador glicoamilase de *Aspergillus niger*. Além disso, pPFJO355 apresenta sequências derivadas de pUC18 para a seleção e propagação em *E. coli*, e um gene *pyrG*, que codifica uma orotidina descarboxilase derivada de *Aspergillus nidulans* para a seleção de um transformante de um mutante *pyrG* da cepa de *Aspergillus*.

H_45178_IFNbam:

5'ACACAAC**TGGGGATCCACCATGGCCAAGAAGCTCCTTCTCACC**
-3' (SEQ ID NO: 11)

H_45178_IFCbgl:

10 5'-GTCACCCTCTAGATCTCT**AAAAAGGAGGGTTGGCGTTGG**-3'
(SEQ ID NO: 12)

As letras em **negrito** representam a sequência codificante. A sequência restante é homóloga aos sítios de inserção de pPFJO355.

Dez picomols de cada um dos oligonucleotídeos iniciadores anteriores foram usados em uma PCR composta de DNA genômico de *Thielavia hyrcania* NN045178, 10 µL de tampão GC 5X (Finnzymes Oy, Espoo, Finlândia), 1,5 µL de DMSO, 2,5 mM de cada um de dATP, dTTP, dGTP, e dCTP, e 1 unidade de DNA polimerase de alta fidelidade PHUSION™ (Finnzymes Oy, Espoo, Finlândia) em um volume final de 50 µL. A amplificação foi realizada usando um termociclador Peltier programado para 1 ciclo a 98 °C por 1 minuto; 5 ciclos de desnaturação a 98 °C por 15 segundos, anelamento a 63 °C por 30 segundos, com um aumento de 1 °C por ciclo, e alongamento a 72 °C por 80 segundos; e outros 25 ciclos cada um com 98 °C por 15 segundos e 72 °C por 80 segundos; extensão final a 72 °C por 10 minutos. O bloco aquecido passou então pra um ciclo de absorção a 4 °C.

Os produtos de reação foram isolados por eletroforese em gel de agarose a 1,0 % usando tampão TBE, onde uma banda de produto de aproximadamente 1,8 kb foi excisada do gel e purificada usando um ilustra

GFX PCR DNA e estojo de purificação de banda em gel, de acordo com as instruções do fabricante.

O plasmídeo pPFJO355 foi digerido com *Bam* I e *Bgl* II, isolado por eletroforese em gel de agarose a 1,0 % usando tampão TBE e purificado usando um ilustra GFX PCR DNA e estojo de purificação de banda em gel, de acordo com as instruções do fabricante.

O fragmento de gene e o vetor digerido foram ligados juntos usando um estojo de clonagem por PCR IN-FUSION® CF Dry-down resultando no pCBHII45178 (Figura 3), em que a transcrição do gene de celobiohidrolase estava no controle do promotor TAKA alfa-amilase de *Aspergillus oryzae*. Em resumo, 30 ng de pPFJO355 digerido com *Bam* HI e *Bgl* II, e 80 ng do produto de PCR purificado a partir do gene de celobiohidrolase da família 6 de *Thielavia hyrcania* NN045178 foram adicionados a uma garrafa de reação e re-suspensos em um volume final de 10 µL com adição de água deionizada. A reação foi incubada a 37 °C por 15 minutos, e então a 50 °C por 15 minutos. Três µL da reação foram usados para transformar células competentes TOP10 de *E. coli* (TIANGEN Biotech (Beijing) Co. Ltd., Beijing, China). Um transformante de *E. coli* contendo pCBHII45178 foi detectado por PCR de colônia e o DNA plasmidial foi preparado usando um estojo QIAprep Spin Miniprep (QIAGEN Inc., Valencia, CA, USA). O inserto do gene de celobiohidrolase da família 6 de *Thielavia hyrcania* NN045178 em pCBHII45178 foi confirmado por seqüenciamento de DNA usando um analisador de DNA 3730XL (Applied Biosystems Inc, Foster City, CA, USA).

O mesmo fragmento de PCR foi clonado no vetor pGEM-T (Promega Corporation, Madison, WI, USA) usando um sistema de vetor pGEM-T (Promega Corporation, Madison, WI, USA) para gerar pGEM-T-CBHII45178 (Figura 4). O inserto do gene de celobiohidrolase da família 6 de *Thielavia hyrcania* NN045178 em T-CBHII45178 foi confirmado por

sequenciamento de DNA da maneira anterior. *E. coli* T-CBHII45178 foi depositada no Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH (DSM) em 20 de maio de 2009 e o número de acesso foi determinado como DSM 22599.

5 **Exemplo 6: Clonagem de um gene de celobiohidrolase da família 6 de *Thielavia hyrcania* NN045097 e construção de um vetor de expressão de *Aspergillus oryzae***

10 A clonagem do gene de celobiohidrolase da família 6 de *Thielavia hyrcania* NN045097 foi realizada de acordo com o mesmo método descrito no exemplo 5.

Os dois oligonucleotídeos iniciadores sintéticos (H_45178_IFNbam e H_45178_IFCbgl) descritos no exemplo 5 foram usados para amplificação por PCR do gene de celobiohidrolase de tamanho completo da família 6 de *Thielavia hyrcania* NN045097.

15 Dez picomols de cada um dos oligonucleotídeos iniciadores anteriores foram usados em uma PCR composta de DNA genômico d e *Thielavia hyrcaniea* NN045097, 10 µL de tampão GC 5X, 1,5 µL de DMSO, 2,5 mM de cada um de dATP, dTTP, dGTP, e dCTP, e 1 unidade de DNA polimerase de alta fidelidade PHUSION™ em um volume final de 50 µL. A
20 amplificação foi realizada usando um termociclador Peltier programado para 1 ciclo a 98 °C por 1 minuto; 10 ciclos de desnaturação a 98 °C por 15 segundos, anelamento a 62 °C por 30 segundos, com um aumento de 1 °C por ciclo e alongamento a 72 °C por 80 segundos; e outros 25 ciclos cada um a 98 °C por 15 segundos e 72 °C por 80 segundos; extensão final a 72 °C por 10
25 minutos. O bloco aquecido passou então pra um ciclo de absorção a 4 °C.

Os produtos de reação foram isolados por eletroforese em gel de agarose a 1,0 % usando tampão TBE, onde uma banda de produto de aproximadamente 1,8 kb foi excisada do gel, e purificados usando um illustra GFX PCR DNA e estojo de purificação de banda em gel, de acordo com as

instruções do fabricante.

O plasmídeo pPFJO355 foi digerido com *Bam* I e *Bgl* II, isolado por eletroforese em gel de agarose a 1,0 % usando tampão de TBE, e purificado usando um illustra GFX PCR DNA e estojo de purificação de
5 banda em gel, de acordo com as instruções do fabricante.

O fragmento de gene e o vetor digerido foram ligados juntos usando um estojo de clonagem por PCR IN-FUSION® CF Dry-down que resulta em pCBHII45097 (Figura 5), em que a transcrição do gene de celobiohidrolase estava no controle do promotor TAKA alfa-amilase de
10 *Aspergillus oryzae*. Em resumo, 30 ng de pPFJO355 digerido com *Bam* I e *Bgl* II e 80 ng do produto de PCR do gene de celobiohidrolase da família 6 de *Thielavia hyrcaniae* NN045097 foram adicionados a uma garrafa de reação e re-suspensos em um volume final de 10 µL com adição de água deionizada. A reação foi incubada a 37 °C por 15 minutos, e então a 50 °C por 15 minutos.
15 Três µL da reação foram usados para transformar células competentes TOP10 de *E. coli*. Um transformante de *E. coli* contendo pCBHII45097 foi detectado por PCR de colônia e o DNA plasmidial foi preparado usando um estojo QIAprep Spin Miniprep. O inserto do gene de celobiohidrolase da família 6 de *Thielavia hyrcaniae* NN045097 em pCBHII45178 foi confirmado por
20 seqüenciamento de DNA usando um analisador de DNA 3730XL.

O mesmo fragmento de PCR foi clonado no vetor pGEM-T usando um sistema de vetor pGEM-T para gerar pGEM-T-CBHII45097 (Figura 6). O inserto do gene de celobiohidrolase da família 6 de *Thielavia hyrcaniae* NN045097 em pGEM-T-CBHII45097 foi confirmado por
25 seqüenciamento de DNA. A *E. coli* T-CBHII45097 foi depositada no Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH (DSM) em 20 de maio de 2009 e o número de acesso foi determinado como DSM 22598.

Exemplo 7: Expressão dos genes de celobiohidrolase da família 6 de

***Thielavia hyrcania* NN045178 e NN45097 em *Aspergillus oryzae* HowB101**

Os protoplastos de *Aspergillus oryzae* HowB101 (WO 95/35385) foram preparados de acordo com o método de Christensen *et al.*, 1988, *Bio/Technology* 6: 1419-1422. Três µg de pCBHII45178 ou
5 pCBHII45097 foram usados para transformar os protoplastos de *Aspergillus oryzae* HowB101.

A transformação de *Aspergillus oryzae* HowB101 com pCBHII45178 ou pCBHII45097 rendeu cerca de 50 transformantes cada. Quatro transformantes de cada transformação foram isolados em placas de
10 meio mínimo individuais.

Os quatro transformantes de cada transformação foram inoculados separadamente em 3 mL de meio YPM em placas de 24 poços e incubados a 30 °C com agitação a 150 rpm. Após 3 dias de incubação, 20 µL de sobrenadante de cada cultura foram analisados em um gel NUPAGE®
15 NOVEX® 4-12% Bis-Tris com MES (Invitrogen Corporation, Carlsbad, CA, USA), de acordo com as instruções do fabricante. O gel resultante foi corado com INSTANTBLUE™ (Expediton Ltd., Babraham Cambridge, UK). Os perfis de SDS-PAGE das culturas mostraram que a maioria dos transformantes produziu uma banda de proteína principal em aproximadamente 70 kDa para
20 ambos os genes. As cepas de expressão que carregam o gene de celobiohidrolase da família 6 de *Thielavia hyrcania* NN045178 e o gene de celobiohidrolase da família 6 de *Thielavia hyrcania* NN045097 foram determinadas como EXP02989 e EXP03004, respectivamente.

Uma amostra para cada cepa de expressão, determinados
25 transformante 4 de EXP02989 e transformante 16 de EXP03004, foi lavada com 10 mL de meio YPM e cada cepa foi inoculada em um frasco de 2 litros contendo 400 mL de meio YPM para gerar caldo para caracterização de cada enzima. As culturas foram coletadas no dia 3 e filtradas usando uma membrana de 0,45 µm DURAPORE® (Millipore, Bedford, MA, USA).

Exemplo 8: Caracterização da celobiohidrolase de *Thielavia hyrcaniae* NN045178 CEL6A

Ensaio de atividade em PASC. Uma solução estoque de lama de PASC foi preparada umedecendo 5 g de AVICEL® (JRS Pharma LP, Patterson, NY, USA) com água, seguido pela adição de 150 mL de *o*-ácido fosfórico gelado 85 %. A suspensão foi agitada lentamente em um banho de gelo por 1 hora. A seguir, 500 mL de acetona gelada foram adicionados durante a agitação. A lama foi filtrada usando MIRACLOTH® (Calbiochem, La Jolla, CA, USA) e então lavada três vezes com 100 mL de acetona gelada (drenagem o mais seca possível após cada lavagem). Finalmente, a lama filtrada foi lavada duas vezes com 500 mL de água, e novamente drenada o mais seca possível após cada lavagem. O PASC foi misturado com água deionizada em um total volume de 500 mL em uma concentração de 10 g/litro, misturado para homogeneidade (usando um homogeneizador ULTRA-TURRAX®, Cole-Parmer, Vernon Hills, IL, USA), e armazenado em um refrigerador por até um mês.

A solução estoque de PASC foi diluída com tampão de acetato de sódio 50 mM pH 5,0 em uma concentração de 2 g/litro e usadas como o substrato. A 150 µL de solução estoque de PASC, 20 µL de amostra de enzima foram adicionados e a mistura de reação foi incubado por 60 minutos com agitação a 850 rpm. No final da incubação, 50 µL de NaOH 2% foram adicionados para finalizar a reação. A mistura de reação foi centrifugada a 1.000 x g. Os açúcares liberados foram medidos misturando primeiro 10 µL da mistura de reação com 90 µL de NaOH 0,4 %, seguido por 50 µL de hidrazida de ácido *p*-hidroxibenzóico 1,5 % em NaOH 2 % (PHBAH, Sigma Chemical Co., St. Louis, MO, USA). A mistura foi fervida a 100 °C por 5 minutos, e a seguir 100 µL foram transferidos para uma placa de microtitulação para uma leitura de absorbância a 410 nm (Spectra Max M2, Molecular devices Sunnyvale, CA, USA). Os brancos foram sintetizados

omitindo o PASC na etapa de hidrólise e substituindo o hidrolisado com tampão na etapa de determinação de açúcar.

Perfil de pH. Em uma placa de microtitulação, 20 μL de cada celobiohidrolase (EXP02989 ou EXP03004) foram misturados com 120 μL de tampão (ácido succínico 100 mM, HEPES, CHES ou CAPSO, e CaCl_2 1 mM, KCl 150 mM, e TRITON® X-100 0,01 %; pH ajustado com NaOH e HCl) em pH 3, 4, 5, 6, 7 ou 8, seguido pela adição de 30 μL de PASC 10 g/litro. As misturas foram colocadas no gelo antes da reação. Os ensaios foram iniciados transferindo as placas de microtitulação para um misturador térmico EPPENDORF® (Eppendorf, Hauppauge, NY, USA), que foi ajustado para 50 °C, e incubado por 60 minutos com agitação a 850 rpm. No final da incubação, 50 μL de NaOH 2 % foram adicionados para parar as reações. As misturas de reação foram centrifugadas a 1.000 x g. Os açúcares liberados foram medidos da maneira descrita anteriormente.

Os resultados mostrados na figura 7 indicaram que as celobiohidrolases possuem atividade em uma ampla faixa de pH, pH 3 a 11. O pH ideal para cada celobiohidrolase foi entorno de pH 5 a 6.

Estabilidade de pH. Uma amostra de 40 μL de cada celobiohidrolase (EXP02989 ou EXP03004) foi misturada com 160 μL de tampão 100 mM (pH 4, 5, 6, 7 ou 8: ácido succínico 100 mM, HEPES, CHES, CAPSO e CaCl_2 1 mM, KCl 150 mM, e TRITON® X-100 0,01%, pH ajustado com NaOH e HCl) e incubada a 50 °C com agitação a 850 rpm. As amostras de 30 μL foram removidas e colocadas no gelo a 0, 10, 30, 60 e 110 minutos. A seguir, 150 μL de PASC 2 g/litro (em acetato de sódio 50 mM pH 5,0) foram adicionados em cada amostra e incubados at 50 °C, 850 rpm por 60 minutos. Trinta μL de cada tampão foram corridos como branco. Os açúcares liberados foram medidos da maneira descrita anteriormente.

Os resultados mostrados nas figuras 8A e 8B indicaram que a celobiohidrolase EXP02989 apresentou atividade residual a pH 4, 5, 6, 7 e 8

de 90 %, 96 %, 96 %, 90 % e 39 %, respectivamente, enquanto a celobiohidrolase EXP03004 apresentou atividade residual a pH 4, 5, 6, 7 e 8 de 89 %, 93 %, 95 %, 75 % e 48 %, respectivamente, após incubação por 110 minutos a 50 °C.

5 Perfil de temperatura. Um volume de 150 µL de 2 g de PASC por litro de tampão de acetato de sódio 50 mM pH 5,0 em um tubo EPPENDORF® foi pré-incubado a 25, 30, 40, 50, 60, 70, 80, 90 e 100 °C. O ensaio iniciou adicionando 20 µL de cada celobiohidrolase (EXP02989 ou EXP03004), seguido pela incubação por 60 minutos com agitação a 850 rpm.
10 No final da incubação, 50 µL de hidróxido de sódio a 2 % foram adicionados para parar a reação e as misturas de reação foram centrifugadas a 1.000 x g. Os açúcares liberados foram medidos da maneira descrita anteriormente.

Os resultados mostrados nas figuras 9A e 9B indicaram que as celobiohidrolases foram ativas em uma ampla faixa de temperatura de 25 a 80 °C e pareceram apresentar atividade ideal em uma temperatura de cerca de 60 °C.

15 Estabilidade de temperatura. Uma amostra de 250 µL de cada celobiohidrolase (EXP02989 ou EXP03004) (diluída com tampão de acetato de sódio 50 mM pH 5) foi incubada a 50, 60 e 70 °C. Uma amostra de 20 µL
20 de cada celobiohidrolase foi removida e colocada no gelo em pontos de tempo de 0, 10, 30, 60 e 110 minutos. A seguir, 150 µL de 2 g de PASC por litro de acetato de sódio 50 mM pH 5,0 foram adicionados em cada amostragem e incubados a 50 °C por 60 minutos. No final da incubação, 50 µL de NaOH 2 % foram adicionados para parar as reações. As misturas de reação foram
25 centrifugadas a 1.000 x g. Os açúcares liberados foram medidos da maneira descrita anteriormente.

Os resultados mostrados nas figuras 10A e 10B indicaram que as celobiohidrolases foram estáveis a 50 °C, menos estáveis a 60 °C, e instáveis a 70 °C. As atividades residuais da celobiohidrolase EXP02989

foram 100 % a 50 °C, 53 % a 60 °C, e 40 % a 70 °C e a celobiohidrolase EXP03004 foram 100 % a 50 °C, 57 % a 60 °C, e 51 % a 70 °C após 60 minutos de incubação.

Depósito de material biológico

- 5 Os materiais biológicos a seguir foram depositados nos termos do Budapest Treaty with the Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH (DSMZ), Mascheroder Weg 1 B, D-38124 Braunschweig, Alemanha, e são fornecidos com os números de acesso a seguir:

Cepa Número de acesso Data

- 10 *E. coli* T-CBHII45097 DSM 22598 20 de maio de 2009

E. coli T-CBHII45178 DSM 22599 20 de maio de 2009

- As cepas foram depositadas em condições que garantem que o acesso à cultura será disponível durante a pendência deste pedido de patente determinada por leis estrangeiras de patente a serem intituladas a esta. Os depósitos representam culturas substancialmente puras das cepas depositadas. Os depósitos são disponíveis da maneira exigida pelas leis de patente estrangeira em países em que as duplicatas do pedido submetido, ou sua progênie, são depositadas. Entretanto, pode-se entender que a disponibilidade de um depósito não constitui uma licença para praticar a invenção submetida em derrogação dos direitos de patente concedidos por ação governamental.
- 15 20

A presente invenção é descrita adicionalmente pelos parágrafos numerados a seguir:

- [1] Um polipeptídeo isolado com atividade celobiohidrolase, selecionado do grupo que consiste em: (a) um polipeptídeo com pelo menos 90 % de identidade de sequência com o polipeptídeo maduro de SEQ ID NO: 2 ou SEQ ID NO: 4; (b) um polipeptídeo codificado por um polinucleotídeo que hibridiza em condições de severidade elevada (i) na sequência que codifica o polipeptídeo maduro de SEQ ID NO: 1 ou SEQ ID NO: 3, (ii) na sequência de DNAc contida na sequência que codifica o polipeptídeo maduro
- 25

de SEQ ID NO: 1 ou SEQ ID NO: 3, ou (iii) na fita complementar de comprimento total de (i) ou (ii); (c) um polipeptídeo codificado por um polinucleotídeo com pelo menos 90 % de identidade de sequência com a sequência que codifica o polipeptídeo maduro de SEQ ID NO: 1 ou SEQ ID NO: 3; (d) um variante que compreende uma substituição, deleção e/ou inserção de um ou mais (vários) aminoácidos do polipeptídeo maduro de SEQ ID NO: 2 ou SEQ ID NO: 4; e (e) um fragmento de um polipeptídeo de (a), (b), (c) ou (d) que apresenta atividade celobiohidrolase.

[2] O polipeptídeo do parágrafo 1, com pelo menos 90 % de identidade de sequência com o polipeptídeo maduro de SEQ ID NO: 2 ou SEQ ID NO: 4.

[3] O polipeptídeo do parágrafo 2, com pelo menos 91 % de identidade de sequência com o polipeptídeo maduro de SEQ ID NO: 2 ou SEQ ID NO: 4.

[4] O polipeptídeo do parágrafo 3, com pelo menos 92 % de identidade de sequência com o polipeptídeo maduro de SEQ ID NO: 2 ou SEQ ID NO: 4.

[5] O polipeptídeo do parágrafo 4, com pelo menos 93 % de identidade de sequência com o polipeptídeo maduro de SEQ ID NO: 2 ou SEQ ID NO: 4.

[6] O polipeptídeo do parágrafo 5, com pelo menos 94 % de identidade de sequência com o polipeptídeo maduro de SEQ ID NO: 2 ou SEQ ID NO: 4.

[7] O polipeptídeo do parágrafo 6, com pelo menos 95 % de identidade de sequência com o polipeptídeo maduro de SEQ ID NO: 2 ou SEQ ID NO: 4.

[8] O polipeptídeo do parágrafo 7, com pelo menos 96 % de identidade de sequência com o polipeptídeo maduro de SEQ ID NO: 2 ou SEQ ID NO: 4.

[9] O polipeptídeo do parágrafo 8, com pelo menos 97 % de identidade de sequência com o polipeptídeo maduro de SEQ ID NO: 2 ou SEQ ID NO: 4.

5 [10] O polipeptídeo do parágrafo 9, com pelo menos 98 % de identidade de sequência com o polipeptídeo maduro de SEQ ID NO: 2 ou SEQ ID NO: 4.

[11] O polipeptídeo do parágrafo 10, com pelo menos 99 % de identidade de sequência com o polipeptídeo maduro de SEQ ID NO: 2 ou SEQ ID NO: 4.

10 [12] O polipeptídeo do parágrafo 1, que é codificado por um polinucleotídeo que hibridiza em condições de severidade elevada (i) na sequência que codifica o polipeptídeo maduro de SEQ ID NO: 1 ou SEQ ID NO: 3, (ii) na sequência de DNAC contida na sequência que codifica o polipeptídeo maduro de SEQ ID NO: 1 ou SEQ ID NO: 3, ou (iii) na fita
15 complementar de comprimento total de (i) ou (ii).

[13] O polipeptídeo do parágrafo 12, que é codificado por um polinucleotídeo que hibridiza em condições de severidade muito elevada (i) na sequência que codifica o polipeptídeo maduro de SEQ ID NO: 1 ou SEQ ID NO: 3, (ii) na sequência de DNAC contida na sequência que codifica o
20 polipeptídeo maduro de SEQ ID NO: 1 ou SEQ ID NO: 3, ou (iii) na fita complementar de comprimento total de (i) ou (ii).

[14] O polipeptídeo do parágrafo 1, que é codificado por um polinucleotídeo com pelo menos 90 % de identidade de sequência com a sequência que codifica o polipeptídeo maduro de SEQ ID NO: 1 ou SEQ ID
25 NO: 3; ou a sequência de DNAC deste.

[15] O polipeptídeo do parágrafo 14, que é codificado por um polinucleotídeo com pelo menos 91 % de identidade de sequência com a sequência que codifica o polipeptídeo maduro de SEQ ID NO: 1 ou SEQ ID NO: 3; ou a sequência de DNAC deste.

[16] O polipeptídeo do parágrafo 15, que é codificado por um polinucleotídeo com pelo menos 92 % de identidade de sequência com a sequência que codifica o polipeptídeo maduro de SEQ ID NO: 1 ou SEQ ID NO: 3; ou a sequência de DNAC deste.

5 [17] O polipeptídeo do parágrafo 16, que é codificado por um polinucleotídeo com pelo menos 93 % de identidade de sequência com a sequência que codifica o polipeptídeo maduro de SEQ ID NO: 1 ou SEQ ID NO: 3; ou a sequência de DNAC deste.

10 [18] O polipeptídeo do parágrafo 17, que é codificado por um polinucleotídeo com pelo menos 94 % de identidade de sequência com a sequência que codifica o polipeptídeo maduro de SEQ ID NO: 1 ou SEQ ID NO: 3; ou a sequência de DNAC deste.

15 [19] O polipeptídeo do parágrafo 18, que é codificado por um polinucleotídeo com pelo menos 95 % de identidade de sequência com a sequência que codifica o polipeptídeo maduro de SEQ ID NO: 1 ou SEQ ID NO: 3; ou a sequência de DNAC deste.

20 [20] O polipeptídeo do parágrafo 19, que é codificado por um polinucleotídeo com pelo menos 96 % de identidade de sequência com a sequência que codifica o polipeptídeo maduro de SEQ ID NO: 1 ou SEQ ID NO: 3; ou a sequência de DNAC deste.

[21] O polipeptídeo do parágrafo 20, que é codificado por um polinucleotídeo com pelo menos 97 % de identidade de sequência com a sequência que codifica o polipeptídeo maduro de SEQ ID NO: 1 ou SEQ ID NO: 3; ou a sequência de DNAC deste.

25 [22] O polipeptídeo do parágrafo 21, que é codificado por um polinucleotídeo com pelo menos 98 % de identidade de sequência com a sequência que codifica o polipeptídeo maduro de SEQ ID NO: 1 ou SEQ ID NO: 3; ou a sequência de DNAC deste.

[23] O polipeptídeo do parágrafo 22, que é codificado por um

polinucleotídeo com pelo menos 99 % de identidade de sequência com a sequência que codifica o polipeptídeo maduro de SEQ ID NO: 1 ou SEQ ID NO: 3; ou a sequência de DNAC deste.

5 [24] O polipeptídeo de qualquer dos parágrafos 1-23, compreendendo ou consistindo em SEQ ID NO: 2 ou SEQ ID NO: 4.

[25] O polipeptídeo de qualquer dos parágrafos 1-23, compreendendo ou consistindo no polipeptídeo maduro de SEQ ID NO: 2 ou SEQ ID NO: 4.

10 [26] O polipeptídeo do parágrafo 25, em que o polipeptídeo maduro são os aminoácidos 18 a 493 de SEQ ID NO: 2 ou aminoácidos 18 a 487 de SEQ ID NO: 4.

[27] O polipeptídeo do parágrafo 1, que é um fragmento de SEQ ID NO: 2 ou SEQ ID NO: 4, em que o fragmento apresenta atividade celobiohidrolase.

15 [28] O polipeptídeo de qualquer dos parágrafos 1-27, que é codificado pelo polinucleotídeo contido no plasmídeo pGEM-T-CBHII45097 que está contido em *E. coli* DSM 22598, ou plasmídeo pGEM-T-CBHII45178 que está contido em *E. coli* DSM 22599.

20 [29] Uma composição compreendendo o polipeptídeo de qualquer dos parágrafos 1-28.

[30] Um polinucleotídeo isolado que codifica o polipeptídeo de qualquer dos parágrafos 1-28.

25 [31] Um construto de ácido nucleico ou vetor de expressão que compreende o polinucleotídeo do parágrafo 30 operavelmente ligado a uma ou mais (várias) sequências controle que direcionam a produção do polipeptídeo em um hospedeiro de expressão.

[32] Uma célula hospedeira recombinante que compreende o polinucleotídeo do parágrafo 30 operavelmente ligado a uma ou mais sequências controle que direcionam a produção do polipeptídeo.

[33] Um método de produzir o polipeptídeo de qualquer dos parágrafos 1-28, compreendendo: (a) cultivar uma célula, que em sua forma tipo selvagem produz o polipeptídeo em condições produtivas para a produção do polipeptídeo; e (b) recuperar o polipeptídeo.

5 [34] Um método de produzir um polipeptídeo com atividade celobiohidrolase, compreendendo: (a) cultivar a célula hospedeira do parágrafo 32 em condições produtivas para a produção do polipeptídeo; e (b) recuperar o polipeptídeo.

10 [35] Uma planta transgênica, parte da planta ou célula de planta transformada por um polinucleotídeo que codifica o polipeptídeo de qualquer dos parágrafos 1-28.

[36] Um método de produzir um polipeptídeo com atividade celobiohidrolase, compreendendo: (a) cultivar a planta transgênica ou célula de planta do parágrafo 35 em condições produtivas para a produção do polipeptídeo; e (b) recuperar o polipeptídeo.

15 [37] Um método de produzir um mutante de uma célula mãe, compreendendo inativar um polinucleotídeo que codifica o polipeptídeo de qualquer dos parágrafos 1-28, que resulta no mutante que produz menos do polipeptídeo do que a célula mãe.

20 [38] Uma célula mutante produzida pelo método do parágrafo 37.

[39] A célula mutante do parágrafo 38, compreendendo adicionalmente um gene que codifica uma proteína natural ou heteróloga.

25 [40] Um método de produzir uma proteína compreendendo: (a) cultivar a célula mutante do parágrafo 39 em condições produtivas para a produção do proteína; e (b) recuperar a proteína.

[41] Uma molécula de RNA inibitório de filamento duplo (RNAs) compreendendo uma subsequência do polinucleotídeo do parágrafo 30, em que opcionalmente o RNAs é uma molécula de RNAsi ou um

RNA_{mi}.

[42] A molécula de RNA inibitório de filamento duplo (RNAds) do parágrafo 41, que apresenta cerca de 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25 ou mais nucleotídeos duplexes em tamanho.

5 [43] Um método de inibir a expressão de um polipeptídeo com atividade celobiohidrolase em uma célula, compreendendo administrar na célula ou expressar na célula a molécula de RNA inibitório de filamento duplo (RNAds) do parágrafo 41.

10 [44] O método do parágrafo 43, em que o RNAds apresenta cerca de 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25 ou mais nucleotídeos duplexes em tamanho.

[45] Uma célula produzida pelo método do parágrafo 43 ou 44.

[46] A célula do parágrafo 45, compreendendo adicionalmente um gene que codifica uma proteína natural ou heteróloga.

15 [47] Um método de produzir uma proteína, compreendendo: (a) cultivar a célula do parágrafo 46 em condições produtivas para a produção da proteína; e (b) recuperar a proteína.

[48] Um polinucleotídeo isolado que codifica um peptídeo sinal compreendendo ou consistindo em aminoácidos 1 a 17 de SEQ ID NO: 2 ou aminoácidos 1 a 17 de SEQ ID NO: 4.

20 [49] Um construto de ácido nucleico ou vetor de expressão que compreende um gene que codifica uma proteína operavelmente ligada ao polinucleotídeo do parágrafo 48, em que o gene é estrangeiro ao polinucleotídeo que codifica o peptídeo sinal.

25 [50] Uma célula hospedeira recombinante compreendendo um gene que codifica uma proteína operavelmente ligada ao polinucleotídeo do parágrafo 48, em que o gene é estrangeiro ao polinucleotídeo que codifica o peptídeo sinal.

[51] Um método de produzir uma proteína, compreendendo:

- (a) cultivar uma célula hospedeira recombinante que compreende um gene que codifica uma proteína operavelmente ligada ao polinucleotídeo do parágrafo 48, em que o gene é estrangeiro ao polinucleotídeo que codifica o peptídeo sinal, em condições produtivas para a produção da proteína; e (b)
- 5 recuperar a proteína.

[52] Um método para degradar ou converter um material celulósico compreendendo: tratar o material celulósico com uma composição de enzima na presença do polipeptídeo com atividade celobiohidrolase de qualquer dos parágrafos 1-28.

- 10 [53] O método do parágrafo 52, em que o material celulósico é pré-tratado.

- [54] O método do parágrafo 52 ou 53, em que a composição de enzima compreende uma ou mais (várias) enzimas celulolíticas selecionadas do grupo que consiste em uma endoglucanase, uma celobiohidrolase, uma
- 15 beta-glicosidase e um polipeptídeo com melhor atividade celulolítica.

- [55] O método de qualquer dos parágrafos 52-54, em que a composição de enzima compreende ou compreende adicionalmente uma ou mais (várias) proteínas selecionadas do grupo que consiste em uma hemicelulase, uma expansina, uma esterase, uma enzima ligninolítica, uma
- 20 pectinase, uma peroxidase, uma protease e uma swolenina.

- [56] O método do parágrafo 55, em que a hemicelulase é uma ou mais (várias) enzimas selecionadas do grupo que consiste em uma acetilmanana esterase, uma acetixilana esterase, uma arabinanase, uma arabinofuranosidase, uma ácido coumárico esterase, uma feruloil esterase,
- 25 uma galactosidase, uma glucuronidase, uma glucuronoil esterase, uma mananase, uma manosidase, uma xilanase e uma xilosidase.

[57] O método de qualquer dos parágrafos 52-56, compreendendo adicionalmente recuperar o material celulósico degradado.

[58] O método do parágrafo 57, em que o material celulósico

degradado é um açúcar.

[59] O método do parágrafo 58, em que o açúcar é selecionado do grupo que consiste em glicose, xilose, manose, galactose e arabinose.

[60] Um método para sintetizar um produto de fermentação compreendendo: (a) sacarificar um material celulósico com uma composição de enzima na presença do polipeptídeo com atividade celobiohidrolase de qualquer dos parágrafos 1-28; (b) fermentar o material celulósico sacarificado com um ou mais microrganismos fermentadores para sintetiza o produto de fermentação; e (c) recuperar o produto de fermentação a partir da fermentação.

[61] O método do parágrafo 60, em que o material celulósico é pré-tratado.

[62] O método do parágrafo 60 ou 61, em que a composição de enzima compreende uma ou mais (várias) enzimas celulolíticas selecionadas do grupo que consiste em uma endoglucanase, uma celobiohidrolase, uma beta-glucosidase e um polipeptídeo com melhor atividade celulolítica.

[63] O método de qualquer dos parágrafos 60-62, em que a composição de enzima compreende ou compreende adicionalmente uma ou mais (várias) proteínas selecionadas do grupo que consiste em uma hemicelulase, uma expansina, uma esterase, uma enzima ligninolítica, uma pectinase, uma peroxidase, uma protease e uma swolenina.

[64] O método do parágrafo 63, em que a hemicelulase é uma ou mais (várias) enzimas selecionadas do grupo que consiste em uma acetilmanana esterase, uma acetixilano esterase, uma arabinanase, uma arabinofuranosidase, uma ácido coumárico esterase, uma feruloil esterase, uma galactosidase, uma glucuronidase, uma glucuronoil esterase, uma mananase, uma manosidase, uma xilanase e uma xilosidase.

[65] O método de qualquer dos parágrafos 60-64, em que as etapas (a) e (b) são realizadas simultaneamente em uma sacarificação e

fermentação simultâneas.

[66] O método de qualquer dos parágrafos 60-65, em que o produto de fermentação é um álcool, um ácido orgânico, uma cetona, um aminoácido, ou um gás.

5 [67] Um método de fermentar um material celulósico compreendendo: fermentar o material celulósico com um ou mais microrganismos fermentadores, em que o material celulósico é sacarificado com uma composição de enzima na presença de um polipeptídeo com atividade celobiohidrolase de qualquer dos parágrafos 1-28.

10 [68] O método do parágrafo 67, em que a fermentação do material celulósico produz um produto de fermentação.

[69] O método do parágrafo 68, compreendendo recuperar adicionalmente o produto de fermentação a partir da fermentação.

15 [70] O método de qualquer dos parágrafos 67-69, em que o material celulósico é pré-tratado antes da sacarificação.

[71] O método de qualquer dos parágrafos 67-70, em que a composição de enzima compreende uma ou mais (várias) enzimas celulolíticas selecionadas do grupo que consiste em uma endoglucanase, uma celobiohidrolase, uma beta-glicosidase e um polipeptídeo com melhor
20 atividade celulolítica.

[72] O método de qualquer dos parágrafos 67-71, em que a composição de enzima compreende ou compreende adicionalmente uma ou mais (várias) proteínas selecionadas do grupo que consiste em uma hemicelulase, uma expansina, uma esterase, uma enzima ligninolítica, uma
25 pectinase, uma peroxidase, uma protease e uma swolenina.

[73] O método do parágrafo 72, em que a hemicelulase é uma ou mais (várias) enzimas selecionadas do grupo que consiste em uma acetilmanana esterase, uma acetixilano esterase, uma arabinanase, uma arabinofuranosidase, uma ácido coumárico esterase, uma feruloil esterase,

uma galactosidase, uma glucuronidase, uma glucuronoil esterase, uma mananase, uma manosidase, uma xilanase e uma xilosidase.

5 [74] O método de qualquer dos parágrafos 67-73, em que o produto de fermentação é um álcool, um ácido orgânico, uma cetona, um aminoácido ou um gás.

10 A invenção e as reivindicações aqui descritas não devem ser limitadas no escopo por aspectos específicos aqui revelados, uma vez que estes aspectos são pretendidos como ilustrações de vários aspectos da invenção. Pretende-se que quaisquer aspectos equivalentes estejam no escopo desta invenção. De fato, várias modificações da invenção, além daquelas mostradas e descritas aqui, se tornarão evidentes aos versados na técnica a partir da descrição precedente. Também pretende-se que tais modificações estejam no escopo das reivindicações em anexo. No case de conflito, a presente revelação, incluindo as definições, controlarão.

REIVINDICAÇÕES

1. Célula hospedeira microbiana transgênica, caracterizada pelo fato de que compreende um construto de ácido nucleico compreendendo um polinucleotídeo que codifica um polipeptídeo tendo atividade celobiohidrolase, em que o polinucleotídeo consiste na SEQ ID NO: 1, nucleotídeos 52 a 1595 de SEQ ID NO: 1, SEQ ID NO: 3, ou nucleotídeos 52 a 1581 de SEQ ID NO: 3.

2. Célula hospedeira microbiana transgênica de acordo com a reivindicação 1, caracterizada pelo fato de que o polipeptídeo tendo atividade celobiohidrolase consiste na SEQ ID NO: 2, aminoácidos 18^a a 493 de SEQ ID NO: 2, SEQ ID NO: 4, ou aminoácidos 18 a 487 de SEQ ID NO: 4.

3. Método para produzir um polipeptídeo tendo atividade celobiohidrolase, caracterizado pelo fato de que compreende:

(a) cultivar uma célula hospedeira microbiana transgênica como definida em qualquer uma das reivindicações 1-2, ou uma planta transgênica ou uma célula de planta transgênica compreendendo um construto de ácido nucleico compreendendo um polinucleotídeo que codifica um polipeptídeo tendo atividade celobiohidrolase, em que o polinucleotídeo consiste na SEQ ID NO: 1, nucleotídeos 52 a 1595 de SEQ ID NO: 1, SEQ ID NO: 3, ou nucleotídeos 52 a 1581 de SEQ ID NO: 3 para a produção do polipeptídeo; e

(b) recuperar o polipeptídeo.

4. Método para produzir um mutante de uma célula mãe, caracterizado pelo fato de que compreende interromper ou deletar um polinucleotídeo que codifica o polipeptídeo tendo atividade celobiohidrolase, em que o polinucleotídeo consiste na SEQ ID NO: 1, nucleotídeos 52 a 1595 de SEQ ID NO: 1, SEQ ID NO: 3, ou nucleotídeos 52 a 1581 de SEQ ID NO: 3, que resulta no mutante que produz menos do polipeptídeo do que a célula mãe.

5. Método para degradar ou converter um material celulósico, caracterizado pelo fato de que compreende: tratar o material celulósico com uma composição de enzima compreendendo um polipeptídeo tendo atividade celobiohidrolase consistindo de SEQ ID NO: 2, aminoácidos 18 a 493 de SEQ ID NO: 2, SEQ ID NO: 4, ou aminoácidos 18 a 487 de SEQ ID NO: 4.

6. Método de acordo com a reivindicação 5, caracterizado pelo fato de que compreende adicionalmente recuperar o material celulósico degradado ou convertido.

7. Método para produzir um produto de fermentação, caracterizado pelo fato de que compreende:

(a) sacarificar um material celulósico com uma composição de enzima compreendendo um polipeptídeo com atividade celobiohidrolase consistindo de SEQ ID NO: 2, aminoácidos 18 a 493 de SEQ ID NO: 2, SEQ ID NO: 4, ou aminoácidos 18 a 487 de SEQ ID NO: 4;

(b) fermentar o material celulósico sacarificado com um ou mais microrganismos fermentadores para produzir o produto de fermentação; e

(c) recuperar o produto de fermentação a partir da fermentação.

8. Método para fermentar um material celulósico, caracterizado pelo fato de que compreende: fermentar o material celulósico com um ou mais microrganismos fermentadores, em que o material celulósico é sacarificado com uma composição de enzima compreendendo um polipeptídeo com atividade celobiohidrolase consistindo de SEQ ID NO: 2, aminoácidos 18 a 493 de SEQ ID NO: 2, SEQ ID NO: 4, ou aminoácidos 18 a 487 de SEQ ID NO: 4.

9. Método de acordo com a reivindicação 8, caracterizado pelo fato de que a fermentação do material celulósico produz um produto de fermentação.

10. Método de acordo com a reivindicação 8, caracterizado pelo fato de que compreende recuperar adicionalmente o produto de fermentação a partir da fermentação.

5 11. Construto de ácido nucleico, caracterizado pelo fato de que compreende um polinucleotídeo consistindo na SEQ ID NO: 1, nucleotídeos 52 a 1595 de SEQ ID NO: 1, SEQ ID NO: 3, ou nucleotídeos 52 a 1581 de SEQ ID NO: 3, operavelmente ligado a uma ou mais sequências controle que direcionam a produção do polipeptídeo em um hospedeiro de expressão.

10 12. Vetor de expressão, caracterizado pelo fato de que compreende um polinucleotídeo consistindo na SEQ ID NO: 1, nucleotídeos 52 a 1595 de SEQ ID NO: 1, SEQ ID NO: 3, ou nucleotídeos 52 a 1581 de SEQ ID NO: 3 operavelmente ligado a uma ou mais sequências controle que direcionam a produção do polipeptídeo em uma célula hospedeira microbiana transgênica.

M A K K L L L T A A L A A T A L A A P I
 1 ATGGCCAAGA AGCTCCTTCT CACCGCCGCC CTCGCGGCTA CTGCCCTGGC TGCTCCCATC
 V E E R Q N C G S V W
 61. GTCGAGGAGC GCCAGAACTG TGGCTCCGTC TGGTAAGTGA CTGCTTGATG ATCTACGGTG
 S Q C G G Q G W S G
 121 AAAGCTGCTC GCTGACACGA CGACTTCCAG GAGTCAATGT GGTGGCCAAG GGTGGTCGGG
 A T C C A S G S T C V A Q N P W Y S Q C
 181 CGCCACCTGC TGCGCGTCTG GAAGCACCTG CGTGGCCCAG AACCCCTGGT ACTCGCAGTG
 L P N S Q V T T T A T S A R A S S S S S
 241 CCTGCCAAC AGCCAGGTGA CCACCACCGC GACTAGCGCT CGCGCCTCGT CGTCTCTGTC
 S S S S T K A S T S S T S R T T V G T T
 301 CTCGTCTCTG TCCACCAAGG CGAGCACCAG CAGCACCAGC CGCACCACCG TGGGCACCAC
 S V G T T T R A S S T T T S P P V V T S
 361 CTCGGTGGGC ACCACCACTC GTGCTAGCTC CACGACCACC TCTCCCCCGG TCGTCTCTCTC
 V V P G G A T A T A S Y S G N P F S G V
 421 GGTCGTTCCT GCGCGTGCCA CGGCCACGGC GAGCTACTCG GGCAACCCCT TCTCCGGCGT
 R L W A N D Y Y A S E V S T L A M P S L
 481 GCGCCTGTGG GCCAACGACT ACTACGCCCTC CGAGGTGTCTG ACTCTCGCCA TGCCTTCCT
 T G A M A T K A A A V A K V P S F Q W L
 541 GACGGGCGCC ATGGCCACCA AGGCGGCCGC TGTCGCCAAG GTGCCCAGCT TCCAGTGGCT
 D R N V T I D T L M V K T L S Q I R A A
 601 GGACCGCAAC GTCACCATCG ACACCCTGAT GGTCAAGACT CTGTCCCAAG TCCGCGCCGC
 N Q A G A N P P Y A
 661 CAACCAGGCC GGTGCCAACC CCCCCTATGC CGGTAAGTTT GAGTTCGATA CTTGTTCACT
 A Q L V V Y D L P D R
 721 TCCAAAAGCT CGGACTAACG ACATGTGCAG CCCAGCTGGT GGTCTACGAC CTCCCCGACC
 D C A A A A S N G E F S I A N N G A A N
 781 GTGACTGCGC TGCCGCCGCC TCCAACGGCG AGTTCCTGAT TGCCAACAAC GGCGCGGCCA
 Y K S Y I D S I R K H L I E Y S D I R T
 841 ACTACAAGTC GTACATTGAC TCGATCCGCA AGCACCTCAT CGAGTACTCG GACATCCGCA
 I L V I E P D S M A N M V T N M N V A K
 901 CCATTCTCGT CATTGAGCCC GACTCGATGG CCAACATGGT CACCAACATG AACGTCGCCA
 C S N A A T T Y R E L T V Y A L K Q L N
 961 AGTCAGCAA CGCGGCCACC ACGTACCGCG AGCTGACCGT CTACGCCCTC AAGCAGCTGA
 L P H V A M Y L D A G H A G W L G W P A
 1021 ACCTGCCCCA CGTCGCCATG TACCTCGACG CCGGCCACGC CGGCTGGCTC GGCTGGCCCG
 N I Q P A A T L F A G I Y S D A G K P A
 1081 CCAACATCCA GCCCCTGCC ACCCTGTTG CCGGCATCTA CAGCGACGCT GGCAAGCCCG
 S V R G L A T N V A N Y N A W S L S S A
 1141 CCTCGGTCCG CGGTTTGGCC ACCAAGTGG CCAACTACAA CGCCTGGAGC CTGTCTGTCG
 P S Y T S P N A N Y D E K H Y V E A F A
 1201 CGCCGTCTGA CACGAGCCCC AACGCCAACT ACGACGAGAA GCCTACGTC GAGGCCTTTG
 P L L Q A A G F P A K F I T D T G R N G
 1261 CCCCCTCCT CCAGCGGCC GGCTTCCCCG CCAAGTTCAT CACCGACACG GGCCGCAACG
 K Q P T G Q S A W G D W C N V K G T G F
 1321 GCAAGCAGCC CACGGGCCAG AGCGCGTGGG GCGACTGGTG CAACGTCAAG GGCACCGCT
 G V R P T S E T G H D L L D A F V W V K
 1381 TCGGTGTCCG CCCGACCTCG GAGACGGGCC ACGACCTCCT CGACGCCTTC GTCTGGGTCA
 P G G E S D G T S D T S A A R Y D Y H C
 1441 AGCCCGGTGG CGAGTCGGAC GGCACCAGCG ACACCAGCGC CGCCCGCTAC GACTACCACT
 G L S D A L Q P A P E A G T W F Q A Y F
 1501 GCGGTCTGTC GGATGCCCTC CAGCCTGCTC CCGAGGCTGG CACGTGGTTC CAGGCCTACT
 E Q L L T N A N P P F *
 1561 TTGAGCAGCT GCTACCAAC GCCAACCTC CTTTTTAG

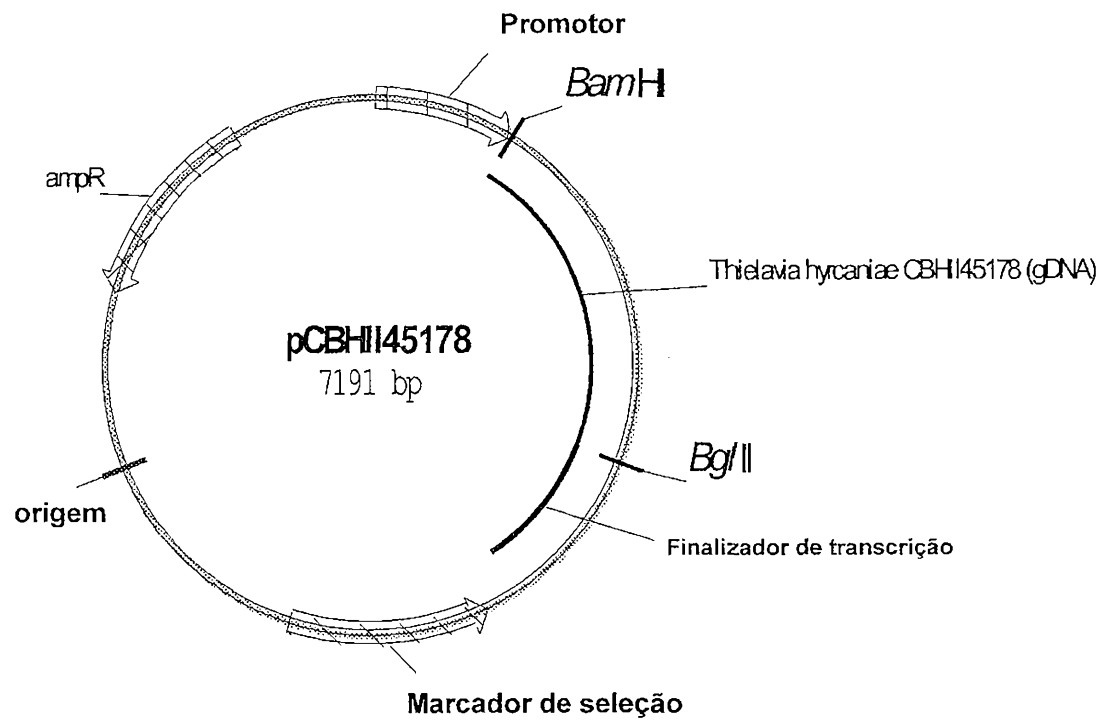
FIG. 1

```

      M A K K L L L T A A L A A S A L A A P I
1  ATGGCCAAGA AGCTCCTTCT CACCGCCGCC CTCGCGGCCT CTGCCCTGGC TGCTCCCATT
      V E E R Q N C G S V W
61  GTCGAGGAGC GCCAGAACTG CGGCTCCGTC TGGTAAGTGA TCGTTCGAGG ATTCCATGGT
      S Q C G G N G W S G
121 GAAAGCTGCT CGCTCACACG ATGCTCTCCA GGAGTCAATG TGGTGGCAAC GGGTGGTCGG
      A T C C A S G S T C V A Q N P W Y S Q C
181 GCGCCACCTG CTGCGCGTCT GGAAGCACCT GCGTAGCCCA GAACCCCTGG TACTCGCAGT
      L P S S Q V T T T T T S A R A S S S S S
241 GCCTGCCCAG CAGCCAGGTG ACCACCACCA CGACTAGCGC TCGCGCCTCG TCGTCGTCTT
      S S T R A S T S S T S R T T S V G T T T
301 CGTCGTCCAC CCGGGCGAGC ACCAGCAGCA CCAGCCGCAC CACCTCGGTG GGCACACCA
      R A S S T T T S A P G V T S S V P G G A
361 CCCGTGCTAG CTCCACGACC ACCTCGGCCC CTGGCGTCAC GTCCAGCGTT CCTGGCGGTG
      T A T A S Y S G N P F S G V R L W A N D
421 CTACGGCCAC GGCCAGCTAC TCGGGCAACC CCTTCTCCGG GGTGCGCCTG TGGGCCAACG
      Y Y A S E V S T L A M P S L T G A M A T
481 ACTACTACGC CTCCGAGGTG TCGACCCCTG CCATGCCTTC CCTGACGGGC GCCATGGCCA
      K A A A V A K V P S F Q W L D R N V T I
541 CCAAGCGGCG CGCCGTCGCC AAGGTCCCCA GCTTCCAGTG GCTGGACCGC AACGTACCA
      D T L M V K T L S Q I R A A N Q A G A N
601 TCGACACCCT GATGGTCAAG ACTCTGTCCC AGATCCGGGC CGCCAACCAG GCCGGTGCCA
      P P Y A
661 ACCCCCCGTA TGCCGGTAAG CTTCGACTCG ATCTTGTCTG CCTGCTCTCA AGTGCTCGAA
      A Q L V V Y D L P D R D C A A
721 CTAAGACGAT GCGCAGCCCA GCTGGTGGTC TACGACCTCC CCGACCGTGA CTGCGCTGCC
      A A S N G E F S I A N N G A A N Y K S Y
781 GCCGCCCTCA ACGGCGAGTT CTCGATTGCC AACAACGGCG CGGCCAACTA CAAGTCGTAC
      I D S I R K H L I E Y S D I R T I L V I
841 ATCGACTCGA TCCGCAAGCA CCTCATCGAG TACTCGGACA TCCGCACCAT CCTGGTTATT
      E P D S M A N M V T N M N V A K C S N A
901 GAGCCCGATT CGATGGCCAA CATGGTCACC AACATGAACG TCGCCAAGTG CAGCAACGCC
      A T T Y R E L T I Y A L K Q L N L P H V
961 GCCACGACCT ACCGCGAGCT GACCATCTAC GCTCTCAAGC AGCTGAACCT GCCCACGTC
      A M Y L D A G H A G W L G W P A N I Q P
1021 GCCATGTACC TCGACGCCGG CCACGCCGGC TGGCTCGGCT GGCCCCGCAA CATCCAGCCC
      A A T L F A G I Y N D A G K P A S V R G
1081 GCTGCTACCC TGTTGCGCGG CATCTACAAC GACGCTGGCA AGCCCGCCTC GGTCCGTGGT
      L A T N V A N Y N A W S L S S A P S Y T
1141 CTGGCCACCA ACGTCGCCAA CTACAACGCC TGGAGCCTGT CCTCGGCCCC GTCGTACAG
      T P N A N Y D E K H Y V E A F A P L L S
1201 ACCCCCAACG CCAACTACGA CGAGAAGCAC TACGTCGAGG CCTTTGCCCC GCTTCTCTCG
      A A G F P A K F I T D T G R N G K Q P T
1261 GCGCTGGCT TCCCCGCCAA GTTCATCACC GACACTGGCC GCAACGGCAA GCAGCCCACC
      G Q S Q W G D W C N V K G T G F G V R P
1321 GGCCAGAGCC AGTGGGGCGA TTGGTGCAAC GTCAAGGGCA CCGGCTTCGG TGTCCGCCCG
      T S E T G H E L L D A F V W V K P G G E
1381 ACCTCCGAGA CGGGCCACGA GCTCCTGGAT GCCTTTGTCT GGGTCAAGCC CCGTGGCGAG
      S D G T S D T S A A R Y D Y H C G L S D
1441 TCCGACGGTA CCAGCGACAC CAGCGCTGCC CGCTACGACT ACCACTGCGG TCTGTGCGAT
      A L Q P A P E A G T W F Q A Y F E Q L L
1501 GCCCTCCAGC CCGCTCCCGA GGCCGGCACG TGGTTCAGG CCTACTTTGA ACAGCTCCTC
      T N A N P P F *
1561 ACCAACGCCA ACCCTCCTTT TTAG

```

FIG. 2

**FIG. 3**

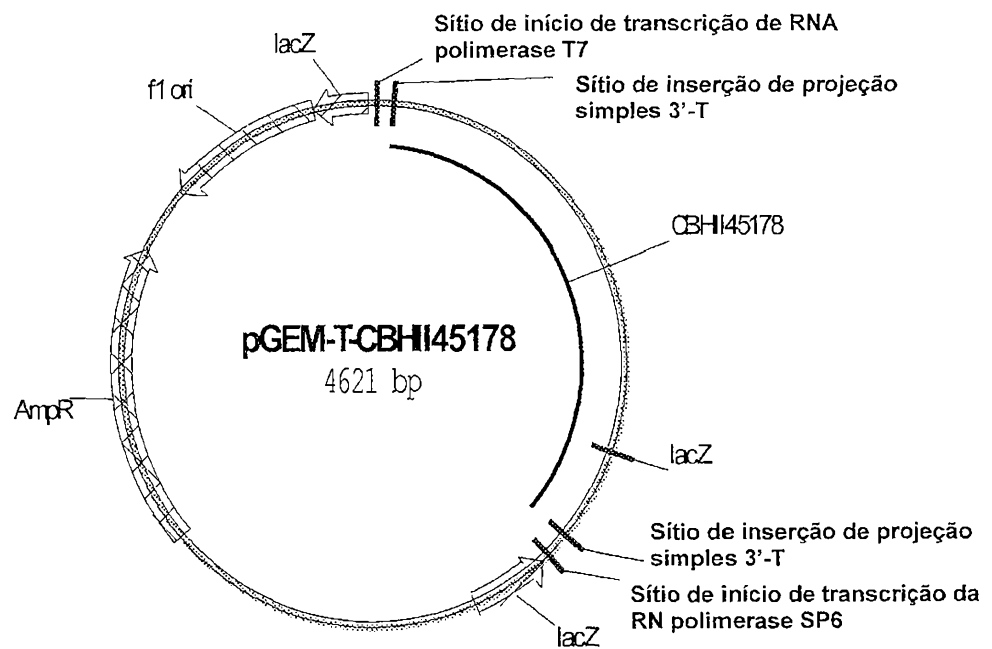
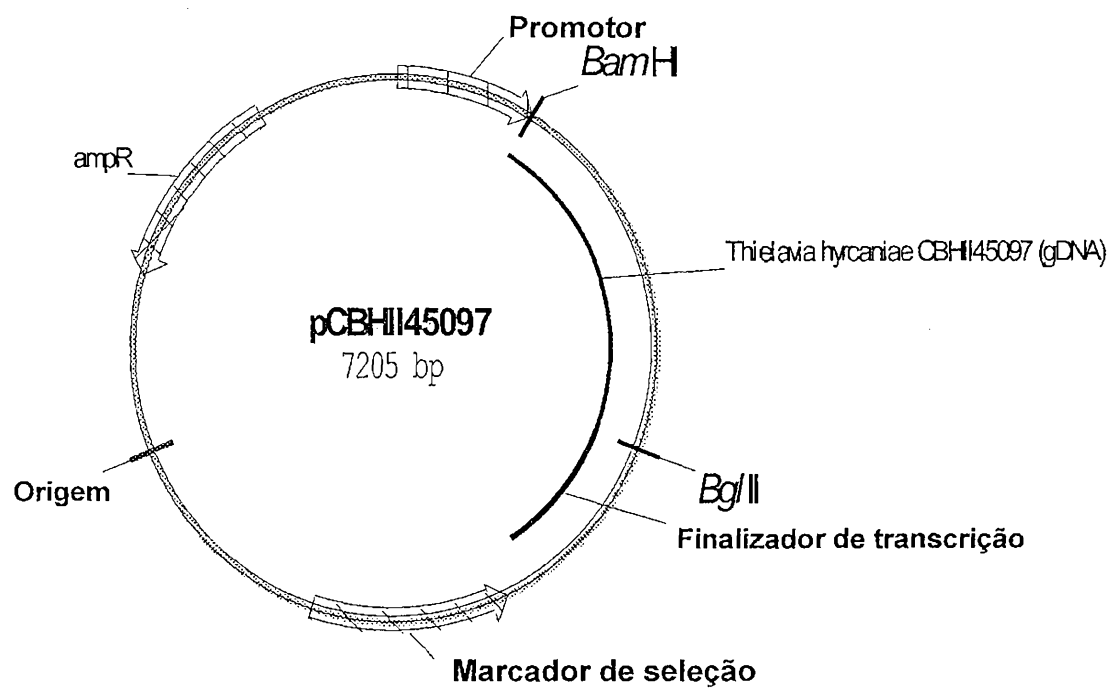
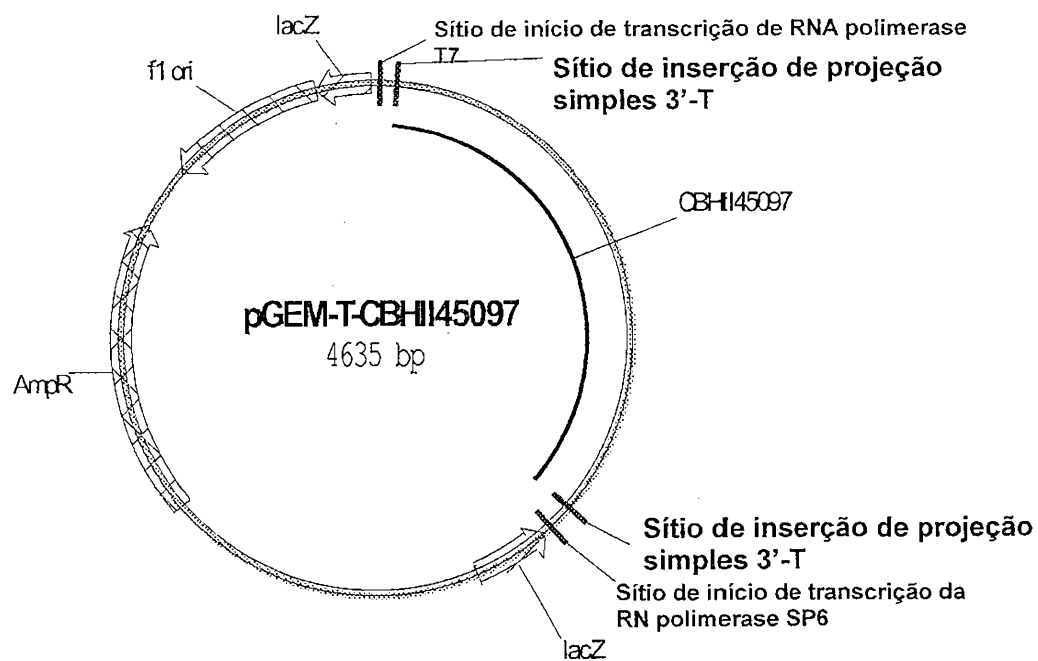


FIG. 4

**FIG. 5**

**FIG. 6**

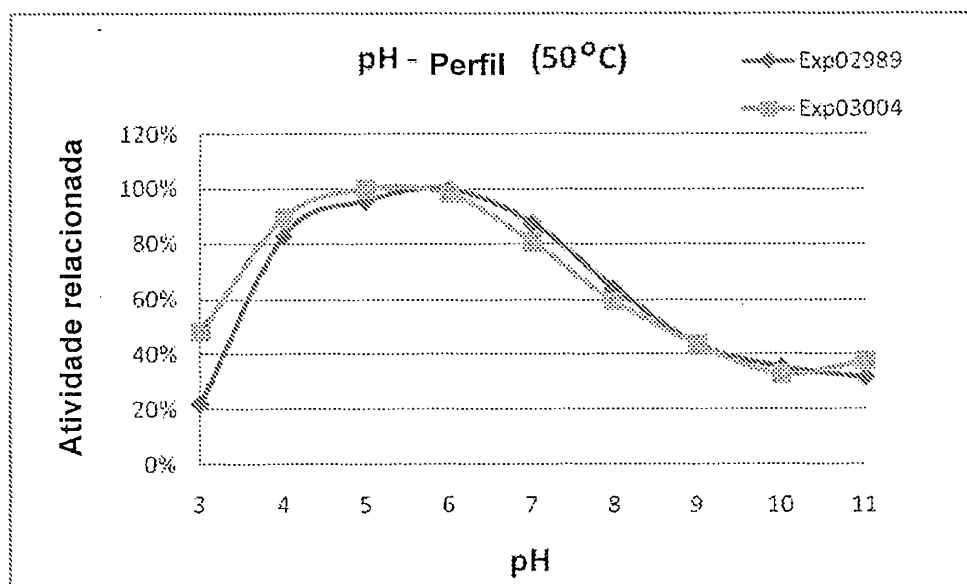
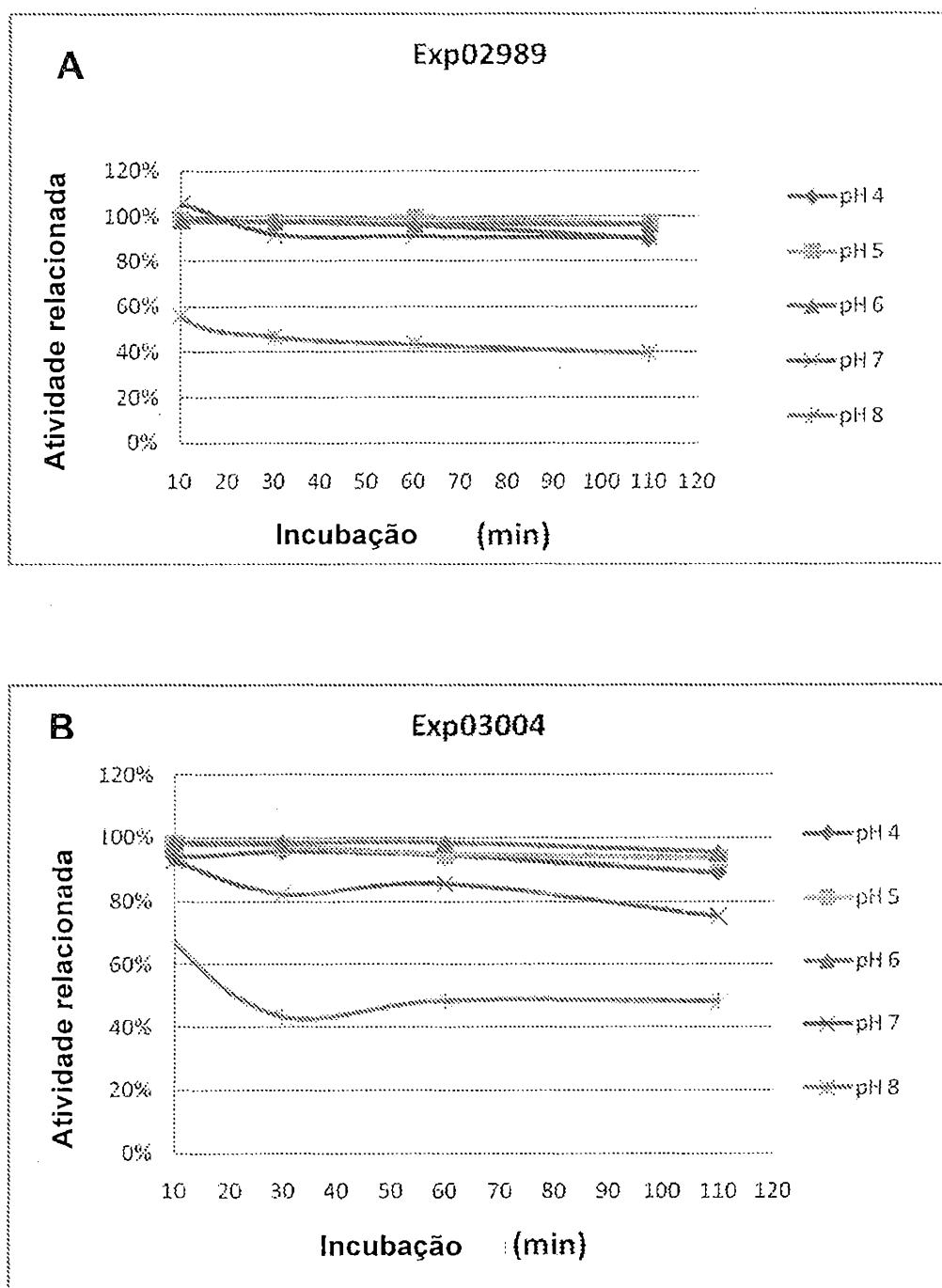
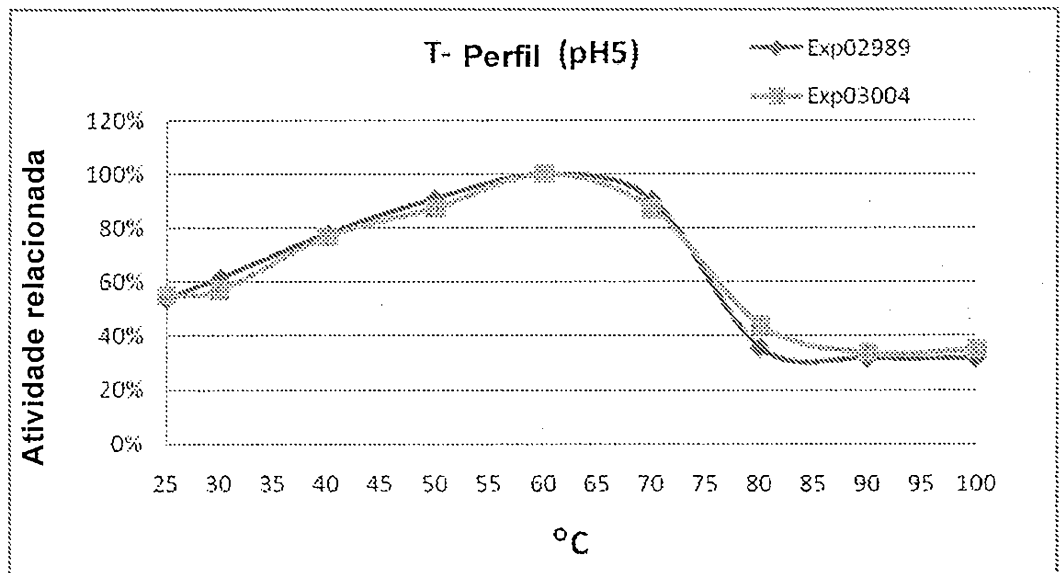
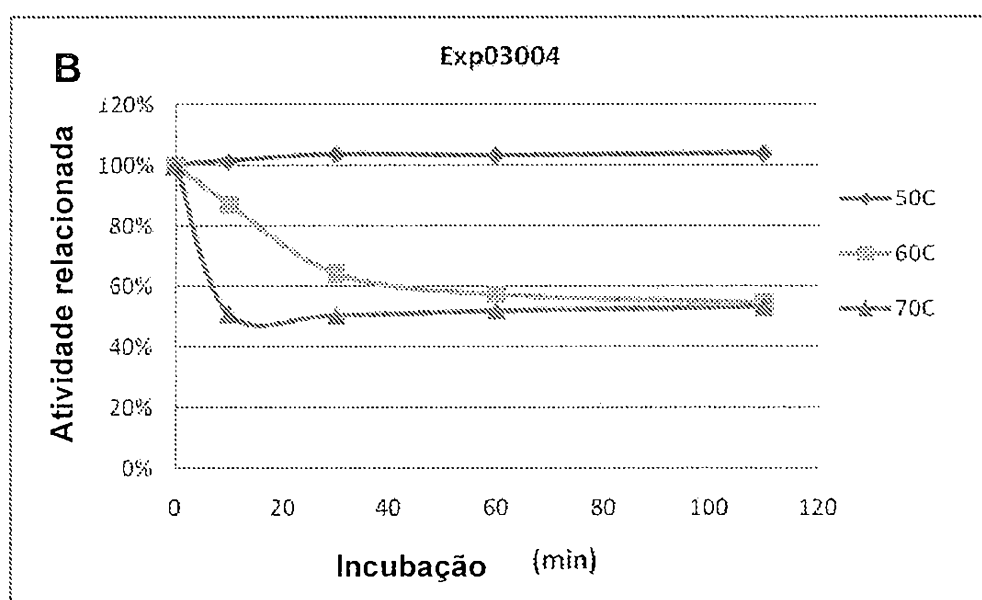
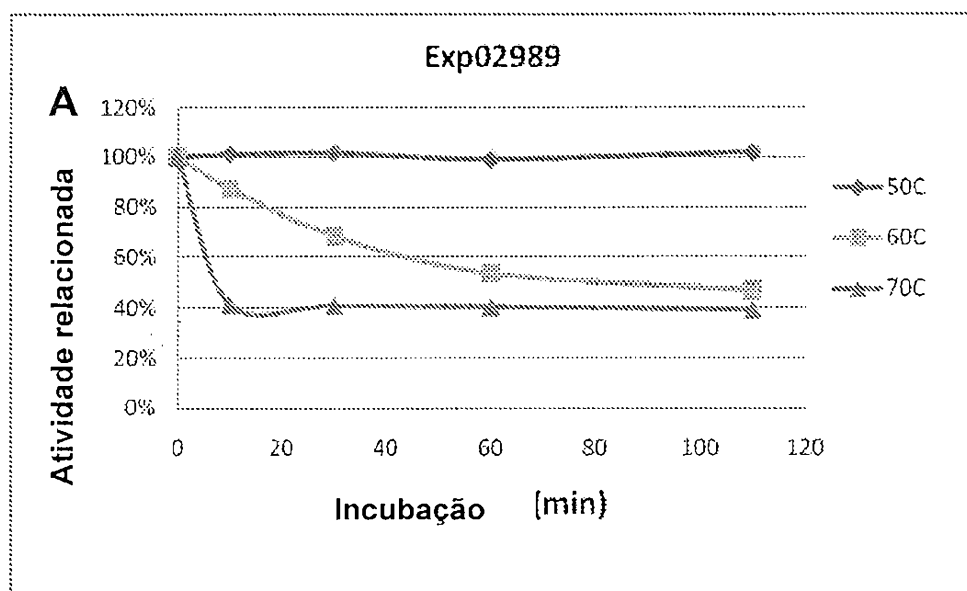


FIG. 7

**FIG. 8**

**FIG. 9**

**FIG. 10**