

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

(21) **PI 0621874-1 A2**

BRPI0621874A2

(22) Data de Depósito: 06/10/2006

(43) **Data da Publicação:** 20/12/2011
(RPI 2137)

(51) Int.Cl.:

C12P 19/04

C12N 9/14

(54) Título: CONSTRUÇÃO DE COMPOSIÇÕES DE CELULASE ALTAMENTE EFICAZES PARA A HIDRÓLISE ENZIMÁTICA DE CELULOSE

(30) Prioridade Unionista: 13/07/2006 US 11/487.547

(73) Titular(es): Dyadic International (USA), Inc.

(72) Inventor(es): Alexander V. Gusakov, Alexey I. Antonov, Arkady P. Sinitsyn, Boris B. Ustinov, Marco Baez, Mark Emalfarb, Oleg N. Okunev, Richard Burlingame, Tatyana N. Salanovich

(74) Procurador(es): Alexandre Ferreira

(86) Pedido Internacional: PCT US2006027347 de 06/10/2006

(87) Publicação Internacional: WO 2008/008070de 17/01/2008

(57) Resumo: CONSTRUÇÃO DE COMPOSIÇÕES DE CELULASE ALTAMENTE EFICAZES PARA A HIDRÓLISE ENZIMÁTICA DE CELULOSE. Esta invenção fornece novas composições enzimáticas usando enzimas de *C. lucknowense* recém identificadas e isoladas, incluindo CBH Ib, CBH IIb, EG II, EG VI, β -glicosidase, e xilanase II em combinação com as enzimas CBH Ia, CBH IIa (previamente descrita como Endo 43), e EG V previamente identificadas. Estas composições enzimáticas demonstram uma capacidade extremamente alta para converter biomassa lignocelulósica (por exemplo, Avicel, algodão, madeira da espécie Douglas fir pré-tratada por organosolvente) em glicose. CBH Ia e IIb, as quais têm um módulo de ligação de celulose (CBM) exibiram um sinergismo considerável com três endoglicanases principais (EG II, EG V, EG VI) provenientes do mesmo fungo na hidrólise de algodão, assim como uma forte sinergia entre si. As composições enzimáticas são eficazes na hidrólise da biomassa lignocelulósica.

"CONSTRUÇÃO DE COMPOSIÇÕES DE CELULASE ALTAMENTE EFICAZES PARA A HIDRÓLISE ENZIMÁTICA DE CELULOSE"

Este pedido é uma prorrogação parcial do Pedido de Patente U.S. 10/394.568, depositado em 21 de Março de 2003, o qual reivindica o benefício do Pedido de Patente U.S. 09/548.938 (agora Patente US 6.573.086), depositado em 13 de Abril de 2000, o qual é uma prorrogação parcial do pedido internacional PCT/NL99/00618, depositado em 6 de Outubro de 1999, o qual é uma prorrogação parcial do pedido internacional PCT/EP98/06496, depositado em 6 de Outubro de 1998. Este pedido também é um pedido de prorrogação parcial do Pedido de Patente U.S. 09/284.152, depositado em 8 de Abril de 1999 o qual é uma prorrogação parcial de 08/731.170 depositado em 10 de Outubro de 1996. Todos os pedidos anteriores aos quais a prioridade é reivindicada são incorporados integralmente neste como referência.

CAMPO DA INVENÇÃO

Esta invenção refere-se a composições e métodos para produzir bioenergia ou outros produtos com valor agregado a partir de biomassa lignocelulósica ou materiais celulósicos. Em particular, a invenção fornece composições enzimáticas capazes de converter uma variedade de substratos celulósicos ou biomassa lignocelulósica em um açúcar fermentável. A invenção também fornece métodos para o uso de tais composições enzimáticas.

INTRODUÇÃO

A bioconversão de biomassa lignocelulósica renovável para um açúcar fermentável que é subsequentemente fermentado para produzir álcool (por exemplo, etanol) como uma alternativa para combustíveis líquidos, tem atraído uma atenção intensiva de pesquisadores desde 1970, quando eclodiu a crise do petróleo por causa da diminuição da produção de petróleo pela OPEC (Bungay H.R., "Energy: the biomass options". NY: Wiley; 1981; Olsson L, Hahn-Hagerdal B. "Fermentation of lignocellulosic hydrolysates for ethanol production", *Enzyme Microb Technol* 1996;18:312-31; Zaldivar J, Nielsen J, Olsson L. "Fuel ethanol production from lignocellulose: a challenge for metabolic engineering and process integration", *Appl Microbiol Biotechnol* 2001;56:17-34; Galbe M, Zacchi G., "A review of the production of ethanol from softwood", *Appl Microbiol Biotechnol* 2002; 59:618-28). O etanol foi amplamente usado como uma combinação de 10 % para a gasolina nos EUA ou como um combustível puro para veículos no Brasil nas últimas duas décadas. A importância do bioetanol combustível aumentará em paralelo com a alta de preços do petróleo e esgotamento gradual de suas fontes. Adicionalmente, açúcares fermentáveis estão sendo usados para produzir plásticos, polímeros e outros produtos com base biológica e, portanto, espera-se que sua indústria cresça substancialmente aumentando a demanda quanto a açúcares fermentáveis de baixo custo abundantes que possam ser usados como um estoque de abastecimento no lugar de matérias-primas com base no petróleo (por exemplo, vide artigo "The Rise Of In-

dustrial Biotech" publicado na Forbes, 24 de Julho de 2006).

Os principais polissacarídeos compreendendo diferentes resíduos lignocelulósicos, que podem ser considerados como uma matéria-prima renovável potencial, são celulose e hemiceluloses (xilanas). A hidrólise enzimática destes polissacarídeos para açúcares solúveis, por exemplo glicose, xilose, arabinose, galactose, manose, e outras hexoses e pentoses ocorre sob a ação de diferentes enzimas que agem em conjunto. Endo-1,4- β -glicanases (EG) e exo-celobiohidrolases (CBH) catalisam a hidrólise de celulose insolúvel para celooligossacarídeos (celbiose como um produto principal), enquanto β -glicosidases (BGL) convertem os oligossacarídeos para glicose. Xilanases junto com outras enzimas adicionais (exemplos não limitantes dos quais incluem α -L-arabinofuranosidases, feruloil e acetilxilana esterases, glicuronidases, e (β -xilosidases) catalisam a hidrólise de hemiceluloses.

Não obstante o tipo de matéria-prima celulósica, o custo e eficiência hidrolítica de enzimas são fatores principais que restringem a comercialização dos processos de bioconversão de biomassa. Os custos de produção de enzimas microbialmente produzidas estão intimamente associados com uma produtividade da cepa produtora de enzima e com a atividade final produzida no caldo de fermentação. A eficiência hidrolítica de um complexo multienzimático no processo de sacarificação de lignocelulose depende tanto das propriedades de enzimas individuais, das sinergias entre elas, quanto de sua razão no coquetel multienzimático.

Chrysosporium lucknowense é um fungo que é conhecido para produzir uma ampla variedade de celulasas, hemicelulasas, e possivelmente outras enzimas adicionais. *C. lucknowense* também secreta pelo menos cinco endoglicanases diferentes, a EG II (51 kDa, Cel5A) sendo a mais ativa. Além disso, cepas mutantes de *C. lucknowense* (incluindo UV18-25) foram desenvolvidas para produzir enzimas para tecido, polpa e papel, detergente e outras aplicações, mas não para a sacarificação enzimática de celulose; estas cepas também podem ser usadas para uma produção de nível elevado de proteínas homólogas e heterólogas. As melhores cepas mutantes de *C. lucknowense* secretam pelo menos 50 a 80 g l⁻¹ de proteína extracelular em fermentações de baixa viscosidade. O genoma fúngico completo de *C. lucknowense* foi sequenciado em 2005 (vide http://WorldWideWeb.dyadic-group.com/wt/dyad/pr_1115654417), e presentemente a anotação do genoma está sendo realizada.

O complexo multienzimático de *C. lucknowense* bruto demonstra resultados moderados na sacarificação da celulose, com apenas uma fração da celulose sendo convertida para glicose sob as condições testadas. Duas celobiohidrolases de *C. lucknowense*, pertencentes às famílias 7 e 6 de glicosídeo hidrolases: CBH Ia (Cel7A) e CBH IIa (Cel6A), foram previamente isoladas e estudadas. CBH Ia foi previamente referida como proteína CBH I, 70(60) kD na Patente US Nº 6.573.086. CBH Ia vive no caldo de cultivo como uma enzima

de tamanho natural (massa molecular observada de 65 kDa, dados SDS-PAGE), que consiste de um domínio catalítico de núcleo e módulo de ligação de celulose (CBM) conectados por um ligador de peptídeo flexível, e sua forma truncada (52 kDa), representando o domínio catalítico da enzima. CBH I (Cel7A) de *C. lucknowense* parece ser levemente menos eficaz na hidrólise de celulose cristalina, porém mais termoestável do que a CBH I de *T. reesei*. CBH IIa foi previamente considerada ser uma endoglicanase e foi referida como Endo e EG6, 43 kD. Vide, por exemplo, Patente US Nº 6.573.086. CBH IIa (43 kDa) não tem CBM, isto é, sua molécula contém apenas o domínio catalítico.

Apesar da pesquisa constante nas últimas décadas para entender a degradação da biomassa lignocelulósica enzimática e produção de celulase, permanece desejável descobrir ou criar novas celulasas e hemicelulasas altamente ativas. Também seria altamente desejável construir composições enzimáticas altamente eficazes capazes de realizar biodegradação rápida e eficaz de materiais lignocelulósicos.

SUMÁRIO DA INVENÇÃO

Esta invenção fornece várias enzimas recém identificadas e isoladas provenientes de *C. lucknowense*. As novas enzimas incluem duas novas celobiohidrolases (CBH Ib e IIb, ou Cel7B e Cel6B), uma endoglicanase (EG VI), (não confundir com CBH IIa, que foi previamente referida como EG 6) uma β -glicosidase (BGL), e uma xilanase (Xyl II). A CBH IIb tem uma alta atividade contra Avicel e algodão e exibiu um sinergismo considerável com outras celulasas de *C. lucknowense*. Com o uso destas novas enzimas, esta invenção fornece composições enzimáticas altamente eficazes para a hidrólise da celulose.

Um objetivo desta invenção é fornecer uma formulação enzimática que inclua pelo menos uma celobiohidrolase isolada obtida a partir de *C. lucknowense*. A celobiohidrolase isolada pode ser tanto CBH Ib quanto IIb. A formulação enzimática pode opcionalmente conter uma endoglicanase e/ou uma β -glicosidase. Além disso, a formulação enzimática pode opcionalmente conter uma hemicelulase.

Um outro objetivo desta invenção é fornecer um método para produzir glicose a partir da celulose. O método inclui produzir uma formulação enzimática que contém pelo menos uma celobiohidrolase isolada obtida a partir de *C. lucknowense*, que pode ser CBH Ib ou IIb. Opcionalmente, a formulação enzimática pode conter uma endoglicanase e/ou uma β -glicosidase. A formulação enzimática é aplicada à celulose para formar glicose.

Ainda um outro aspecto desta invenção é fornecer um método de produzir etanol. O método inclui fornecer uma formulação enzimática que contém pelo menos uma celobiohidrolase isolada obtida a partir de *C. lucknowense*, que pode ser CBH Ib ou IIb. A formulação enzimática opcionalmente pode conter uma endoglicanase e/ou uma β -glicosidase. Além disso, a formulação enzimática pode opcionalmente conter uma hemicelulase. O método ainda inclui aplicar a formulação enzimática à celulose para produzir glicose e subsequen-

temente fermentar a glicose para produzir etanol.

Esta invenção também fornece um método de produzir energia a partir do etanol. O método inclui fornecer uma formulação enzimática que contém pelo menos uma celobiohidrolase isolada obtida a partir de *C. lucknowense*, que pode ser CBH Ib ou IIb. A formulação enzimática opcionalmente pode conter uma endoglicanase e/ou uma β -glicosidase. Além disso, a formulação enzimática pode opcionalmente conter uma hemicelulase. O método ainda inclui aplicar a formulação enzimática à celulose para produzir glicose, fermentar a glicose para produzir etanol, e queimar o dito etanol para produzir energia.

Um outro aspecto desta invenção é fornecer uma cepa de *Chrysosporium lucknowense* mutante capaz de expressar pelo menos uma celobiohidrolase e pelo menos uma endo-1,4- β -glicanase em níveis mais altos do que a cepa não mutante correspondente sob as mesmas condições. A celobiohidrolase é selecionada a partir do grupo que consiste de CBH Ia, CBH IIa, CBH Ib, e CBH IIb; e a endo-1,4- β -glicanase é selecionada a partir do grupo que consiste de EG II, EG V, e EG VI.

Ainda um outro aspecto desta invenção é fornecer proteínas que exibem pelo menos 65 % da identidade do aminoácido conforme determinado pelo algoritmo BLAST com as sequências de aminoácido de CBH Ib, CBH IIb, EG VI, BGL, e Xyl II de SEQ ID NOs. 2, 4, 16, 12, e 18, respectivamente, ou uma parte destas tendo pelo menos 20 aminoácidos contíguos. Esta invenção também considera as sequências de ácido nucléico correspondentes que codificam uma tal proteína.

Um aspecto desta invenção fornece uma formulação enzimática compreendendo pelo menos uma enzima selecionada a partir do grupo que consiste de CBH Ib, CBH IIb, EG II, EG VI, BGL, e Xyl II.

Um outro aspecto desta invenção fornece um método de produzir açúcares fermentáveis a partir de material lignocelulósico. O método compreende (a) fornecer uma formulação enzimática compreendendo pelo menos uma enzima selecionada a partir do grupo que consiste de CBH Ib, CBH IIb, EG II, EG VI, BGL, e Xyl II; e (b) aplicar a formulação enzimática ao material lignocelulósico para produzir açúcares fermentáveis.

A invenção também fornece um método de produzir um produto de fermentação ou um material de partida para um produto de fermentação a partir de um açúcar fermentável. Este método compreende (a) fornecer uma formulação enzimática, em que a formulação enzimática contém pelo menos uma enzima selecionada a partir do grupo que consiste de CBH Ib, CBH IIb, EG II, EG VI, BGL, e Xyl II; (b) aplicar a formulação enzimática ao material lignocelulósico para produzir um açúcar fermentável; e (c) fermentar o dito açúcar fermentável para produzir um produto de fermentação.

Em um outro aspecto, a invenção fornece um método de produzir energia a partir de um açúcar fermentável. O método compreende (a) fornecer uma formulação enzimática,

em que a formulação enzimática compreende pelo menos uma enzima selecionada a partir do grupo que consiste de CBH Ib, CBH IIb, EG II, EG VI, BGL, e Xyl II; (b) aplicar a formulação enzimática ao material lignocelulósico para produzir um açúcar fermentável; (c) fermentar o açúcar fermentável para produzir um produto de fermentação combustível; e (d) queimar o dito produto de fermentação combustível para produzir energia.

Um objetivo da invenção é fornecer uma cepa de *Chrysosporium lucknowense* mutante capaz de expressar pelo menos uma celobiohidrolase e pelo menos uma endo-1,4- β -glicanase em níveis mais altos do que a cepa não mutante correspondente sob as mesmas condições. A celobiohidrolase é selecionada a partir do grupo que consiste de CBH Ia, CBH Ib, CBH IIa e CBH IIb; e a endo-1,4- β -glicanase é selecionada a partir do grupo que consiste de EG II, EG V, e EG VI.

A invenção também fornece uma proteína que exhibe pelo menos 65 % da identidade do aminoácido conforme determinado pelo algoritmo BLAST com as sequências de aminoácido de CBH Ib, IIb, EG VI, BGL, Xyl II como definido neste relatório ou uma parte destas tendo pelo menos 20 aminoácidos contíguos.

Um outro aspecto desta invenção fornece uma sequência de ácido nucléico tendo pelo menos 80 % de homologia com a sequência de ácido nucléico que codifica CBH Ib, CBH IIb, EG II, EG VI, BGL, ou Xyl II, como definido neste relatório.

A invenção também fornece um método para degradar um material lignocelulósico em açúcares fermentáveis. O método inclui contatar o material lignocelulósico com uma quantidade eficaz de um produto multienzimático derivado de um microorganismo, para produzir pelo menos um açúcar fermentável. Pelo menos uma enzima no produto multienzimático é selecionada a partir do grupo que consiste de CBH Ia, CBH Ib, CBH IIa, CBH IIb, EG II, EG V, EG VI, BGL, e Xyl II.

Em um outro aspecto, a invenção fornece um microorganismo ou planta capaz de expressar uma ou mais de uma enzima selecionada a partir do grupo que consiste de CBH Ia, CBH Ib, CBH IIa, CBH IIb, EG II, EG V, EG VI, BGL, e Xyl II.

BREVE DESCRIÇÃO DOS DESENHOS

FIG. 1: SDS/PAGE (A) e focalização isoelétrica (B) de celobiohidrolases purificadas provenientes de *C. lucknowense*. Colunas: 1, marcadores com massas moleculares diferentes; 2 e 5, CBH Ib; 3 e 6, CBH IIb; 4, marcadores com pI diferente.

FIG. 2: Cinéticas de progresso da hidrólise de Avicel (5 mg ml⁻¹) por celobiohidrolases purificadas (0,1 mg ml⁻¹) na presença de BGL de *A. japonicus* purificada (0,5 U ml⁻¹), 40 °C, pH 5,0.

FIG. 3: Sinergismo entre CBH IIb e outras enzimas de *C. lucknowense* purificadas durante a hidrólise da celulose de algodão (5 mg ml⁻¹) na presença da BGL de *A. japonicus* purificada (0,5 U ml⁻¹), 40 °C, pH 5,0. A concentração de CBH e EG foi de 0,15 e 0,05 mg ml⁻¹.

¹, respectivamente. Dados experimentais para os pares de enzimas são mostrados com símbolos vazios (curvas contínuas); as somas teóricas de concentrações de glicose obtidas sob a ação de enzimas individuais são mostradas com símbolos cheios (linhas pontilhadas).

FIG. 4: Cinéticas de progresso da hidrólise do algodão (25 mg ml⁻¹) por combinação #1 de enzimas de *C. lucknowense* purificadas e NCE L-600, uma preparação de celulase multienzimática de *C. lucknowense* comercial na proteína com carga de 0,5 mg ml⁻¹, 50 °C, pH 5,0 (vide texto e Tabela 4 para detalhes).

FIG. 5: Cinéticas de progresso da hidrólise de Avicel (50 mg ml⁻¹) por combinação #1 de enzimas de *C. lucknowense* purificadas e NCE-L, uma preparação de celulase multienzimática de *C. lucknowense* comercial na proteína com carga de 0,5 mg ml⁻¹, 50 °C, pH 5,0 (vide texto e Tabela 4 para detalhes).

FIG. 6: Cinéticas de progresso da hidrólise de madeira da espécie Douglas fir pré tratada (50 mg ml⁻¹) por combinação #1 de enzimas de *C. lucknowense* purificadas e NCE-L 600, um *C. lucknowense* comercial na proteína com carga de 0,5 mg ml⁻¹, 50 °C, pH 5,0 (vide texto e Tabela 4 para detalhes).

FIG. 7: Cinéticas de progresso da hidrólise de madeira da espécie Douglas fir pré tratada (50 mg ml⁻¹) por combinações diferentes de enzimas de *C. lucknowense* purificadas na proteína com carga de 0,5 mg ml⁻¹, 50 °C, pH 5,0 (vide texto e Tabela 5 para detalhes).

FIG. 8: gene de cbh2 que codifica CBH IB.

FIG. 9: gene de cbh4 que codifica CBH.

FIG. 10: gene de cbh1 que codifica CBH Ia

FIG. 11: gene de EG6 que codifica CBH lia

FIG. 12: gene de eg2 que codifica EG II

FIG. 13: gene de bgl1 que codifica BGL

FIG. 14: gene de eg5 que codifica EGV

FIG. 15: gene de eg7 que codifica EG VI

FIG. 16: gene de xyl2 que codifica Xyl II

DESCRIÇÃO DETALHADA DA INVENÇÃO

A presente invenção fornece métodos e composições para a conversão de biomassa da planta em açúcares fermentáveis que podem ser convertidos em produtos úteis. Os métodos incluem métodos para degradar material lignocelulósico usando misturas enzimáticas para liberar açúcares. As composições da invenção incluem combinações enzimáticas que decompõem lignocelulose. Como usado neste relatório, os termos “biomassa” ou “material lignocelulósico” incluem materiais contendo celulose e/ou hemicelulose. Geralmente, estes materiais também contêm xilana, lignina, proteína, e carboidratos, tais como amido e açúcar. Lignocelulose é geralmente encontrada, por exemplo, nos caules, folhas, cálices, cascas, e sabugos de plantas ou folhas, galhos, e madeira de árvores. O processo de con-

verter um carboidrato complexo (tal como amido, celulose, ou hemicelulose) em açúcares fermentáveis também é referido neste relatório como “sacarificação.” Açúcares fermentáveis, como usado neste relatório, referem-se a açúcares simples, tais como glicose, xilose, arabinose, galactose, manose, ramnose, sacarose e frutose.

5 A biomassa pode incluir biomassa virgem e/ou biomassa não virgem tal como biomassa agrícola, resíduos orgânicos comerciais, fragmentos de construção e demolição, resíduos sólidos urbanos, papel usado e sobras orgânicas. Formas comuns de biomassa incluem árvores, arbustos e gramas, trigo, palha de trigo, bagaço de cana de açúcar, milho, palhas de milho, grão de milho incluindo fibra dos grãos, produtos e subprodutos provenientes da moagem dos grãos tais como milho, trigo e cevada (incluindo moagem úmida e moagem a seco) assim como resíduos sólidos urbanos, papel usado e sobras orgânicas. A biomassa também pode ser, mas não é limitada a, material herbáceo, resíduos agrícolas, resíduos florestais, resíduos sólidos urbanos, papel usado, e resíduos da fabricação de polpa e papel. “Biomassa agrícola” inclui galhos, matos, canas, milho e palhas de milho, safras energéticas, florestas, frutos, flores, grãos, gramas, safras herbáceas, folhas, casca de árvore, acículas, lenhas, raízes, mudas, safras de madeira de curta rotação, arbustos, gramíneas do tipo switchgrass, árvores, vegetais, cascas de frutos, videiras, polpa de beterraba, triguihos, cascas de aveia, e madeiras duras e moles (não incluindo madeiras com materiais deletérios). Além disso, a biomassa agrícola inclui materiais de resíduos orgânicos gerados a partir de processos agrícolas incluindo atividades de agricultura e silvicultura, especificamente incluindo resíduos de madeira florestal. A biomassa agrícola pode ser qualquer um dos citados acima estabelecidos individualmente ou em qualquer combinação ou mistura dos mesmos.

Os açúcares fermentáveis podem ser convertidos em produtos de fermentação com valor agregado úteis, exemplos não limitantes dos quais incluem aminoácidos, vitaminas, produtos farmacêuticos, suplementos de ração animal, especialidades químicas, matérias-primas químicas, plásticos, solventes, combustíveis, ou outros polímeros orgânicos, ácido láctico, e etanol, incluindo etanol combustível. Produtos com valor agregado específicos que podem ser produzidos pelos métodos da invenção incluem, mas não são limitados a, biocombustíveis (incluindo etanol e butanol); ácido láctico; plásticos; em especial produtos químicos; ácidos orgânicos, incluindo ácido cítrico, ácido succínico e ácido maléico; solventes; suplementos de ração animal; produtos farmacêuticos; vitaminas; aminoácidos, tais como lisina, metionina, triptofano, treonina, e ácido aspártico; enzimas industriais, tais como proteases, celulasas, amilases, glicanases, lactases, lipases, liases, oxidoredutases, transferases e xilanasas; e matérias-primas químicas.

Como usado neste relatório, um produto multienzimático pode ser obtido a partir de ou derivado de uma fonte microbiana, planta, ou outra fonte ou combinação das mesmas, e

conterá enzimas capazes de degradar material lignocelulósico. Exemplos de enzimas compreendendo os produtos multienzimáticos da invenção incluem celulasas (tais como celobiohidrolases, endoglicanase, β -glicosidases, hemicelulasas (tais como xilanases, incluindo endoxilanases, exoxilanase, e β -xilosidase), ligninases, amilases, α -arabinofuranosidases, α -glicuronidases, α -glicuronidases, arabinases, glicuronidases, proteases, esterases (incluindo ácido ferúlico esterase e acetilxilana esterase), lipases, glicomananases, e xilogliconases.

Em algumas modalidades, o produto multienzimático compreende uma hemicelulase. Hemicelulose é um polímero complexo, e sua composição frequentemente varia amplamente de organismo para organismo, e de um tipo de tecido para um outro. No geral, um componente principal de hemicelulose é xilose ligada a beta-1,4, um açúcar de cinco carbonos. Entretanto, esta xilose é frequentemente ramificada como ligações beta-1,3, e pode ser substituída com ligações para arabinose, galactose, manose, ácido glicurônico, ou por esterificação para ácido acético. Hemicelulose também pode conter glicana, que é um termo geral para açúcares de seis carbonos ligados à beta. Essas hemiceluloses incluem xiloglicana, glicomanana, e galactomanana.

A composição, natureza de substituição, e grau de ramificação da hemicelulose é muito diferente em plantas dicotiledonêas (dicotiledôneos, isto é, planta cujas sementes têm dois cotilédones ou folhas seminais tais como feijões-de-lima, amendoins, amêndoas, ervilhas, feijões) em comparação às plantas monocotiledôneas (monocotiledôneos; isto é, plantas tendo um cotilédone ou folha seminal único tal como milho, trigo, arroz, gramas, cevada). Em dicotiledôneos, a hemicelulose é compreendida principalmente de xiloglicanas que são cadeias de glicose ligadas a 1,4-beta com cadeias laterais de xilosil ligadas a 1,6-beta. Em monocotiledôneos, incluindo a maioria das safras de grão, os principais componentes de hemicelulose são heteroxilanas. Estas são principalmente compreendidas de polímeros de cadeia principal de xilose ligada a 1,4-beta com ligações 1,3-beta para arabinose, galactose e manose assim como xilose modificada por ácidos acéticos ligados ao éster. Também presentes estão beta-glicanas ramificadas compreendidas de cadeias de glicosil ligadas a 1,3- e 1,4-beta. Em monocotiledôneos, celulose, heteroxilanas e beta-glicanas estão presentes em quantidades aproximadamente iguais, cada uma compreendendo cerca de 15 a 25 % da matéria seca de paredes celulares.

Enzimas hemicelulósicas, isto é hemicelulasas, incluem tanto enzimas exohidrolíticas quanto endohidrolíticas, tais como xilanase, β -xilosidase e esterases, que clivam ativamente o material hemicelulósico através da hidrólise. Estas enzimas xilanase e esterase clivam as cadeias laterais de xilana e acetyl da xilana e os xilo-oligomêros remanescentes são não substituídos e podem, dessa forma, ser hidrolisados apenas com Pxilosidase. Além disso, várias atividades secundárias menos conhecidas foram descobertas nas preparações

enzimáticas que hidrolisam a hemicelulose. Embora o produto multienzimático possa conter muitos tipos de enzimas, misturas compreendendo enzimas que aumentam ou realçam a liberação de açúcar da biomassa são preferidas, incluindo hemicelulases. Em uma modalidade, a hemicelulase é uma xilanase, uma arabinofuranosidase, uma acetil xilana esterase, uma glicuronidase, uma endo-galactanase, uma mananase, uma endo arabinase, uma exo arabinase, uma exo-galactanase, uma ácido ferúlico esterase, uma galactomananase, uma xilogliconase, ou misturas de qualquer uma destas. Em particular, as enzimas podem incluir glicoamilase, β -xilosidase e/ou β -glicosidase. As enzimas do produto multienzimático podem ser fornecidas por uma variedade de fontes. Em uma modalidade, as enzimas podem ser produzidas pelo crescimento de microorganismos ou plantas que produzem as enzimas naturalmente ou em virtude de serem geneticamente modificados para expressarem a enzima ou enzimas. Em uma outra modalidade, pelo menos uma enzima do produto multienzimático está comercialmente disponível.

Uma modalidade da presente invenção refere-se a uma enzima isolada para catalisar a conversão de material lignocelulósico em açúcares fermentáveis como descrito neste relatório, um homólogo da mesma, e/ou um fragmento da mesma. Também incluídas na invenção estão moléculas de ácido nucléico isoladas que codificam qualquer uma de tais proteínas, homólogos ou fragmentos das mesmas. De acordo com a presente invenção, uma proteína ou polipeptídeo isolado é uma proteína que foi removida de seu ambiente natural (isto é, que foi submetida à manipulação humana) e pode incluir, por exemplo, proteínas purificadas, proteínas parcialmente purificadas, proteínas recombinantemente produzidas, e proteínas sinteticamente produzidas. Como tal, "isolada" não reflete a extensão a que a proteína foi purificada. De preferência, uma proteína isolada da presente invenção é produzida recombinantemente. Um peptídeo isolado pode ser produzido sinteticamente (por exemplo, quimicamente, tal como por síntese de peptídeo) ou recombinantemente. Uma proteína isolada também pode ser fornecida como um produto de fermentação bruto, ou uma preparação de proteína que foi parcialmente purificada ou purificada (por exemplo, a partir de um microorganismo) usando procedimentos de purificação de proteína conhecidos na técnica. Além disso, e somente por via de exemplo, uma proteína referenciada como sendo derivada de, ou proveniente de um organismo particular, tal como uma celulase e/ou hemicelulase de "*Chrysosporium lucknowense*" refere-se a uma celulase e/ou hemicelulase (geralmente incluindo um homólogo de uma celulase e/ou hemicelulase que ocorre naturalmente) proveniente de um microorganismo *Chrysosporium lucknowense*, ou a uma celulase e/ou hemicelulase que foi de outro modo produzida a partir do conhecimento da estrutura (por exemplo, sequência), e possivelmente da função, de uma celulase e/ou hemicelulase que ocorre naturalmente proveniente de *Chrysosporium lucknowense*. Em outras palavras, referência geral a uma celulase e/ou hemicelulase de *Chrysosporium lucknowense* ou uma

celulase e/ou hemicelulase derivada de *Chrysosporium lucknowense* inclui qualquer celulase e/ou hemicelulase que tem estrutura e função substancialmente similares a uma celulase e/ou hemicelulase que ocorre naturalmente proveniente de *Chrysosporium lucknowense* ou que é um homólogo biologicamente ativo (isto é, tem atividade biológica) de uma celulase e/ou hemicelulase que ocorre naturalmente proveniente de *Chrysosporium lucknowense* como descrito em detalhe neste relatório. Como tal, uma celulase e/ou hemicelulase de *Chrysosporium lucknowense* pode incluir proteínas purificadas, parcialmente purificadas, recombinantes, mutantes/modificadas e sintéticas. A mesma descrição aplica-se para fazer referência a outras proteínas ou peptídeos descritos neste relatório e a outras fontes microbianas para tais proteínas ou peptídeos.

Uma modalidade da presente invenção refere-se a moléculas de ácido nucléico isoladas compreendendo, que consistem essencialmente de, ou que consistem de sequências de ácido nucléico que codificam qualquer uma das enzimas descritas neste relatório, incluindo um homólogo ou fragmento de qualquer uma de tais enzimas, assim como as sequências de ácido nucléico que são completamente complementares a estas. De acordo com a presente invenção, uma molécula de ácido nucléico isolada é uma molécula de ácido nucléico que foi removida de seu ambiente natural (isto é, que foi submetida à manipulação humana), seu ambiente natural sendo o genoma ou cromossomo em que a molécula de ácido nucléico é encontrada na natureza. Como tal, "isolada" não necessariamente reflete a extensão a que a molécula de ácido nucléico foi purificada, mas indica que a molécula não inclui um genoma total ou um cromossomo total em que a molécula de ácido nucléico é encontrada na natureza. Uma molécula de ácido nucléico isolada pode incluir um gene. Uma molécula de ácido nucléico isolada que inclui um gene não é um fragmento de um cromossomo que inclui tal gene, mas particularmente inclui a região de codificação e regiões reguladoras associadas com o gene, mas nenhum dos genes adicionais que são naturalmente encontrados no mesmo cromossomo. Uma molécula de ácido nucléico isolada também pode incluir uma sequência de ácido nucléico especificada flanqueada por ácidos nucléicos adicionais (isto é, na extremidade 5' e/ou a 3' da sequência) que normalmente não flanqueiam a sequência de ácido nucléico especificada na natureza (isto é, sequências heterólogas). Molécula de ácido nucléico isolada pode incluir DNA, RNA (por exemplo, mRNA), ou derivados tanto de DNA quanto de RNA (por exemplo, cDNA). De preferência, uma molécula de ácido nucléico isolada da presente invenção é produzida usando tecnologia do DNA recombinante (por exemplo, reação em cadeia da polimerase (PCR) amplificação, clonagem) ou síntese química. Um homólogo de molécula de ácido nucléico pode ser produzido usando vários métodos conhecidos àqueles habilitados na técnica (vide, por exemplo, Sambrook *et al.*, Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Labs Press (1989)). Por exemplo, moléculas de ácido nucléico podem ser modificadas usando uma variedade de

técnicas incluindo, mas não limitadas a, técnicas de mutagênese e técnicas de DNA recombinante clássicas, tais como mutagênese sítio-dirigida, tratamento químico de uma molécula de ácido nucléico para induzir mutações, clivagem por enzima de restrição de um fragmento de ácido nucléico, ligação de fragmentos de ácido nucléico, amplificação por PCR e/ou mutagênese de regiões selecionadas de uma sequência de ácido nucléico, síntese de misturas de oligonucleotídeo e ligação de grupos de mistura para “construir” uma mistura de moléculas de ácido nucléico e combinações das mesmas. Homólogos de molécula de ácido nucléico podem ser selecionados a partir de uma mistura de ácidos nucléicos modificados por triagem para a função da proteína codificada pelo ácido nucléico e/ou por hibridização com um gene do tipo selvagem.

Uma outra modalidade da presente invenção inclui uma molécula de ácido nucléico recombinante compreendendo um vetor recombinante e uma sequência de ácido nucléico que codifica proteína ou peptídeo tendo pelo menos uma atividade enzimática útil para catalisar a conversão de material lignocelulósico em açúcares fermentáveis. De acordo com a presente invenção, um vetor recombinante é uma molécula de ácido nucléico projetada (isto é, artificialmente produzida) que é usada como uma ferramenta para manipular uma sequência de ácido nucléico de escolha e para introduzir uma tal sequência de ácido nucléico em uma célula hospedeira. O vetor recombinante é, portanto, adequado para o uso em clonagem, sequenciação, e/ou de outro modo manipular a sequência de ácido nucléico de escolha, tal como pela expressão e/ou liberação da sequência de ácido nucléico de escolha em uma célula hospedeira para formar uma célula recombinante. Um tal vetor tipicamente contém sequências heterólogas de ácido nucléico, isto é, sequências de ácido nucléico que não são naturalmente encontradas adjacentes à sequência de ácido nucléico a ser clonada ou liberada, embora o vetor também possa conter sequências de ácido nucléico reguladoras (por exemplo, promotoras, não traduzidas) que são naturalmente encontradas adjacentes às moléculas de ácido nucléico da presente invenção ou que são úteis para a expressão das moléculas de ácido nucléico da presente invenção (debatido em detalhe abaixo). O vetor pode ser tanto RNA quanto DNA, tanto procariótico quanto eucariótico, e tipicamente é um plasmídeo. O vetor pode ser mantido como um elemento extracromossômico (por exemplo, um plasmídeo) ou ele pode ser integrado no cromossomo de um organismo recombinante (por exemplo, um micróbio ou uma planta). O vetor total pode permanecer dentro de uma célula hospedeira, ou sob certas condições, o DNA de plasmídeo pode ser suprimido, deixando para trás a molécula de ácido nucléico da presente invenção. A molécula de ácido nucléico integrada pode estar sob controle de promotor cromossômico, sob controle de promotor natural ou plasmídeo, ou sob uma combinação de vários controles de promotor. Cópias únicas ou múltiplas da molécula de ácido nucléico podem ser integradas no cromossomo. Um vetor recombinante da presente invenção pode conter pelo menos um marcador

selecionável.

Tipicamente, uma molécula de ácido nucléico recombinante inclui pelo menos uma molécula de ácido nucléico da presente invenção operativamente ligada a uma ou mais sequências de controle de expressão. De acordo com a presente invenção, a frase “operativamente ligada” refere-se à ligação de uma molécula de ácido nucléico a uma sequência de controle de expressão (por exemplo, uma sequência de controle da transcrição e/ou uma sequência de controle da tradução) de uma maneira tal que a molécula possa ser expressada quando transfectada (isto é, transformada, transferida, transfectada, conjugada ou conduzida) em uma célula hospedeira. Sequências de controle da transcrição são sequências que controlam a iniciação, prolongamento, ou terminação da transcrição. Sequências de controle da transcrição particularmente importantes são aquelas que controlam a iniciação da transcrição, tal como sequências promotoras, realçadoras, operadoras e repressoras.

Sequências de controle da transcrição adequadas incluem qualquer sequência de controle da transcrição que pode funcionar em uma célula hospedeira ou organismo em que a molécula de ácido nucléico recombinante seja introduzida.

Enzimas e Ácidos Nucléicos Que codificam as Enzimas

Como descrito nos exemplos, esta invenção fornece várias enzimas purificadas, incluindo duas celobiohidrolases, (CBH Ib, SEQ ID NO. 2; CBH IIb, SEQ ID NO. 4), uma endoglicanase (EG VI, SEQ ID NO. 16), uma β -glicosidase (BGL, SEQ ID NO. 12), e uma xilanase (Xyl SEQ ID NO. 18). Esta invenção também considera variantes de tais enzimas, incluindo variantes tendo sequência de aminoácido com pelo menos 65 %, 70 %, ou 75 % da identidade do aminoácido com estas enzimas, conforme determinado pelo algoritmo BLAST convencionalmente usado.

Adicionalmente, a invenção fornece os ácidos nucléicos que codificam estas sequências, incluindo o gene de *cbh2* (SEQ ID NO. 1, que codifica CBH Ib), o gene de *cbh4* (SEQ ID NO. 3, que codifica CBH IIb); o gene de *eg7* (SEQ ID NO. 15, que codifica EG VI), o gene *bgl1* (SEQ ID NO. 11, que codifica BGL), e o gene de *xyl2* (SEQ ID NO. 17, que codifica Xyl II). Esta invenção também considera variantes destes ácidos nucléicos, incluindo variantes que têm pelo menos 80 %, 85 % ou 90 % de homologia com estes ácidos nucléicos.

Como descrito neste relatório, as enzimas recém identificadas e isoladas de acordo com a invenção podem ser usadas em combinação com pelo menos uma outra enzima que promove sacarificação de materiais celulósicos. Em modalidades preferidas, esta enzima adicional é derivada de *C. lucknowense*. Por exemplo, a enzima pode ser CBH Ia (SEQ ID NO. 6), CBH IIa (SEQ ID NO. 8), EG II (SEQ ID NO. 10) ou EG V (SEQ ID NO. 14). Observa-se, entretanto, que em certas modalidades preferidas, CBH Ia, CBH IIa EG II, e EG V podem ser obtidas pela modificação genética de um microorganismo ou planta para expres-

sar cbh1 (SEQ ID NO. 5, que codifica CBH Ia), EG6 (SEQ ID NQ. 7, que codifica CBH IIa), eg2 (SEQ ID NO. 9, que codifica EG II), e/ou EG5 (SEQ ID NO. 13, que codifica EG V). Uma combinação particularmente útil para a sacarificação é CBH Ia, CBH Ib, CBH IIb, EG II, EG V, BGL, e Xyl II.

Em certas modalidades, os polinucleotídeos e polipeptídeos da invenção são desenvolvidos usando técnicas de evolução molecular para criar e identificar novas variantes com características estruturais, funcionais, e/ou físicas desejadas. Técnicas de evolução molecular podem ser "Rearranjo do DNA", ou "PCR sexual" (WPC, Stemmer, PNAS, 91:10747, (1994)), também referidas como "evolução molecular dirigida", "rearranjo de éxon", "evolução enzimática dirigida", "evolução *in vitro*" e "evolução artificial". Tais termos de referência são conhecidos na técnica e são abrangidos pela invenção. Características tais como a atividade, a cinética enzimática da proteína, a K_i , K_{cat} , K_m , V_{max} , K_d da proteína, termoestabilidade, pH ideal, e semelhantes podem ser modificadas. Em certas modalidades, os polinucleotídeos e/ou polipeptídeos da invenção podem ser desenvolvidos para conferir propriedades que são vantajosas para a sacarificação e fermentação enzimática *in situ*. Por exemplo, enzimas podem ser desenvolvidas para atuarem da melhor forma em um ambiente que é adequado para a fermentação de açúcares. Em um exemplo, as enzimas são desenvolvidas para ter atividade máxima em um ambiente com temperatura elevada e alto teor de álcool no ambiente, tal como um ambiente onde um organismo tal como levedura está fermentando os açúcares. Desta forma, a sacarificação de lignocelulose e fermentação ocorrem em uma etapa de processo único. Em um outro exemplo, as enzimas são desenvolvidas para resistir a produtos químicos severos ou ambientes térmicos, tais como aqueles que podem ser experimentados durante pré tratamentos lignocelulósicos, como descrito neste relatório. Nestas modalidades, não é necessário pré tratar química ou termicamente a lignocelulose antes de adicionar as enzimas. Particularmente, o tratamento e a sacarificação enzimática podem ser realizados simultaneamente. Certamente, esta invenção também considera processos que envolvem etapas múltiplas para produzir açúcares a partir de lignocelulose, tais como aqueles onde enzimas desenvolvidas primeiro sacarificam lignocelulose, que é subsequentemente fermentada por um organismo, tal como levedura, por exemplo.

Em outras modalidades, a capacidade de realçar características específicas de uma proteína também pode ser aplicável para mudar a atividade caracterizada de uma enzima para uma atividade completamente não relacionada a sua atividade inicialmente caracterizada. Outros aperfeiçoamentos desejáveis da invenção seriam especificar cada proteína individual, e seriam assim bem conhecidos na técnica e considerados pela invenção.

Expressão de Enzimas

Os microorganismos úteis na presente invenção e/ou como uma fonte de enzimas úteis na presente invenção incluem qualquer microorganismo que produz uma enzima capaz

de degradar material lignocelulósico, incluindo bactérias, levedura, e fungos filamentosos. Em virtude da simplicidade e conveniência, microorganismos fúngicos filamentosos serão debatidos neste relatório; entretanto, uma pessoa habilitada na técnica reconhecerá que outros microorganismos serão úteis na presente invenção. Fungos filamentosos foram am-

plamente usados na indústria para a produção de proteínas. Estes fungos são, de maneira única, adaptados para a produção e secreção de proteínas devido ao seu nicho biológico como descontaminantes microbianos. Em ambientes ricos em polímeros biológicos, tais como solos de floresta, os fungos competem por secretar enzimas que degradam esses polí-

meros, produzindo monômeros que podem ser prontamente utilizados como nutrientes para cultivo. A capacidade natural dos fungos em produzir proteínas foi amplamente explorada, principalmente para a produção de enzimas industriais. Níveis de produção de proteína em isolatos naturais podem ser aumentados em cepas aperfeiçoadas por ordens de grandeza; rendimentos de produção de dezenas de gramas de proteína por litro de cultura de fermentação são comuns.

Cepas fúngicas, incluindo, mas não limitadas a, várias espécies de *Talaromyces*, *Aspergillus*, *Trichoderma*, *Neurospora*, *Penicillium*, *Fusarium*, *Humicola*, *Myceliophthora*, *Corynascus*, *Chaetomium*, *Tolypocladium*, *Thielavia*, *Acremonium*, *Sporotrichum*, *Thermoascus*, e *Chrysosporium*, são consideradas na presente invenção. Estes são um pequeno número de muitos gêneros possíveis de fungos que serão fontes úteis de enzimas e/ou seriam adequados como organismos hospedeiros para produzir tais misturas enzimáticas. Tais fungos podem ser obtidos, por exemplo a partir de vários depositários tais como a American Type Culture Collection (ATCC), a All Russian Collection of Microorganisms of the Russian Academy of Sciences (VKM), e Centraalbureau voor Schimmelcultures.

Cepas mutantes de *C. lucknowense*

Cepas particulares de *Chrysosporium* expressam proteínas em quantidades extremamente grandes e sequências que regulam expressão natural a partir destas cepas são de particular interesse. Estas cepas foram designadas como cepa de *Chrysosporium* CI, cepa UV 13-6, cepa NG7C-19 e cepa UV18-25. Elas foram depositadas de acordo com o Budapest Treaty com o instituto depositário All Russian Collection (VKM) em Moscou. A cepa tipo CI selvagem foi depositada de acordo com o Budapest Treaty com o número VKM F-3500 D, data do depósito em 29 de Agosto de 1996, CI UV13-6 mutante foi depositada com número VKM F-3632 D, e a data do depósito em 9 de Fevereiro de 1998, CI NG7c-19 mutante foi depositada com número VKM F-3633 D e a data do depósito em 9 de Fevereiro de 1998 e CI UV18-25 mutante foi depositada com número VKM F-3631 D e data do depósito em 9 de

Fevereiro de 1998.

De preferência, uma região reguladora de expressão permitindo alta expressão no hospedeiro selecionado é aplicada. Esta também pode ser uma região reguladora de alta

expressão derivada de um hospedeiro heterólogo, tal como são bem conhecidos na técnica. Exemplos específicos de proteínas conhecidas a serem expressadas em grandes quantidades e assim fornecendo sequências reguladoras de expressão adequadas para a invenção apresentam-se sem ser limitados a estas: hidrofobina, protease, amilase, xilanase, pectinase, esterase, beta-galactosidase, celulase (por exemplo, endo-glicanase, celobiohidrolase) e poligalacturonase. A alta produção foi confirmada tanto no estado sólido quanto nas condições de fermentação submersas. Ensaios para avaliar a presença ou produção de tais proteínas são bem conhecidos na técnica.

Sequências reguladoras de expressão heterólogas também trabalham eficazmente em *Chrysosporium* como sequências de *Chrysosporium* naturais. Isto possibilita que construções e vetores bem conhecidos sejam usados na transformação de *Chrysosporium* assim como oferecendo outras numerosas possibilidades para a construção de vetores que possibilitam boas taxas de expressão neste novo hospedeiro na expressão e secreção. Como taxas de expressão extremamente altas para celulase foram confirmadas para cepas de *Chrysosporium*, as regiões reguladoras de expressão de tais proteínas são particularmente preferidas.

Uma construção de ácido nucléico compreendendo uma região reguladora de expressão de ácido nucléico proveniente de *Chrysosporium lucknowense* ou um derivado desta, forma uma modalidade separada da invenção como faz a cepa de *Chrysosporium* mutante compreendendo tais regiões operativamente ligadas a um gene que codifica um polipeptídeo a ser expressado. Em modalidades preferidas, uma tal construção de ácido nucléico será uma região reguladora de expressão proveniente de *Chrysosporium* associada com a expressão de celobiohidrolase, endoglicanase, β -glicosidase, e/ou xilanase.

A invenção também abrange cepas de *Chrysosporium* geneticamente projetadas em que a sequência que é introduzida pode ser de origem de *Chrysosporium*. Uma tal cepa pode, entretanto, ser distinguida de cepas que ocorrem naturalmente em virtude de, por exemplo, sequências heterólogas estando presentes na sequência de ácido nucléico usada para transformar ou transfectar o *Chrysosporium*, em virtude do fato de que cópias múltiplas da sequência que codifica o polipeptídeo de interesse estão presentes, ou em virtude do fato de que estas são expressadas em uma quantidade que excede aquela cepa não projetada sob condições idênticas ou em virtude do fato de que a expressão ocorre sob condições normalmente de não expressão. A última pode ser o caso, se um promotor induzível regula a sequência de interesse contrário à situação não recombinante ou se um outro fator induz a expressão no caso da cepa não projetada. A invenção, como definido nas modalidades precedentes não é intencionada a abranger cepas de *Chrysosporium* que ocorrem naturalmente. A invenção é dirigida a cepas derivadas através da engenharia usando tecnologias genéticas clássicas ou metodologias de engenharia genética.

Um método de produção de um microorganismo ou planta recombinante também é parte da invenção objeto. O método compreende introduzir de maneira estável uma sequência de ácido nucléico que codifica um polipeptídeo heterólogo ou homólogo em uma cepa microbiana ou planta, a sequência de ácido nucléico sendo operativamente ligada a uma região reguladora de expressão. Tais procedimentos para transformar fungos filamentosos foram previamente relatados. Em uma modalidade preferida, o *Chrysosporium lucknowense* mutante é derivado de UV18-25 (Acc. Nº VKM F-3631 D) que foi projetada para superexpressar o gene de Xyl II.

Organismos Geneticamente Modificados

Como usado neste relatório, um microorganismo geneticamente modificado pode incluir uma bactéria, levedura, fungo, ou outro micróbio geneticamente modificado. Um tal microorganismo geneticamente modificado tem um genoma que é modificado (isto é, mutante ou modificado) de sua forma normal (isto é, tipo selvagem ou que ocorre naturalmente) tal que um resultado desejado seja obtido (por exemplo, atividade e/ou produção aumentada ou modificada de pelo menos uma enzima ou um produto multienzimático para a conversão de material lignocelulósico em açúcares fermentáveis). A modificação genética de um microorganismo pode ser realizada usando-se desenvolvimento de cepa e/ou técnicas de genética molecular clássicas. Tais técnicas conhecidas no ramo são geralmente divulgadas para microorganismos, por exemplo, em Sambrook *et al.*, 1989, Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Labs Press. A referência Sambrook *et al.*, *ibid.*, é incorporada integralmente como referência neste relatório. Um microorganismo geneticamente modificado pode incluir um microorganismo em que moléculas de ácido nucléico foram inseridas, suprimidas ou modificadas (isto é, mutante; por exemplo, por inserção, supressão, substituição, e/ou inversão de nucleotídeos), de uma tal maneira que tais modificações fornecem o efeito desejado dentro do microorganismo.

Em um aspecto da invenção, um microorganismo geneticamente modificado pode endogenamente conter e expressar uma enzima ou um produto multienzimático para a conversão de material lignocelulósico em açúcares fermentáveis, e a modificação genética pode ser uma modificação genética de uma ou mais de tais enzimas endógenas, por meio das quais a modificação tem algum efeito na capacidade do microorganismo converter material lignocelulósico em açúcares fermentáveis.

Em um outro aspecto da invenção, um microorganismo geneticamente modificado pode endogenamente conter e expressar uma enzima ou um produto multienzimático para a conversão de material lignocelulósico em açúcares fermentáveis, e a modificação genética pode ser uma introdução de pelo menos uma sequência de ácido nucléico exógena (por exemplo, uma molécula de ácido nucléico recombinante), em que a sequência de ácido nucléico exógena codifica pelo menos uma enzima adicional útil para a conversão de material

lignocelulósico em açúcares fermentáveis e/ou uma proteína que melhora a eficiência da enzima ou produto multienzimático para a conversão de material lignocelulósico em açúcares fermentáveis. Neste aspecto da invenção, o microorganismo também pode ter pelo menos uma modificação para um gene ou genes compreendendo sua(s) enzima(s) endógena(s) para a conversão de material lignocelulósico em açúcares fermentáveis.

Ainda em um outro aspecto da invenção, o microorganismo geneticamente modificado não necessariamente endógeno (naturalmente) contém uma enzima ou um produto multienzimático para a conversão de material lignocelulósico em açúcares fermentáveis, mas é geneticamente modificado para introduzir pelo menos uma molécula de ácido nucléico recombinante que codifica pelo menos uma enzima, uma multiplicidade de enzimas, ou um produto multienzimático para a conversão de material lignocelulósico em açúcares fermentáveis. Um tal microorganismo pode ser usado em um método da invenção, ou como um microorganismo de produção para produtos de fermentação brutos, enzimas recombinantes parcialmente purificadas, e/ou enzimas recombinantes purificadas, qualquer das quais depois podem ser usadas em um método da presente invenção.

Plantas Geneticamente Modificadas

A invenção também considera plantas geneticamente modificadas compreendendo tais genes. As plantas podem ser usadas para a produção das enzimas, ou como o material lignocelulósico usado como um substrato nos métodos da invenção. Métodos para gerar plantas recombinantes são conhecidos na técnica. Por exemplo, métodos numerosos para a transformação da planta foram desenvolvidos, incluindo protocolos de transformação biológica e física. Vide, por exemplo, Miki *et al.*, "Procedures for Introducing Foreign DNA in Plants" em *Methods in Plant Molecular Biology and Biotechnology*, Glick, B.R. e Thompson, J.E. Eds. (CRC Press, Inc., Boca Raton, 1993) páginas 67-88. Além disso, vetores e métodos de cultura *in vitro* para transformação e regeneração de célula ou tecido de planta de vegetais estão disponíveis. Vide, por exemplo, Gruber *et al.*, "Vectors for Plant Transformation" em *Methods in Plant Molecular Biology and Biotechnology*, Glick, B.R. e Thompson, J.E. Eds. (CRC Press, Inc., Boca Raton, 1993) páginas 89-119.

Em certas modalidades da invenção, plantas geneticamente modificadas que expressam as enzimas desta invenção são obtidas introduzindo-se um vetor de expressão em plantas com base no sistema de transformação natural de *Agrobacterium*. Vide, por exemplo, Horsch *et al.*, *Science*, 227:1229 (1985). *A. tumefaciens* e *A. rhizogenes* são bactérias do solo patogênicas da planta que geneticamente transformam as células da planta. Os plasmídeos Ti e Ri de *A. tumefaciens* e *A. rhizogenes*, respectivamente, carregam genes responsáveis pela transformação genética da planta. Vide, por exemplo, Kado, C.I., *Crit. Rev. Plant. Sci.* 10:1 (1991). Descrições de sistemas e métodos do vetor *Agrobacterium* para transferência genética mediada por *Agrobacterium* são fornecidas por numerosas refe-

rências, incluindo Gruber *et al.*, *supra*, Miki *et al.*, *supra*, Moloney *et al.*, *Plant Cell Reports* 8:238 (1989), e Patentes U.S. Nºs 4.940.838 e 5.464.763, incorporadas neste relatório integralmente como referência.

Em outras modalidades, plantas geneticamente modificadas são obtidas por transformação mediada por microprojétil em que DNA é carregado na superfície de microprojéteis. O vetor de expressão é introduzido nos tecidos da planta com um dispositivo biolístico que acelera os microprojéteis em velocidade suficiente para penetrar nas paredes celulares e membranas da planta. Sanford *et al.*, *Part. Sci. Technol.* 5:27 (1987), Sanford, J.C., *Trends Biotech.* 6:299 (1988), Sanford, J.C., *Physiol. Plant* 79:206 (1990), Klein *et al.*, *Biotechnology* 10:268 (1992).

Um outro método para liberação física de DNA para plantas considerado por esta invenção é sonicação de células alvo. Zhang *et al.*, *Bio Technology* 9:996 (1991). Alternativamente, a fusão de lipossomo ou esferoplasto foi usada para introduzir vetores de expressão em plantas. Deshayes *et al.*, *EMBO J.*, 4:2731 (1985), Christou *et al.*, *Proc Natl. Acad. Sci. USA* 84:3962 (1987). A absorção direta de DNA em protoplastos usando precipitação de CaCh, álcool polivinílico ou poli-L-ornitina também foi relatada. Hain *et al.*, *Mol. Gen. Genet.* 199:161 (1985) e Draper *et al.*, *Plant Cell Physiol.* 23:451 (1982). A eletroporação de protoplastos e células e tecidos totais também foram descritos. Donn *et al.*, In *Abstracts of VIIIth International Congress on Plant Cell and Tissue Culture IAPTC*, A2-38, p. 53 (1990); D'Halluin *et al.*, *Plant Cell* 4:1495-1505 (1992) e Spencer *et al.*, *Plant Mol. Biol.* 24:51-61 (1994).

Métodos de Usar as Enzimas e Cepas Mutantes de *C. lucknowense*

Esta invenção também fornece métodos de sacarificação enzimática de materiais celulósicos. Qualquer material contendo celulose pode ser tratado pelas enzimas desta invenção, exemplos não limitantes dos quais incluem podas de pomar, chaparral, resíduos de moagem, resíduos urbanos de madeira, sobras orgânicas, resíduos urbanos, resíduos de corte de madeira, desbastes florestais, safras de madeira de curta rotação, resíduos industriais, palha de trigo, palha de aveia, palha de arroz, palha de cevada, palha de centeio, palha do linho, cana-de-açúcar, forragem de milho, hastes do milho, sabugos de milho, palhas de milho, pradarias, gamagrass, capim rabo-de-raposa; polpa de beterraba, polpa de frutas cítricas, cascas de semente, resíduos animais celulósicos, cortes de grama, algodão, e algas.

Em certas modalidades preferidas, os materiais lignocelulósicos são pré tratados antes de serem expostos às enzimas ou misturas enzimáticas da invenção. De modo geral, o pré tratamento pode ser qualquer procedimento que torna a sacarificação enzimática subsequente dos materiais lignocelulósicos mais eficaz (isto é, menos demorada ou menos cara). Por exemplo, o material lignocelulósico pode ser pré tratado por métodos incluindo, mas

não limitados a, exposição a ácidos, bases, solventes, calor, peróxidos, ozônio, ou alguma combinação dos mesmos, antes da sacarificação enzimática. Estes pré tratamentos também podem ser combinados com outras formas de processamento, tais como trituração mecânica, trituração, moagem, ou depressurização rápida (por exemplo, explosão a vapor).

5 Geralmente, a sacarificação enzimática de acordo com a invenção envolve o uso de CBH Ia, CBH Ib, EG VI, BGL, Xyl II, ou misturas das mesmas. Uma ou mais destas enzimas podem ser ainda combinadas com outras enzimas capazes de promover a sacarificação enzimática, que pode ser derivada de *C. lucknowense*, uma cepa mutante, ou um outro organismo. Por exemplo, em uma modalidade, a sacarificação enzimática envolve uma mistu-
10 ra enzimática compreendendo CBH Ia, CBH Ib, CBH IIa, EG II, EG V, BGL, e Xyl II. Em outras modalidades preferidas, a mistura enzimática contém uma celobiohidrolase, que pode ser CBH Ia, CBH Ib, CBH IIa, CBH IIb, e misturas da mesma, com uma β -glicosidase tal como BGL.

Em certas modalidades, as composições enzimáticas são composições enzimáticas artificiais que contêm formas purificadas de CBH Ia, CBH Ib, CBH IIa, EG II, EG VI, BGL, ou Xyl II. As formas purificadas destas enzimas podem ser usadas sozinhas ou misturadas entre si. Em certas modalidades preferidas, as enzimas purificadas selecionadas estão presentes em quantidades relativas mais altas do que seria o caso para as secreções enzimáticas do *C. lucknowense* do tipo selvagem.

20 Em certas modalidades, a invenção fornece uma cepa mutante de *C. lucknowense* que é capaz de expressar CBH Ia, CBH Ib, CBH IIa, CBH IIb, EG II, EG V, EG VI, BGL, ou Xyl II, ou misturas das mesmas em proporções mais altas do que encontradas nas secreções enzimáticas do organismo do tipo selvagem. As enzimas secretadas de uma tal cepa mutante de *C. lucknowense* podem servir como uma fonte bruta a partir da qual formas puri-
25 ficadas de CBH Ia, CBH Ib, CBH IIa, CBH IIb, EG II, EG V, EG VI, BGL, ou Xyl II, podem ser produzidas. Alternativamente, as enzimas secretadas de uma tal cepa mutante também podem ser aplicadas diretamente aos materiais celulósicos a serem sacarificados. Em modalidades particularmente preferidas, os materiais celulósicos são expostos diretamente à cepa mutante de *C. lucknowense* em um ambiente conducente à proliferação da cepa mutante de
30 *C. lucknowense*, tal como em um biorreator. As secreções *in situ* de CBH Ia, CBH Ib, CBH IIa, CBH IIb, EG II, EG V, EG VI, BGL, ou Xyl II, ou misturas das mesmas, pela cepa mutante de *C. lucknowense*, em proporções mais altas do que encontradas nas secreções enzimáticas do organismo do tipo selvagem, levam à sacarificação *in situ* realçada do material celulósico.

35 A seguir do tratamento enzimático pelas composições enzimáticas inventivas da invenção, o açúcar fermentável que é produzido, pode ser exposto a microorganismos, que ocorrem naturalmente ou geneticamente projetados, que são capazes de fermentar o açúcar

para produzir etanol ou algum outro produto de fermentação com valor agregado. De preferência, substancialmente toda a glicose é convertida em etanol, que pode ser subsequentemente usado como um combustível, solvente, ou reagente químico. Em modalidades preferidas, o etanol é usado como um combustível para fornecer energia para veículos de transporte, exemplos não limitantes dos quais incluem carros, caminhões, ônibus, bicis motorizados e motocicletas. Outros produtos de fermentação potenciais provenientes da glicose incluem, mas não são limitados a, biocombustíveis (incluindo etanol); ácido láctico; plásticos; em especial produtos químicos; ácidos orgânicos, incluindo ácido cítrico, ácido succínico e ácido maléico; solventes; suplementos de ração animal; produtos farmacêuticos; vitaminas; aminoácidos, tais como lisina, metionina, triptofano, treonina, e ácido aspártico; enzimas industriais, tais como proteases, celulasas, amilases, glicanases, lactases, lipases, liases, oxidoredutases, e transferases; e matérias-primas químicas.

EXEMPLOS

Exemplo 1: Isolamento Enzimático

Os filtrados de cultura produzidos pelas cepas mutantes de *C. lucknowense* foram usados para o isolamento de enzimas individuais. A preparação comercial NCE-L600 (*C. lucknowense*) foi proveniente da Dyadic International, Inc., USA.

BGL altamente purificada (celobiase) de *Aspergillus japonicus* foi obtida a partir de uma preparação comercial, tendo 50 U mg^{-1} de proteína de atividade de celobiase específica (pH 5,0, 40°C), e foi usada nos experimentos em hidrólise de celulose insolúvel.

Exemplo 2: Purificação Enzimática

A purificação enzimática foi realizada por cromatografia em um sistema FPLC da Pharmacia (Suécia). Celobiohidrolases e endoglicanases BGL e Xyl II foram isoladas de um filtrado de cultura de UV18-25 de *C. lucknowense*. BGL e Xyl II (xilanase II) foram isoladas dos filtrados de cultura produzidos pelas cepas mutantes UV18 Δ Cbh1#10 e Xyl2-18 de *C. lucknowense*, respectivamente.

Em todos os casos, o primeiro estágio de purificação foi a cromatografia de troca aniônica em uma coluna Source 15Q (volume de 40 ml). A coluna foi equilibrada com tampão Bis-Tris-HCl 0,02 M, pH 6,8. O filtrado de cultura inicial foi preliminarmente dessalinizado e transferido no tampão primário por filtração em gel em Acrylex P4 (Reanal, Hungria). A amostra (400 mg de proteína) foi aplicada à coluna Source 15Q, e a eluição foi realizada com um gradiente de NaCl 0 a 1 M em uma taxa de fluxo de 10 ml min^{-1} .

A primeira fração da proteína depois da Source 15Q, eluída em NaCl 0,05 M e tendo alta atividade de Avicelase, foi submetida à cromatografia de interação hidrofóbica em uma coluna Isopropil Source 15 (Pharmacia, Suécia). A coluna foi equilibrada com sulfato de amônio 1,7 M em tampão Na-acetato 50 mM, pH 5,0. As proteínas foram eluídas com um gradiente linear reverso de sulfato de amônio 1,7 a 0 M em uma taxa de fluxo de 4 ml min^{-1} .

A fração da proteína com a atividade mais alta contra Avicel (eluindo em uma concentração de sal de 0,30 a 0,35 M) continha a proteína homogênea com uma massa molecular de 70 kDa (CBH IIb, vide Fig. 1).

A fração da proteína depois da Source 15Q, eluída em NaCl 0,22 M e tendo a atividade contra Avicel e *p*-NP- β -D-celobiosídeo, foi ainda purificada por cromatofocalização em uma coluna Mono P HR 5/20 (Pharmacia, Suécia). A coluna foi equilibrada com tampão Na-formiato 0,025 M, pH 4,0. As proteínas foram eluídas com um gradiente de pH 4,5 a 3,0 (usando Polybuffer 74) em uma taxa de fluxo de 0,5 ml min⁻¹. CBH Ib 60 kDa homogênea foi obtida como um resultado da cromatofocalização (FIG. 1).

As duas celobiohidrolases recém isoladas são homogêneas de acordo com os dados de SDS-PAGE e isoeletrofocalização (FIG. 1), descobriu-se que suas massas moleculares foram de 60 e 70 kDa, pI 3,8 e 5,6, respectivamente. O mapeamento genético da massa do peptídeo usando espectrometria de massa MALDI-TOF (dados não mostrados) indicou que estas proteínas foram diferentes das celobiohidrolases acima mencionadas (Cel6A e Cel7A) assim como de outras enzimas de *C. lucknowense* previamente isoladas. A sequência *de novo* subsequente de peptídeos tripticos a partir de novas celobiohidrolases, usando espectrometria de massa TOF/TOF (MS/MS) em conjunto, seguido pela pesquisa BLAST no banco de dados SWISS-PROT (UniProtKB) mostrou que as proteínas de 60 kDa e 70 kDa exibem similaridade de sequência para celobiohidrolases a partir de famílias 7 e 6 de GH (Tabela 1, vide classificação nas famílias em HyperTextTransferProtocol://afinb.cnrs-mrs.fr/CAZY/). Desse modo, elas foram classificadas como Cel7B (CBH Ib) e Cel6B (CBH IIb), respectivamente. Assim, o fungo *C. lucknowense* secreta pelo menos quatro celobiohidrolases codificadas por genes diferentes, dois deles pertencentes à família glicosil hidrolase 6 (GH6) e duas outras enzimas à família GH7 (Tabela 2). As moléculas da CBH Ia (Cel7A) e CBH IIb (Cel6B) representam celulasas típicas que consistem de um domínio catalítico e CBM conectados por um ligador de peptídeo flexível. As moléculas de CBH Ib (Cel7B) e CBH IIa (Cel6A) consistem apenas dos domínios catalíticos (elas não têm CBM). Deve ser observado que a maioria dos fungos *T. reesei* estudados tem apenas duas celobiohidrolases: I (Cel7A) e II (Cel6A). Outros fungos, tais como *Humicola insolens*, também segregam duas celobiohidrolases (Cel7A e Cel6A), enquanto que *Phanerochaete cluysosporium* produz pelo menos sete celobiohidrolases diferentes, das quais seis enzimas pertencem à família GH7. Todas as enzimas mencionadas, exceto a CBH I-1 de *P. chrysosporium* (Cel7A), possuem CBM.

A BGL foi isolada a partir da fração da proteína depois da Source 15Q (eluída em NaCl 0,10 M) contendo a atividade mais alta contra *p*-NP- β -D-glicopiranosídeo e celobiose. A fração foi submetida à cromatografia de interação hidrofóbica como descrito acima, a BGL homogênea com uma massa molecular de 106 kDa e pI 4,8 foi eluída a 1,3 M de sulfato de

amônio. Descobriu-se que a atividade específica da BGL para *p*-NP- β -D-glicopiranosídeo e celobiose foi de 11 e 26 U mg⁻¹ de proteína, respectivamente (40 °C, pH 5,0). A BGL purificada teve atividade ideal ao pH 4,0 e reteve >50 % de atividade na faixa do pH 2,5 a 6,5. A temperatura ideal foi de 40 °C. Depois do aquecimento durante três horas, a enzima reteve 10 % de atividade a 60 °C, 64 % a 50 °C, e 100 % a 40 °C. A enzima foi altamente ativa contra celobiose, gentiobiose, e laminarobiose como substratos. A atividade fraca também foi observada usando soforose, celotriose, celotetraose, celopentaose, e celohexaose como substratos. Nenhuma atividade foi observada com lactose ou tregalose como substratos.

A Xyl II homogênea (24 kDa, pI 7,9) foi obtida depois da cromatografia de troca aniônica seguido por cromatografia de interação hidrofóbica como descrito acima e filtração em gel em uma coluna Superose 12 HR 10/30 (Pharmacia, Suécia). A eluição no último estágio cromatográfico foi realizada com tampão Na-acetato 0,1 M, pH 5,0, em uma taxa de fluxo de 0,3 ml min⁻¹. A Xyl II teve atividade de xilanase específica de 395 U mg⁻¹ de proteína (50 °C, pH 5,0, xilana birchwood como um substrato). A enzima teve um pH ideal de 6,0 e uma temperatura ideal de 70 °C. A Xyl II foi altamente específica para xilana como substrato, sem nenhuma atividade contra carboximetilcelulose (CMC) ou β -glicana de cevada.

As CBH Ia (65 kDa), CBH IIa (43 kDa), EG II (51 kDa), EG V (25 kDa), EG VI (47 kDa) de *C. lucknowense* foram purificadas como descrito em outra parte deste relatório (vide, Gusakov AV, Sinitsyn AP, Salanovich TN, Bukhtojarov FE, Markov AV, Ustinov BB, van Zeijl C, Punt P, Burlingame R. "Purification, cloning and characterisation of two forms of thermostable and highly active cellobiohydrolase I (Cel7A) produced by the industrial strain of *Chrysosporium lucknowense*" *Enzyme Microb Technol* 2005;36:57-69; Bukhtojarov FE, Ustinov BB, Salanovich TN, Antonov AI, Gusakov AV, Okunev ON, Sinitsyn AP. "Cellulase complex of the fungus *Chrysosporium lucknowense*: isolation and characterization of endoglucanases and cellobiohydrolases", *Biochemistry* (Moscou) 2004;69:542-51.

A pureza da enzima foi caracterizada por SDS-PAGE e isoeletrofocalização. SDS-PAGE foi realizado em 12 % de gel usando um equipamento Mini Protean II (Bio-Rad Laboratories, USA). Isoeletrofocalização foi realizada em um Modelo 111 Mini IEF Cell (Bio-Rad Laboratories, USA). A marcação da proteína foi realizada com Coomassie Blue.

Exemplo 3: MALDI-TOF e espectrometria de massa TOF/TOF em conjunto de peptídeos

A digestão triptica em gel das tiras de proteína depois do SDS-PAGE foi realizada essencialmente como descrito por Smith (Smith BE. Protein sequencing protocols. Totowa: Humana Press; 1997). Tripsina (Promega, modificada, 5 µg/mL) em NH₄HCO₃ 50 mM foi usada para uma digestão de proteína. Os peptídeos resultantes foram extraídos de um gel com 20 % de acetonitrila aquosa contendo 0,1 % de ácido trifluoroacético e submetida à MS MALDI-TOF (vide, James P. (Ed.) Proteome research: mass spectrometry. Berlin: Springer-

Verlag; 2001.) Os peptídeos selecionados a partir de espectros de massa dos digestos tripticos da CBH Ib e IIb foram analisados por espectrometria de massa em conjunto de modo a determinar suas sequencias *de novo*. O espectrômetro de massa TOF/TOF Ultraflex (Bruker Daltonik GmbH, Alemanha) foi usado nos experimentos da MS.

5 Exemplo 4: Ensaios de atividade enzimática

A atividade de CMCase foi medida pelo ensaio de açúcares redutores liberados depois de 5 mM de reação enzimática com 0,5 % de carboximetilcelulose (CMC, viscosidade do meio, Sigma, USA) ao pH 5,0 e 50 °C (Sinitsyn AP, Chemoglazov VM, Gusakov AV. "Methods of investigation and properties of cellulolytic enzymes" (in Russian), *Biotechnology Series*, v. 25. Moscou: VINITI Press; 1990). As atividades enzimáticas contra β -glicana de cevada (Megazyme, Austrália) e xilana birchwood (Sigma, USA) foram determinadas do mesmo modo como a atividade de CMCase, exceto o tempo de incubação que foi de 10 mM. A atividade de avicelase foi determinada analisando-se açúcares redutores liberados depois de 60 mM de reação enzimática com 5 mg ml⁻¹ de Avicel PH 105 (Serva, Alemanha) 15 ao pH 5,0 e 40 °C. Os açúcares redutores foram analisados pelo método de Somogyi-Nelson (Sinitsyn AP, Chemoglazov VM, Gusakov AV, "Methods of investigation and properties of cellulolytic enzymes" (in Russian); *Biotechnology Series*, v. 25. Moscou: VINITI Press; 1990; Somogyi M., "Notes on sugar determination" *J Biol Chem* 1952;195:19-23. A atividade sobre papel de filtro (FPA) foi determinada conforme recomendado por Ghose (Ghose TK. 20 "Measurement of cellulase activities", *Pure Appl Chem* 1987;59:257-68).

As atividades contra *p*-NP- β -D-glicopiranosídeo, *p*-NP- β -D-celobiosídeo e *p*-NP- β -D-lactosídeo (Sigma, USA) foram determinadas ao pH 5,0 e 40 °C como descrito em outra parte deste relatório (Gusakov AV, Sinitsyn AP, Salanovich TN, Bukhtojarov FE, Markov AV, Ustinov BB, van Zeijl C, Punt P, Burlingame R. "Purification, cloning and characterisation of 25 two forms of thermostable and highly active cellobiohydrolase I (Cel7A) produced by the industrial strain of *Chrysosporium lucknowense*", *Enzyme Microb Technol* 2005;36:57-69).

A atividade de celobiase foi avaliada ao pH 5,0 e 40 °C medindo-se a taxa inicial de liberação de glicose a partir de celobiose 2 mM pelo método da glicose oxidase-peroxidase (Sinitsyn AP, Chemoglazov VM, Gusakov AV, "Methods of investigation and properties of 30 cellulolytic enzymes" (in Russian), *Biotechnology Series*, v. 25. Moscow: VINITI Press; 1990).

Todas as atividades foram expressadas em Unidades Internacionais, isto é, uma unidade de atividade correspondeu à quantidade de enzima hidrolisando um μ mol de substrato ou liberando um μ mol de açúcares redutores (em equivalentes de glicose) por um minuto. 35

Exemplo 5: Hidrólise enzimática de substratos celulósicos

A hidrólise enzimática de substratos celulósicos foi realizada ao pH 5,0 sob agita-

ção magnética. Avicel PH 105 (Serva, Alemanha), pré tratado de algodão com mistura de acetona-etanol (1:1) durante dois dias de modo a remover a cera da superfície de fibras de celulose, e madeira da espécie Douglas fir pré tratada por organosolv foram usadas como substratos.

- 5 Os experimentos em cinéticas de progresso da hidrólise de Avicel por celobiohidrolases individuais purificadas e experimentos em interação sinérgica entre celulasas de *C. lucknowense* (com algodão como um substrato) foram realizados a 40 °C. A concentração de substrato nesses experimentos foi de 5 mg ml⁻¹. De modo a eliminar o efeito da inibição do produto (celobiose) nas cinéticas e converter todos os celooligossacarídeos em glicose, a hidrólise foi realizada na presença de BGL purificada (celobiase) de *A. japonicus*, que foi extra adicionada ao sistema de reação em quantidade excessiva (0,5 U ml⁻¹).

- 10 Os experimentos em sacarificação enzimática de Avicel, algodão, e madeira da espécie Douglas fir pré tratada por combinações de enzimas de *C. lucknowense* purificadas e preparações multienzimáticas brutas foram realizadas a 50 °C. A concentração de Avicel e madeira pré tratada nesses experimentos foi de 50 mg ml⁻¹, enquanto que a concentração de algodão foi de 25 mg ml⁻¹.

- Um experimento típico foi realizado da seguinte maneira. Uma quantidade pesada de substrato celulósico seco foi colocada em um tubo de ensaio plástico 2 ml, depois 0,5 a 1 ml de tampão Na-acetato 0,05 M, contendo NaN₃ 1 mM para prevenir contaminação microbiana, foi adicionado, e o substrato foi embebido no tampão durante 1 h. Depois, o tubo foi colocado em um banho de água termo-estabelecido, localizado em um agitador magnético, e solução enzimática adequadamente diluída no mesmo tampão foi adicionada à suspensão de substrato de modo a ajustar o volume total do sistema de reação a 2 ml e iniciar a hidrólise. O tubo foi hermeticamente fechado com uma tampa, e a hidrólise foi realizada com agitação magnética. Em tempos definidos na reação, uma alíquota da suspensão (0,05 a 0,1 ml) foi tomada, diluída, centrifugada durante 3 min a 15000 rpm, e as concentrações de glicose e açúcares redutoras no sobrenadante foram determinadas pelos métodos da glicose oxidase-peroxidase e de Somogyi-Nelson. Nesses casos, quando a glicose foi um produto único da reação, o grau de conversão de substrato (para Avicel e algodão, que representou substratos celulósicos puros) foi calculado usando a seguinte equação:

30
$$\text{Conversão (\%)} = \frac{\text{Concentração de glicose (mg ml}^{-1}\text{)} \times 100 \text{ \%}}{\text{Concentração de substrato inicial (mg ml}^{-1}\text{)}} \times 1,11$$

- Os experimentos cinéticos foram realizados em duplicatas. A concentração de proteína foi a medição de carga de enzima no sistema de reação. No caso de enzimas purificadas, a concentração de proteína foi calculada a partir da absorção de UV a 280 nm usando coeficientes de extinção de enzima prognosticados pela ferramenta ProtParam (<http://WorldWideWeb.expasy.ch/tools/protparam.HyperTextMarkupLanguage>). Para as pre-

parações multienzimáticas brutas, a concentração de proteína foi determinada pelo método de Lowry usando albumina de soro bovino como um padrão.

A CBH Ib e IIb exibiram atividade máxima ao pH 4,7 e 5,0. Ambas enzimas foram estáveis durante a incubação de 24 h ao pH 5,0 e 50 °C. O estudo da adsorção da enzima em Avicel, realizado ao pH 5,0 e 6 °C, revelou que apenas a CBH IIb tem CBM. Depois da incubação da CBH Ib e IIb (1 mg ml⁻¹) com Avicel (25 mg ml⁻¹) durante 30 min em agitação, o grau de adsorção de proteína foi de 65 e 99 %, respectivamente. Deve ser observado que o grau de adsorção do domínio catalítico da CBH Ia de *C. lucknowense* foi de 59 % sob as mesmas condições, enquanto que para a CBH Ia de *C. lucknowense* de tamanho natural (uma enzima com CBM) foi de 89 %.

A CBH IIb teve uma alta atividade contra Avicel e atividade de CMCase muito baixa, enquanto que a atividade para derivados de *p*-nitrofenil sintéticos de dissacarídeos foi completamente ausente (Tabela 2). A CBH Ib exibiu atividade de Avicelase mais baixa, mas *p*-NP- β -D-celobiosídeo e *p*-NP- β -D-lactosídeo hidrolisados, que é típica para celulasas da família 7. Para uma comparação, atividades específicas de celobiohidrolases de *C. lucknowense* previamente isoladas (agora chamadas como CBH Ia e CBH IIa) também são fornecidas na Tabela 2.

A FIG. 2 mostra as cinéticas de progresso da hidrólise de Avicel para todas as celobiohidrolases de *C. lucknowense* purificadas, onde as enzimas foram equalizadas por concentração de proteína (0,1 mg ml⁻¹). De modo a eliminar o efeito da inibição do produto (celobiose) nas cinéticas, a hidrólise foi realizada na presença de BGL purificada (celobiase) de *A. japonicus*, adicionada ao sistema de reação em quantidade excessiva (0,5 U ml⁻¹).

A taxa da hidrólise mais alta entre um número reduzido de celobiohidrolases testadas, incluindo três outras enzimas de *C. lucknowense* (CBH Ia, Ib, IIa) foi observada no caso de CBH IIb de *C. lucknowense*: 3,2 mg ml⁻¹ de glicose, isto é, 58 % de conversão de celulose foi obtida depois de 5 dias da hidrólise (vide FIG. 2). A CBH Ia de *C. lucknowense* (que tem um CBM) foi notavelmente menos eficaz (o rendimento de glicose depois de 5 dias foi de 2,5 mg ml⁻¹, que correspondeu ao grau de conversão de celulose de 46 %, respectivamente). Como esperado, as celobiohidrolases de *C. lucknowense* sem CBM (CBH Ib e IIa) tiveram a capacidade mais baixa para hidrolisar Avicel: apenas 23 e 21 % de conversão de celulose foi obtida depois do mesmo tempo de reação.

Ambas celobiohidrolases de *C. lucknowense* tendo um CBM (Ia e IIb) exibiram um sinergismo considerável com três endoglicanases principais a partir do mesmo fungo (EG II, EG V, EG VI) na hidrólise de algodão assim como uma sinergia forte entre si (Tabela 3). Nestes estudos, a concentração de algodão foi de 5 mg ml⁻¹, a concentração de CBH foi de 0,15 mg ml⁻¹ em todos casos, enquanto que a concentração de EG foi sempre de 0,05 mg ml⁻¹. De modo a eliminar o efeito da inibição do produto nas cinéticas e converter os oligos-

sacarídeos intermediários em glicose, a hidrólise foi realizada na presença de BGL purificada de *A. japonicus*, adicionada ao sistema de reação em quantidade excessiva (0,5 U ml⁻¹). Os experimentos foram realizados ao pH 5,0 e 40 °C durante 140 h.

Como observado na Tabela 3, celobiohidrolases individuais, CBH Ia e CBH Ib, e as endoglicanases individuais, não hidrolisam completamente o algodão sob as condições testadas. A CBH Ib forneceu o rendimento de glicose mais alto depois de 140 h da hidrólise: 1,18 mg ml⁻¹, que correspondeu ao grau de conversão de substrato de 21 %. Entretanto, quando qualquer celobiohidrolase foi incubada com endogluacanase, um sinergismo considerável foi observado. Os rendimentos de glicose mais altos (4,1 a 4,7 mg ml⁻¹) foram obtidos com combinações de CBH Ia ou CBH Ib com EG II, o coeficiente de sinergismo sendo variado na faixa de 2,6 a 2,8. Um sinergismo forte ($K_{\text{syn}} = 2,75$) também foi observado entre CBH Ia e CBH Ib. De fato, a combinação de duas celobiohidrolases (1:1 em peso) com BGL forneceu conversão praticamente completa (98,6 %) de celulose de algodão em glicose depois de 140 h da hidrólise.

Como um exemplo, as cinéticas de progresso da hidrólise de algodão por combinações de CBH Ib com outras enzimas de *C. lucknowense* são mostradas na FIG. 3, onde os dados experimentais reais são mostrados com símbolos vazios (curvas contínuas) enquanto que as somas teóricas de concentrações de glicose obtidas sob a ação de enzimas individuais são mostradas com símbolos cheios (linhas pontilhadas). Os rendimentos de glicose obtidos depois de 140 h da hidrólise de algodão sob a ação de celobiohidrolases e endoglicanases individuais e suas combinações são resumidos na Tabela 3. O coeficiente de sinergismo (K_{syn}) foi calculado como uma razão de concentração de glicose experimental (coluna 2 da Tabela 3) para a soma teórica de concentrações de glicose (coluna 3).

Usando quatro enzimas de *C. lucknowense* purificadas (CBH Ia e Ib, EG II, BGL), um complexo de celulase artificial foi construído (combinação #1 de *C.l.*), o qual demonstrou uma capacidade extremamente alta para converter substratos celulósicos diferentes em glicose (FIGS. 4 a 6). Esta composição multienzimática foi notavelmente mais eficaz na hidrólise de celulose cristalina pura (algodão e Avicel) do que a preparação multienzimática de *C. lucknowense* bruta NCE-L600. Em 72 h da hidrólise de um substrato lignocelulósico (madeira da espécie Douglas fir pré tratada por organosolv), a combinação #1 de *C.l.* também foi muito eficaz na hidrólise da celulose.

Na combinação #1 de *C. lucknowense*, a enzima consistiu das duas celobiohidrolases CBH Ia e CBH Ib, e da endoglicanase EG II, as enzimas com capacidade de adsorção forte em celulose cristalina (as moléculas destas enzimas têm CBM). A atividade de celulases severamente adsorvidas é gradualmente diminuída durante o curso da hidrólise de celulose insolúvel como um resultado da mobilidade limitada da enzima ao longo da superfície do substrato ou da ligação improdutiva (desse modo chamada pseudoinativação). Sem de-

seja estar ligado por teoria, acredita-se que possa existir um sinergismo entre celulasas severamente e livremente adsorvidas em que celulasas de ligação livre (enzimas sem CBM) podem destruir obstáculos impedindo a ação processiva das celobiohidrolases severamente adsorvidas, assim auxiliando para que elas se movam para os próximos sítios reativos de
 5 celulose. A concentração de proteína total no sistema de reação foi de $0,5 \text{ mg ml}^{-1}$. A composição da composição multienzimática (combinação #1 de *C.l.*) foi a seguinte: $0,2 \text{ mg ml}^{-1}$ de CBH Ia + $0,2 \text{ mg ml}^{-1}$ de CBH Iib + $0,08 \text{ mg ml}^{-1}$ de EG II + $0,02 \text{ mg ml}^{-1}$ de BGL. Avicel (50 mg ml^{-1}) e algodão (25 mg ml^{-1}) foram usados como substratos representando celulose cristalina pura nestes experimentos. A amostra de madeira da espécie Douglas fir pré trata-
 10 da por organosolv (50 mg ml^{-1}) foi tomada como um exemplo de matéria-prima lignocelulósica real que pode ser usada para bioconversão em etanol. Uma preparação de celulase multienzimática de *C. lucknowense* bruta NCE L-600 (diluída de modo que a concentração de proteína no sistema de reação também seria de $0,5 \text{ mg ml}^{-1}$) foi tomada para uma comparação nestes estudos. Os experimentos da hidrólise com esta, também foram realizados na
 15 presença de BGL de *A. japonicus* extra adicionada ($0,5 \text{ U ml}^{-1}$).

As cinéticas de progresso da hidrólise de algodão, Avicel e Douglas fir por preparações multienzimáticas de celulase diferentes são mostradas nas FIGS. 4 a 6. Deve ser observado que em todos os casos, as concentrações de glicose e açúcares redutores depois de 24 a 72 h da hidrólise em um experimento concreto foram praticamente as mesmas, isto
 20 é, glicose composta de >96 % dos açúcares solúveis totais. Desse modo, o rendimento da glicose pode ser tomado como critério confiável em comparação à eficiência hidrolítica de amostras multienzimáticas diferentes.

Na hidrólise do algodão (FIG. 4), a combinação #1 de enzimas de *C. lucknowense* purificadas forneceu rendimento de glicose muito maior depois de 72 h da reação ($23,4 \text{ mg ml}^{-1}$, isto é, 84 % de grau de conversão de substrato) do que o de $4,2 \text{ mg ml}^{-1}$ exibido por
 25 (NCE-L600). Na hidrólise de Avicel (FIG. 5), a combinação #1 de *C.l.* também foi superior ($45,0 \text{ mg ml}^{-1}$ de glicose, ou 81 % de conversão de substrato depois de 72 h da hidrólise). No caso de Douglas fir pré tratada (FIG. 6), a combinação #1 de *C.l.* também foi eficaz (glicose $28,8 \text{ mg ml}^{-1}$, 63 % de conversão depois de 72 horas).

Diferente de Avicel e algodão, a amostra de madeira pré tratada continha não apenas celulose (~85 %) mas também lignina (13 %) e hemicelulose (2 %). A combinação #1 de quatro enzimas de *C. lucknowense* artificiais foi composta de apenas celulasas; todas elas, exceto para a BGL, tendo CBM. Todas as outras amostras multienzimáticas possuíam não apenas celulase mas também xilanase e outros tipos de atividade de carbohidrase, isto é,
 30 elas continham enzimas não-celulase adicionais. Isto pode esclarecer a eficiência relativamente menor da combinação #1 de *C.l.* em Douglas fir pré tratada comparada à preparação de *P. verruculosum* #151 (FIG. 6).

Em um conjunto de experimentos (FIG. 7), a amostra de madeira pré tratada foi hidrolisada por composições diferentes de enzimas de *C. lucknowense* purificadas, para que celulases carentes de um CBM fossem incluídas (EG V ou EG V em combinação com CBH Ib). A concentração de proteína total no sistema de reação foi mantida no mesmo nível de 0,5 mg ml⁻¹ (Tabela 5). De fato, duas combinações (#3 e #4) de *C.l.*, contendo enzimas fracamente adsorvidas, forneceram um aperfeiçoamento notável do rendimento de glicose depois de 72 h da reação enzimática em comparação com a combinação #1 de *C.l.*.

Em dois experimentos, a Xyl II de *C. lucknowense* altamente ativa (Xynl 1A) foi adicionada às quatro enzimas acima mencionadas (combinações #2 e #4 de *C.l.*). Visto que um sinergismo entre celulases severamente e livremente adsorvidas foi descrito [38], EG V ou EG V junto com CBH Ib (ambas as enzimas têm carência de CBM) foram usadas nas combinações #3 e #4 de *C.l.*.

Como pode ser observado na FIG. 7, a taxa inicial de formação de glicose diminuiu sequencialmente a partir da combinação #1 para combinação #4 de *C.l.*, entretanto, o rendimento de glicose depois de 2 a 3 dias da hidrólise aumentou na mesma sequência. A Xyl II demonstrou apenas leve efeito positivo no rendimento de glicose, enquanto que a EG V ou EG V junto com CBH Ib forneceu um aumento notável na concentração do produto depois de 72 h da hidrólise da madeira (37 e 41 mg ml⁻¹, respectivamente) comparado à combinação #1 de *C.l.* (29 mg ml⁻¹), isto é, as combinações #3 e #4 tiveram melhor desempenho do que todas as amostras multienzimáticas brutas (FIG. 6).

O baixo desempenho da preparação de *C. lucknowense* bruta (NCE-L600) na hidrólise de substratos celulósicos diferentes (FIGS. 4 a 6) merece uma atenção especial. Sem desejar estar ligado pela teoria, pode ser esclarecido o teor total baixo de celobiohidrolases diferentes na NCE-L600 (35 a 40 % do teor de proteína total). Além disso, duas das quatro celobiohidrolases de *C. lucknowense* (Ib e IIa) carecem de CBM, enquanto que duas outras enzimas (CBH Ia e IIb) também parcialmente perdem a CBM durante o curso de fermentação. A ausência de CBM na maior parte de celobiohidrolases a partir da NCE-L600 pode levar à atividade mais baixa da preparação bruta para celulose cristalina.

Tabela 1

Identificação de peptídeos nas proteínas de <i>C. lucknowense</i> isoladas usando MALDI-TOF MS/MS				
Enzima	m/z	Peptídeo ^a	Identificação ^b BLAST	Nº UniProtKB
Proteína 60 kDa	1133,6	HEYGTNIGSR	118 HEYGTNIGSR 127 (cbh1,2 <i>Humicola grisea</i> – GH7)	O94093
	1829,9	MGNQDFYGPGLTVDTSK	291 LGNTDFYGPGLTVDT 305 (cbhB <i>Aspergillus niger</i> – GH7)	Q9UVS8
Proteína 70 kDa	1061,4	YPANDYYR	127 ANYYR 132 (Avicelase 2 <i>Humicola insolens</i> – GH6)	Q9C1S9
	1990,0	HYIEAFSPLLNSAGFPAR	367 KYIEAFSPLLNAAGFPA 383 (CBH II <i>Neurospora crassa</i> – GH6)	Q872J7
	2073,5	LWQPTGQQQWGDWCNVK	381 QPTGQQQWGDWCNV 394 (CBH II <i>T. reesei</i> – GH6)	P07987

^a Visto que a MS/MS não pode distinguir entre os resíduos de Leu e Ile (eles têm as mesmas massas), pode haver ambiguidade nas posições apropriadas dos peptídeos identificados.

^b Resíduos conservados nas enzimas de *C. lucknowense* são mostrados em negrito.

Tabela 2

Atividades específicas (U mg ⁻¹ de proteína) de celobiohidrolases purificadas de <i>C. lucknowense</i> para substratos diferentes ao pH 5,0 e 40 °C								
Enzima	Massa mol. (kDa)	Designação do domínio cat.	Pre-sença de CBM	Avicel	CMC ^a	β -glicana ^a de celvada	p-NP- β -D-celobiosídeo	p-NP- β -D-lactosídeo
CBH Ia	65	Cel7A	Sim	0,21	0,1	<0,1	0,021	0,12
CBH Ib	60	Cel7B	Não	0,12	0,3	<0,1	0,020	0,09
CBH IIa	43	Cel6A	Não	0,08	1,1	2,0	0	0
CBH IIb	70	Cel6B	Sim	0,22	0,2	0,2	0	0

^a Atividade foi determinada a 50 °C.

Tabela 3

Sinergismo entre celulasas de *C. lucknowense* na hidrólise de celulose de algodão (5 mg ml⁻¹) ao pH 5,0 e 40 °C na presença de 0,5 U ml⁻¹ de BGL de *A. japonicus*. Em todos os casos, a concentração de CBH foi de 0,15 mg ml⁻¹, a concentração de EG foi de 0,05 mg ml⁻¹.

Enzima	Concentração de glicose depois de 140 h, experimental (mg ml ⁻¹)	Concentração de glicose depois de 140 h, teórico ^a (mg ml ⁻¹)	K _{syn}
CBH Ia	0,81	—	—
CBH IIb	1,18	—	—
EG II	0,64	—	—
EG V	0,70	—	—
EG VI	0,40	—	—
CBH Ia + EG II	4,05	1,45	2,79
CBH Ia + EG V	3,68	1,51	2,44
CBH Ia + EG VI	3,93	1,21	3,25
CBH IIb + EG II	4,72	1,82	2,59
CBH IIb + EG V	3,81	1,88	2,03
CBH IIb + EG VI	4,05	1,58	2,56
CBH Ia + CBH IIb	5,47	1,99	2,75

^a Calculado como uma soma de concentrações de glicose obtidas sob a ação de enzimas individuais.

Tabela 4

Atividades específicas (U mg⁻¹ de proteína) de preparações multienzimáticas para substratos diferentes ao pH 5,0 e 50 °C

Preparação	Proteína (mg ml ⁻¹ ou mg g ⁻¹)	Papel de filtro	CMC	Xilana	Celobiose ^a
NCE-L600	45	0,25	12,2	4,8	0,07
Combinação #1 de C.I.	1000	1,10	6,6	0	1,05

^a Atividade foi determinada a 40 °C.

Tabela 5

Composição de combinações multienzimáticas artificiais com base em enzimas de *C. lucknowense* purificadas e rendimentos de glicose depois de 72 h da hidrólise de madeira

da espécie Douglas fir pré tratada (50 mg ml^{-1}), pH 5,0, 50°C . A concentração de proteína total no sistema de reação foi de $0,5 \text{ mg ml}^{-1}$, a concentração de cada componente e rendimentos de glicose são fornecidos em mg ml^{-1} .

Combinação	CBH Ia	CBH Ib	CBH IIb	EG II	EG V	BGL	Xyl II	Rendimento de glicose
#1	0,2	0	0,2	0,08	0	0,02	0	28,8
#2	0,2	0	0,2	0,07	0	0,02	0,01	30,1
#3	0,2	0	0,2	0,04	0,04	0,02	0	37,3
#4	0,1	0,1	0,2	0,03	0,04	0,02	0,01	41,0

Listagem de Sequência

<110> Dyadic International, Inc.
GUSAKOV, Alexander V.
SALANOVICH, Tatyana N.
ANTONOV, Alexey I.
USTINOV, Boris B.
OKUNEV, Oleg N.
BURLINGAME, Richard P.
EMALFARB, Mark A.
BAEZ, Marco
SINITSYN, Arkady P.

<120> "CONSTRUÇÃO DE COMPOSIÇÕES DE CELULASE ALTAMENTE EFICAZES PARA A HIDRÓLISE ENZIMÁTICA DE CELULOSE"

<130> 3123-4008PC

<140> TBD

<141> 2006-07-13

<160> 18

<170> PatentIn versão 3.3

<210> 1

<211> 6360

<212> DNA

<213> Chrysosporium lucknowense

<400> 1

ctcagattct aggggtaggg cgggagcaga ggcgaaaatt gggttgtaga atatgaggag	60
ctagggttgt taaactcaaa gaacttcttg ctcttggtct tagtcttctc tcctgggaaa	120
agggggtttt tccgaaagcg gcgctatacg aagccagagg ctactttcct tgctttggat	180
ggcccttgtc caccgttctt gtttcccgtt tgtcaattgc gacgttgccg gcaacctagg	240
tcctaataat taggtagata tttcggtaga ggtagtttaa ttatgcttca gtagagaaat	300
cgttgtctcc acgtctcgca accttgcgaa acttcgccac attgaagata gcattgtctg	360
agttgatttt aaccctttcc agagacgata taatagtgcg agtttctttg atcggaatca	420
tcgacattcg gattttccct taattatatg aagtattcgg cccacggaac cgggccccga	480
gcagggttgaa ccgcgcaaaa cctcaaccga gtcacctcgc gtccatgttt gtcattggaat	540
caggctccga atcccgtcag atcagtcagt tctggtggct atggacgcgg gagttacggc	600
cagtcgtccc gttgttcttg ggggttgatc aacaggagga agagatctga gatcgaacta	660
cacccattga tttatcgacg cataatcaag ttttaataaaa accaaacagc gtgtttggtg	720
ctaccaccga atgcgagatc cgggctagcc cgcggaagga tgatggccac agatctagcg	780
tcattgtatga ttattaccta tgcattctatc ttcgtatctg cctcggggtg gcaacacctg	840
accgagagac gactcgacaa cctgacactt ggcaaaagac atttcggttg acagcgggag	900
aactccagcg aggaagtcgc ccagagatgc ggatgagaag acaacgccga gacgtgccgg	960

cgttggctct	ccacgaatcg	gagccgactc	ttccgtttgg	ccaatctccg	ggataaatcc	1020
cagcggcggg	tcacgtcacg	tttcatgggg	aggcgcggac	agccatccca	gccaggccat	1080
ggaagagaac	aattcttggg	ggtagcgacc	gagccaaaag	gggggggggg	gaagcgggag	1140
gggaagaagt	ggtattagag	cacgcaccgg	aaaacgcatt	tgggcccttg	ccaacaaaca	1200
ccacaccccg	cgtcctggga	gcaagacatc	caggatgcaa	cccagtaggg	gatgccaaga	1260
agcatctacg	gcaccatctg	ccggcgcctc	gcctgttaga	gtcccggcac	ccgccaatgg	1320
ggccgtgctg	ggccctgccc	ggcaatgctg	gcgcagcggc	atcaacaaca	ttgctcgggg	1380
aggggcccga	ttttattgat	tagcaaaaaa	acaattaaat	tacccttcca	ttccagcaga	1440
gcttctcctc	cacgcggcgg	cgggaccgct	tgtggacggc	ggtacactac	aaccgcgggg	1500
ctccagtctc	cgtgctgggc	gtgcagatca	cgacccggaa	gagaaatgat	cgcggtctga	1560
cgccgggtac	ggagtactga	gccgccaacc	acagccgatg	gaccgtgata	tctcaatgcg	1620
ttcaagcaac	acagcaacac	cctggacgag	tctctcctcc	cctaccaccc	cctccccccc	1680
tgccctggcc	gcgaacgggg	cgcgtacccc	agatttctac	tccgtactga	caccccaatc	1740
tattcccgtc	ggcgtcggcc	agtctggggc	ggtccggcca	agactctcgg	tgcacgatac	1800
cgcgacgaaa	tcggattaac	cgttggctga	tcaattccaa	gtcaagggag	aagtggatatg	1860
gaaagtgcgc	tcagttttcc	actgcccccg	acaggcaggt	tccggatctg	gacagcagtc	1920
ttccgaatct	ttggcagaga	ctcatgataa	tataaaaagg	caaatgaggc	ggcgccttgg	1980
acagggtccat	tctcccaccg	ctcaaccagc	ctccaattcc	tcagaagtct	gttgctctct	2040
cgcagtcgca	gtcaagatga	agcagtacct	ccagtacctc	gcggcgaccc	tgcccctggg	2100
gggcctggcc	acggcccagc	aggcgggtaa	cctgcagacc	gagactcacc	ccaagctcac	2160
ttggtcgaag	tgcacggccc	cgggatcctg	ccaacaggtc	aacggcgagg	tcgtcatcga	2220
ctccaactgg	cgctgggtgc	acgacgagaa	cgcgcagaac	tgctacgacg	gcaaccagtg	2280
gaccaacgct	tgcagctctg	ccaccgactg	cgccgagaat	tgcgcgctcg	agggtgccga	2340
ctaccagggc	acctatggcg	cctcgaccag	cggcaatgcc	ctgacgctca	ccttcgtcac	2400
taagcacgag	tacggcacca	acattgggtc	gcgcctctac	ctcatgaacg	gcgcgaacaa	2460
gtaccagatg	ttcacccctc	agggcaacga	gctggccttc	gacgtcgacc	tctcggccgt	2520
cgagtgcggc	ctcaacagcg	ccctctactt	cgtggccatg	gaggaggatg	gcggtgtgtc	2580
gagctacccg	accaacacgg	ccggtgctaa	gttcggcact	ggggtaagtt	caacgacccg	2640
agacgggtgc	ccttattatc	tgctgcgaaa	acggacgggtc	cccttttgct	aactaccctc	2700
ctccaaacag	tactgcgacg	cccaatgcgc	acgcgacctc	aagttcgtcg	gcggcaaggg	2760
caacatcgag	ggctggaagc	cgtccaccaa	cgatgccaat	gccggtgtcg	gtccttatgg	2820
cgggtgctgc	gctgagatcg	acgtctggta	agttttgttg	cctgggcagc	aatggtatat	2880

tagctcgagt ggttcccgtc gttgctgacc ctctcttacc agggagtcga acaagtatgc	2940
tttcgctttc accccgcacg gttgcgagaa ccctaaatac cacgtctgcg agaccaccaa	3000
ctgcggtggc acctactccg aggaccgctt cgctggtgac tgcgatgcca acggctgcca	3060
ctacaacccc taccgcatgg gcaaccagga cttctacggt cccggcttga cggctcgatac	3120
cagcaagaag ttcacgtgag tacaccgtgc ttgaagcccc ctccccccc cccccaaaa	3180
aaaaaaagaa aaaagaagtc aaatgattga tgctaaccaa atcaaataac agcgtcgtca	3240
gccagttcga ggagaacaag ctcaccaggt tcttcgtcca ggacggcaag aagattgaga	3300
tccccggccc caaggctgag ggcacgatg cggacagcgc cgctatcacc cctgagctgt	3360
gcagtgccct gttcaaggcc ttcgatgacc gtgaccgctt ctcggagggtt ggcggcttcg	3420
atgccatcaa cacggccctc agcactccca tggctcctgt catgtccatc tgggatgatg	3480
tacgttacct aacccccccc cctttttttt ttcccgttc tctccccgaa actgccacta	3540
cttatatacg tcccgcgtcc atgatgctta ctttttctcc ttccagcact acgccaatat	3600
gctctggctc gactcgagct acccccctga gaaggctggc cagcctggcg gtgaccgtgg	3660
cccgtgtcct caggactctg gcgtcccggc cgacgttgag gctcagtacc ctaatgcgtg	3720
agtcgaaacc gtaaaatgtc gggcaaaaaa aagatcgctc aagctaacga aataatatga	3780
ttagcaaggt catctgggtc aacatccgct tcggccccat cggctcgact gtcaacgtct	3840
aaactgcaac ctgaccgggc cttttctctc cacccccacc cctctcaagt tctctctggt	3900
ggagccctcg tgtccttctt ttccctagggt cgcgaaacct tgagcttggt tatcgtaggg	3960
tcattgtgta catacacaaa aacttaacat ctgctaccaa gatcttggcg ctttgccagg	4020
tcttctcaaa cctcgaagca ctgagccttt gtctcccgag tgaagtagga tgactattta	4080
cgttgcaaga ctacgcggtg aaggggacgg agcagacctg ccacagatat tcgtttggtt	4140
gcttgattta tagcagagtc cgaacgtaga catggccccct gaagggtgcca accctagata	4200
gccagaagcc ttgttttacg aaagggtggt caaccaacgg tgctcctcgc tcagcgaatc	4260
taccgcacg caatgtatcg taagaatgtg aactaaaggg aacgacgagg catagggaaa	4320
cgtcaatgtg gcttgaataa cagagttaaa tacctaatag aagaaattag catgccaaga	4380
ttgagccagc aacacatggt agaatagcca gcaaaggacg cttgttcgct tgatctcgaa	4440
ccgtccaacc tgattcgaag gaggagggaa aagttgaaga atacgggcaa taattactcg	4500
aggttcctat gccctgcaga gtctaattaa tattaaaggc accaccgcga tgattccgca	4560
attataagca taataagctc gcgggccccca cacgtgcctt caccctccca tgtgtataca	4620
atctgtacct cgttattgtc gaatcgctat tccgatagcg aaggtctggc actcatcaga	4680
taccgtgaca tcgattgaga tttggccggg ccaccggtag taagcgatga gttggtcatc	4740

```

aattatcaac aatgcgctca atcagcgata atcagcctat caaccgcaa atcatacgcg 4800
catcaacgaa ttgtccatca tgcacgtagc ttgtcggcag tgccgcatac cctccagagc 4860
atcatagccg ggatagaaaag ctcgctttca gccgtcccag agtccgagat gcaggtagca 4920
agccttcaag accagttata tgtgacccgg gtaaaatact tggtgagatg caatgggagc 4980
agcttcgggc acttataagc ttacttagat attatctcaa ggtttctttt tgaactcctc 5040
ctagacatctt actataaaact accgagcttc aatgctagac gccctccttc tgtaaaatag 5100
tcttttcctt ctaagagcat ctgccttttt tcccttaggc ttagaggata gggcccctcc 5160
atcttgctgc gacggcctta gccttgggga gtaattattg gtatccgcgt acctgtttcc 5220
cagacagccg aagtttcgac gacaaagtaa ttattgcgac aataccaccg ccatatgcta 5280
ttccgagtgg gtgagccccg aaaacatcgc ttaccgcac gccatcccag acgacagagg 5340
gcgactttga tgtcttgctc cagatcgccg cacctaacac ggtgggatgg gctggtatcg 5400
tatgggacgg catcatgggc aacaaccccc tgacgggggtg ttggggcaat ggaaacacca 5460
ccgttgcttc gagccggatc gcaaggtaag ccgaagagga caaatgacga tgagactttc 5520
tttctttttt attttatttt tttttaaatt tcttttttaa gcgtaatgaa aagagctaca 5580
tatctgtggg tcgttcctca atttcagcga cctctccacc gaagcatcgt caaataagaa 5640
gttgctcgaa acaaaggggtg tcagaagcta tagagcttct aaggatatta gccacatata 5700
tgccatagct gtataaggct atttaacgct ttggccagct cctttgtcta taaatattag 5760
tcgttttgct tcctttgtag ataattttta caaggcactc ttttccttta tatagccacc 5820
tactatagac tgctttcaac gctcccggaa gcttattact acgttcggca gttataagcc 5880
tggcgccttg actactcctc tgccgacgta tctttaatat tagtagtagc ttcttctatt 5940
acgaactctc ttaccctgct ttaatacgtc ttcgacgacg tgtctattat atctaagatc 6000
ctagtcgaga cttctatatg cttactagg cctagttctt agaacttgta gtatattaaa 6060
ctatagttat aggctaaatt tgctagtata tagagatttg ttaaccttaa tagtaattat 6120
aaactagatc tagaagtttt atagtgccta acctataaat aagctagaga taaccttatt 6180
ttagcttcct aggagtaatt cctagaagga gtattacctt taatatctat agatttgata 6240
ccttctaata tagctatcat agctaaattt atataattat aagattcctt ttataaaaat 6300
attatatata ctatagatat tagtaagtag ataggatagc tataatacta gctagtatat 6360

```

```

<210> 2
<211> 450
<212> PRT
<213> Chrysosporium lucknowense

```

```

<400> 2

```

```

Met Lys Gln Tyr Leu Gln Tyr Leu Ala Ala Thr Leu Pro Leu Val Gly

```

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

Cys Glu Thr Thr Asn Cys Gly Gly Thr Tyr Ser Glu Asp Arg Phe Ala
 260 265 270
 Gly Asp Cys Asp Ala Asn Gly Cys Asp Tyr Asn Pro Tyr Arg Met Gly
 275 280 285
 Asn Gln Asp Phe Tyr Gly Pro Gly Leu Thr Val Asp Thr Ser Lys Lys
 290 295 300
 Phe Thr Val Val Ser Gln Phe Glu Glu Asn Lys Leu Thr Gln Phe Phe
 305 310 315 320
 Val Gln Asp Gly Lys Lys Ile Glu Ile Pro Gly Pro Lys Val Glu Gly
 325 330 335
 Ile Asp Ala Asp Ser Ala Ala Ile Thr Pro Glu Leu Cys Ser Ala Leu
 340 345 350
 Phe Lys Ala Phe Asp Asp Arg Asp Arg Phe Ser Glu Val Gly Gly Phe
 355 360 365
 Asp Ala Ile Asn Thr Ala Leu Ser Thr Pro Met Val Leu Val Met Ser
 370 375 380
 Ile Trp Asp Asp His Tyr Ala Asn Met Leu Trp Leu Asp Ser Ser Tyr
 385 390 395 400
 Pro Pro Glu Lys Ala Gly Gln Pro Gly Gly Asp Arg Gly Pro Cys Pro
 405 410 415
 Gln Asp Ser Gly Val Pro Ala Asp Val Glu Ala Gln Tyr Pro Asn Ala
 420 425 430
 Lys Val Ile Trp Ser Asn Ile Arg Phe Gly Pro Ile Gly Ser Thr Val
 435 440 445
 Asn Val
 450

<210> 3
 <211> 3900
 <212> DNA
 <213> Chrysosporium lucknowense

<400> 3
 ccgcaagtga atatgtaatt actcaatgga agttctcgaa acggagtcca gaaatgatgt 60
 ggttctgtgg gaatgcggca agaggcgacg ttgccgtgaa tgcgtgaaca ttccgcctc 120
 ttcttcttct cctcttcttc cttcttcttc tttcgggtcg cggatggttg acggccagcg 180

tgcgcacggc tgcgtgttat cgagcgtcgg tacgtctagc caacatcccc tagacacgac	240
gaccaagcgt cttgagaatg caacaacgtc tcggaacctg gcacgcatct tccgccgcag	300
gtcggcagac gccgcctggg caataccacc cctgtccagg ccctttcccc gcaggcagag	360
ccgcgctctt cctttcatgg ttattcagga acgtggcttc cgagattctc gcctgttctc	420
ccccagtcaa cctgccgacc gtaaccgggt tccaccaccg cggactgtcc gcaaaacctg	480
gttcgccccga gattaatat ctatttccgg actaagtga caacacaaa gcaccccttc	540
cgcctcgcgc tctagaatct gctttctaac ccggttctcg ggcccttccc ttctcgcgacg	600
cctccgctct ccttaccagg caccatccgc aataggtaag gtagccaacc gttttggagc	660
gtgattctgc caaggaccgc atccttgcat tcgccatctg gtcaaggacc cctctttccc	720
gtccattctt ggtggctcta tcgggacggc gttcccatg gctctccagg agagtgatgt	780
gcgagtctgg agagccgggg ttggcgtcac gatgctgcc acctagggcc ggccagcccc	840
gcactgcgt cccgttgatc cgtctatccc cgtcaagagc accagccccg gcgctcgtga	900
atcttcgact tgcttcgactt gctacagggt ataaagagga tgcacgccgc cctcgatcgg	960
cctgtgtggt ttctctccct cgtgccaaac cactcccacc tcccggcccc agatagttgc	1020
ttgtttcgct ccgtgagagg gacacacacc aatggccaag aagcttttca tcaccgccgc	1080
gcttgcggt gccgtgttg cgcccccg cattgaggag cgccagaact gcggcgctgt	1140
gtggttaagaa agcccgggtc gagtctcca tgattttctc gtcgagtaat ggcataaggg	1200
ccaccccttc gactgaccgt gagaatcgat caaatccagg actcaatgcg gcggtaacgg	1260
gtggcaagggt cccacatgct gcgcctcggg ctcgacctgc gttgcgcaga acgagtggta	1320
ctctcagtgc ctgcccaaca gccagggtgac gagttccacc actccgtcgt cgacttccac	1380
ctcgcagcgc agcaccagca cctccagcag caccaccagg agcggcagct cctcctcctc	1440
ctccaccacg ccccgccccg tctccagccc cgtgaccagc attcccggcg gtgcgacctc	1500
cacggcgagc tactctggca accccttctc gggcgtccgg ctcttcgcca acgactacta	1560
cagggtccgag gtccacaatc tcgccattcc tagcatgact ggtactctgg cggccaaggc	1620
ttccgccgtc gccgaagtcc ctagcttcca gtggctcgac cggaacgtca ccatcgacac	1680
cctgatggtc cagactctgt cccagggtcc ggctctcaat aaggccgggtg ccaatcctcc	1740
ctatgctggt gagttacatg gcgacttgcc ttctcgtccc ctacctttct tgacgggatc	1800
ggttacctga cctggaggca aaacaacaac agcccaactc gtcgtctacg acctccccga	1860
ccgtgactgt gccgccgtg cgtccaacgg cgagttttcg attgcaaacg gcggcgccgc	1920
caactacagg agctacatcg acgctatccg caagcacatc attgagtact cggacatccg	1980
gatcatcctg gttatcgagc ccgactcgat ggccaacatg gtgaccaaca tgaacgtggc	2040

caagtgcagc aacgccgcgt cgacgtacca cgagttgacc gtgtacgcgc tcaagcagct	2100
gaacctgccc aacgtcgcca tgtatctcga cgccggccac gccggctggc tcggctggcc	2160
cgccaacatc cagcccgcg ccgagctggt tgccggcatc tacaatgatg ccggcaagcc	2220
ggctgccgtc cgcggcctgg ccactaacgt cgccaactac aacgcctgga gcatcgcttc	2280
ggccccgtcg tacacgtcgc ctaaccctaa ctacgacgag aagcactaca tcgaggcctt	2340
cagcccgtc ttgaactcgg ccggcttccc cgcacgcttc attgtcgaca ctggccgcaa	2400
cggcaaacia cctaccggtg tgtttttttt tcttttgtct ctgtcccccc cttttctccc	2460
ccttcagttg gcgtccacaa ggtctcttag tctgtcttca tctgtgacca acctcccccc	2520
ccccggcacc gccacaaacc gtttgactct atactcttgg gaatgggagc cgaaactgac	2580
cgttccacag gccaaacaaca gtggggtgac tgggtgcaatg tcaagggcac cggctttggc	2640
gtgcgcccga cggccaacac gggccacgag ctggctgatg cttttgtctg ggtcaagccc	2700
ggcggcgagt ccgacggcac aagcgacacc agcgccgccc gctacgacta ccactgcggc	2760
ctgtccgatg ccctgcagcc tgccccgag gctggacagt ggtccaggc ctacttcgag	2820
cagctgctca ccaacgcaa cccgcccttc taaacctcgt cataaagaga gagagatggc	2880
gggcatgggc ctgattgggt tcattgacca tgcggctctt ctgggggtac atattttacc	2940
tacctaccta taaataaggc ggcctatcgg gctctcgctt cgttttattag gtacttgctc	3000
ttgtacatac tttgtttata catacagcag ttagcatcca ctattcgttt cgacaaagcg	3060
gaactttcca gaaaaaaaaa ggttgtagat aattagtctt taggcttcga ttctttgtgc	3120
ctttcttttt ggtaaaaaaa aaattttttt tgaggcatga ttaccttagg tacgttcgtc	3180
gttgatttgg tccccctgca ttttggcgcg agagcagctc agccccttgc aaatccctca	3240
acgggcgttc aattccctcc actcgggtct tcagcgagac cagccgtcca gagtatccca	3300
gcgtgtagtt gccccacgaa ccagtcgtcc tcgtaagcct cgtcaaagtg tccaagagca	3360
gtatagaagc aacgacctcc gtcaaaagtc tggcaccatg cgatcgggtg gtcctccccg	3420
tgcgccccgc cctcgtagga cttctcatcc acgccaagga gcacgtgcag gccgtcggac	3480
gtcgcgccgc ggtgcgcctt gaagttgtac cattcgtcct tccagacgcg ctccagctgc	3540
gcctgcttgg gttcctgcgg ttctctgcgg tcctgcgctg gccggtcggc gccgccgtct	3600
tggtcacacg cccgcagcga catgactggg tgtttcgggt cgagcagctt gacgagccccg	3660
acctgggggt ccgggtgggt gtcgaacacg gcgccaatga ggtggccgta ccattcggtat	3720
gactgcatgg cgaagctggc gcagtgtacc gccacgatcc cgccgccccg ctggacgaaa	3780
ccccgcaggg cgcccagctg cgcgccgtcc aggaactcgc ccgagcactg caggaggagc	3840
atgacgcgat acgccgagag ggagccgggg ctgaacacgg cgggatcctc gctgtcgtcc	3900

<210> 4
 <211> 481
 <212> PRT
 <213> Chrysosporium lucknowense

<400> 4

Met Ala Lys Lys Leu Phe Ile Thr Ala Ala Leu Ala Ala Ala Val Leu
 1 5 10 15

Ala Ala Pro Val Ile Glu Glu Arg Gln Asn Cys Gly Ala Val Thr Gln
 20 25 30

cys Gly Gly Asn Gly Trp Gln Gly Pro Thr cys cys Ala Ser Gly Ser
 35 40 45

Thr cys Val Ala Gln Asn Glu Trp Tyr Ser Gln Cys Leu Pro Asn Ser
 50 55 60

Gln Val Thr Ser Ser Thr Thr Pro Ser Ser Thr Ser Thr Ser Gln Arg
 65 70 75 80

ser Thr Ser Thr Ser Ser Ser Thr Thr Arg Ser Gly ser ser Ser Ser
 85 90 95

ser Ser Thr Thr Pro Pro Pro Val Ser Ser Pro Val Thr Ser Ile Pro
 100 105 110

Gly Gly Ala Thr Ser Thr Ala Ser Tyr Ser Gly Asn Pro Phe ser Gly
 115 120 125

Val Arg Leu Phe Ala Asn Asp Tyr Tyr Arg Ser Glu Val His Asn Leu
 130 135 140

Ala Ile Pro Ser Met Thr Gly Thr Leu Ala Ala Lys Ala Ser Ala Val
 145 150 155 160

Ala Glu Val Pro Ser Phe Gln Trp Leu Asp Arg Asn Val Thr Ile Asp
 165 170 175

Thr Leu Met Val Gln Thr Leu Ser Gln Val Arg Ala Leu Asn Lys Ala
 180 185 190

Gly Ala Asn Pro Pro Tyr Ala Ala Gln Leu Val Val Tyr Asp Leu Pro
 195 200 205

Asp Arg Asp Cys Ala Ala Ala Ala Ser Asn Gly Glu Phe Ser Ile Ala
 210 215 220

Asn Gly Gly Ala Ala Asn Tyr Arg Ser Tyr Ile Asp Ala Ile Arg Lys
 225 230 235 240
 His Ile Ile Glu Tyr Ser Asp Ile Arg Ile Ile Leu Val Ile Glu Pro
 245 250 255
 Asp Ser Met Ala Asn Met Val Thr Asn Met Asn Val Ala Lys Cys Ser
 260 265 270
 Asn Ala Ala Ser Thr Tyr His Glu Leu Thr Val Tyr Ala Leu Lys Gln
 275 280 285
 Leu Asn Leu Pro Asn Val Ala Met Tyr Leu Asp Ala Gly His Ala Gly
 290 295 300
 Trp Leu Gly Trp Pro Ala Asn Ile Gln Pro Ala Ala Glu Leu Phe Ala
 305 310 315 320
 Gly Ile Tyr Asn Asp Ala Gly Lys Pro Ala Ala Val Arg Gly Leu Ala
 325 330 335
 Thr Asn Val Ala Asn Tyr Asn Ala Trp Ser Ile Ala Ser Ala Pro Ser
 340 345 350
 Tyr Thr Ser Pro Asn Pro Asn Tyr Asp Glu Lys His Tyr Ile Glu Ala
 355 360 365
 Phe Ser Pro Leu Leu Asn Ser Ala Gly Phe Pro Ala Arg Phe Ile Val
 370 375 380
 Asp Thr Gly Arg Asn Gly Lys Gln Pro Thr Gly Gln Gln Gln Trp Gly
 385 390 395 400
 Asp Trp Cys Asn Val Lys Gly Thr Gly Phe Gly Val Arg Pro Thr Ala
 405 410 415
 Asn Thr Gly His Glu Leu Val Asp Ala Phe Val Trp Val Lys Pro Gly
 420 425 430
 Gly Glu Ser Asp Gly Thr Ser Asp Thr Ser Ala Ala Arg Tyr Asp Tyr
 435 440 445
 His Cys Gly Leu Ser Asp Ala Leu Gln Pro Ala Pro Glu Ala Gly Gln
 450 455 460
 Trp Phe Gln Ala Tyr Phe Glu Gln Leu Leu Thr Asn Ala Asn Pro Pro
 465 470 475 480

Phe

<210> 5
 <211> 1648
 <212> DNA
 <213> Chyrsosporium lucknowense

<220>
 <221> misc_feature
 <222> (812)..(812)
 <223> n is a, c, g, or t

<220>
 <221> misc_feature
 <222> (1162)..(1162)
 <223> n is a, c, g, or t

<400> 5
 atgtacgccca agttcgcgac cctcgccgcc cttgtggctg gcgccgctgc tcagaacgcc 60
 tgcactctga ccgctgagaa ccacccctcg ctgacgtggc ccaagtgcac gtctggcggc 120
 agctgcacca gcgtccaggg ttccatcacc atcgacgccca actggcgggtg gactcaccgg 180
 accgatagcg ccaccaactg ctacgagggc aacaagtggg atacttcgta ctgcagcgat 240
 ggtccttctt gcgcctccaa gtgctgcac gacggcgctg actactcgag cacctatggc 300
 atcaccacga gcggtaactc cctgaacctc aagtctgtca ccaagggcca gtactcgacc 360
 aacatcggct cgcgtaacta cctgatggag agcgacacca agtaccagag taagttcctc 420
 tcgcacccgg ccgcccgggag atgatggcgc ccagcccgtg gacgcgaatg acacagtgtt 480
 ccagctcctc ggcaacgagt tcaccttcga tgtcgacgtc tccaacctcg gctgcggcct 540
 caatggcgcc ctctacttcg tgtccatgga tgccgatggc ggcattgtcca agtactcggg 600
 caacaaggca ggtgccaaagt acggtaccgg ctactgtgat tctcagtgcc cccgcgacct 660
 caagttcatc aacggcgagg ccaacgtaga gaactggcag agctcgacca acgatgccaa 720
 cgccggcacg ggcaagtacg gcagctgctg ctccgagatg gacgtctggg aggccaacaa 780
 catggccgcc gccttcactc cccacccttg cnccgtgatc ggccagtcgc gctgcgaggg 840
 cgactcgtgc ggcggtacct acagcaccga ccgctatgcc ggcattctgcg accccgacgg 900
 atgcgacttc aactcgtacc gccagggcaa caagaccttc tacggcaagg gcatgacggc 960
 cgacacgacc aagaagatca cggtcgtcac ccagttcctc aagaactcgg ccggcgagct 1020
 ctccgagatc aagcggttct acgtccagaa cggcaaggct atccccaaact ccgagtccac 1080
 catcccgggc gtcgagggca actccatcac ccaggactgg tgcgaccgcc agaaggccgc 1140
 cttcggcgac gtgaccgact tncaggacaa gggcggcatg gtccagatgg gcaaggccct 1200
 cgcgggggccc atggtcctcg tcatgtccat ctgggacgac cacgccgtca acatgctctg 1260

```

gctcgactcc acctggccca tcgacggcgc cggcaagccg ggcgccgagc gcggtgcctg 1320
ccccaccacc tcggggcgctc ccgctgaggt cgaggccgag gcccccaact ccaacgtcat 1380
cttctccaac atccgcttcg gcccacatcg ctccaccgtc tccggcctgc cgcacggcgg 1440
cagcggcaac cccaaccgc ccgtcagctc gtccaccccg gtccctcctt cgtccaccac 1500
atcctccggt tcctccggcc cgactggcgg cacgggtgtc gctaagcact atgagcaatg 1560
cggaggaatc gggttcactg gccctaccca gtgcgagagc ccctacactt gcaccaagct 1620
gaatgactgg tactcgcagt gcctgtaa 1648

```

```

<210> 6
<211> 526
<212> PRT
<213> Chrysosporium lucknowense

```

```

<220>
<221> misc_feature
<222> (249)..(249)
<223> Xaa pode ser qualquer aminoácido que ocorre naturalmente

```

```

<220>
<221> misc_feature
<222> (365)..(365)
<223> Xaa pode ser qualquer aminoácido que ocorre naturalmente

```

```

<400> 6

```

```

Met Tyr Ala Lys Phe Ala Thr Leu Ala Ala Leu Val Ala Gly Ala Ala
1          5          10          15

```

```

Ala Gln Asn Ala Cys Thr Leu Thr Ala Glu Asn His Pro Ser Leu Thr
          20          25          30

```

```

Trp Ser Lys Cys Thr Ser Gly Gly Ser Cys Thr Ser Val Gln Gly Ser
          35          40          45

```

```

Ile Thr Ile Asp Ala Asn Trp Arg Trp Thr His Arg Thr Asp Ser Ala
          50          55          60

```

```

Thr Asn Cys Tyr Glu Gly Asn Lys Trp Asp Thr Ser Tyr Cys Ser Asp
          65          70          75          80

```

```

Gly Pro Ser Cys Ala Ser Lys Cys Cys Ile Asp Gly Ala Asp Tyr Ser
          85          90          95

```

```

Ser Thr Tyr Gly Ile Thr Thr Ser Gly Asn Ser Leu Asn Leu Lys Phe
          100          105          110

```

```

Val Thr Lys Gly Gln Tyr Ser Thr Asn Ile Gly Ser Arg Thr Tyr Leu

```

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

Gly Gly Met Val Gln Met Gly Lys Ala Leu Ala Gly Pro Met Val Leu
370 375 380

Val Met Ser Ile Trp Asp Asp His Ala Val Asn Met Leu Trp Leu Asp
385 390 395 400

Ser Thr Trp Pro Ile Asp Gly Ala Gly Lys Pro Gly Ala Glu Arg Gly
405 410 415

Ala Cys Pro Thr Thr Ser Gly Val Pro Ala Glu Val Glu Ala Glu Ala
420 425 430

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

Ser Thr Val Ser Gly Leu Pro Asp Gly Gly Ser Gly Asn Pro Asn Pro
450 455 460

Pro Val Ser Ser Ser Thr Pro Val Pro Ser Ser Ser Thr Thr Ser Ser
465 470 475 480

Gly Ser Ser Gly Pro Thr Gly Gly Thr Gly Val Ala Lys His Tyr Glu
485 490 495

Gln Cys Gly Gly Ile Gly Phe Thr Gly Pro Thr Gln Cys Glu Ser Pro
500 505 510

Tyr Thr Cys Thr Lys Leu Asn Asp Trp Tyr Ser Gln Cys Leu
515 520 525

<210> 7

<211> 4376

<212> DNA

<213> Chrysosporium lucknowense

<400> 7

ggatccacac ctaccatacc ggatagtatg ctaccaagt gacatagggt tggtaaagta	60
atacgagaac tcagagagca ctgcccatat ggctcgccaa tgacctcaag tgccagggtca	120
gctttgcgag acagacctga ggcgctcgga tgtgtgacat ggaacgcgcc ggatcgccctt	180
gttgattaat tatagggaag tagcgaggaa ggtttcagca attgacgtga gcgtacatta	240
aaagctgtat gatttcagga agacgagcca tggaccaggt ttcaaggctg aatggcttga	300
cgacttaagc accgaacgag gaatgaaaga atgaaaagtg ggggatcatt ctggcccctc	360
ctcgtatgtc gagtgttaaa gaaggcggtt ctacggagga cctaaagagc tccaatttgc	420
tctgttgagc ttaagccaca tatctcaaga tgaatacatg tcaggcatag tcaccctgat	480
cttggttcac agtccacaca cttttcagtt cagcatgttg attcctcatc catatcactt	540

tccattacta tctcttttatg tccttgggtca agactccaag gaaccgatag gtgagcatcg	600
gtgagggtcc ctcaagggtac caaagtagcc atcatcaccg aggtctggga atggcgccgt	660
gcccgatctg agtcctccaa ctccacggta cgacgacagc acgtcacatt gacgcaccac	720
ggttgaacaa gcagagaggg acacgtcttg ctacgcgaat cctggcactg gatggagacg	780
cgtgtgagca ggtttccgga accatgacgg cctgggtccgg cttctcgaac aaagaagtgg	840
aacacaaaaa gaaccgaaac ggaaacgcag gcacggcatc gacgaccgga ttgtcccacg	900
gggacctcgg ccagtcaagc gttgccctgg ccgtcagctc cctggcgacg gggattcagc	960
acatctcacg ttataggcga cctcatcccc cttccgtctt gtgcggtcgt tgctccgtgc	1020
cgagtacca ggcgtgccgg ggcctttagc cggggcgga tcagagtcaa gatgcggccg	1080
aattggacgg cagacgaagt ttcgtagagg gtcattgatc gactgacga caccaccccc	1140
tgcgtgatcc cgtggccctg ggctgggaat tgccggctaa taatctacgg cttaatagat	1200
atgcactttg cacgcggtgc agataaataa gctgtggttt caaacactgg cctccgtact	1260
ttaccaccca actgccgctt agcgcgggga cctgagtctt gggagtgcgc ggagcggcag	1320
ccacctcggg ttagcgtaca cacgacggct gcatgcgggg atgccgcgtg catggcttca	1380
tagtgtacga cagaccgtca agtccaaatc tgggtgatgc ttgatgagat gacagcgagc	1440
cccgtcggcg gcaccccggc tatgcatcgc gaattgacaa cactctcagc tctattgcga	1500
cccatcggat aaaagaagaa gaaaaaatg gaccttgagt acgggcgtca gaaacaaaaa	1560
aaaaactccg gaaccaaata tgtcgggcat ggccgggggtg aacgaccgct actccccgtt	1620
cccttcttcg caaacagaac gctacagagg gttttctggt ttgtcaaaga gttcggaggt	1680
cctctgctcc gcgaatgcgt ggtgaacca ccagcagcca ttgttcttgc atgcgtggcg	1740
gaccgttagc cgctgatcga catggcgagc ttccacctc agacctggag cagacggttg	1800
cgaggagcaa ggggtgccc tccccctgac ggtcggaccc caatgacttc cccaaacggg	1860
gacatcgagg gtcgtgcatg atggtggaaa gtagttgcag tatgggaagt accccggggt	1920
gccaggaacc gttgttcggc cccccacatt ttctctctgc catgtcaact gtgtgtcgtt	1980
cgagagttcc tggtccggc cccccgtcca attccctaac gggaccgcgg ggcacgcct	2040
gtaactaact tccaaatgaa gccggatatg agggagggag attggatctg gcaagccagc	2100
cattcgctgc gatcggcact cgtccgtcag cccgcagtc catatcccca aaggcaactg	2160
ctcggcgcgg ctcaagtctt cttcggaacg tccagcccga aggcgcgcgc cagcaccggc	2220
cctatgttcc tgattgcgat cctcgatctc cagagacggg tcacctcgcc tcgaggacgg	2280
tgcaggggca tcggcttcgc ttcctagagc tccgggctgt gtgtggtcaa ggggagaagg	2340
cggcggcgcc aaggtgcgtc tcggcgact caccatcgc ctttaccccc ctcccccca	2400

gtatataaaa gatggccatc gtctcctcgt ctgcttgagg	2460
gcaccacagc ctagctctaa cccagcttgt cgtgtgttgt	2520
cagtccgcca ccctggcggt cgccgccacg gccctcgctg	2580
cagaagcccc gccaggcctc ggcgggctgc gcgtcggccg	2640
aacgtgttcc agcagtacac gctgcacccc aacaacttct	2700
gccgccgagg ccatctccga ctcggcgctg gccgagaagg	2760
ggtaccttcc tgtggctcga caccatcgag aacattggcc	2820
gacgtgccct gcgagaacat cgtgggtctc gtcatctacg	2880
gcggccaagg cctccaacgg cgagctcaag gtcggcgagc	2940
tacatcgaca gtgagttaac cttttgtggc cccttctttt	3000
tgagtgggggt tgtgagagag aaaatggggc gagcttaaag	3060
agatcgccga gatcctcaag gccactcca acacggcctt	3120
actcgctccc caacctggtc accaatagcg acctgcagac	3180
gctaccgcca ggggtgtcgc tatgccctca agcagctcaa	3240
acatcgatgc cggccacggg ggctggctcg gctgggacgc	3300
aggagctcgc cagcgtctac aagtctgctg gttcgccctc	3360
ccaacgtggc tggttggaac gcctggtaag acactctatg	3420
cgagcggaat ggcgtgaaat gcatggtgct gacctttgat	3480
accaggagcc cgggtgagttc tcggacgcct cggatgccca	3540
agaagatcta catcaacacc tttggcgctg agctcaagtc	3600
ccatcatcga cactggccgc aacgggtgtca ccggctctccg	3660
gcaacgtcaa cggcgccggc ttcgggtgtgc gcccgactgc	3720
ccgacgcctt cgtgtgggtc aagcccgggt gcgagtcgga	3780
cggcgcgcta cgacagcttc tcgggcaagc ccgacgcctt	3840
gtacctggaa ccaggcctac ttcgagatgc tcctcaagaa	3900
ctcctcgacg gcttcttgct gtcagtcgct ctgacgggtg	3960
tcctgctgct gctgctccgc ggggagggga ggcaacgaaa	4020
aaaacagaaa caagcgaggc gcgggtgcaat ggtcgtgcgt	4080
ttctagtgtg tagttttgat agtcgtacat aaggggtttc	4140
gtctttttgc gagttgttgc gactcgtgat tatggccttt	4200
agaaccacag cgtgttgggg tagcagcttg ctccgtagga	4260
ctctggaatt gcagtcagcc tgcgtcgccc ctctaggaaa	4320

ggctgcagct tacaaacgcg agcatggtga acatctccga gaaaaggag ggatcc

4376

<210> 8
 <211> 395
 <212> PRT
 <213> Chrysosporium lucknowense

<400> 8

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

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

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

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

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

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

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

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

Ser Asn Gly Glu Leu Lys Val Gly Glu Leu Asp Arg Tyr Lys Thr Glu
 130 135 140

Tyr Ile Asp Lys Ile Ala Glu Ile Leu Lys Ala His Ser Asn Thr Ala
 145 150 155 160

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

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

Val Ala Tyr Ala Leu Lys Gln Leu Asn Leu Pro Asn Val Val Met Tyr
 195 200 205

Ile Asp Ala Gly His Gly Gly Trp Leu Gly Trp Asp Ala Asn Leu Lys
210 215 220

Pro Gly Ala Gln Glu Leu Ala Ser Val Tyr Lys Ser Ala Gly Ser Pro
225 230 235 240

Ser Gln Val Arg Gly Ile Ser Thr Asn Val Ala Gly Trp Asn Ala Trp
245 250 255

Asp Gln Glu Pro Gly Glu Phe Ser Asp Ala Ser Asp Ala Gln Tyr Asn
260 265 270

Lys Cys Gln Asn Glu Lys Ile Tyr Ile Asn Thr Phe Gly Ala Glu Leu
275 280 285

Lys Ser Ala Gly Met Pro Asn His Ala Ile Ile Asp Thr Gly Arg Asn
290 295 300

Gly Val Thr Gly Leu Arg Asp Glu Trp Gly Asp Trp Cys Asn Val Asn
305 310 315 320

Gly Ala Gly Phe Gly Val Arg Pro Thr Ala Asn Thr Gly Asp Glu Leu
325 330 335

Ala Asp Ala Phe Val Trp Val Lys Pro Gly Gly Glu Ser Asp Gly Thr
340 345 350

Ser Asp Ser Ser Ala Ala Arg Tyr Asp Ser Phe Cys Gly Lys Pro Asp
355 360 365

Ala Phe Lys Pro Ser Pro Glu Ala Gly Thr Trp Asn Gln Ala Tyr Phe
370 375 380

Glu Met Leu Leu Lys Asn Ala Asn Pro Ser Phe
385 390 395

<210> 9

<211> 5777

<212> DNA

<213> Chrysosporium lucknowense

<400> 9

tgctgctctg atgtgctgat gcacagcttc ccctcgcgat tgccggcagg atctccaacc 60

ctctggatcg gagcagacga tcagcgggca caatggccag cttgccagcg ttcaactcca 120

agttgacccg cttttatcac gcccaagctg gacatgcaca ggcttggtt ctcgtgttcc 180

tacgatctgc acagtaggtt tgactgctga tcttcgcttt cctgtgcgcc ctccccctcc 240

ctcacgggta ccttatcctt gcctgtaacc ccgcgttatg tcaaacttga gtttgaccaa 300

tgctagcgca aaagtaccta catagtacta tgtaataagg taggtacata catcagtagg	360
cgtttatcta gtaaatTTTtg gctTTTTgaa actcaattgc tcctctcctc gcctccacct	420
ctgcttggca attgacaacc ctggctgtgc ctagaggtag catcgacgat caatcaaadc	480
taaagtattc gagattgacc tttctgtctt aattatatta attatccgca caatgctgta	540
gtcattgact ctcttttcaa gttgccttct cgtttatgta tgtacaatgg gcgggtcatgc	600
ttcatgccaa cagatgggtc tatcggaaca atgtttgact ttctggtcgc cccgtcgaac	660
tgTTTTgatt tcgcacggga agtgttctta ccaaagctaa gtcgactcgt ggagcttcgt	720
aacggccagt gatcgttgat cgcttttggga ggagttgcga tggagcgaga ccggctacga	780
gcacgttcgc aaaggcagca cgatagacga ccctccgtgg cgccattcgg gagatgcaca	840
tgacataagc atatcaatac tcacctgaac tcatcggccg atgcctcgca ggtagttaca	900
agacatatTT gtgtgggtat attatcccaa cccgtacctt tgtcgcgtca ttctggatatg	960
tgctgatgcc tacttaggga gcaaagacgc ctctcctcac ctgcgggtta cttacttact	1020
gtgcagcatg gccttatgtt ctcccgggtc ttgcttgcgc gaatgaacaa aaacgcccga	1080
agaaaagccg cttcttcgag ttgtgtctac ccgaacataa gaggttattg tcgcagaccg	1140
ccagcaaatg tcaacaaccc acccacggcg ttccagaacc ttcgaaatat catctagttt	1200
aagtttaaat gacggcccga gtcccagccg agattcccat attggccgat accagcggtc	1260
ccttgTTTT ccaaggttgt ctctgcaact ggcgcatctg cctacaacga gatataatta	1320
ccgTTTTctt ttgcaaaagg gcatgcatgg atgtatatta tttatgcctg cagaacgaga	1380
agcaatcatg gtgtaggttt tgtgcggtat ggagctaata atattgaacg gatctctggt	1440
ccgtcctaaa tcgttgaaac gctaggccca ggaggacctg ctcgacttgg cgaacggaga	1500
tttccaggat gaaaggtcgg aacatgtcca tccgcggcca gcctgaacac ttttgctcgt	1560
ttccggacca tcgaccacg aaaacagtgc ggttgctggc acagtcagca ctacgatgg	1620
cgatggtcca gcccgttccc gcccgatgcc cacttgcagc gcaactctcc ttcattcggc	1680
ggcccggcgg tgtctggcct attagtacga ttttgatac cggttggtc gccgccgcgg	1740
tttttcttgg ccgatacggg aatctcggtg gtcccaactc cacctgggca cgctctggtg	1800
ccaacatgga acttcgggat gccgctccgg gcacagtcaa gcgctttaa atacgacttt	1860
acccacaag aatcgaggcg taaccggaa ttagggacac ctggacggcg caaccctgg	1920
accgaagggc ctgctaacc gggttcctgg agccgcatgc gcggctgccc gcttgcccgc	1980
tcttgagatg acacttcttt tcagcgaggg atggctgggc agggaaatga tgtattataa	2040
gaagcgagcc gattccgaag gactcgacct cctctctcgc cctgtgtccg ccagctaatt	2100
acagcactcc ttctcgactt gaaacgccc agatgaagtc ctccatcctc gccagcgtct	2160

tcgccacggg	cgccgtggct	caaagtggtc	cgtaggcagca	atgtggtggc	atcggatggc	2220
aaggatcgac	cgactgtgtg	tcgggttacc	actgctgcta	ccagaatgat	tggtagagcc	2280
agtgcgtgcc	tggcgcggcg	tcgacaacgc	tccagacatc	taccacgtcc	aggcccaccg	2340
ccaccagcac	cgcccctccg	tcgtccacca	cctcgcctag	caagggcaag	ctcaagtggc	2400
tcggcagcaa	cgagtcgggc	gccgagttcg	gggagggcaa	ctaccccggc	ctctggggaa	2460
agcacttcat	cttcccgtcg	acttcggcga	ttcaggtacg	ggccaataat	aatatattat	2520
tatagcaggc	aggagggagc	aggagaagaa	gggaggggca	ggtggccaac	aatcgggaaga	2580
agaccgggag	gcactgaccg	ttgattcctt	tgtgtaatag	acgctcatca	atgatggata	2640
caacatcttc	cggatcgact	tctcgatgga	gcgtctgggtg	cccaaccagt	tgacgtcgtc	2700
cttcgacgag	ggctacctcc	gcaacctgac	cgaggtggtc	aacttcgtga	cgaacgcggg	2760
caagtacgcc	gtcctggacc	cgcacaacta	cggccggtag	tacggcaacg	tcatacagga	2820
cacgaacgcg	ttccggacct	tctggaccaa	cctggccaag	cagttcgcct	ccaactcgct	2880
cgtcatcttc	gacaccaaca	acgagtacaa	cacgatggac	cagaccctgg	tgctcaacct	2940
caaccaggcc	gccatcgacg	gcatccgggc	cgccggcgcg	acctcgcagt	acatcttcgt	3000
cgagggcaac	gcgtggagcg	gggcctggag	ctggaacacg	accaacacca	acatggccgc	3060
cctgacggac	ccgcagaaca	agatcgtgta	cgagatgcac	cagtacctcg	actcggacag	3120
ctcgggcacc	cacgccgagt	gcgtcagcag	caacatcggc	gcccagcgcg	tcgtcggagc	3180
caccagtggt	ctccgcgcca	acggcaagct	cggcgtcctc	ggcgagttcg	ccggcggcgc	3240
caacgccgtc	tgccagcagg	ccgtcaccgg	cctcctcgac	cacctccagg	acaacagcga	3300
ggtctggctg	ggtgccctct	ggtgggcccgc	cggccctggg	tggggcgact	acatgtactc	3360
gttcggtaag	tttctccctt	gttcttggtt	ttccccccag	taaggagtc	aggcaacatg	3420
cccaagaccg	gctcggcttc	gcttcaaggc	gttcgttgta	cacactgaag	agttccaact	3480
tccaaccctg	ttcgtgtcct	ccgatcagct	tcgacggggt	gaagggggaa	gggatttggg	3540
agtgaggtgg	aggtcaaaaag	gagggatata	cccagatctc	cacaaacggc	cctgagccaa	3600
caacagcctc	tgggggtcaaa	atgggcgcca	accatacggg	cattcactca	ggacacctgc	3660
taacgcgtct	cttttttttg	tttcagagc	ctccttcggg	caccggctat	gtcaactaca	3720
actcgatcct	aaagaagtac	ttgccgtaag	gggcatgcag	caaggtcgag	cgagcattat	3780
tcagggccat	ctgcttgtgt	cggcaggcat	cacgtcaacc	catcgaatcg	gacagcggaa	3840
tgctccgaga	tgccatacac	taagtctggt	gatgacgtga	gaatgctggc	cctggtcggg	3900
ggttaccgcc	aacaaaaagc	acccggacgc	tgccgcgccc	ggataccatg	gtttcatgta	3960
catattgggt	ctttgctttc	ttacgggggg	gggggggggg	gggggctctg	cagcgttgct	4020
gaqcgattcg	tttccaagta	tatactttgt	ctggaattga	attttgagtg	acattgaccc	4080

```

aatcaaccag ctcggtgtgc tcacctcccg ttaccccccc tcttctcccc ctgctcggct 4140
tggcttttct ctccggtgtg gagcacggcc acggcggtcc caatccatat aagatcgatg 4200
gtatactatg gtatacacta gcttggaat aaactaatcc atacgctaac taatggacgg 4260
attatcctaa gggtcaccgg ctcaccgttg gatataacac ctaggatacg ggagagctga 4320
tagaaaggga tgtactccgt attgtactgt acaatacaaa gtacagatag cacacgaagt 4380
acggtaggtg gtcccgcta gtccggacca acaatagaac atgcgttcct ggggacctgc 4440
aggaaagaag gggggggggg ttgccaagac gcccggggtt caaagaaagc cccgggcccgc 4500
cgatgagatg agacggacgc cggcccaagg agaggccggt ggtcgatcct gcaaagcca 4560
gcaaaaaaaaa tccataccat aatccagtca actttcgtca cactcctgtg aaacgagctg 4620
gagggactgc tggaaggtt ttgcaggta atcactgtat gtggagcatg ccgtacctac 4680
tgtgcttcgt taacagatag agttccagtt gaacacacaa agttctgccc cgcctgccag 4740
acgtgaaaag aagctcctcc gggggagctt taggcaactg ggagggctct ctcccagggt 4800
catggtgtct gctcttcttc aaatttttat gctgccacc catttgacag aggtgtgcac 4860
accgttgcca ggtcttgcca tccggcaaaa agcagaaaag tcgaccatc gcctaagaaa 4920
ggcggtcgga aggggatcgg atgctcattg cggcttagcg tctgcccatt ctgacgctgc 4980
ccattgtttt gtgtcgcat cgtcttcgga tgtcggatca agagtcccgg attttttccc 5040
ctgtgcttcc agcctaattc gagcgggagc tggctcgggt tcgagtggag ttgccttgtt 5100
ggtggagcag caaccagcca attcactccc ccgcattttc gcggccgccc aggcattccc 5160
ggcatgcgtt tgggcggtaa ctactccgta ctggggtagg tgaaattgggt tctcccgtcg 5220
caggaggctc gtgctcggtc aggggagaac aaagtccaac tgctccttcc tggcaacaat 5280
gagagggggg tctattgcca acgttgacg aaaggagcag ccacaaaacc caaaagcagg 5340
ttaccttact gtacctgagc ttgaacgtcg cgtagcattg gagctctcgt ctaccggcgg 5400
cgtcacactc cattggcagg tcaaggcagt cagtggcagc gaccaacaa cgtcaatgct 5460
tgttacccca gaattacccc gggctgcaac actgcagggg ccgccgccga tgttgatcac 5520
cggttgatta cttctcgcc cgcaaccggg agatgagaag cagaactttg ttctccttcc 5580
aaaaaggacc tgacttgagg ggaacgcact gccggcagtg gagtggatgc acgctagtta 5640
tatgtttccc gccatcccca gtccgccgt cgcgctcgtg aggctcagtt tggcttcccg 5700
tgccgccgac aaacgagcgg tgcataatta catttcgctc catgtaccgt gcaccctccc 5760
cgttcgcgac cgtagta 5777

```

```

<210> 10
<211> 389
<212> PRT

```

<213> Chrysosporium lucknowense

<400> 10

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

Gln Ser Gly Pro Trp Gln Gln Cys Gly Gly Ile Gly Trp Gln Gly Ser
20 25 30

Thr Asp Cys Val Ser Gly Tyr His Cys Val Tyr Gln Asn Asp Trp Tyr
35 40 45

Ser Gln Cys Val Pro Gly Ala Ala Ser Thr Thr Leu Gln Thr Ser Thr
50 55 60

Thr Ser Arg Pro Thr Ala Thr Ser Thr Ala Pro Pro Ser Ser Thr Thr
65 70 75 80

Ser Pro Ser Lys Gly Lys Leu Lys Trp Leu Gly Ser Asn Glu Ser Gly
85 90 95

Ala Glu Phe Gly Glu Gly Asn Tyr Pro Gly Leu Trp Gly Lys His Phe
100 105 110

Ile Phe Pro Ser Thr Ser Ala Ile Gln Thr Leu Ile Asn Asp Gly Tyr
115 120 125

Asn Ile Phe Arg Ile Asp Phe Ser Met Glu Arg Leu Val Pro Asn Gln
130 135 140

Leu Thr Ser Ser Phe Asp Glu Gly Tyr Leu Arg Asn Leu Thr Glu Val
145 150 155 160

Val Asn Phe Val Thr Asn Ala Gly Lys Tyr Ala Val Leu Asp Pro His
165 170 175

Asn Tyr Gly Arg Tyr Tyr Gly Asn Val Ile Thr Asp Thr Asn Ala Phe
180 185 190

Arg Thr Phe Trp Thr Asn Leu Ala Lys Gln Phe Ala Ser Asn Ser Leu
195 200 205

Val Ile Phe Asp Thr Asn Asn Glu Tyr Asn Thr Met Asp Gln Thr Leu
210 215 220

Val Leu Asn Leu Asn Gln Ala Ala Ile Asp Gly Ile Arg Ala Ala Gly
225 230 235 240

Ala Thr Ser Gln Tyr Ile Phe Val Glu Gly Asn Ala Trp Ser Gly Ala
245 250 255

Trp Ser Trp Asn Thr Thr Asn Thr Asn Met Ala Ala Leu Thr Asp Pro
260 265 270

Gln Asn Lys Ile Val Tyr Glu Met His Gln Tyr Leu Asp Ser Asp Ser
275 280 285

Ser Gly Thr His Ala Glu Cys Val Ser Ser Asn Ile Gly Ala Gln Arg
290 295 300

Val Val Gly Ala Thr Gln Trp Leu Arg Ala Asn Gly Lys Leu Gly Val
305 310 315 320

Leu Gly Glu Phe Ala Gly Gly Ala Asn Ala Val Cys Gln Gln Ala Val
325 330 335

Thr Gly Leu Leu Asp His Leu Gln Asp Asn Ser Glu Val Trp Leu Gly
340 345 350

Ala Leu Trp Trp Ala Ala Gly Pro Trp Trp Gly Asp Tyr Met Tyr Ser
355 360 365

Phe Glu Pro Pro Ser Gly Thr Gly Tyr Val Asn Tyr Asn Ser Ile Leu
370 375 380

Lys Lys Tyr Leu Pro
385

<210> 11
<211> 6060
<212> DNA
<213> Chrysosporium lucknowense

<400> 11
ccggcctcca gttccaggag cttggctctg ccgacatact gtgtacacta ggaattctct 60
tatgcgggggt gtgcgcgggg aaatgttggg gaactcgagt tgggtcatgt ggacaagacc 120
aatgggagct gacatcattg tgcgaccctg taaaccggaa gctacaacaa cattctggat 180
tctacactag tggaagaggt aagtaattga cgacaagcaa gaagcattgc catgttctgc 240
gaaggatgcg ggtgtttttg catgagcagg aagctgtggc tttttagtgc tcctttgtgc 300
tcgcccggcg cgcagaacac taccgaaacg caggggactg cgtgcctctg gggtcgaatg 360
ccgatcccca tcttcacatt cccaccatcg tgttctgtta acgaagccgg agcggcgggg 420
actcgaagct ccactacgta tggatacttg ggaccgtacg gagtgtgttg gtacggatgc 480

ctgcacaagt gttgtgcttc ctacgaagac gccaacccac ataatacaca aaagctgttg	540
taagtcgagt tacctcaggc acgttcgggc aactcgggca acctgacgag atttccccgc	600
cattccgcca agaggccggc gcctgccctg attaggcagc tcttggaaca atactatgta	660
gaatggaagc tccatccata gtcagctcca ttggcggtcc cagtgatctc gatggctgga	720
tggctgctct gtacggtaca tacatagtaa gttctcgctt tgagagccca attcgctgca	780
atagcatctt tccccgcagt gcgccggccg ccctgggtcc cgctccacaa tgaccttgct	840
tctggagctt ctcgacgaac agatcggccc gtttcttctc cacaccaatc cgaaccagtc	900
gggagcatgg ctgcggatgc gacgcagcct tccttcgcgc tgtacaaaca gctccgggaa	960
cgctgactgg tatgtacgga ctacagtaag tacactacga gtgcacatac tgacgaatac	1020
cggcctcaga ggaacctggc aggaccctac cccacacgaa accacagcga gaaagcgcaa	1080
tggatcagta actactgcga agtaaccgtg gtcccgggca aaggatctga gggccgatcg	1140
ctcgtggggc tgcgaggcga gggagagcaa acaagccagt cctcccgca acctggaaaa	1200
tcacttataa acacacgtca ccggcgccgg ggtgcgcgcc atgtgtcacc tccaggctcc	1260
tcccgggcga tgatctctgc cgggtccatc aatcatctcg gttcgccgca gctgcttctt	1320
tctgtgcagt gaacgctctc aaactgcaac gacgctgtcc gacatgaagg ctgctgcgt	1380
ttcctgcctc ttcggcagta cccttgccgt tgcaggcgcc attgaatcga gaaaggtatg	1440
gacgggcttt cgtcaaagac tcgctccccg atcaacttcc cttttcatcc agaccacccc	1500
aaccttccca gtcctgcttc gagcacgata tcttcgggca gcacccacc cacatccact	1560
cagattagcg gcgacaccgt tgactgttgc aatccgcaat cgacatgcaa cttccagccg	1620
cagcccaatg gctgctcacg cttcccgcga aagcctcact tgctgacaat catcgtcagg	1680
ttcaccagaa gcccctcgcg agatctgaac ctttttacct gtcgccatgg atgaatccca	1740
acgccgacgg ctgggcggag gcctatgccc aggccaagtc ctttgtctcc caaatgactc	1800
tgctagagaa ggtcaacttg accacgggag tcgggtaagt tttgtcattt tgtccaggta	1860
acatgcaaat ggttctgcta acaataactt accgtagctg gggggctgag cagtgcgtcg	1920
gccaagtggg cgcgatccct cgccttggac ttcgcagtct gtgcatgcat gactccccctc	1980
tcggcatccg aggagccgac tacaactcag cgttccccctc tggccagacc gttgctgcta	2040
cctgggatcg cggctctgatg taccgtcgcg gctacgcaat gggccaggag gccaaaggca	2100
agggcatcaa tgtccttctc ggaccagtcg ccggccccctc tggccgcatg cccgagggcg	2160
gtcgtaactg ggaaggcttc gctccggatc ccgtccttac cggcatcggc atgtccgaga	2220
cgatcaaggg cattcaggat gctggcgta tcgcttgtgc gaagcacttt attggaaacg	2280
agcaggggtga gtagtcaaag acgggcccgtc tcggacccgc ggcttcaagc tgctgactct	2340
gctgcagagc acttcagaca ggtgccagaa gccaggggat acgggttaca catcagcgaa	2400

accctctcct ccaacattga cgacaagacc atgcacgagc tctacctttg gccgtttgcc	2460
gatgccgtcc gggccggcgt cggctctgtc atgtgctcgt accagcaggt caacaactcg	2520
tacgcctgcc agaactcgaa gctgctgaac gacctcctca agaacgagct tgggtttcag	2580
ggcttcgtca tgagcgactg gcaggcacag cacactggcg cagcaagcgc cgtggctggt	2640
ctcgatatgt ccatgccggg cgacacccag ttcaacactg gcgtcagttt ctggggcgcc	2700
aatctcaccc tcgccgtcct caacggcaca gtccctgcct accgtctcga cgacatggcc	2760
atgcgcatca tggccgcctt cttcaaggtc accaagacca cccacctgga acccatcaac	2820
ttctccttct ggaccgacga cacttatggc ccgatccact gggccgcaa gcatggctac	2880
cagaagatta attcccacgt tgacgtccgc gccgaccacg gcaacctcat ccgggagatt	2940
gccgccaagg gtacggtgct gctgaagaat accggctctc taccctgaa caagccaaag	3000
ttcgtggccg tcatcggcga ggatgctggg tcgagcccca acgggcccaa cggctgcagc	3060
gaccgcggct gtaacgaagg cacgctcgcc atgggctggg gatccggcac agccaactat	3120
ccgtacctcg tttccccga cgccgcgtc caggccccgg ccatccagga cggcacgagg	3180
tacgagagcg tcctgtccaa ctacgccgag gaaaagacaa aggtctctggt ctcgaggcc	3240
aatgcaaccg ccatcgtctt cgtcaatgcc gactcaggcg agggctacat caacgtggac	3300
ggtaacgagg gcgaccgtaa gaacctgact ctctggaaca acggtgatac tctggtcaag	3360
aacgtctcga gctgggtgcag caacaccatc gtcgtcatcc actcggtcgg cccggtcctc	3420
ctgaccgatt ggtacgacaa ccccaacatc acggccattc tctgggctgg tcttcgggc	3480
caggagtcgg gcaactccat caccgacgtg ctttacggca aggtcaaccc cgccgcccgc	3540
tcgcccttca cttggggcaa gaccgcgaa agctatggcg cggacgtcct gtacaagccg	3600
aataatggca atggtgccc ccaacaggac ttcaccgagg gcgtcttcat cgactaccgc	3660
tacttcgaca aggttgacga tgactcggtc atctacgagt tcggccacgg cctgagctac	3720
accaccttcg agtacagcaa catccgcgtc gtcaagtcca acgtcagcga gtaccggccc	3780
acgacgggca ccacggccca ggccccgacg tttggcaact tctccaccga cctcgaggac	3840
tatctcttcc ccaaggacga gttcccctac atctaccagt acatctaccc gtacctcaac	3900
acgaccgacc cccggagggc ctcggccgat cccactacg gccagaccgc cgaggagttc	3960
ctcccgcccc acgccaccga tgacgacccc cagccgctcc tccggtcctc gggcggaac	4020
tccccggcg gcaaccgcca gctgtacgac attgtctaca caatcacggc cgacatcacg	4080
aatacgggct ccgttgtagg cgaggaggtta ccgcagctct acgtctcgct gggcgggtccc	4140
gaggatccca aggtgcagct gcgcgacttt gacaggatgc ggatcgaacc cggcgagacg	4200
aggcagttca ccggccgcct gacgcgcaga gatctgagca actgggacgt cacggtgcag	4260

gactgggtca	tcagcaggta	tccaagacg	gcatatgttg	ggaggagcag	ccggaagttg	4320
gatctcaaga	ttgagcttcc	ttgaatgagt	ttcatcaggg	gctgcagagg	gatggtaaca	4380
cgttcttaat	cagaagtatg	atggagaaaa	gcacttggca	agttccggtg	agcaaaaaga	4440
aggcacttat	taagtgtagg	gcggtgttct	atgtttaata	ggtgctatgt	ttacatataa	4500
ttagtatata	atgatttaat	aattatgttt	agcagttgct	aatgtcgtaa	atttcggcgt	4560
gtgatgactg	ctacaacact	ggttctgtct	tctagtcgcc	attgttaatt	atgaaggtta	4620
ttgtctacaa	tttctaatac	cttatggatg	attgccacgc	tggtttcaaa	ctcgttacgc	4680
gcaaattggt	cgattgaggt	attattcatt	gtaagtacct	ccgtacagcg	tccccaacta	4740
tttcattca	cgagatgcct	cgcttttcgg	tgctttcgga	acagggctgg	cagcggatca	4800
tggcgcgatc	aaaacatggc	gagcagctgt	ccaggacgga	ggacaggttg	gggactgatg	4860
cctcccggac	gcattaaggt	cagaagatag	acacgtttta	cacagcgttg	agaccgacaa	4920
gccacattag	gcagcgccgg	ttgcaccacc	gccgtcacgg	gcaacgggtc	aatcaatcga	4980
caacagtgga	agacaaagta	ctgaagatca	ggtattaata	gtgtgagaga	gaaacagacg	5040
gtggaactag	ggtgctaata	tttctcttga	tttcgggtgtc	catggtagta	cagaacacaa	5100
gaaaaagaag	gaggagttag	cggagaagga	ggagggggaa	gccagaaaaa	agaacatgaa	5160
aaagcataca	cattggagtc	ggtcagtcgg	ttgattgggt	tggtagagag	cgaaaaagca	5220
agcgtcacct	gtaggattcg	aacctacgct	cccgaaggaa	ctgcctaaga	acgctaagca	5280
aggtttagcag	ggcagcgcg	taaccactcc	gccaaagtga	ctgtcgttga	tcatggtcga	5340
attcaagtag	cttataggag	ttcaaccaga	tcacaaatgc	ataggtgctc	gtagaacggt	5400
ctaagtatga	gttgattata	agcaaccgaa	tggctctcag	cggcaacacc	gtagctgaag	5460
taacaaaacg	cacctttggt	tactttctga	ctataaaaat	gggatatttg	gaaatgacca	5520
cccgataagg	tgtcaaattc	taaatgactg	tctgggtgtg	aagatgttac	tgtggttcca	5580
ccacgaacca	gttttagtat	ccgcatgctt	cagtctctgc	gcctcgacag	gcggaggggtg	5640
tgtgttagat	cagaatcgat	gtgacgctgt	gaccgcgagg	ctctcgagcc	taggtgcggt	5700
agttctgttc	aaaaagaagt	gtgtggccgg	gtttgggcgc	ccttatagcc	taccatcctg	5760
gctgtgggtc	ccgagcgga	gccggttctc	cgttttgggt	ccgataaagt	gtcataatctg	5820
cctcccgtt	tcgcatctaa	tttctgactt	cgttcgggac	ctctggagac	gtagggatag	5880
gtatgggata	tgcccggcat	ttcgtaaatg	tccatagtct	ctttcgggac	gaggcggcaa	5940
gctctcagag	ctatctaagc	ttaaccaacc	cctgatcctt	aaccctccca	gaccacacct	6000
cctgggagaa	taaaccgggc	tccaagatcg	aaatcgaaat	cagtgcgcga	acttgaaatc	6060

<210> 12
<211> 871

<212> PRT

<213> Chrysosporium lucknowense

<400> 12

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

Ala Ser Leu Ala Asp Asn His Arg Gln Val His Gln Lys Pro Leu Ala
 20 25 30

Arg Ser Glu Pro Phe Tyr Pro Ser Pro Trp Met Asn Pro Asn Ala Asp
 35 40 45

Gly Trp Ala Glu Ala Tyr Ala Gln Ala Lys Ser Phe Val Ser Gln Met
 50 55 60

Thr Leu Leu Glu Lys Val Asn Leu Thr Thr Gly Val Gly Trp Gly Ala
 65 70 75 80

Glu Gln Cys Val Gly Gln Val Gly Ala Ile Pro Arg Leu Gly Leu Arg
 85 90 95

Ser Leu Cys Met His Asp Ser Pro Leu Gly Ile Arg Gly Ala Asp Tyr
 100 105 110

Asn Ser Ala Phe Pro Ser Gly Gln Thr Val Ala Ala Thr Trp Asp Arg
 115 120 125

Gly Leu Met Tyr Arg Arg Gly Tyr Ala Met Gly Gln Glu Ala Lys Gly
 130 135 140

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

Met Pro Glu Gly Gly Arg Asn Trp Glu Gly Phe Ala Pro Asp Pro Val
 165 170 175

Leu Thr Gly Ile Gly Met Ser Glu Thr Ile Lys Gly Ile Gln Asp Ala
 180 185 190

Gly Val Ile Ala Cys Ala Lys His Phe Ile Gly Asn Glu Gln Glu His
 195 200 205

Phe Arg Gln Val Pro Glu Ala Gln Gly Tyr Gly Tyr Asn Ile Ser Glu
 210 215 220

Thr Leu Ser Ser Asn Ile Asp Asp Lys Thr Met His Glu Leu Tyr Leu
 225 230 235 240

Trp	Pro	Phe	Ala	Asp 245	Ala	Val	Arg	Ala	Gly 250	Val	Gly	Ser	Val	Met 255	Cys
Ser	Tyr	Gln	Gln 260	Val	Asn	Asn	Ser	Tyr 265	Ala	Cys	Gln	Asn	Ser 270	Lys	Leu
Leu	Asn	Asp 275	Leu	Leu	Lys	Asn	Glu 280	Leu	Gly	Phe	Gln	Gly 285	Phe	Val	Met
Ser	Asp 290	Trp	Gln	Ala	Gln	His 295	Thr	Gly	Ala	Ala	Ser 300	Ala	Val	Ala	Gly
Leu 305	Asp	Met	Ser	Met	Pro 310	Gly	Asp	Thr	Gln	Phe 315	Asn	Thr	Gly	Val	Ser 320
Phe	Trp	Gly	Ala	Asn 325	Leu	Thr	Leu	Ala	Val 330	Leu	Asn	Gly	Thr	Val 335	Pro
Ala	Tyr	Arg	Leu 340	Asp	Asp	Met	Ala	Met 345	Arg	Ile	Met	Ala	Ala 350	Leu	Phe
Lys	Val	Thr 355	Lys	Thr	Thr	His	Leu 360	Glu	Pro	Ile	Asn	Phe 365	Ser	Phe	Trp
Thr	Asp 370	Asp	Thr	Tyr	Gly	Pro 375	Ile	His	Trp	Ala	Ala 380	Lys	His	Gly	Tyr
Gln 385	Lys	Ile	Asn	Ser	His 390	Val	Asp	Val	Arg	Ala 395	Asp	His	Gly	Asn	Leu 400
Ile	Arg	Glu	Ile	Ala 405	Ala	Lys	Gly	Thr	Val 410	Leu	Leu	Lys	Asn	Thr 415	Gly
Ser	Leu	Pro	Leu 420	Asn	Lys	Pro	Lys	Phe 425	Val	Ala	Val	Ile	Gly 430	Glu	Asp
Ala	Gly	Ser 435	Ser	Pro	Asn	Gly	Pro 440	Asn	Gly	Cys	Ser	Asp 445	Arg	Gly	Cys
Asn	Glu 450	Gly	Thr	Leu	Ala	Met 455	Gly	Trp	Gly	Ser	Gly 460	Thr	Ala	Asn	Tyr
Pro 465	Tyr	Leu	Val	Ser	Pro 470	Asp	Ala	Ala	Leu	Gln 475	Ala	Arg	Ala	Ile	Gln 480
Asp	Gly	Thr	Arg	Tyr	Glu	Ser	Val	Leu	Ser	Asn	Tyr	Ala	Glu	Glu	Lys

485					490					495					
Thr	Lys	Ala	Leu 500	Val	Ser	Gln	Ala	Asn 505	Ala	Thr	Ala	Ile	Val 510	Phe	Val
Asn	Ala	Asp 515	Ser	Gly	Glu	Gly	Tyr 520	Ile	Asn	Val	Asp	Gly 525	Asn	Glu	Gly
Asp	Arg 530	Lys	Asn	Leu	Thr	Leu 535	Trp	Asn	Asn	Gly	Asp 540	Thr	Leu	Val	Lys
Asn 545	Val	Ser	Ser	Trp	Cys 550	Ser	Asn	Thr	Ile	Val 555	Val	Ile	His	Ser	Val 560
Gly	Pro	Val	Leu	Leu 565	Thr	Asp	Trp	Tyr	Asp 570	Asn	Pro	Asn	Ile	Thr 575	Ala
Ile	Leu	Trp	Ala 580	Gly	Leu	Pro	Gly	Gln 585	Glu	Ser	Gly	Asn	Ser 590	Ile	Thr
Asp	Val	Leu 595	Tyr	Gly	Lys	Val	Asn 600	Pro	Ala	Ala	Arg	Ser 605	Pro	Phe	Thr
Trp	Gly 610	Lys	Thr	Arg	Glu	Ser 615	Tyr	Gly	Ala	Asp	Val 620	Leu	Tyr	Lys	Pro
Asn 625	Asn	Gly	Asn	Gly	Ala 630	Pro	Gln	Gln	Asp	Phe 635	Thr	Glu	Gly	Val	Phe 640
Ile	Asp	Tyr	Arg	Tyr 645	Phe	Asp	Lys	Val	Asp 650	Asp	Asp	Ser	Val	Ile 655	Tyr
Glu	Phe	Gly	His 660	Gly	Leu	Ser	Tyr	Thr 665	Thr	Phe	Glu	Tyr	Ser 670	Asn	Ile
Arg	Val	Val 675	Lys	Ser	Asn	Val	Ser 680	Glu	Tyr	Arg	Pro	Thr 685	Thr	Gly	Thr
Thr	Ala 690	Gln	Ala	Pro	Thr	Phe 695	Gly	Asn	Phe	Ser	Thr 700	Asp	Leu	Glu	Asp
Tyr 705	Leu	Phe	Pro	Lys	Asp 710	Glu	Phe	Pro	Tyr	Ile 715	Tyr	Gln	Tyr	Ile	Tyr 720
Pro	Tyr	Leu	Asn	Thr 725	Thr	Asp	Pro	Arg	Arg 730	Ala	Ser	Ala	Asp	Pro 735	His

Tyr Gly Gln Thr Ala Glu Glu Phe Leu Pro Pro His Ala Thr Asp Asp
740 745 750

Asp Pro **Gln** Pro Leu Leu Arg Ser Ser Gly Gly Asn Ser Pro Gly Gly
755 760 765

Asn Arg Gln Leu Tyr Asp Ile Val Tyr Thr Ile Thr Ala Asp Ile Thr
770 775 780

Asn Thr Gly Ser Val Val Gly Glu Glu Val Pro Gln Leu Tyr Val Ser
785 790 795 800

Leu Gly Gly Pro Glu Asp Pro Lys Val Gln Leu Arg Asp Phe Asp Arg
805 810 815

Met Arg Ile Glu Pro Gly Glu Thr Arg Gln Phe Thr Gly Arg Leu Thr
820 825 830

Arg Arg Asp Leu Ser Asn Trp Asp Val Thr Val Gln Asp Trp Val Ile
835 840 845

Ser Arg Tyr Pro Lys Thr Ala Tyr Val Gly Arg Ser Ser Arg Lys Leu
850 855 860

Asp Leu Lys Ile Glu Leu Pro
865 870

<210>	13
<211>	998
<212>	DNA
<213>	Chrysosporium lucknowense

<400> 13							
atggcccatc	tctccgccac	caccggggttc	ctcgcctctc	cggccctggc	cctggcccag		60
ctctcgggca	gcggccagac	gacccggtac	tgggactgct	gcaagccgag	ctgcgcctgg		120
cccggcaagg	gcccctcgtc	tccggtgcag	gcctgcgaca	agaacgacaa	cccgtcaac		180
gacggcggtc	ccacccggtc	cggctgcgac	gcggggcggca	gcgcctacat	gtgctcctcc		240
cagagccctt	gggccgtcag	cgacgagctg	tcgtacggct	gggcggccgt	caagctcgcc		300
ggcagctccg	agtcgcagtg	gtgctgcgcc	tgctacgagc	tgaccttcac	cagcggggccg		360
gtcgcgggca	agaagatgat	tgtgcaggcg	accaacaccg	gtggcgacct	gggcgacaac		420
cactttgacc	tggccgtgag	ttgcctcccc	ttctccccgg	accgctcaga	ttagatgaga		480
ttagactttg	ctcgtaaadc	ggtccaagat	tcccttgact	gaccaacaaa	catcatacgg		540
gcagatcccc	ggtggcggtg	tcggtatttt	caacggtaag	ctggtgcccc	cggacccctc		600
cccggacccc	tccccctttt	cctccagcga	gccgagttgg	gatcgccgag	atcgagaact		660

cacacaactt ctctctcgac agcctgcacc gaccagtacg gcgctcccc gaacggctgg 720
 ggcgaccgct acggcgcat ccattccaag gaagagtgcg aatccttccc ggaggccctc 780
 aagccccggt gcaactggcg cttcgactgg tacgttgctt tgacataccg gaaccaatt 840
 cctccaaccc cccccctttt ctcccccaac tccgggggta gtcggaatgt cgcgactgac 900
 cctatttcag gttcctaaaac gccgacaacc cgtcgggtcac cttccaggag gtggcctgcc 960
 cgtcggagct cacgtccaag agcggctgct cccgttaa 998

<210> 14
 <211> 225
 <212> PRT
 <213> Chrysosporium lucknowense

<400> 14

Met His Leu Ser Ala Thr Thr Gly Phe Leu Ala Leu Pro Ala Leu Ala
 1 5 10 15

Leu Ala Gln Leu Ser Gly Ser Gly Gln Thr Thr Arg Tyr Trp Asp Cys
 20 25 30

Cys Lys Pro Ser Cys Ala Trp Pro Gly Lys Gly Pro Ser Ser Pro Val
 35 40 45

Gln Ala Cys Asp Lys Asn Asp Asn Pro Leu Asn Asp Gly Gly Ser Thr
 50 55 60

Arg Ser Gly Cys Asp Ala Gly Gly Ser Ala Tyr Met Cys Ser Ser Gln
 65 70 75 80

Ser Pro Trp Ala Val Ser Asp Glu Leu Ser Tyr Gly Trp Ala Ala Val
 85 90 95

Lys Leu Ala Gly Ser Ser Glu Ser Gln Trp Cys Cys Ala Cys Tyr Glu
 100 105 110

Leu Thr Phe Thr Ser Gly Pro Val Ala Gly Lys Lys Met Ile Val Gln
 115 120 125

Ala Thr Asn Thr Gly Gly Asp Leu Gly Asp Asn His Phe Asp Leu Ala
 130 135 140

Ile Pro Gly Gly Gly Val Gly Ile Phe Asn Ala Cys Thr Asp Gln Tyr
 145 150 155 160

Gly Ala Pro Pro Asn Gly Trp Gly Asp Arg Tyr Gly Gly Ile His Ser
 165 170 175

Lys Glu Glu Cys Glu Ser Phe Pro Glu Ala Leu Lys Pro Gly Cys Asn
 180 185 190

Trp Arg Phe Asp Trp Phe Gln Asn Ala Asp Asn Pro Ser Val Thr Phe
 195 200 205

Gln Glu Val Ala Cys Pro Ser Glu Leu Thr Ser Lys Ser Gly Cys Ser
 210 215 220

Arg
 225

<210> 15
 <211> 4190
 <212> DNA
 <213> Chrysosporium lucknowense

<400> 15
 gcgcttccgg cctgggagag taaaatgacg gaagccgggc cccgtccgac tgcgtttgtc 60
 ccaactcggg agcaggcatc gttttttggg cgggaggaag cggtgcaaca cgcactatcg 120
 ccaaggtgga ctccggcga tctggagggt cggcccgcgg aggacggaat ccgggctgaa 180
 tctgcgcaaa ggctgaccct gcgatgggtg gaaaatgtaa atatgtgaag ttataggcat 240
 ataggactca gcgatgacat ggaaattgca gaggcattgt ggatttcagc gtttggcatg 300
 cattggctcg atctctcgcc ttgtctgatg tgatcccgcc ggagggtgtt cgggtctctgg 360
 ggaaggggacc cccctgggccc cccacactgc cccgcattcat gcctcgccac gactcccgcg 420
 cgccgaggaa gaacttcggg tctttgtgac gggagattcc actgagtgag cattggccaa 480
 ccaagcacac aattactccg tacatacaca gtacttctga ctccgtaaag taaaccgtgt 540
 gtttcaaaga tcggtaatcc gtaacaggta ctccgtatct aaggtaaatt taccctgtgc 600
 acggagcaga acctgaactt cttccccctt cttactcgag tagtcaccct actccaacca 660
 gcggcttttc aactcgcaaa gtcttggtta taacagtgc tatacctgca tttcgtatct 720
 cgctagtgtg aagacgacca cacgcggaca aagaaagaaa aatccaattg ccgatggct 780
 cttagtttga ggacagcagc gaaggactac actgcgccgt agtgaccagg ccaagaaacg 840
 cgaatcgtat attaacggca aatcaaaatg gattatatgc catttcgctt ccgggttgcg 900
 tgctcgtccg aagtctgggt ccgatcgatt gcgaaccccc ggaatcgcg gatgattcct 960
 acagccgccg aaaggggggg ggggggaggg ggggtctggac gggacgtgca taacttcgaa 1020
 tttctagaat attgctgggt ggggttccct cagccctgcg agcgcgcccc cttctggaac 1080
 cgcacccttc accggttcca cacacagagg acatgggtgg aaatgtgtac ctgacgggtg 1140
 cccctttggg acagtggaga ggcggatgtt cggataacca tccggagccg cagtgtcgac 1200

caagatcttg gcttaccatc gacaccaaca tgcggactcg tccctcagtc atggagcctt	1260
ggctcgcgga gcctccgttc gaagcggcta tcccgtcctg ccagcggagg atctcgtacc	1320
gcttccgcga actgtgaatg tcctgggtat aagagcatgg cgcgaccttg tctcgtcagg	1380
aacggggagg aggagggctt ggtagggtc gcgttcgttt ggagattgct gagctctgag	1440
ccttcggtcc ttggatccct gcggtccccg gtctcctctc tctctctctc tctctctctc	1500
tctctctctt cttcccacgc tcgttcgaca gacgcctccc cttcttcgct ctcttttccc	1560
tcgcacgtag cacactaata gtgcaccatg cgcgtctcta gtttggtcgc ggcccttgct	1620
accggtggtc ttgtcgccgc cagcctaag cccaaggggt cgtcgcccc tggggccgtg	1680
gacgcgaacc ctttcaagg caagacgcag ttcgtcaacc cggcatgggc ggccaagctg	1740
gaacagacca aaaaggcgtt cctggccagg aacgacaccg tcaatgccgc caagacggag	1800
aaggtccagc agaccagctc gtctgtctgg gtctcgagga tcgccgagct ctccaacatc	1860
gacgacgcca tcgcggctgc ccgcaaggcg cagaagaaga cgggcaggag gcagatcgtc	1920
ggcctgggtg tctacaacct tccggaccgc gactgcagcg cgggcgagag cgcgggcgag	1980
ctcagcagcg acaagaacgg gctcgagatc tacaagactg agttcgtcaa gcccttcgcc	2040
gacaaggtgg cggccgcaaa ggacctcgac ttcgccatcg tcctggagcc cgactcgtg	2100
gccaacctgg tcaccaacct gggcatcgag ttctgcgcca acgccgcccc cgtctaccgc	2160
gagggcatcg cctatgccat ctccagcctt cagcagccaa acgtgcactt gtacatcgat	2220
gctgcccacg gcggctggct cggctgggac gacaacctgc cgctggccgc caaggagttt	2280
gccgaggtgg tcaagcttgc cggcgagggc aagaagatcc gcggcttcgt caccaacgtg	2340
tccaactaca accccttcca cgccgtcgtg cgcgagaact ttaccgagtg gagcaactcg	2400
tgggacgagt ctactacgc ctctcgtc acaccgttcc tcgagaaaga ggggctgccg	2460
gcacgcttca tcgtcgacca gggtcgcgtt gccctcccgg gagcccgcaa ggagtgggtga	2520
gtttcgacca gattgacct cgacccatgc gaccgagatt gctgacgatt gaattgcgtg	2580
tcccgtcccc caggggtgaa tgggtgcaacg tggcaccgc cggatttggc cccgcgcccc	2640
cgaccagggc caacaacacc gtcgtcgatg ctctcgtctg ggtcaagcct ggcggcgaga	2700
gcgacggcga gtgtggcttg gctggcgccc ccaaggccgg ccagtgggtc gacgagtacg	2760
cccagatgct ggtcgagaat gcccaccgt ctgtcgtcca caagtggtag ataaattttg	2820
gagtccgaga aggggtcccag atagactttt gttttaaaac aaaatgcaag gtgtcgacag	2880
atactggctt aacattaacc aagcaccatg aacatgactt gtcaacatat tgatacatc	2940
cgtgctttc ccatacgtgc tctcaggtct cagggatcaa atggataggt cggtaatgca	3000
aaacgatcca ttggatatcc agaagagaga aaaaaaaaaa gacatgcatg ctttgtctgt	3060

catcatgagg aaacaaagga aaaacaaacg atcgtcgtgt tccaacaagc tttccaagac 3120
cacaagaccc atccaccaac acaaccaaac gacaagcaat acgatggacc gccgttggtc 3180
catctctcaa gagctgacta aacgaacagt cgttgaaatc atcctacatg agtacgccgc 3240
accacctgtt atcgtgtaaa ccaaatcgcc tgttaaagtg catcatctct taggtatgat 3300
cgtaagtcc ggtcacggtc acggatcagg gatggttctc aattcgtgtg tcgcgtagcc 3360
gccgccgtat ctggacaaga cttcttgat tgcctcgaaa ccgcttttgc cgccctaata 3420
atctgtagcc ttcttacctg gtggtgcctt gaaagacgcg gcaggcaaca cttcgcaggt 3480
ctgtggcgca ccagcaccag gctgtggtga tgccccgaa ccggtcgtcg acttgctcgc 3540
ggtgtcctcg gctggtgggg atgggggtga tgagggcttg gaggggtgtg ttgcgcccgc 3600
aacatccggc tccggctccg gaccgtccac agacattgga cctgcgagca tgactcgtgc 3660
cttcagccag accaaagcca tgccatcatc gcctctgccg acgctgttga gcgggaggct 3720
gatgttctca gccagaactg cgggctgtac ggccatgacc atgggctgtt cggctctggc 3780
gtcttgccgc gggttctccc tgccagcttg ttgtgcgcgc tgcctgcgag attcgacttc 3840
gacctgggcg tggcagaggg tgacgagggc cgttgacgcc ttgatctcct tgctcccat 3900
gtccttccac ccgtacaggc ggacgggtgc catacgcgtc cacagcctgc acgagaacct 3960
cagggcgctc tcaatgagtt ctgtcaactt gctctccagc ctctctatgc cgcgagcatc 4020
ctgatcctgg agcagaaacc gtgccgagcc tccgaggaaa cgctccttca gcttccgcgc 4080
gtagtttagg cgtgattcaa caaacgtccg gcgggactcg ttgttgcccg cagcagcgac 4140
gtccttgatg ctgaagccgc cgtcggcgaa caggcgcac atctgggccc 4190

<210> 16
<211> 381
<212> PRT
<213> Chrysosporium lucknowense

<400> 16

Met Arg Val Ser Ser Leu Val Ala Ala Leu Ala Thr Gly Gly Leu Val
1 5 10 15

Ala Ala Thr Pro Lys Pro Lys Gly Ser Ser Pro Pro Gly Ala Val Asp
20 25 30

Ala Asn Pro Phe Lys Gly Lys Thr Gln Phe Val Asn Pro Ala Trp Ala
35 40 45

Ala Lys Leu Glu Gln Thr Lys Lys Ala Phe Leu Ala Arg Asn Asp Thr
50 55 60

Val Asn Ala Ala Lys Thr Glu Lys Val Gln Gln Thr Ser Ser Phe Val

65		70		75		80
Trp Val Ser Arg	Ile 85	Ala Glu Leu Ser	Asn 90	Ile Asp Asp Ala	Ile 95	Ala
Ala Ala Arg	Lys 100	Ala Gln Lys Lys	Thr 105	Gly Arg Arg Gln	Ile 110	Val Gly
Leu Val	Leu 115	Tyr Asn Leu Pro	Asp 120	Arg Asp Cys Ser	Ala 125	Gly Glu Ser
Ala Gly	Glu 130	Leu Ser Ser	Asp 135	Lys Asn Gly Leu	Glu 140	Ile Tyr Lys Thr
Glu 145	Phe Val	Lys Pro	Phe 150	Ala Asp Lys Val	Ala 155	Ala Ala Lys Asp Leu 160
Asp Phe Ala	Ile Val 165	Leu Glu Pro Asp	Ser 170	Leu Ala Asn Leu	Val 175	Thr
Asn Leu Gly	Ile 180	Glu Phe Cys Ala	Asn 185	Ala Ala Pro Val	Tyr 190	Arg Glu
Gly Ile	Ala 195	Tyr Ala Ile Ser	Ser 200	Leu Gln Gln Pro	Asn 205	Val His Leu
Tyr Ile	Asp 210	Ala Ala His	Gly 215	Gly Trp Leu Gly	Trp 220	Asp Asp Asn Leu
Pro 225	Leu Ala Ala	Lys Glu 230	Phe Ala Glu Val	Val 235	Lys Leu Ala Gly	Glu 240
Gly Lys Lys	Ile Arg 245	Gly Phe Val Thr	Asn 250	Val Ser Asn Tyr	Asn 255	Pro
Phe His Ala	Val 260	Val Arg Glu Asn	Phe 265	Thr Glu Trp Ser	Asn 270	Ser Trp
Asp Glu	Ser 275	His Tyr Ala Ser	Ser 280	Leu Thr Pro Phe	Leu 285	Glu Lys Glu
Gly Leu	Pro 290	Ala Arg Phe	Ile 295	Val Asp Gln Gly	Arg 300	Val Ala Leu Pro
Gly 305	Ala Arg Lys	Glu Trp 310	Gly Glu Trp Cys	Asn 315	Val Ala Pro Ala	Gly 320

Phe Gly Pro Ala Pro Thr Thr Arg Val Asn Asn Thr Val Val Asp Ala
 325 330 335

Leu Val Trp Val Lys Pro Gly Gly Glu Ser Asp Gly Glu Cys Gly Leu
 340 345 350

Ala Gly Ala Pro Lys Ala Gly Gln Trp Phe Asp Glu Tyr Ala Gln Met
 355 360 365

Leu Val Glu Asn Ala His Pro Ser Val Val His Lys Trp
 370 375 380

<210> 17
 <211> 3000
 <212> DNA
 <213> Chrysosporium lucknowense

<400> 17
 cgcgccccg tctttgaacg cttgagaagc gcacggtgaa gaaccatcaa ctccgattcc 60
 gctcctcatc ctcccacgaa gccgattgaa atagccacag cggctatgta cggattactc 120
 tgctccgttt gcacatccat acacagcgct atttttaaaa gttcaggacg gccaagccccg 180
 gttcttgga cggacgaccc ggattccgaa agctccagcg ctcaatgcgg tcagtcgtgg 240
 cgctgatact gctgatctgc tgatctcata aaccgcgaac ttcaactttt cactttgaag 300
 cgtatacacg cagcgctctt ttcaccggcg cattcatact cgcaaattaa ccgctaatat 360
 cctcgcactt ggataatgtg tagccgacac ggaggagggg gggtgggggg ggggtggggg 420
 gagacatgat ggtctgcca acggatatta ttattttgtt gttttgtata attactgcgg 480
 caacattctc aaagggggcg tgcctcgcgg cgggaaagcc catgacagag aattggacag 540
 ctccaagctc gcgatatact ctaacaacgg cgtgactcgg caatgaaggc ctgccgctcg 600
 agtgataggg cgaagtaaaa cggacgttac atgcggcact tagccggctg atgccggaga 660
 atacgggatt caacgataca atcacacgat gcgacacacc tcggcgactt ggcgctctat 720
 ggaagaaggc tgggttaaag ctggcgtaga ttttgcgct cttggtttct taaccggggt 780
 atttctattt ctcatatgcc gcgagcgaat gcgggggtgca gagcgcccgg gagtcgatgg 840
 tcctatcaga caagagcctg gccccggaac ctgggataat agaagccaaa ttaagccatg 900
 ggagtatcgt ccgggggtag gaaccgcacg ggcaactaga ggaggaagaa tttggtataa 960
 agggaggacg gcggaacagg cttgatggac atgaatcaga agacgacact gggcaactaa 1020
 acagcttgca gcagagtttt gtgccttgca taggccctcg atatcatggt ctcgttcact 1080
 ctctcctca cggtcatcgc cgctgcggtg acgacggcca gccctctcga ggtggtcaag 1140
 cgcgccatcc agccgggcac gggcaccac gaggggtact tctactcggt ctggaccgac 1200
 ggccgtggct cggtcgactt caaccccggg ccccgcggt cgtacagcgt cacctggaac 1260

aacgtcaaca actgggttgg cggcaagggc tggaaacccgg gcccgccgcg caagattgcg	1320
tacaacggca cctggaacaa ctacaacgtg aacagctgtg cgttgctctc ctctttctcc	1380
ctttcgcttg ttttccttga tgattgggat ccatttttaa agagaaggaa aaaaaaaca	1440
aaggaaaata gaagataact aacgccaagc tctggcagac ctcgccctgt acggctggac	1500
tcgcaacccg ctggtcgagt attacatcgt ggaggcatac ggcacgtaca acccctcgtc	1560
gggcacggcg cggctgggca ccatcgagga cgacggcggc gtgtacgaca tctacaagac	1620
gacgcggtac aaccagccgt ccatcgaggg gacctccacc ttcgaccagt actggtccgt	1680
ccgccgccag aagcgcgtcg gcggcactat cgacacgggc aagcactttg acgagtggaa	1740
gcgccagggc aacctccagc tcggcacctg gaactacatg atcatggcca ccgaggggcta	1800
ccagagctct ggttcggcca ctatcgaggt ccgggaggcc taaagaagcc aggcgccttt	1860
cttttgtttt gcaggagggg gtagaggggg ggggggaggg aaaacgaaaa gtagcagggt	1920
ggttttatgc cggcagccgt gggccattcg agtgcaacct gtatctctct ctctcccaag	1980
tctccgggct ctttctcaga gaacttcaat atgtctgggg acaaaccacc ttgtgaaata	2040
caacggtaat tatctaagtt tgagtgccct atcgtatgct tctgaaaatt tcctgctcct	2100
tgatacaagt cggtttgagc cgagccaatg agactgtgtc gattgataga ggccctgaag	2160
gatcaagcgc gatgcaacaa ttaagcatga ctacgtgcct agctgcagat aaatggaagc	2220
cactcaccaa ggtcaacccc gcatactggc acgtaagaac cttccgtgta caaggcccaa	2280
ccgactcaca tatctatctg cttgggtttt gggatgcggt tttttacca caaaacaaat	2340
ttgatacaat gctctgctgt gcccggggtg ctgagaccaa gccgtaatca gcgggcaggg	2400
aatcgagtag gtcacgcctg ttgcttggtc tagaacaac taatattaaa aagccttggtg	2460
ctcggcacac atacagaact cgacctgagg catgttcttg gaaggcggct agccagtcaa	2520
gtctggcacc aggccttggt ctcgtcgagg ataccgaggg cgaggaggat gaggaagacc	2580
tctttctcgc ctcagatctc ttaggggacg aagaagacaa cgccggagcc acacaataat	2640
taggtctcat atcagacgtt tcggcctggc cgagctaata tgtctaatta tgcccatcag	2700
ccgtatgtcg aggcagggtg caccgatacg ctcgccgcgc cgcctcattc atctccgact	2760
gggcacaatg tcgccatctc ggccgtcaag gtggtgcaag atacctatta tgcaagcaga	2820
ggatcagatg gcgggccgat acgagcggct gctccggctt gcgagaaagc cgcttcgcag	2880
caaggtatcg tggcaggccg ccattttcgg ttgggtattc tttgtcttgt ttgcttcgta	2940
attatgtcct ggctggcatt gtgggaaggg gcgaacctct tgatttccga tgggggtcga	3000

<210> 18
 <211> 218
 <212> PRT

<213> Chrysosporium lucknowense

<400> 18

Met Val Ser Phe Thr Leu Leu Leu Thr Val Ile Ala Ala Ala Val Thr
1 5 10 15

Thr Ala Ser Pro Leu Glu Val Val Lys Arg Gly Ile Gln Pro Gly Thr
20 25 30

Gly Thr His Glu Gly Tyr Phe Tyr Ser Phe Trp Thr Asp Gly Arg Gly
35 40 45

Ser Val Asp Phe Asn Pro Gly Pro Arg Gly Ser Tyr Ser Val Thr Trp
50 55 60

Asn Asn Val Asn Asn Trp Val Gly Gly Lys Gly Trp Asn Pro Gly Pro
65 70 75 80

Pro Arg Lys Ile Ala Tyr Asn Gly Thr Trp Asn Asn Tyr Asn Val Asn
85 90 95

Ser Tyr Leu Ala Leu Tyr Gly Trp Thr Arg Asn Pro Leu Val Glu Tyr
100 105 110

Tyr Ile Val Glu Ala Tyr Gly Thr Tyr Asn Pro Ser Ser Gly Thr Ala
115 120 125

Arg Leu Gly Thr Ile Glu Asp Asp Gly Gly Val Tyr Asp Ile Tyr Lys
130 135 140

Thr Thr Arg Tyr Asn Gln Pro Ser Ile Glu Gly Thr Ser Thr Phe Asp
145 150 155 160

Gln Tyr Trp Ser Val Arg Arg Gln Lys Arg Val Gly Gly Thr Ile Asp
165 170 175

Thr Gly Lys His Phe Asp Glu Trp Lys Arg Gln Gly Asn Leu Gln Leu
180 185 190

Gly Thr Trp Asn Tyr Met Ile Met Ala Thr Glu Gly Tyr Gln Ser Ser
195 200 205

Gly Ser Ala Thr Ile Glu Val Arg Glu Ala
210 215

REIVINDICAÇÕES

1. Formulação enzimática, a dita formulação **CARACTERIZADA** pelo fato de que compreende pelo menos uma enzima selecionada a partir do grupo que consiste de CBH Ib (SEQ ID NO. 2), CBH Iib (SEQ ID NO. 4), EG II (SEQ ID NO. 10), EG VI (SEQ ID NO. 16),
5 BGL (SEQ ID NO 12) e Xyl II (SEQ ID NO. 18).

2. Método de produzir açúcares fermentáveis a partir de material lignocelulósico, o dito método **CARACTERIZADO** pelo fato de que compreende:

(a) fornecer uma formulação enzimática, em que a dita formulação enzimática compreende pelo menos uma enzima selecionada a partir do grupo que consiste de CBH Ib
10 (SEQ ID NO. 2), CBH Iib (SEQ ID NO. 4), EG II (SEQ ID NO. 10), EG VI (SEQ ID NO. 16), BGL (SEQ ID NO 12) e Xyl II (SEQ ID NO. 18);

(b) aplicar a dita formulação enzimática ao material lignocelulósico para produzir açúcares fermentáveis.

3. Método, de acordo com a reivindicação 2, **CARACTERIZADO** pelo fato de que
15 os açúcares fermentáveis compreendem pelo menos um açúcar do grupo que consiste de glicose, xilose, arabinose, galactose, manose, ramnose, sacarose e frutose.

4. Método, de acordo com a reivindicação 2, **CARACTERIZADO** pelo fato de que o material lignocelulósico é selecionado a partir do grupo que consiste de podas de pomar, chaparral, resíduos de moagem, resíduos urbanos de madeira, resíduos urbanos, resíduos
20 de corte de madeira, desbastes florestais, safras de madeira de curta rotação, resíduos industriais, palha de trigo, palha de aveia, palha de arroz, palha de cevada, palha de centeio, palha do linho, cascas de soja, cascas de arroz, palha de arroz, alimentação de glúten de milho, cascas de aveia, cana-de-açúcar, forragem de milho, hastes do milho, sabugos de milho, palhas de milho, pradarias, gamagrass, capim rabo-de-raposa; polpa de beterraba,
25 polpa de frutas cítricas, cascas de semente, resíduos animais celulósicos, cortes de grama, algodão, algas, árvores, arbustos, gramas, trigo, palha de trigo, bagaço de cana de açúcar, milho, palhas de milho, grão de milho, fibra dos grãos, produtos e subprodutos da moagem úmida ou moagem a seco de grãos, resíduos sólidos urbanos, papel usado, sobras orgânicas, material herbáceo, resíduos agrícolas, resíduos florestais, resíduos sólidos urbanos,
30 papel usado, resíduos da fabricação de polpa e papel, galhos, matos, canas, milho, palhas de milho, safras energéticas, florestas, frutos, flores, grãos, gramas, safras herbáceas, folhas, casca de árvore, acículas, lenhas, raízes, mudas, arbustos, gramíneas do tipo switchgrass, árvores, vegetais, cascas de frutos, videiras, polpa de beterraba, triguilhos, cascas de aveia, madeiras duras e moles, materiais de resíduos orgânicos gerados a partir de processos agrícolas, resíduos da madeira florestal, ou combinações dos mesmos.
35

5. Método de produzir um produto de fermentação ou um material de partida para um produto de fermentação a partir de um açúcar fermentável, o dito método

CARACTERIZADO pelo fato de que compreende:

(a) fornecer uma formulação enzimática, em que a dita formulação enzimática compreende pelo menos uma enzima selecionada a partir do grupo que consiste de CBH Ib (SEQ ID NO. 2), CBH Iib (SEQ ID NO. 4), EG II (SEQ ID NO. 10), EG VI (SEQ ID NO. 16),
5 BGL (SEQ ID NO 12), e Xyl II (SEQ ID NO. 18);

(b) aplicar a dita formulação enzimática ao material lignocelulósico para produzir um açúcar fermentável; e

(c) fermentar o dito açúcar fermentável para produzir um produto de fermentação.

6. Método, de acordo com a reivindicação 5, **CARACTERIZADO** pelo fato de que o
10 açúcar fermentável é selecionado a partir do grupo que consiste de glicose, xilose, arabinose, galactose, manose, ramnose, sacarose e frutose.

7. Método, de acordo com a reivindicação 5, **CARACTERIZADO** pelo fato de que o material lignocelulósico é selecionado a partir do grupo que consiste de podas de pomar, chaparral, resíduos de moagem, resíduos urbanos de madeira, resíduos urbanos, resíduos
15 de corte de madeira, desbastes florestais, safras de madeira de curta rotação, resíduos industriais, palha de trigo, palha de aveia, palha de arroz, palha de cevada, palha de centeio, palha do linho, cascas de soja, cascas de arroz, palha de arroz, alimentação de glúten de milho, cascas de aveia, cana-de-açúcar, forragem de milho, hastes do milho, sabugos de milho, palhas de milho, pradarias, gamagrass, capim rabo-de-raposa; polpa de beterraba,
20 polpa de frutas cítricas, cascas de semente, resíduos animais celulósicos, cortes de grama, algodão, algas, árvores, arbustos, gramas, trigo, palha de trigo, bagaço de cana de açúcar, milho, palhas de milho, grão de milho, fibra dos grãos, produtos e subprodutos da moagem úmida ou moagem a seco de grãos, resíduos sólidos urbanos, papel usado, sobras orgânicas, material herbáceo, resíduos agrícolas, resíduos florestais, resíduos sólidos urbanos,
25 papel usado, resíduos da fabricação de polpa e papel, galhos, matos, canas, milho, palhas de milho, safras energéticas, florestas, frutos, flores, grãos, gramas, safras herbáceas, folhas, casca de árvore, acículas, lenhas, raízes, mudas, arbustos, gramíneas do tipo switchgrass, árvores, vegetais, cascas de frutos, videiras, polpa de beterraba, triguilhos, cascas de aveia, madeiras duras e moles, materiais de resíduos orgânicos gerados a partir de processos agrícolas, resíduos da madeira florestal, ou combinações dos mesmos.
30

8. Método, de acordo com a reivindicação 5, **CARACTERIZADO** pelo fato de que o dito produto de fermentação é um biocombustível.

9. Método, de acordo com a reivindicação 5, **CARACTERIZADO** pelo fato de que o dito produto de fermentação é selecionado a partir do grupo que consiste de ácido láctico;
35 plásticos; especialidades químicas, ácidos orgânicos, solventes, suplementos de ração animal, produtos farmacêuticos, vitaminas; aminoácidos, enzimas industriais, e matérias-primas químicas.

10. Método de produzir energia a partir de um açúcar fermentável, o dito método **CARACTERIZADO** pelo fato de que compreende:

(a) fornecer uma formulação enzimática, em que a dita formulação enzimática compreende pelo menos uma enzima selecionada a partir do grupo que consiste de CBH Ib (SEQ ID NO. 2), CBH IIb (SEQ ID NO. 4), EG II (SEQ ID NO. 10), EG VI (SEQ ID NO. 16), BGL (SEQ ID NO 12), e Xyl II (SEQ ID NO. 18);

(b) aplicar a dita formulação enzimática ao material lignocelulósico para produzir um açúcar fermentável;

(c) fermentar o dito açúcar fermentável para produzir um produto de fermentação combustível;

(d) queimar o dito produto de fermentação combustível para produzir energia.

11. Método, de acordo com a reivindicação 10, **CARACTERIZADO** pelo fato de que o açúcar fermentável é selecionado a partir do grupo que consiste de glicose, xilose, arabinose, galactose, manose, ramnose, sacarose e frutose.

12. Método, de acordo com a reivindicação 10, **CARACTERIZADO** pelo fato de que o material lignocelulósico é selecionado a partir do grupo que consiste de podas de pomar, chaparral, resíduos de moagem, resíduos urbanos de madeira, resíduos urbanos, resíduos de corte de madeira, desbastes florestais, safras de madeira de curta rotação, resíduos industriais, palha de trigo, palha de aveia, palha de arroz, palha de cevada, palha de centeio, palha do linho, cascas de soja, cascas de arroz, palha de arroz, alimentação de glúten de milho, cascas de aveia, cana-de-açúcar, forragem de milho, hastes do milho, sabugos de milho, palhas de milho, pradarias, gamagrass, capim rabo-de-raposa; polpa de beterraba, polpa de frutas cítricas, cascas de semente, resíduos animais celulósicos, cortes de grama, algodão, algas, árvores, arbustos, gramas, trigo, palha de trigo, bagaço de cana de açúcar, milho, palhas de milho, grão de milho, fibra dos grãos, produtos e subprodutos da moagem úmida ou moagem a seco de grãos, resíduos sólidos urbanos, papel usado, sobras orgânicas, material herbáceo, resíduos agrícolas, resíduos florestais, resíduos sólidos urbanos, papel usado, resíduos da fabricação de polpa e papel, galhos, matos, canas, milho, palhas de milho, safras energéticas, florestas, frutos, flores, grãos, gramas, safras herbáceas, folhas, casca de árvore, acículas, lenhas, raízes, mudas, arbustos, gramíneas do tipo switchgrass, árvores, vegetais, cascas de frutos, videiras, polpa de beterraba, triguilhos, cascas de aveia, madeiras duras e moles, materiais de resíduos orgânicos gerados a partir de processos agrícolas, resíduos da madeira florestal, ou combinações dos mesmos.

13. Método, de acordo com a reivindicação 10, **CARACTERIZADO** pelo fato de que o dito produto de fermentação combustível é um álcool.

14. Método, de acordo com as reivindicações 2, 5, ou 10, **CARACTERIZADO** pelo fato de que o material lignocelulósico é submetido a um pré-tratamento antes de ser exposto

às enzimas.

15. Método, de acordo com a reivindicação 14, **CHARACTERIZADO** pelo fato de que o pré-tratamento compreende expor a biomassa lignocelulósica a um ácido, base, solvente, calor, peróxido, ozônio, trituração mecânica, trituração, moagem, despressurização rápida, ou uma combinação dos mesmos.

16. Método, de acordo com a reivindicação 15, **CHARACTERIZADO** pelo fato de que o dito solvente é uma mistura de acetona/etanol ou organosolvente.

17. Cepa de *Chrysosporium lucknowense* mutante, **CHARACTERIZADA** pelo fato de que é capaz de expressar pelo menos uma celobiohidrolase e pelo menos uma endo-1,4- β -glicanase em níveis mais altos do que a cepa não mutante correspondente sob as mesmas condições;

em que pelo menos a dita uma celobiohidrolase é selecionada a partir do grupo que consiste de CBH Ia, CBH Ib, CBH Ha e CBH Hb; e

em que pelo menos a dita uma endo-1,4- β -glicanase é selecionada a partir do grupo que consiste de EG II, EG V, e EG VI.

18. Cepa de *Chrysosporium lucknowense* mutante, de acordo com a reivindicação 17, a dita cepa **CHARACTERIZADA** pelo fato de que é capaz de expressar uma β -glicosidase e/ou uma xilanase em níveis mais altos do que a cepa não mutante correspondente sob as mesmas condições.

19. Cepa de *Chrysosporium lucknowense* mutante, de acordo com a reivindicação 18, **CHARACTERIZADA** pelo fato de que a dita xilanase é Xyl II.

20. Cepa de *Chrysosporium lucknowense* mutante, de acordo com a reivindicação 17, **CHARACTERIZADA** pelo fato de que a dita mutante é derivada de uma cepa mutante de *Chrysosporium lucknowense* selecionada a partir do grupo que consiste de cepa de *C. lucknowense* C1 (VKM F-3500 D), UV13-6 (VKM F-3632 D), NG7C-19 (VKM F-3633 D), e UV18-25 (VKM F-3631 D).

21. Proteína, **CHARACTERIZADA** pelo fato de que exibe pelo menos 65 % da identidade de aminoácido conforme determinado pelo algoritmo BLAST com a sequência de aminoácido de CBH Ib da SEQ ID NO. 2 ou uma parte desta tendo pelo menos 20 aminoácidos contíguos.

22. Proteína, de acordo com a reivindicação 21, **CHARACTERIZADA** pelo fato de que a proteína exibe pelo menos 70 % da identidade de aminoácido conforme determinado pelo algoritmo BLAST com a sequência de aminoácido de CBH Ib da SEQ ID NO. 2 ou uma parte desta tendo pelo menos 20 aminoácidos contíguos.

23. Proteína, de acordo com a reivindicação 21, **CHARACTERIZADA** pelo fato de que a dita proteína exibe pelo menos 75 % da identidade de aminoácido conforme determinado pelo algoritmo BLAST com a sequência de aminoácido de CBH Ib da SEQ ID NO. 2 ou

uma parte desta tendo pelo menos 20 aminoácidos contíguos.

24. Sequência de ácido nucléico, **CHARACTERIZADA** pelo fato de que codifica uma proteína conforme definida na reivindicação 21.

5 25. Sequência de ácido nucléico, **CHARACTERIZADA** pelo fato de que codifica uma proteína conforme definida na reivindicação 22.

26. Sequência de ácido nucléico, **CHARACTERIZADA** pelo fato de que codifica uma proteína conforme definida na reivindicação 23.

10 27. Proteína, **CHARACTERIZADA** pelo fato de que exibe pelo menos 65 % da identidade do aminoácido conforme determinado pelo algoritmo BLAST com a sequência de aminoácido de CBH IIb da SEQ ID NO. 4 ou uma parte desta tendo pelo menos 20 aminoácidos contíguos.

15 28. Proteína, de acordo com a reivindicação 27, **CHARACTERIZADA** pelo fato de que a proteína exibe pelo menos 70 % da identidade de aminoácido conforme determinado pelo algoritmo BLAST com a sequência de aminoácido de CBH IIb da SEQ ID NO. 4 ou uma parte desta tendo pelo menos 20 aminoácidos contíguos.

29. Proteína, de acordo com a reivindicação 28, **CHARACTERIZADA** pelo fato de que a dita proteína exibe pelo menos 75 % da identidade de aminoácido conforme determinado pelo algoritmo BLAST com a sequência de aminoácido de CBH IIb da SEQ ID NO. 4 ou uma parte desta tendo pelo menos 20 aminoácidos contíguos.

20 30. Sequência de ácido nucléico, **CHARACTERIZADA** pelo fato de que codifica uma proteína conforme definida na reivindicação 27.

31. Sequência de ácido nucléico, **CHARACTERIZADA** pelo fato de que codifica uma proteína conforme definida na reivindicação 28.

25 32. Sequência de ácido nucléico, **CHARACTERIZADA** pelo fato de que codifica uma proteína conforme definida na reivindicação 29.

33. Proteína, **CHARACTERIZADA** pelo fato de que exibe pelo menos 65 % da identidade do aminoácido conforme determinado pelo algoritmo BLAST com a sequência de aminoácido de EG VI da SEQ ID NO. 16 ou uma parte desta tendo pelo menos 20 aminoácidos contíguos.

30 34. Proteína, de acordo com a reivindicação 33, **CHARACTERIZADA** pelo fato de que a proteína exibe pelo menos 70 % da identidade de aminoácido conforme determinado pelo algoritmo BLAST com a sequência de aminoácido de EG VI da SEQ ID NO. 16 ou uma parte desta tendo pelo menos 20 aminoácidos contíguos.

35 35. Proteína, de acordo com a reivindicação 34, **CHARACTERIZADA** pelo fato de que a dita proteína exibe pelo menos 75 % da identidade de aminoácido conforme determinado pelo algoritmo BLAST com a sequência de aminoácido de EG VI da SEQ ID NO. 16 ou uma parte desta tendo pelo menos 20 aminoácidos contíguos.

36. Sequência de ácido nucléico, **CHARACTERIZADA** pelo fato de que codifica uma proteína conforme definida na reivindicação 33.

37. Sequência de ácido nucléico, **CHARACTERIZADA** pelo fato de que codifica uma proteína conforme definida na reivindicação 34.

5 38. Sequência de ácido nucléico, **CHARACTERIZADA** pelo fato de que codifica uma proteína conforme definida na reivindicação 35.

10 39. Proteína, **CHARACTERIZADA** pelo fato de que exibe pelo menos 65 % da identidade de aminoácido conforme determinado pelo algoritmo BLAST com a sequência de aminoácido de BGL da SEQ ID NO. 12 ou uma parte desta tendo pelo menos 20 aminoácidos contíguos.

40. Proteína, de acordo com a reivindicação 39, **CHARACTERIZADA** pelo fato de que a proteína exibe pelo menos 70 % da identidade de aminoácido conforme determinado pelo algoritmo BLAST com a sequência de aminoácido de BGL da SEQ ID NO. 12 ou uma parte desta tendo pelo menos 20 aminoácidos contíguos.

15 41. Proteína, de acordo com a reivindicação 42, **CHARACTERIZADA** pelo fato de que a dita proteína exibe pelo menos 75 % da identidade de aminoácido conforme determinado pelo algoritmo BLAST com a sequência de aminoácido de BGL da SEQ ID NO. 12 ou uma parte desta tendo pelo menos 20 aminoácidos contíguos.

20 42. Sequência de ácido nucléico, **CHARACTERIZADA** pelo fato de que codifica uma proteína conforme definida na reivindicação 39.

43. Sequência de ácido nucléico, **CHARACTERIZADA** pelo fato de que codifica uma proteína conforme definida na reivindicação 40.

44. Sequência de ácido nucléico, **CHARACTERIZADA** pelo fato de que codifica uma proteína conforme definida na reivindicação 41.

25 45. Proteína, **CHARACTERIZADA** pelo fato de que exibe pelo menos 65 % da identidade de aminoácido conforme determinado pelo algoritmo BLAST com a sequência de aminoácido de Xyl II da SEQ ID NO. 18 ou uma parte desta tendo pelo menos 20 aminoácidos contíguos.

30 46. Proteína, de acordo com a reivindicação 45, **CHARACTERIZADA** pelo fato de que a proteína exibe pelo menos 70 % da identidade de aminoácido conforme determinado pelo algoritmo BLAST com a sequência de aminoácido de Xyl II da SEQ ID NO. 18 ou uma parte desta tendo pelo menos 20 aminoácidos contíguos.

35 47. Proteína, de acordo com a reivindicação 46, **CHARACTERIZADA** pelo fato de que a dita proteína exibe pelo menos 75 % da identidade de aminoácido conforme determinado pelo algoritmo BLAST com a sequência de aminoácido de Xyl II da SEQ ID NO. 18 ou uma parte desta tendo pelo menos 20 aminoácidos contíguos.

48. Sequência de ácido nucléico, **CHARACTERIZADA** pelo fato de que codifica uma

proteína conforme definida na reivindicação 45.

49. Sequência de ácido nucléico, **CARACTERIZADA** pelo fato de que codifica uma proteína conforme definida na reivindicação 46.

50. Sequência de ácido nucléico, **CARACTERIZADA** pelo fato de que codifica uma proteína conforme definida na reivindicação 47.

51. Sequência de ácido nucléico, **CARACTERIZADA** pelo fato de que tem pelo menos 80 % de homologia com a sequência de ácido nucléico que codifica CBH Ib como mostrado na SEQ ID NO. 1.

52. Sequência de ácido nucléico, **CARACTERIZADA** pelo fato de que tem pelo menos 80 % de homologia com a sequência de ácido nucléico que codifica CBH IIb como mostrado na SEQ ID NO. 3.

53. Sequência de ácido nucléico, **CARACTERIZADA** pelo fato de que tem pelo menos 80 % de homologia com a sequência de ácido nucléico que codifica EG VI como mostrado na SEQ ID NO. 15.

54. Sequência de ácido nucléico, **CARACTERIZADA** pelo fato de que tem pelo menos 80 % de homologia com a sequência de ácido nucléico que codifica BGL como mostrado na SEQ ID NO. 11.

55. Sequência de ácido nucléico, **CARACTERIZADA** pelo fato de que tem pelo menos 80 % de homologia com o ácido nucléico que codifica Xyl II como mostrado na SEQ ID NO. 17.

56. Método para degradar um material lignocelulósico em açúcares fermentáveis, o dito método **CARACTERIZADO** pelo fato de que compreende

contatar o material lignocelulósico com uma quantidade eficaz de um produto multi-enzimático derivado de um microorganismo, para produzir pelo menos um açúcar fermentável

em que pelo menos uma enzima no produto multienzimático é selecionada a partir do grupo que consiste de CBH Ia (SEQ ID NO. 6), CBH Ib (SEQ ID NO. 2), CBH IIa (SEQ ID NO 8), CBH IIb (SEQ ID NO. 4), EG II (SEQ ID NO. 10), EG V (SEQ ID NO. 14), EG VI (SEQ ID NO. 16), BGL (SEQ ID NO. 12), e Xyl II (SEQ ID NO. 18).

57. Microorganismo ou planta, **CARACTERIZADO** pelo fato de que é capaz de expressar uma ou mais de uma enzima selecionada a partir do grupo que consiste de CBH Ia (SEQ ID NO. 6), CBH Ib (SEQ ID NO. 2), CBH IIa (SEQ ID, NO 8), CBH IIb (SEQ ID NO. 4), EG II (SEQ ID NO. 10), EG V (SEQ ID NO. 14), EG VI (SEQ ID NO. 16), BGL (SEQ ID NO. 12), e Xyl II (SEQ ID NO. 18).

58. Microorganismo, de acordo com a reivindicação 57, **CARACTERIZADO** pelo fato de que o dito microorganismo é um fungo.

59. Microorganismo, de acordo com a reivindicação 58, **CARACTERIZADO** pelo fa-

to de que o dito fungo é selecionado a partir do grupo que consiste de *Talaromyces*, *Aspergillus*, *Trichoderma*, *Neurospora*, *Penicillium*, *Fusarium*, *Humicola*, *Myceliophthora*, *Corynascus*, *Chaetomium*, *Tolypocladium*, *Thielavia*, *Acremonium*, *Sporotrichum*, *Thermoascus*, e *Chrysosporium*.

5 60. Microorganismo, de acordo com a reivindicação 59, **CHARACTERIZADO** pelo fato de que o dito fungo é *Chrysosporium lucknowense*.

 61. Microorganismo, de acordo com a reivindicação 60, **CHARACTERIZADO** pelo fato de que o dito *Chrysosporium* é uma cepa selecionada a partir do grupo que consiste da cepa do tipo selvagem, Acc. Nº VKM F-3500 D; CI UV13-6, Acc. Nº VKM F-3632 D; cepa C1
10 NG7C-19, Acc. Nº VKM F-3633 D; e CI UV18-25, Acc. Nº VKM F-3631 D.

 62. Sequência de ácido nucléico, **CHARACTERIZADA** pelo fato de que tem pelo menos 80 % de homologia com a sequência de ácido nucléico que codifica EG II como mostrado na SEQ ID NO. 9.

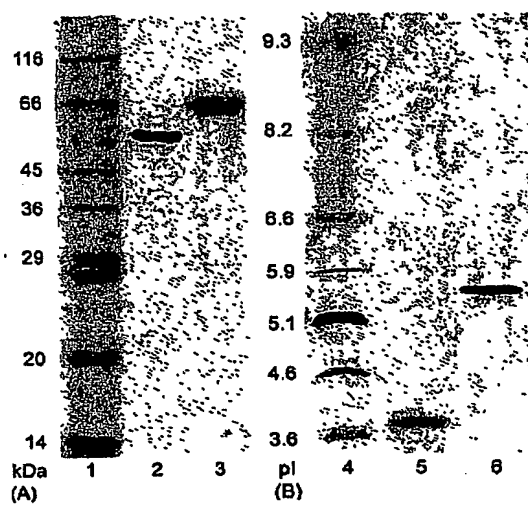


FIGURA 1

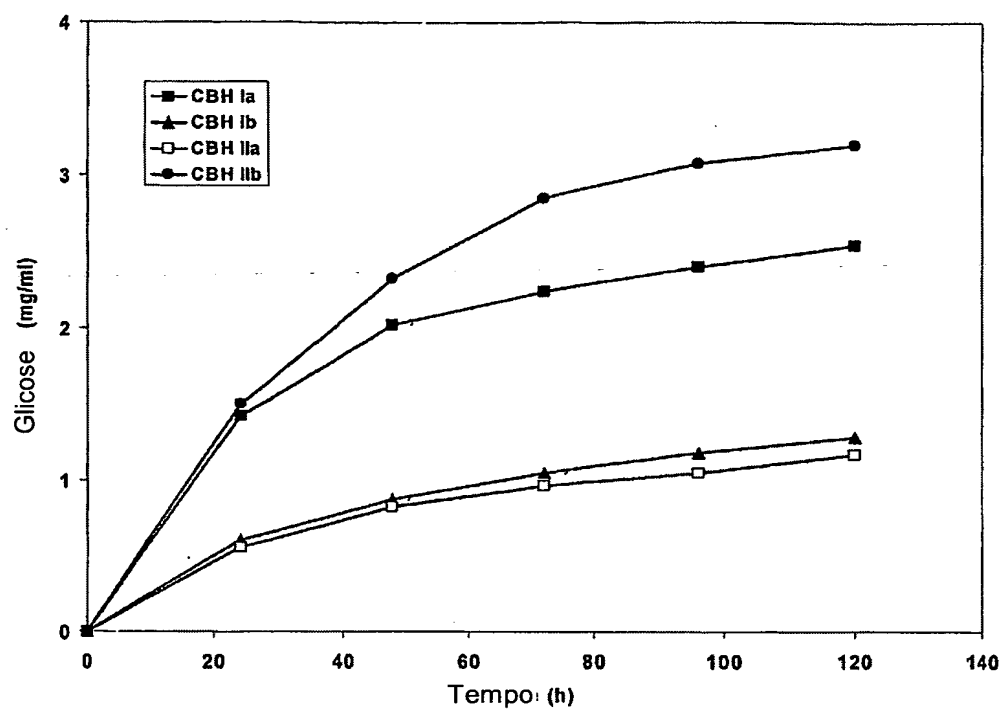


FIGURA 2

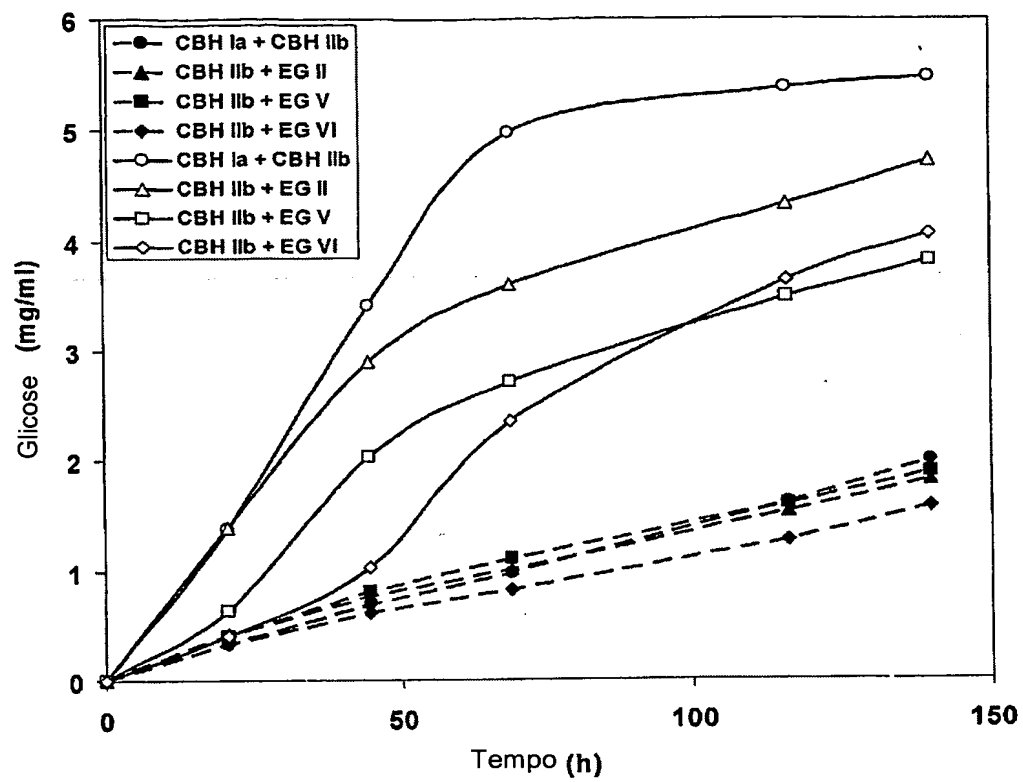


FIGURA 3

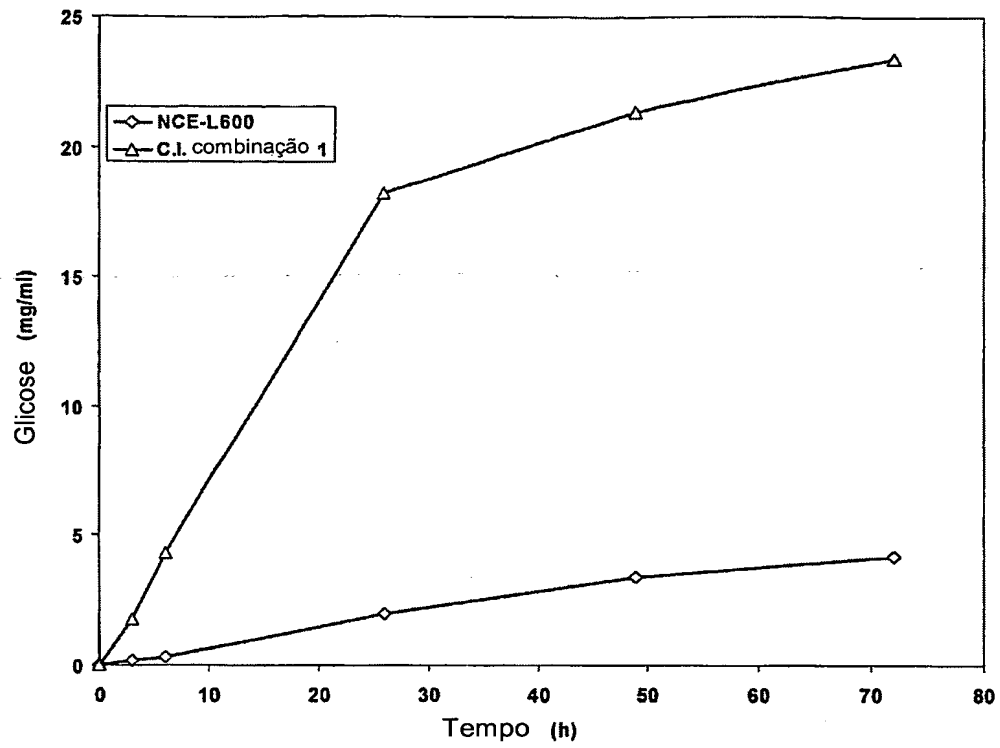


FIGURA 4

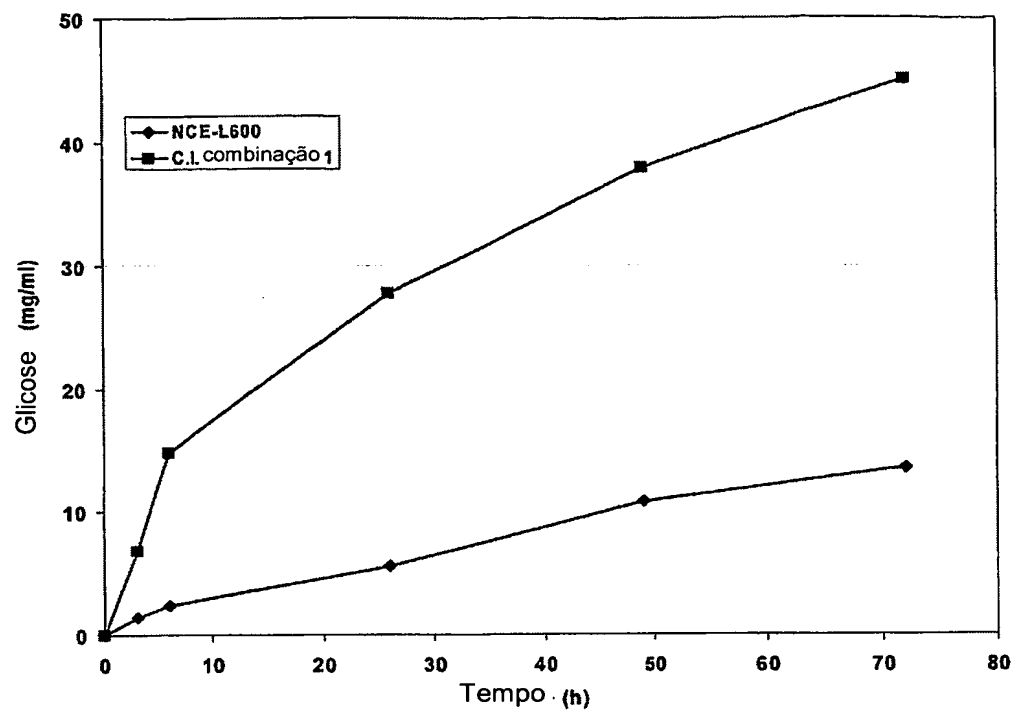


FIGURA 5

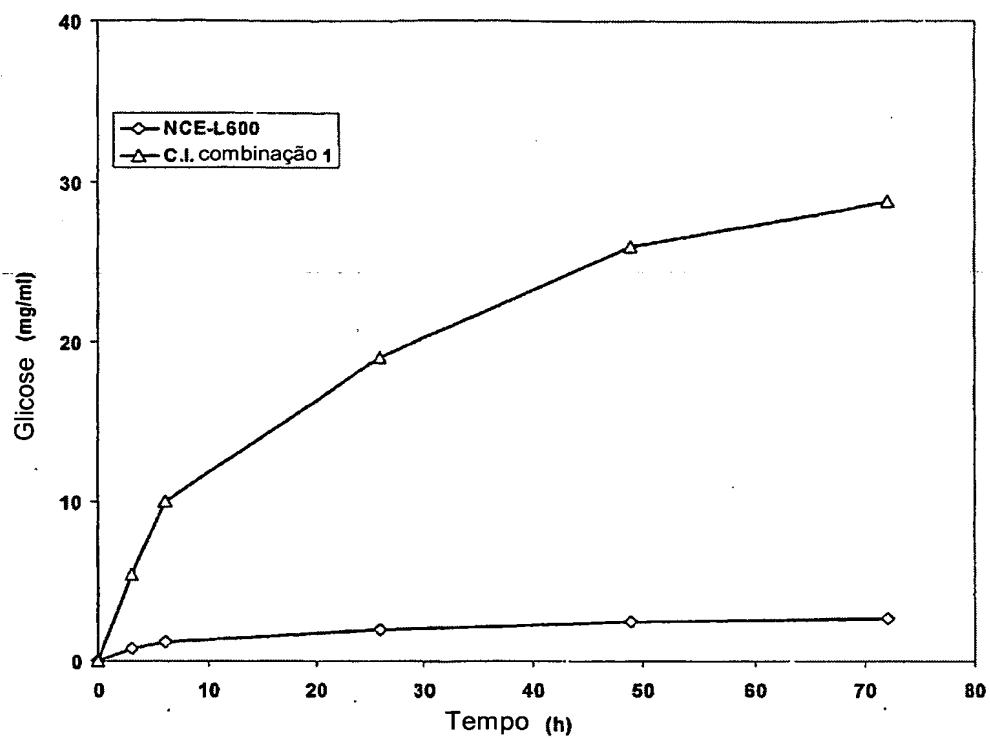


FIGURA 6

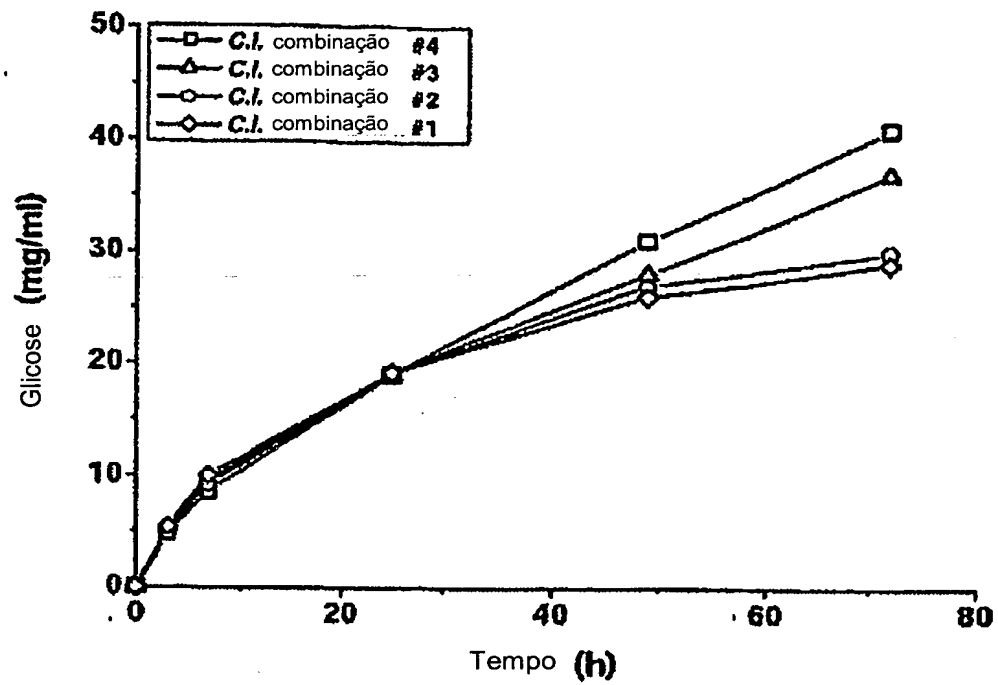


FIGURA 7

FIGURA 8

Tradução de Contig 2370 14521-20840 cbh2 (1-6360)

```

1      ctccagattcttaggggtagggcgaggagcagaggcgaaaattgggttgtagaatatgaggag
61     ctagggttggttaaactcaaagaacttcttgccttcttgccttcttctcctctgggaaa
121    aggggggtttttccgaaagcggcgtatagcaagccagagggtacttctccttgccttggat
181    ggcccttgccaccgttcttgccttcccggttgcaattgcgacggttgcgggcaacctagg
241    tcctaataattaggtagatatctcggtagaggtaggttaattatgcttcagtagagaaat
301    cgttgctcaccgctcgcgaaccttgcgaaacttgcgcacattgaagatagcattgtctg
361    agttgattttaaccctttccagagacgatataatagtgcaggttcttcttgatcggaatca
421    tcgacattcggttttcccttaattatatgaagtattcggcccaacggaacggggcccgga
481    gcagggtgaaccgcgcaaaacctcaaccgagtcacctcgcgtccatggttctcatggaat
541    caggctccgaatcccgtcagatcagtcagttctggtggctatggacgcgggagttacggc
601    cagtcgctccggtgttctgggggggttgatcaacaggagggaagagatctgagatcgaaacta
661    caccattgatttatcgacgcataatcaagtttaataaaaaacaaacagcgtggttgggtg
721    ctaccaccgaatgcgagatccgggctagcccgcggaaggatgatggccacagatctagcg
781    tcatgtatgattattacctatgcatctatcttctgctatctgcctcgggttggcaacacctg
841    accgagagacgactcgacaacctgacacttggcaaaagacatttccggttgacagcgggag
901    aactccagcgagggaagtgcgccagagatgcggatgagaagacaacgcggagcgtgccgg
961    cgttggctctccacgaatcggagccgactcttccggttggccaatctccgggataaatcc
1021   cagcggcggtcacgtcacgttctatggggaggcgcggaacagccatccagccagggccat
1081   ggaagagaacaattcttgggggttagcgaccgagccaaaaggggggggggggaagcgggag
1141   ggaagaagtgggtatttagagcacaacgcgcaacggaaaacgcatttgggcccttgccaacaaca
1201   ccacaccccgctcctgggagcaagacatccaggatgcaacccagtaggggatgccaaaga
1261   agcatctacggcaccatctgcccgcgcctcgcctgttagagtcggcgaccccgccaatgg
1321   ggccgtgctggggccctgcccgggcaatgctggcgagcggcatcaacaacattgctcgggg
1381   aggggcccgtattttattgattagcaaaaaaacaaattaaattacccttccattccagcaga
1441   gcttctcctccacgcggcgggcgggacgcgttggtagcggcggtacactacaaccgcgggg
1501   ctccagtcctccgtgctggggcggtgcagatcacgaccgggaagagaaatgatcgcggtctga
1561   cgccgggtacggagtagtgagccgccaaccacagccgatggaccgtgatattctcaatgcg
1621   ttcaagcaacacagcaacacccctggacgagtcctcctccctaccacccctccccc
1681   tggcctggcgcggaacggggcgcgtagccccagatttctactcgtactgacaccccaatc
1741   tattcccgttggcgctgcggcagtcctggggcggtccggccaagactctcgggtgcacgatac
1801   cgcgacgaaatcggattaaacggttgggtgatcaattccaagtcaagggagaagtggtagt
1861   gaaagtgcgctcagtttccactgccccgacaggcaggttccggatctggacagcagtc
1921   ttccgaatctttggcagagactcatgataatataaaaaggcaaatgagggcgcccttgg
1981   acaggtccattctcccaccgctcaaccagcctccaattcctcagaagtctggtgctctct
2041   cgcagtcgcagtcgaagTGAAGCAGTACCTCCAGTACCTCGCGGCGACCCTGCCCTGGT
      M K Q Y L Q Y L A A T L P L V
2101   GGGCCTGGCCACGGCCACGAGCGGGTAACCTGCAGACCGAGACTCACCCCAAGCTCAC
      G L A T A Q Q A G N L Q T E T H P K L T
2161   TTGTCGAAGTGCACGGCCCCGGGATCCTGCCAACAGGTCAACGGCGAGGTCGTTCATCGA
      W S K C T A P G S C Q Q V N G E V I D
2221   CTCCAAGTGGCGCTGGGTGCACGACGAGAACGCGCAGAAGTGTACGACGGCAACCAAGTG
      S N W R W V H D E N A Q N C Y D G N Q W
2281   GACCAACGCTTGCAGCTCTGCCACGACTGCGCGGAGAATTGCGCGCTCGAGGGTGCCGA
      T N A C S S A T D C A E N C A L E G A D
2341   CTACCAGGGCACCTATGGCGCTCGACCGGCAATGCCCTGACGCTACCTTCGTAC
      Y Q G G T Y G A S T S G N A L T L T F V T
2401   TAAGCAGGATACGGCACCAACATGGTTTCGGCGCTCTACCTCATGAACGGCGCGAACAA
      K H E Y G T N I G S R L Y L M N G A N K
2461   GTACCAGATGTTACCTCAAGGGCAACGAGCTGGCCTTCGACGTCGACCTCTCGGCCGT
      Y Q M F T L K G N E L A F D V D L S A V
2521   CGAGTGGCGCTCAACAGCGCCCTCTACTTCGTGGCCATGGAGGAGGATGGCGGTGTGTC
      E C G L N S A L Y F V A M E E D G G V S
2581   GAGCTACCCGACCAACACGGCCGGTGCTAAGTTCGGCACTGGGgtaagttcaacgacccg
      S Y P T N T A G A K F G T G
2641   agacgggtgccccttattatctgctgcgaaaacggaggtccccttttgcctaactaccctc
2701   ctccaaacagTACTGCGACGCCCCAATGCGCACGCGACCTCAAGTTCGTGGCGGCAAGGG
      Y C D A Q C A R D L K F V G G K G
2761   CAACATCGAGGGCTGGAAGCCGTCCACCAACGATGCCAATGCCGGTGTGCGTCTTATGG
      N I E G W K P S T N D A N A G V P Y G
2821   CGGGTGCTGCGCTGAGATCGACGCTCTGgtaagttttgttgctgggcagcaatgggtatat
      G C C A E I D V W
2881   tagctcgagtggttcccgtcgttgcgtgaccctctcttaccagGGAGTCGAACAAGTATGC
      E S N K Y A
2941   TTTCCGCTTTACCCCCGACGGTTGCGGAGAACCCTAAATACCACGTCGTGCGAGACCACCAA
      F A F T P H G C E N P K Y H V C E T T N

```

(FIGURA 8, continuada)

3001 CTGCGGTGGCACCTACTCCGAGGACCGCTTCGCTGGTGAAGTGCATGCCAACGGCTGCCGA
 C G G T Y S E D R F A G D C D A N G C D
 3061 CTACAACCCCTACCGCATGGGCAACGAGACTTCTACGGTCCCGGCTTGACGGTTCGATAC
 Y N P Y R M G N Q D F Y G P G L T V D T
 3121 CAGCAAGAAGTTACGtgtagtacacggtgcttgaagccccctccccccccccccccaaaa
 S K K F T
 3181 aaaaaaagaaaaaagaagtcaaatgattgatgctaaccaaatcaaataacagCGTCGTCA
 V V
 3241 GCCAGTTCGAGGAGAACAAGCTCACCCAGTTCCTTCGTCCAGGACGGCAAGAAGATTGAGA
 S Q F E E N K L T Q F F V Q D G K I E
 3301 TCCCGGGCCCCAAGGTGAGGGGTCGATGCGGACAGCGCCGCTATCACCCCTGAGCTGT
 I P G P K V E G I D A D S A A I T P E L
 3361 GCAGTGCCTGTTCAGGCGCTTCGATGACCGTGACCGCTTCCTCGGAGGTTGGCGGCTTCG
 C S A L F K A F D D R D R F S E V G G F
 3421 ATGCCATCAACACGGCCCTCAGCAGCTCCCATGGTCTCTCGTCATGTCCATCTGGGATGATg
 D A I N T A L S T P M V L V M S I W D D

 3481 tacgttacctaaccccccccccctttttttttcccgcttctctccccgaaactgcccacta
 3541 cttatatacgtccccgctccatgatgcttaccttttctccttccagCACTACGCCAATAT
 H Y A N M
 3601 GCTCTGGCTCGACTCGAGCTACCCCTGAGAAGGCTGGCCAGCCTGGCGGTGACCGTGG
 L W L D S S Y P P E K A G Q P G D R G
 3661 CCGGTGTCTCAGGACTCTGGCGTCCCGGCCGACGTTGAGGCTCAGTACCTTAATGCGTG
 P C P Q D S G V P A D V E A Q Y P N A *
 3721 agtcgaaccgtaaaatgtcgggcaaaaaaagatcgctcaagctaacgaaataatatga
 3781 ttagCAAGGTCATCTGGTCAACATCCGCTTCGGCCCCATCGGCTCGACTGTCAACGCTCT
 K V I W S N I R F G F I G S T V N V
 3841 AAactgcaacctgaccgggcccccttctctccacccccacccctctcaagttctctctggt
 3901 ggagccctcggtgctcttcttctcctaggttcgcaacctttgagcttggtgtatcgtaggg
 3961 tcattgtgtacatacacaaaaacttaacatctgctaccaagatcttggcgctttgcccagg
 4021 tcttctcaaacctcgaagcactgagcctttgtctccgagtgaaagtgaggtgactattta
 4081 cgttgcaagactacgcggttaaaggggacggagcagacctgccacagatattcggttggtt
 4141 gcttgatttatagcagagtcggaacgttagacatggccccctgaaggtgccaaccttagata
 4201 gccagaagccttggtttacgaaaggggtggtcaaccaacgggtgctcctcgctcagcgaatc
 4261 taccgcacgcaatgtatcgtaagaatgtgaactaaagggaaacgagcgaggtataggga
 4321 cgtcaatgtggcttgaataacagagttaaatacctaataagaagaatttagcatgccaga
 4381 ttgagccagcaacacatggttagaatagccagcaaaaggacgcttggttcgcttgatctcgaa
 4441 ccgtccaacctgattcgaaggaggagggaaggtgaagaataccggcaataattactcg
 4501 aggttccctatgcccctcagagtcgaacttaataatgaaggacaccccgatctccgca
 4561 attataagcataataagctcgccgggccccacacgtgcttccacctcccattggtatata
 4621 atctgtacctcggttattgtcgaatcgctattccgatagcgaaggtctggcactcatcaga
 4801 taccgtgacatcgattgagatttgccggggccaccggtagtaagcgatgagttggtcatc
 4681 aattatcaacaatcgctcaatcagcgataatcagcctatcaaccgcggaatcatcagcg
 4741 catcaacgaattgtccatcatgcacgtagcttggtcggcagtgccgcataccctccagagc
 4861 atcatagccgggatagaaagctcgcttccagcgtcccagagtcgagatgcaggttagca
 4921 agccttcaagaccagttatatgtgaccgggtaaaataacttggtgagatgcaatggcgct
 4981 agcttcgggcaactataagcttctactagatattatctcaaggttctttttgaactctc
 5041 ctagacatttactataaaactaccgagcttcaatgcttagacgcccctccttctgttaaatag
 5101 tcttttcttctaagagcatctgcctttttcccttaggcttagaggataggccccctcc
 5161 atcttgctgcgacggccttagccttggggagtaattatttggtatcccggtacctgtttcc
 5221 cagacagccgaagtttcgacgacaaaagtaattatttgcgacaataccaccgccatagcta
 5281 ttccgagtggtgagccccgaaaacatcgcttaccgcatcgccatcccagacgacagagg
 5341 gcgactttgatgtcttgctccagatcgccgcacctaacacgggtgggatgggctgggtatcg
 5401 tatgggacggcatcatggtcaacaacccccctgacgggggtgtggggcaatggaaacacca
 5461 ccgttgctctcgagccggatcgcaaggtgaagccgaagaggacaaatgacgatgagactttc
 5521 tttctttttttattttatttttttttaaatctttttttaagcgtaatagaaaagagctaca
 5581 tatctgtggttcggttcctcaatttcagcgacctctccaccgaagcatcgtaaaataagaa
 5641 gttgtcggaacaaaaggggtgcagaagctatagagcttcttaaggatattagccacatata
 5701 tgccatagctgtataaggtattttaacgcttttgccagctcctttgtctataaatattag
 5761 tcggtttgtctcctttgttagataattttaacaaggcaactcttttctttatagccacc
 5821 tactatagactgctttcaacgctcccgaagcttattactacgttcggcagttataagcc
 5881 tggcgcttgactactcctctgcgacgtatctttaatatagtagtagcttcttctatt
 5941 acgaactctcttaccctgctttaatacgttttcgacgacgtgtctattatcttaagatc
 6001 ctagtgcgacactctatatgccttactaggcctagtctcttagaacttgtagtatattaaa
 6061 ctatagttataggctaaatttgctagtatatagagatttggttaaccttaatagtaattat
 6121 aaactagatctagaagttttatagtgccataacctataaataagctagagataaccttatt
 6181 ttagcttcttaggagtaattctagaaggagattacctttaatatctatagatttgata
 6241 ccttctaataatagctatcatagctaaatttatataattataagatttcttttataaaaaat

6301	attatatactatagatattagtaagtagataggatagctataatactagctagtatat
------	--

FIGURA 9

SEQ ID NOs: 3 e 4 gene de cbh4 que codifica CBH IIb

```

1      cgcgaagtgaatatgtaattactcaatggaagtctctgaaacggagtcagaaatgatgt
61     gggtctgtgggaatgcggcaagaggcgacgttgccgtgaatgcgtgaacattcccgcctc
121    ttcttcttctcgtcttcttcttcttcttcttcttctcggtcgcggttggttgacggccagcg
181    tgcgcacggctgcgtgttatcgagcgtcggtacgtctagccaacatcccgtagacacgac
241    gaccaagcgtcttgagaatgcaacaacgtctcggaacctggcacgcattctccgccgcag
301    gtccggcagacgcgcctgggcaataaccacccctgtccaggccctttcccgcaggcagag
361    ccgcgctcttcttcttcatgggtatttcaggaacgtggcttccgagattctcgcctgttctc
421    ccccagtcacctgacgacgttaaccgggtccaccacgcggactgtccgcaaaacctg
481    gttcgcccgagattaatatgctatttccggactaagtgcacaacacacaagcacccttc
541    cgcctcgcgctctagaatctgctttctaaacccggttctcgggcccttcccttctcgcaag
601    cctccgctctccttaccaggcaccatccgcaataggtaaggtagccaacctgtttggagc
661    gtgattctgccaaggacgcgcatccttgcattegccatctggtcaaggacccctctttccc
721    gctccattctggtggctctatcgggacggcgttccccatggctctccaggagagtgatgt
781    gcgagctctggagagccggggttggcgtcacgatgctgcccacctagggccggccagcccg
841    gcactgcgctcccggttgatcgtctatccccgtcaagagcaccagccccggcgctcgtga
901    attttcgacttggttcgacttgctacagggtgataaaggagatgcacgccgcctcgcgcg
961    cctgtgtggtttctctccctcgtgcaaaaccactcccacctcccgcggcgagatagttgc
1021   ttgtttcgctccgtgagagggacacacaccaATGGCCAAGAAGCTTTTCATCACCGCCGC
                                     M A K K L F I T A A

1081   GCTTGCGGCTGCCGTGTTGGCGGCCCCCGTCATTGAGGAGCGCCAGAACTGCGGCGCTGT
361     L A A A V L A A P V I E E R Q N C G A V

1141   gtggaagaaagcccgggtccgagtcctccatgattttctcgtcgagtaatggcataaggg

1201   ccaccccttcgactgaccgtgagaatcgatcaaatccagGACTCAATGCGGCGGTAACGG
                                     T Q C G G N G

1261   GTGGCAAGGTCCCACATGCTGCGCCTCGGGCTCGACCTGCGTTGCGCAGAACGAGTGGTA
      W Q G P T C C A S G S T C V A Q N E W Y

```

FIGURA 9 (CONTINUADA)

SEQ ID NOs: 3 e 4 gene de cbh4 que codifica CBH IIb

1321 CTCTCAGTGCCTGCCCCAACAGCCAGGTGACGAGTTCACCACTCCGTCGTCGACTTCCAC
 S Q C L P N S Q V T S S T T P S S T S T

1381 CTCGCAGCGCAGCACCAGCACCTCCAGCAGCACCACCAGGAGCGGCAGCTCCTCCTCCTC
 S Q R S T S T S S S T T R S G S S S S S

1441 CTCCACCACGCCCCCGCCGTCCTCCAGCCCCGTGACCAGCATTCGCGGCGGTGCGACCTC
 S T T P P P V S S P V T S I P G G A T S

1501 CACGGCGAGCTACTCTGGCAACCCCTTCTCGGGCGTCCGGCTCTTCGCCAACGACTACTA
 T A S Y S G N P F S G V R L F ██████████

1561 CAGGTCCGAGGTCCACAATCTCGCCATTCTAGCATGACTGGTACTCTGGCGGCCAAGGC
 ██████ S E V H N L A I P S M T G T L A A K A

1621 TTCCGCCGTCGCCGAAGTCCCTAGCTTCCAGTGGCTCGACCGGAACGTCACCATCGACAC
 S A V A E V P S F Q W L D R N V T I D T

1681 CCTGATGGTCCAGACTCTGTCCCAGGTCCGGGCTCTCAATAAGGCCGGTGCCAATCCTCC
 L M V Q T L S Q V R A L N K A G A N P P

1741 CTATGCTGgtgagttacatggcgacttgcccttctcgccccctacctttcttgacgggac
 Y A

1801 gggtacctgacctggaggcaaaacaacaacagCCCAACTCGTCGTCTACGACCTCCCCGA
 A Q L V V Y D L P D

1861 CCGTGA CTGTGCCGCCGCTGCGTCCAACGGCGAGTTTTTCGATTGCAAACGGCGGCGCCGC
 R D C A A A A S N G E F S I A N G G A A

1921 CAACTACAGGAGCTACATCGACGCTATCCGCAAGCACATCATTGAGTACTCGGACATCCG

FIGURA 9 (CONTINUADA)

SEQ ID NOs: 3 e 4 gene de cbh4 que codifica CBH IIb

N Y R S Y I D A I R K H I I E Y S D I R
 1981 GATCATCCTGGTTATCGAGCCCGACTCGATGGCCAACATGGTGACCAACATGAACGTGGC
 I I L V I E P D S M A N M V T N M N V A
 2041 CAAGTGCAGCAACGCCCGCTCGACGTACCACGAGTTGACCGTGACGCGCTCAAGCAGCT
 K C S N A A S T Y H E L T V Y A L K Q L
 2101 GAACCTGCCCAACGTGCCCATGTATCTCGACGCCGCCACGCCGGCTGGCTCGGCTGGCC
 N L P N V A M Y L D A G H A G W L G W P
 2161 CGCCAACATCCAGCCCCGCCGAGCTGTTTGCCGGCATCTACAATGATGCCGGCAAGCC
 A N I Q P A A E L F A G I Y N D A G K P
 2221 GGCTGCCGTCGCGGCCCTGGCCACTAACGTCGCCAACTACAACGCCTGGAGCATCGCTTC
 A A V R G L A T N V A N Y N A W S I A S
 2281 GGCCCCGTCGTACACGTCGCCTAACCCCTAACTACGACGAGAAGCACTACATCGAGGCCTT
 A P S Y T S P N P N Y D E K [REDACTED]
 2341 CAGCCCGCTCTTGAACCTCGGCCGGCTTCCCCGCACGCTTCATTGTGACACTGGCCGCAA
 [REDACTED] F I V D T G R N
 2401 CGGCAAACAACCTACCGgtatggttttttttcttttgtctctgtcccccttttctccc
 G K [REDACTED]
 2461 ccttcagttggcgtccacaaggtctcttagtctctgcttcattctgtgaccaacctcccccc
 2521 ccccggcaccgcccacaaccgtttgactctatactcttgggaatggcgccgaaactgac
 2581 cgttcacagGCCAACAACAGTGGGGTGACTGGTGCAATGTCAAGGGCACCGGCTTTGGC

FIGURA 9 (CONTINUADA)

SEQ ID NOs: 3 e 4 gene de cbh4 que codifica CBH IIb


G T G F G

2641 GTGCGCCCGACGGCCAACACGGGGCCACGAGCTGGTCGATGCCTTTGTCTGGGTCAAGCCC
 V R P T A N T G H E L V D A F V W V K P

2701 GGCGGCGAGTCCGACGGCACAAGCGACACCAGCGCGCCCGCTACGACTACCACTGCGGC
 G G E S D G T S D T S A A R Y D Y H C G

2761 CTGTCCGATGCCCTGCAGCCTGCCCCGAGGCTGGACAGTGGTTCCAGGCCTACTTCGAG
 L S D A L Q P A P E A G Q W F Q A Y F E

2821 CAGCTGCTCACCAACGCCAACCCGCCCTTCTAAacctcgtcataaagagagagagatggc
 Q L L T N A N P P F *

2881 gggcatgggcctgattgggttcattgaccatgcggtccttctgggggtacatatatttacc
 2941 tacctacctataaataaggcggcctatcggtcctcgcttcttattaggtactgttc
 3001 ttgtacatactttgtttatacatagcagttagcatccactattcggttcgacaaagcg
 3061 gaactttccagaaaaaaaaagggtgtacataattagtccttttaggcttcgattccttgtgc
 3121 ctttcttttggtaaaaaaaaaatttttttgaggcatgattaccttaggtacgttcgtc
 3181 gttgtattggccccctgcattttggcgagagagcagctcagcccccttgcaaaccctca
 3241 acgggcgttcaattccctccactcgggtcttcagcgagaccagccgtccagagtatcca
 3301 gcgtgtagttgccccacgaaccagtcgtcctcgtaagcctcgtaaagtgtccaagagca
 3361 gtatagaagcaacgacctccgtcaaaagtctggcaccatcgatcggtggtcctccccg
 3421 tgcgccccgcccctcgtaggacttctcatccacgccaaggagcagctgcaggccgtcggac
 3481 gtcgccccggggtgcgccttgaagttgtaccattcgctcctccagacgcgctccagctgc
 3541 gcctgcttgggttcctgcggttcctgcggttcctgcgctggccggtcggcgcgcgctct
 3601 tggtcacacgcccgcagcgacatgactgggtgtttcgggtcgagcagcttgacgagcccg
 3661 acctggggttcgggtggttgcgaacacggcgccaatgaggtggccgtaccattcggtat
 3721 gactgcatggcgaagctggcgagtgtagccacgatcccgccgcccgcctggacgaaa
 3781 ccccgagggcgcccagctgcgcgcgtccaggaactcgccgagcactgcaggaggacg
 3841 atgacgcgatacgccgagaggagccggggctgaacacggcgggatcctcgctgtcgctcc

FIGURA 10

SEQ ID NOs: 5 e 6 gene de cbh1 que codifica CBH Ia

ATGTACGCCAAGTTCGCGACC 1800
 M Y A K F A T

CTCGCCGCCCTTGTGGCTGGCGCCGCTGCTCAGAACGCCTGCACTCTGACCGCTGAGAAC 1860
 L A A L V A G A A A Q N A C T L T A E N

CACCCCTCGCTGACGTGGTCCAAGTGACGCTCTGGCGGCAGCTGCACCAGCGTCCAGGGT 1920
 H P S L T W S K C T S G G S C T S V Q G

TCCATCACCATCGACGCCAACTGGCGGTGGACTCACCGGACCGATAGCGCCACCAACTGC 1980
 S I T I D A N W R W T H R T D S A T N C

TACGAGGGCAACAAGTGGGATACTTCGTACTGCAGCGATGGTCCTTCTTGGCGCTCCAAG 2040
 Y E G N K W D T S Y C S D G P S C A S K

TGCTGCATCGACGGCGCTGACTACTCGAGCACCTATGGCATCACCACGAGCGGTAACCTCC 2100
 C C I D G A D Y S S T Y G I T T S G N S

CTGAACCTCAAGTTCGTACCAAGGGCCAGTACTCGACCAACATCGGCTCGCGTACCTAC 2160
 L N L K F V T K G Q Y S T N I G S R T Y

CTGATGGAGAGCGACACCAAGTACCAGagtaagttcctctcgacccggccgcccgggaga 2220
 L M E S D T K Y Q M

tgatggcgcccagcccgtgacgcgaatgacacaGTGTTCCAGCTCCTCGGCAACGAGTT 2280
 F Q L L G N E F

CACCTTCGATGTCGACGTCTCCAACCTCGGCTGCGGCCTCAATGGCGCCCTCTACTTCGT 2340
 T F D V D V S N L G C G L N G A L Y F V

GTCCATGGATGCCGATGGTGGCATGTCCAAGTACTCGGGCAACAAGGCAGGTGCCAAGTA 2400
 S M D A D G G M S K Y S G N K A G A K Y

CGGTACCGGCTACTGTGATTCTCAGTGCCCCGCGACCTCAAGTTCATCAACGGCGAGGC 2460
 G T G Y C D S Q C P R D L K F I N G E A

CAACGTAGAGAACTGGCAGAGCTCGACCAACGATGCCAACGCCGGCACGGGCAAGTACGG 2520
 N V E N W Q S S T N D A N A G T G K Y G

CAGCTGCTGCTCCGAGATGGACGTCTGGGAGGCCAACAACATGGCCGCCGCTTCACTCC 2580
 S C C S E M D V W E A N N M A A A F T P

CCACCCCTTGCNCCGTGATCGGCCAGTCGCGCTGCGAGGGCGACTCGTGCGGCGGTACCTA 2640
 H P C ? V I G Q S R C E G D S C G G T Y

CAGCACCGACCGCTATGCCGGCATCTGCGACCCCGACGGATGCGACTTCAACTCGTACCG 2700
 S T D R Y A G I C D P D G C D F N S Y R

CCAGGGCAACAAGACCTTCTACGGCAAGGGCATGACGGTCGACACGACCAAGAAGATCAC 2760
 Q G N K T F Y G K G M T V D T T K K I T

FIGURA 10 (CONTINUADA)

SEQ ID NOs: 5 e 6 gene de cbh1 que codifica CBH Ia

GGTCGTCACCCAGTTCTCTCAAGAACTCGGCCGCGAGCTCTCCGAGATCAAGCGGTTCTA 2820
V V T Q F L K N S A G E L S E I K R F Y
CGTCCAGAACGGCAAGGTCATCCCCAACTCCGAGTCCACCATCCCGGGCGTCGAGGGCAA 2880
V Q N G K V I P N S E S T I P G V E G N
CTCCATCACCCAGGACTGGTSCGACCGCCAGAAGGCCGCTTCGGCGACGTGACCGACTT 2940
S I T Q D W C D R Q K A A F G D V T D ?
NCAGGACAAGGGCGGCATGGTCCAGATGGGCAAGGCCCTCGCGGGGGCCCATGGTCCTCGT 3000
Q D K G G M V Q M G K A L A G P M V L V
CATGTCCATCTGGGACGACCACGCCGTC AACATGCTCTGGCTCGACTCCACCTGGCCCAT 3060
M S I W D D H A V N M L W L D S T W P I
CGACGGCGCCGGCAAGCCGGGCGCCGAGCGCGGTGCTGCCCCACCACCTCGGGCGTCCC 3120
D G A G K P G A E R G A C P T T S G V P
CGCTGAGGTGAGGGCCGAGGCCCACTCCAACGTCATCTTCTCCAACATCCGCTTCGG 3180
A E V E A E A P N S N V I F S N I R F G
CCCCATCGGCTCCACCGTCTCCGGCCTGCCCGACGGCGGCAGCGGCAACCCCAACCCGCC 3240
P I G S T V S G L P D G G S G N P N P P
CGTCAGCTCGTCCACCCCGGTCCCCTCCTCGTCCACCACATCCTCCGGTTCCTCCGGCCC 3300
V S S S T P V P S S S T T S S G S S G P
GACTGGCGGCACGGGTGTCGCTAAGCACTATGAGCAATGCGGAGGAATCGGGTTCACTGG 3360
T G G T G V A K H Y E Q C G G I G F T G
CCCTACCCAGTSCGAGAGCCCCTACACTTGACCAAGCTGAATGACTGGTACTCGCAGTG 3420
P T Q C E S P Y T C T K L N D W Y S Q C
CCTGTAA
L *

FIGURA 11

SEQ ID NOs: 7 e 8 gene de eg6 que codifica CBH IIa

```

ggatccacac ctaccataacc ggatagtatg ctaccaagt gacatagggg tggtaaagta 60
atacgagaac tcagagagca ctgcccataat ggctcgccaa tgacctcaag tgccagggtca 120
gctttgagag acagacctga gcgcgtcgga tgtgtgacat ggaacgcgcc ggatcgccctt 180
gttgattaat tatagggaag tagcgaggaa ggtttcagca attgacgtga gcgtacatta 240
aaagctgtat gatttcagga agacgagcca tggaccagggt ttcaaggctg aatggcttga 300
cgacttaagc accgaacgag gaatgaaaga atgaaaagtg ggggatcatt ctggcccttc 360
ctcgtatgtc gagtggttaa gaaggcgggt ctacggagga cctaaagagc tccaatttgc 420
tctgttgagc ttaagccaca tatctcaaga tgaatacatg tcaggcatag tcacctgat 480
cttgttcac cagtcacaca cttttcagtt cagcatgttg attcctcatc catatcactt 540
tccattacta tctctttatg tccttggtca agactccaag gaaccgatag gtgagcatcg 600
gtgaggtccc ctcaaggtac caaagtagcc atcatcaccg aggtctggga atggcgccgt 660
gcccgatctg agtcctccaa ctccacggta cgacgacagc acgtcacatt gacgcaccac 720
gggtgaacaa gcagagaggg acacgtcttg ctacgcgaat cctggcactg gatggagacg 780
cgtgtgagca gggttcggga accatgacgg cctggtcggg cttctcgaa aaagaagtgg 840
aacacaaaaa gaaccgaaac ggaaacgcag gcacggcatc gacgaccgga ttgtccacg 900
gggacctcgg ccagtcaagc gttgccctgg ccgtcagctc cctggcgagc gggattcagc 960
acatctcacg ttataggcga cctcatcccc ctccgtctt gtgcggctgt tgctccgtgc 1020
cgagtacca ggctgcccgg ggccttttagc cggggcgga tcagagtcaa gatgcggccg 1080
aattggacgg cagacgaagt ttcgtagagg gtcgatcg gcactgacga caccacccc 1140
tgctgatcc cgtggccctg ggctgggaat tgccggctaa taatctacgg cttaatagat 1200
atgcactttg cagcggtgtc agataaataa gctgtggtt caaacactgg cctccgtact 1260
ttaccacca actgcccgtt agcgcgggga cctgagctt gggagtgcgc ggagcggcag 1320
ccacctcggg ttagcgtaca cagacgggt gcatgcgggg atgcccgtg catggcttca 1380
tagtgtacga cagaccgtca agtccaaatc tgggtgatgc ttgatgagat gacagcgagc 1440
cccgtcgggg gcaccccggc tatgcatcgc gaattgacaa cactctcagc tctattgcga 1500
cccatcggat aaaagaagaa gaaaaaatg gaccttgagt acgggcgtca gaaacaaaa 1560
aaaaactccg gaaccaaata tgtcgggcgt ggcgggggtg aacgaccgct actccccgtt 1620
cccttcttog caaacagaac gctacagagg gttttctggt ttgtcaaaga gttcggaggt 1680
cctctgtcc gcgaatgctt ggtgaaccca ccagcagcca ttgttcttgc atgctggcg 1740
gaccgttagc cgtgatcga catggcgagc tcccacctc agacctggag cagacgggtg 1800
cgaggagcaa ggggctgccc tccccctgac ggtcggaccc caatgacttc cccaaacggg 1860
gacatcgagg gtcgtgcatg atggtggaaa gtagttgcag tatgggaagt accccgggtt 1920
gccaggaacc gttgttcggc cccccacatt ttctctctgc catgtcaact gtgtgtcgtt 1980
cgagagttcc tggctcgggc ccccggtcca attccctaac gggaccgcgg ggcatcgcc 2040

```

FIGURA 11 (CONTINUADA)

SEQ ID NOs: 7 e 8 gene de eg6 que codifica CBH IIa

```

gtaactaact tccaaatgaa gccggatatg agggagggag attggatctg gcaagccagc 2100
cattcgctgc gatcggcact cgtccgtcag ccccgagtc catatcccca aaggcaactg 2160
ctcggcgcgcg ctcaagtctt cttcgggaacg tccagcccga aggcgcgcgc cagcaccggc 2220
cctatgttcc tgattgcgat cctcgatctc cagagacggg tcacctcgcc tcgaggacgg 2280
tgcaggggca toggcttcgc ttcctagagc tccgggctgt gtgtggtcaa ggggagaagg 2340
cggcgcgcgcc aaggtgcgtc tcggcgact caccatcgc ctttaccccc ctcccccca 2400
gtatataaaa gatggccatc gtctcctcgt ctgcttggga agaaaggatc tctcgaccat 2460
gcaccacagc ctagctctaa cccagcttgt cgtgtgttgt tgcccagc atg aag ttc 2517
                                     Met Lys Phe
                                     1
gtg cag tcc gcc acc ctg gcg ttc gcc gcc acg gcc ctc gct gcg ccc 2565
Val Gln Ser Ala Thr Leu Ala Phe Ala Ala Thr Ala Leu Ala Ala Pro
      5                      10                      15
tcg cgc acg act ccc cag aag ccc cgc cag gcc tcg gcg ggc tgc gcg 2613
Ser Arg Thr Thr Pro Gln Lys Pro Arg Gln Ala Ser Ala Gly Cys Ala
      20                      25                      30                      35
tcg gcc gtg acg ctc gat gcc agc acc aac gtg ttc cag cag tac acg 2661
Ser Ala Val Thr Leu Asp Ala Ser Thr Asn Val Phe Gln Gln Tyr Thr
                      40                      45                      50
ctg cac ccc aac aac ttc tac cgt gcc gag gtc gag gct gcc gcc gag 2709
Leu His Pro Asn Asn Phe Tyr Arg Ala Glu Val Glu Ala Ala Ala Glu
                      55                      60                      65
gcc atc tcc gac tcg gcg ctg gcc gag aag gcc cgc aag gtc gcc gac 2757
Ala Ile Ser Asp Ser Ala Leu Ala Glu Lys Ala Arg Lys Val Ala Asp
                      70                      75                      80
gtc ggt acc ttc ctg tgg ctc gac acc atc gag aac att ggc cgg ctg 2805
Val Gly Thr Phe Leu Trp Leu Asp Thr Ile Glu Asn Ile Gly Arg Leu
                      85                      90                      95
gag ccc gcg ctc gag gac gtg ccc tgc gag aac atc gtg ggt ctc gtc 2853
Glu Pro Ala Leu Glu Asp Val Pro Cys Glu Asn Ile Val Gly Leu Val
      100                      105                      110                      115
atc tac gac ctc ccg ggc cgt gac tgc gcg gcc aag gcc tcc aac ggc 2901
Ile Tyr Asp Leu Pro Gly Arg Asp Cys Ala Ala Lys Ala Ser Asn Gly
                      120                      125                      130

```

FIGURA 11 (CONTINUADA)

SEQ ID NOs: 7 e 8 gene de eg6 que codifica CBH IIa

gag ctc aag gtc ggc gag ctc gac agg tac aag acc gag tac atc gac a	2950
Glu Leu Lys Val Gly Glu Leu Asp Arg Tyr Lys Thr Glu Tyr Ile Asp	
135 140 145	
gtgagttaac cctttgtggc cccttctttt ccccgagag agcgtctggt tgagtggggt	3010
tgtgagagag aaaatggggc gagcttaaag actgacgtgt tggctcgag ag atc	3065
Lys Ile	
gcc gag atc ctc aag gcc cac tcc aac acg gcc ttc gcc ctc gtc atc	3113
Ala Glu Ile Leu Lys Ala His Ser Asn Thr Ala Phe Ala Leu Val Ile	
150 155 160 165	
gag ccc gac tcg ctc ccc aac ctg gtc acc aat agc gac ctg cag acg	3161
Glu Pro Asp Ser Leu Pro Asn Leu Val Thr Asn Ser Asp Leu Gln Thr	
170 175 180	
tgc cag cag agc gct tcc ggc tac cgc gag ggt gtc gcc tat gcc ctc	3209
Cys Gln Gln Ser Ala Ser Gly Tyr Arg Glu Gly Val Ala Tyr Ala Leu	
185 190 195	
aag cag ctc aac ctc ccc aac gtg gtc atg tac atc gat gcc ggc cac	3257
Lys Gln Leu Asn Leu Pro Asn Val Val Met Tyr Ile Asp Ala Gly His	
200 205 210	
ggg ggc tgg ctc ggc tgg gac gcc aac ctc aag ccc ggc gcc cag gag	3305
Gly Gly Trp Leu Gly Trp Asp Ala Asn Leu Lys Pro Gly Ala Gln Glu	
215 220 225	
ctc gcc agc gtc tac aag tct gct ggt tcg ccc tcg caa gtc cgc ggt	3353
Leu Ala Ser Val Tyr Lys Ser Ala Gly Ser Pro Ser Gln Val Arg Gly	
230 235 240 245	
atc tcc acc aac gtg gct ggt tgg aac gcc tg gtaagacact ctatgtcccc	3405
Ile Ser Thr Asn Val Ala Gly Trp Asn Ala Trp	
250 255	
ctcgtcggtc aatggcgagc ggaatggcgt gaaatgcatg gtgctgacct ttgatctttt	3465
ccccctocta tag g gac cag gag ccc ggt gag ttc tcg gac gcc tcg gat	3515
Asp Gln Glu Pro Gly Glu Phe Ser Asp Ala Ser Asp	
260 265	
gcc cag tac aac aag tgc cag aac gag aag atc tac atc aac acc ttt	3563
Ala Gln Tyr Asn Lys Cys Gln Asn Glu Lys Ile Tyr Ile Asn Thr Phe	
270 275 280	

FIGURA 11 (CONTINUADA)

SEQ ID NOs: 7 e 8 gene de eg6 que codifica CBH IIa

ggc gct gag ctc aag tct gcc ggc atg ccc aac cac gcc atc atc gac	3611
Gly Ala Glu Leu Lys Ser Ala Gly Met Pro Asn His Ala Ile Ile Asp	
285 290 295 300	
act ggc cgc aac ggt gtc acc ggt ctc cgc gac gag tgg ggt gac tgg	3659
Thr Gly Arg Asn Gly Val Thr Gly Leu Arg Asp Glu Trp Gly Asp Trp	
305 310 315	
tgc aac gtc aac ggc gcc ggc ttc ggt gtg cgc ccg act gcc aac act	3707
Cys Asn Val Asn Gly Ala Gly Phe Gly Val Arg Pro Thr Ala Asn Thr	
320 325 330	
ggc gac gag ctc gcc gac gcc ttc gtg tgg gtc aag ccc ggt ggc gag	3755
Gly Asp Glu Leu Ala Asp Ala Phe Val Trp Val Lys Pro Gly Gly Glu	
335 340 345	
tcc gac ggc acc agc gac tcg tcg gcg gcg cgc tac gac agc ttc tgc	3803
Ser Asp Gly Thr Ser Asp Ser Ser Ala Ala Arg Tyr Asp Ser Phe Cys	
350 355 360	
ggc aag ccc gac gcc ttc aag ccc agc ccc gag gcc ggt acc tgg aac	3851
Gly Lys Pro Asp Ala Phe Lys Pro Ser Pro Glu Ala Gly Thr Trp Asn	
365 370 375 380	
cag gcc tac ttc gag atg ctc ctc aag aac gcc aac ccg tcc ttc	3896
Gln Ala Tyr Phe Glu Met Leu Leu Lys Asn Ala Asn Pro Ser Phe	
385 390 395	
taagctcctc gacggcttct tgctgtcagt cgctctgacg gtggtgtgct ggtggtgccc	3956
ctgctcctgc tgctgtctgct ccgcggggag gggagggaac gaaaatgaag tcctgcttca	4016
aaacaaaaca gaaacaagcg aggcgcgggtg caatggtcgt gcgttcgtct tttttcatgt	4076
tcccttctag tgtagtagtt tgatagtcgt acataagggg tttcagaacc gtctctctgt	4136
ctcggctcttt ttgcgagttg ttgcgactcg tgattatggc ctttgttgct cgttgcgcca	4196
gagtagaacc acagcgtgtt ggggtagcag cttgctccgt aggacgtagg gaaacaacct	4256
gagactctgg aattgcagtc agcctgcgtc gccctctag gaaacgaagg ggagaaccag	4316
tagtggtctgc agcttacaaa cgcgagcatg gtgaacatct ccgagaaaag ggagggatcc	4376

FIGURA 12

SEQ ID NOs: 9 e 10 gene de eg2 que codifica EG II

```

1      tgctgctctgatgtgetgatgcacagcttccccctcgcgattgcccggcaggatctccaacc
61     ctctggatcggagcagacgatcagcgggcacaaatggccagcttgccagcgttcaactcca
121    agttgacccgcttttatcacgcccagctggacatgcacaggcttggtctctcgtgttcc
181    tacgatctgcacagtaggtttgactgctgatcttcgctttcctgtgcgccccctccccctcc
241    ctcacgggtaccttatccttgctgtaacccccgcgttatgtcaaaacttgagtttgaccaa
301    tgctagcgcaaaagtacctacatagtactatgtaataaggtaggtacatacatcagtagg
361    cgtttatctagtaaattttggctttttgaaactcaattgctcctctcctcgccctccacct
421    ctgcttggaattgacaacctggctgtgcctagaggtagcatcgacgatcaatcaaatc
481    taaagtattcgagattgacctttctgctctaatttatattaattatccgcacaatgctgta
541    gtcattgactctcctttcaagttgccttctcggttatgtatgtacaatggcggtcatgc
601    ttcatgccaacagatggttctatcggaacaatgtttgactttctggctcgccccgctcgac
661    tgttttgatctcgacgggaagtgttcttaccaaagctaagtcgactcgtggagcttcgt
721    aacggccagtgatcgttgatcgcttttgaggaggtgcatggagcgagaccggctacga
781    gcacgttcgcaaaggcagcagcatagacgacccctccgtggcgccattcgggagatgcaca
841    tgacataagcatatcaatactcacctgaactcatcggccgatgcctcgaggttagttaca
901    agacataatttggtggtggtatattatcccaacccgtacctttgtcgcgctcatttcggtatg
961    tgctgatgcctacttagggagcaaagacgcctctcctcacctgcgggttacttacttact
1021   gtgcagcatggccttatgttctccgggtcttgcttgcgcgaaatgaacaaaaacgcccga
1081   agaaaagccgcttcttcgagttgtgtctacccgaacataagaggttattgtcgagaccg
1141   ccagcaaatgtcaacaacccacccacggcggtccagaaccttcgaaatatcatctagttt
1201   aagtttaaatgacggcccgagtcaccagccgagattcccatattggccgataccagcggtc
1261   ccttggtttttccaaggttgctctcgtaactggcgcatctgcctacaacgagatataatta
1321   ccgttttcttttgcaaaagggcatgcatggatgtatattatttatgcctgcagaacgaga
1381   agcaatcatggtgtaggttttggtcggtatggagctaataatattgaacggatctctggt
1441   ccgtcctaaatcgttgaaacgctaggcccaggaggacctgctcgacttggcgaacggaga
1501   tttccaggatgaaaggctcggaacatgtccatccgcggccagcctgaacacttttgctcgt
1561   ttccggaccatcgacccacgaaaaacagtgcggttgctggcacagtacgactcacgatgg
1621   cgatggtccagcccgttcccgcccgatgccacttgacgcgcaactctccttcattcggc
1681   ggccccggcggtgtctggcctattagtagcgttttggtatccggcttggtcgccgcgcgg
1741   ttttcttgccgatacgggaatctcggtggtcccaactccacctgggcacgctctggtg
1801   ccaacatggaaacttcgggatgccgctccgggcacagtcaagcgctttaaatacgaacttt
1861   accccacaagaatcgaggcgtaacccggaattagggacacctggacggcgcaacccctgg
1921   accgaagggcctcgctaaccgggttcctggagccgcatgcgcggctgcccgttgccgcg
1981   tcttgagatgacacttcttttcagcgagggatggtcgggcagggaaatgatgtattataa

```

FIGURA 12 (CONTINUADA)

SEQ ID NOs: 9 e 10 gene de eg2 que codifica EG II

2041 gaagcgagccgattccgaaggactcgacccccctctctcgccctgtgtccgccagctaatt
 2101 acagcactccttctcgacttgaaacgccccgagATGAAGTCCTCCATCCTCGCCAGCGTCT
 700 M K S S I L A S V

2161 TCGCCACGGGCGCCGTGGCTCAAAGTGGTCCGTGGCAGCAATGTGGTGGCATCGGATGGC
 720 F A T G A V A Q S G P W Q Q C G G I G W

2221 AAGGATCGACCGACTGTGTGTCTGGGTTACCACTGCGTCTACCAGAATGATTGGTACAGCC
 740 Q G S T D C V S G Y H C V Y Q N D W Y S

2281 AGTGCCTGCCTGGCGCGCGTCGACAACGCTCCAGACATCTACCACGTCCAGGCCCACCG
 760 Q C V P G A A S T T L Q T S T T S R P T

2341 CCACCAGCACCGCCCTCCGTGCTCCAACCTCGCCTAGCAAGGGCAAGCTCAAGTGGC
 780 A T S T A P P S S T T S P S K G K L K W

2401 TCGGCAGCAACGAGTCGGGCGCCGAGTTCGGGGAGGGCAACTACCCCGCCTCTGGGGAA
 800 L G S N E S G A E F G E G N Y P G L W G

2461 AGCACTTCATCTTCCCGTCGACTTCGGCGATTTCAGgtacggggccaataataatatattat
 820 K H F I F P S T S A I Q
 2521 tatagcaggcaggagggagcaggagaagaaggaggggcaggtggccaacaatcggaaga
 2581 agaccgggaggcactgaccgttgattcctttgtgtaatagACGCTCATCAATGATGGATA
 861 T L I N D G Y

2641 CAACATCTTCCGGATCGACTTCTCGATGGAGCGTCTGGTGCCCAACCAAGTTGACGTCGTC
 881 N I F R I D F S M E R L V P N Q L T S S

2701 CTTTCGACGAGGGCTACCTCCGCAACCTGACCGAGGTGGTCAACTTCGTGACGAACGCGGG
 901 F D E G Y L R N L T E V V N F V T N A G

2761 CAAGTACGCCGTCCTGGACCCGCACAACTACGGCCGGTACTACGGCAACGTCATCACGGA
 921 K Y A V L D P H N Y G R Y Y G N V I T D

2821 CACGAACGCGTTCCGGACCTTCTGGACCAACCTGGCCAAGCAGTTCGCCTCCAACCTCGCT

FIGURA 12 (CONTINUADA)
SEQ ID NOs: 9 e 10 gene de eg2 que codifica EG II

```

941      T N A F R T F W T N L A K Q F A S N S L

2881    CGTCATCTTCGACACCAACAACGAGTACAACACGATGGACCAGACCCTGGTGCTCAACCT
961      V I F D T N N E Y N T M D Q T L V L N L

2941    CAACCAGGCCGCCATCGACGGCATCCGGGGCCGCCGGCGGACCTCGCAGTACATCTTCGT
981      N Q A A I D G I R A A G A T S Q Y I F V

3001    CGAGGGCAACGCGTGGAGCGGGGCTGGAGCTGGAACACGACCAACACCAACATGGCCGC
1001     E G N A W S G A W S W N T T N T N M A A

3061    CCTGACGGACCCGCGAACAAGATCGTGTACGAGATGCACCAGTACCTCGACTCGGACAG
1021     L T D P Q N K I V Y E M H Q Y L D S D S

3121    CTCGGGCACCCACGCCGAGTGCCTCAGCAGCAACATCGGCGCCCAGCGCGTCGTGCGAGC
1041     S G T H A E C V S S N I G A Q R V V G A

3181    CACCCAGTGGCTCCGCGCCAACGGCAAGCTCGGCGTCTCGGCGAGTTCGCCGGCGGCGC
1061     T Q W L R A N G K L G V L G E F A G G A

3241    CAACGCCGTCTGCCAGCAGGCGCTCACCGGCCTCCTCGACCACCTCCAGGACAACAGCGA
1081     N A V C Q Q A V T G L L D H L Q D N S E

3301    GGTCTGGCTGGGTGCCCTCTGGTGGGCGCGCGGTCCCTGGTGGGGCGACTACATGTACTC
1101     V W L G A L W W A A G P W W G D Y M Y S

3361    GTTCGgtaagtttctcccttggtcttggctttccccccagtaagggagtcagggaacat
1121     F

3421    gcccaagaccggctcggcttcgcttcaaggcggttcggtgtacacactgaagagttccaac
3481    ttccaaccctggtcgtgtcctccgatcagcttcgacggggtgaaggggaagggatttgg
3541    gagtgaggtggaggtcaaaaggagggatatccccagatctccacaaacggccctgagcca
3601    acaacagcctctgggggtcaaaatgggcgccaaccatacgggtcattcactcaggacacctg
3661    ctaacgcgtctcttttttttggttccagAGCCTCCTTCGGGCACCGGCTATGTCAACTAC
1221                                     E P P S G T G Y V N Y

3721    AACTCGATCCTAAGAAGTACTTGCCGTAAggggcatgcagcaaggtcgagcgagcatta

```

FIGURA 12 (CONTINUADA)

SEQ ID NOs: 9 e 10 gene de eg2 que codifica EG II

```

1241      N S I L K K Y L P *
3781      ttcagggccatctgcttgtgtcggcaggcatcacgtcaacccatcgaatcggacagcggg
3841      atgctccgagatgccatacactaagtctggtgatgacgtgagaatgctggccctggtcgg
3901      gggttaccgccaacaaaaagcaccggacgtgccgcgcccggataccatggtttcatgt
3961      acatattggttctttgctttcttacgggggggggggggggggggctctgcagcgttgc
4021      tgagcgattcgtttccaagtataactttgtctggaattgaattttgagtacattgacc
4081      caatcaaccagctcgggtgtgtcacctcccgttacccccctcttctccccctgctcggc
4141      ttggctttctctccggtgtggagcacggccacggcgggtcccaatccatataagatcgat
4201      ggtatactatggtatacactagcttgggaataaaactaatccatacgctaactaatggacg
4261      gattatcctaaggggtcaccggctcaccgttggatataacacctaggatacgggagagctg
4321      atagaaagggatgtactccgtattgtactgtacaatacaaagtacagatagcacacgaag
4381      tacggtaggtggtcccgcctagtcgggaccaacaatagaacatgcgttcttggggacctg
4441      caggaaagaaggggggggggttgccaagacgcccggggttcaaagaaagcccgggcgg
4501      ccgatgagatgagacggacgcccggccaaggagaggccgggtggtcgatcctgcaaagcc
4561      agcaaaaaaatccataccataatccagtcactttcgtcacactcctgtgaaacgagct
4621      ggagggactgctggaaagggttttgacaggttaatcactgtatgtggagcatgccgtacct
4681      ctgtgcttcgttaacagatagagttccagttgaacacacaaaagtcttgcgccgcctgcc
4741      gacgtgaaaagaagctcctccgggggagctttaggcaactgggagggctctctcccaggt
4801      tcatggtgtctgctcttcttcaaattttatgctgccaccccatttgacagaggtgtgca
4861      caccgttgccaggtcttgccatccggcaaaaagcagaaaagtcgacccatcgccctaagaa
4921      aggcgggtcggaaggggatcgatgctcattgagggttagcgtctgccattctgacgctg
4981      cccattgttttgtgtgcattcgtcttcggatgtcggatcaagagtcgccgatttttcc
5041      cctgtgcttccagcctaattctgagcgggagctgggtcgggttcgagtgagggtgccttgt
5101      tgggtggagcagcaaccagccaattcactccccgcattttcgcgggccgccaggcatccc
5161      cggcatgcgtttgggcggttaactactccgtactggggtagggtgaaattggttctcccgtc
5221      gcaggaggctcgtgctcggtcaggggagaaacaaagtccaactgctccttctggcaacaa
5281      tgagaggggggttctattgccaacgttgacgaaaggagcagccacaaaacccaaaagcag
5341      gttaccttactgtacctgagcttgaacgtcgcgtagcattggagctctcgtctaccggcg
5401      gogtcacactccattggcaggtcaaggcagtcagtggcagcgacccaacaacgtcaatgc
5461      ttgttaccacagaattacccggggtgcaacactgcagggggccgcgcgcatggtgatca
5521      ccggttgattacttctcgcccgcaaccgggagatgagaagcagaactttgttctccttt
5581      caaaaaggacctgacttgcggggaacgcactgccggcagtgagggtggatgcacgctagtt
5641      atatgtttccgcccatcccagtcgcgccgtcgcgctcgtgagggtcagtttggttccc
5701      gtgccgcgacaaaagagcgggtgcataattacatttcgctccatgtaccgtgcaccctcc
5761      ccgttcgcgaccgtagta

```

FIGURA 13

SEQ ID NOs: 11 e 12 gene de bgl1 que codifica BGL

```

1      cggcctccagttccaggagcttggctctgccgacatactgtgtacactaggaattctct
61     tatgcggggtgtgcgcggggaaatgttggggaactcgagttgggtcatgtggacaagacc
121    aatgggagctgacatcattgtgcgacccgttaaaccggaagctacaacaacattctggat
181    tctacactagtgggaagaggttaagtaattgacgacaagcaagaagcattgccatgttctgc
241    gaaggatgcggggtgtttttgcatgagcaggaagctgtggctttttagtgtcctttgtgc
301    tcgccccggcgcgagaacactaccgaaacgcaggggaactgcgctgcctctggggtcgaatg
361    ccgatccccatcttcacattcccaccatcgtgttctgttaacgaagccggagcggcgga
421    actcgaagctccactacgtatggatacttgggaccgtacggagtgtgttggtacggatgc
481    ctgcacaagtgttgtgtctcctacgaagacgccaacccacataatacaciaaagctgttg
541    taagtcgagttacctcaggcacgttcgggcaactcgggcaacctgacgagatttccccgc
601    cattccgccaagaggccggcgccctgcctgattaggcagctcttggaacaatactatgta
661    gaatggaagctccatccatagtccagctccattggcgggtcccagtgatctcgatggctgga
721    tggctgctctgtacgggtacatacatagtaagttctcgcttgagagcccaattcgctgca
781    atagcatctttccccgcagtgcgccggcgccctgggtcccgcctccacaatgacctgtct
841    tctggagcttctcgacgaacagatcggcccggttcttctccacaccaatccgaaccagtcc
901    gggagcatggctgcggtgacgacgcagccttccttcgcgctgtacaaaacagctccgggaa
961    cgtcgactggtatgtacggactacagtaagtaactacgagtgcacatactgacgaatac
1021   cggcctcagaggaacctggcaggacctacccacacgaaaccacagcgagaaagcgcaa
1081   tggatcagtaactactgcgaagtaaacgtgggtcccgggcaaaggatctgagggccgatcg
1141   ctcggtggggctgcgaggcgaggagagcaaacaagccagtcctcccgcgaacctggaaaa
1201   tcaattataaacacacgtcaccggcgccgggtgcgcgccatgtgtcacctccaggctcc
1261   tcccgggcgatgatctctgcgggtgccatcaatcatctcggttcgcccgcagctgcttctt
1321   tctgtgcagtgaacgctctcaaaactgcaacgcagcgtgtccgacatgaaggctgctgcgct
1381   ttctgcctcttcggcagtaaccttgccgttgaggcgccattgaatcgagaaaggatg
1441   gacgggctttcgtcaaagactcgctccccgatcaacttcccccttccatccagaccacccc
1501   aacctcccagtcctgcttcgagcacgatctcttcgggcagcaccacccacatccact
1561   cagattagcggcgacaccgttgactgttgcaatccgcaatcgacATGCAACTTCCAGCCG
                                     M Q L P A
1621   CAGCCCAATGGCTGCTCACGCTTCCCGCGAAAGCCTCACTTGCTGACAATCATCGTCAGG
A A Q W L L T L P A K A S L A D N H R Q
1681   TTCACCAGAAGCCCCTCGCGAGATCTGAACCTTTTTACCCGTCGCCATGGATGAATCCCA
V H Q K P L A R S E P F Y P S P W M N P
1741   ACGCCGACGGCTGGGCGGAGGCTATGCCCAGGCCAAGTCCTTTGTCTCCCAAATGACTC
N A D G W A E A Y A Q A K S F V S Q M T

```

FIGURA 13 (CONTINUADA)

SEQ ID NOs: 11 e 12 gene de bgl1 que codifica BGL

1801 TGCTAGAGAAGGTCAACTTGACCACGGGAGTCGGgtaagttttgtcattttgtccaggta
 L L E K V N L T T G V G

C1 Bgl1 236 for

1861 acatgcaaagtgttctgctaacaataaacttaccgtagCTGGGGGGCTGAGCAGTGCCTCG
 W G A E Q C V

1921 GCCAAGTGGGCGCGATCCCTCGCCTTGGACTTCGAGTCTGTGCATGCATGACTCCCCTC
 G Q V G A I P R L G L R S L C M H D S P

1981 TCGGCATCCGAGGAGCCGACTACAACCTCAGCGTTCCCCTCTGGCCAGACCGTTGCTGCTA
 L G I R G A D Y N S A F P S G Q T V A A

2041 CCTGGGATCGCGGTCTGATGTACCGTCGCGGCTACGCAATGGGCCAGGAGGCCAAAGGCA
 T W D R G L M Y R R G Y A M G Q E A K G

2101 AGGGCATCAATGTCTTCTCGGACCAGTCGCGGCCCCCTTGGCCGCATGCCCGAGGGCG
 K G I N V L L G P V A G P L G R M P E G

2161 GTCGTAACCTGGGAAGGCTTCGCTCCGATCCCGTCTTACCGGCATCGGCATGTCCGAGA
 G R N W E G F A P D P V L T G I G M S E

C1Bgl1 682 rev

2221 CGATCAAGGGCATTTCAGGATGCTGGCGTCATCGCTTGTGCGAAGCACTTTATTGGAAACG
 T I K G I Q D A G V I A C A K H F I G N

2281 AGCAGGgtgagtagtcaaagacggggccgtctcggacccgcggcttcaagtgctgactct
 E Q

2341 gctgcagAGCACTTCAGACAGGTGCCAGAAGCCCAGGGATACGGTTACAACATCAGCGAA
 E H F R Q V P E A Q G Y G Y N I S E

2401 ACCCTCTCCTCCAACATTGACGACAAGACCATGCACGAGCTCTACCTTTGGCCGTTTGCC
 T L S S N I D D K T M H E L Y L W P F A

2461 GATGCCGTCCGGGCCGGCGTCCGCTCTGTCTGTGCTCGTACCAGCAGGTCAACAACTCG
 D A V R A G V G S V M C S Y Q Q V N N S

2521 TACGCCTGCCAGAACTCGAAGCTGCTGAACGACCTCCTCAAGAACGAGCTTGGGTTTCAG
 Y A C Q N S K L L N D L L K N E L G F Q

2581 GGCTTCGTCTATGAGCGACTGGCAGGCACAGCACACTGGCGCAGCAAGCGCCGTGGCTGGT
 G F V M S D W Q A Q H T G A A S A V A G

2641 CTCGATATGTCCATGCCGGGCGACACCCAGTTCAACACTGGCGTCAGTTTCTGGGGCGCC
 L D M S M P G D T Q F N T G V S F W G A

2701 AATCTCACCCCTCGCCGTCCTCAACGGCACAGTCCCTGCCTACCGTCTCGACGACATGGCC
 N L T L A V L N G T V P A Y R L D D M A

FIGURA 13 (CONTINUADA)

SEQ ID NOs: 11 e 12 gene de bgl1 que codifica BGL

2761 ATGCGCATCATGGCCGCCCTCTTCAAGGTCACCAAGACCACCCACCTGGAACCCATCAAC
 M R I M A A L F K V T K T T H L E P I N
 2821 TTCTCCTTCTGGACCGACGACACTTATGGCCCGATCCACTGGGCCGCCAAGCATGGCTAC
 F S F W T D D T Y G P I H W A A K H G Y
 2881 CAGAAGATTAATTCCACGTTGACGTCCGCGCCGACCACGGCAACCTCATCCGGGAGATT
 Q K I N S H V D V R A D H G N L I R E I
 2941 GCCGCCAAGGGTACGGTGCTGCTGAAGAATACCGGCTCTCTACCCCTGAACAAGCCAAAG
 A A K G T V L L K N T G S L P L N K P K
 3001 TTCGTGGCCGTCATCGGCGAGGATGCTGGGTCGAGCCCCAACGGGCCAACGGGTGCAGC
 F V A V I G E D A G S S P N G P N G C S
 3061 GACCGCGGCTGTAACGAAGGCACGCTCGCCATGGGCTGGGGATCCGGCACAGCCAACTAT
 D R G C N E G T L A M G W G S G T A N Y
 3121 CCGTACCTCGTTTCCCCGACGCGCGCTCCAGGCCCGGGCCATCCAGGACGGCACGAGG
 P Y L V S P D A A L Q A R A I Q D G T R
 3181 TACGAGAGCGTCTGTCCAACCTACGCCGAGGAAAAGACAAAGGCTCTGGTCTCGCAGGCC
 Y E S V L S N Y A E E K T K A L V S Q A
 3241 AATGCAACCGCCATCGTCTTCGTCAATGCCGACTCAGGCGAGGGCTACATCAACGTGGAC
 N A T A I V F V N A D S G E G Y I N V D
 3301 GGTAACGAGGGCGACCGTAAGAACCTGACTCTCTGGAACAACGGTGATACTCTGGTCAAG
 G N E G D R K N L T L W N N G D T L V K
 3361 AACGTCTCGAGCTGGTGCAGCAACACCATCGTCGTATCCACTCGGTCCGCCCGGTCTCTC
 N V S S W C S N T I V V I H S V G P V L
 3421 CTGACCGATTGGTACGACAACCCCAACATCACGGCCATTCTCTGGGCTGGTCTTCCGGGC
 L T D W Y D N P N I T A I L W A G L P G
 3481 CAGGAGTCGGGCAACTCCATCACCGACGTGCTTTACGGCAAGGTCAACCCCGCCGCCCGC
 Q E S G N S I T D V L Y G K V N P A A R
 3541 TCGCCCTTCACTTGGGGCAAGACCCGCGAAAGCTATGGCGCGACGTCTGTACAAGCCG
 S P F T W G K T R E S Y G A D V L Y K P
 3601 AATAATGGCAATGGTGCGCCCAACAGGACTTCACCGAGGGCGTCTTCATCGACTACCGC
 N N G N G A P Q Q D F T E G V F I D Y R
 3661 TACTTCGACAAGGTTGACGATGACTCGGTCTCTACGAGTTTCGGCCACGGCCTGAGCTAC
 Y F D K V D D D S V I Y E F G H G L S Y
 3721 ACCACCTTCGAGTACAGCAACATCCGCGTCGTCAAGTCCAACGTGACGAGTACCGGGCC
 T T F E Y S N I R V V K S N V S E Y R P
 3781 ACGACGGGCACACGGGCCAGGCCCGACGTTTGGCAACTTCTCCACCGACCTCGAGGAC

FIGURA 13 (CONTINUADA)

SEQ ID NOs: 11 e 12 gene de bgl1 que codifica BGL

```

      T T G T T A Q A P T F G N F S T D L E D
3841 TATCTCTTCCCAAGGACGAGTTCCCCTACATCTACCACTACATCTACCCGTACCTCAAC
      Y L F P K D E F P Y I Y Q Y I Y P Y L N
3901 ACGACCGACCCCGGAGGGCCTCGGCCGATCCCCACTACGGCCAGACCGCCGAGGAGTTC
      T T D P R R A S A D P H Y G Q T A E E F
3961 CTCCCGCCCCACGCCACCGATGACGACCCCCAGCCGCTCCTCCGGTCCTCGGGCGGAAAC
      L P P H A T D D D P Q P L L R S S G G N
4021 TCCCCCGGCGGCAACCGCCAGCTGTACGACATTGTCTACACAATCACGGCCGACATCACG
      S P G G N R Q L Y D I V Y T I T A D I T
4081 AATACGGGCTCCGTTGTAGCGGAGGAGGTACCGCAGCTCTACGTCTCGCTGGGCGGTCCC
      N T G S V V G E E V P Q L Y V S L G G P
4141 GAGGATCCCAAGGTGCAGCTGCGCGACTTTGACAGGATGCGGATCGAACCCGGCGAGACG
      E D P K V Q L R D F D R M R I E P G E T
4201 AGGCAGTTTCACCGCGCGCTGACGCGCAGAGATCTGAGCAACTGGGACGTCACGGTGCGAG
      R Q F T G R L T R R D L S N W D V T V Q
4261 GACTGGGTGTCAGCAGGTATCCCAAGACGGCATATGTTGGGAGGAGCAGCCGGAAGTTG
      D W V I S R Y P K T A Y V G R S S R K L
4321 GATCTCAAGATTGAGCTTCCTTGAATGAGTTTCATCAGGGGCTGCAGAGGGATGGTAACA
      D L K I E L P *
4381 CGTTCTTAATCAGAAGTATGATGGAGAAAAGCACTTGGCAAGTTCGGTGAGCAAAAAGA
4441 AGGCACTTATTAAGTGTAGGGCGGTGTTCTATGTTTAATAGGTGCTATGTTTACATATAA
4501 TTAGTATATAATGATTTAATAATTATGTTTAGCAGTTGCTAATGTCGTAAATTTCCGGCGT
4561 GTGATGACTGCTACAACACTGGTTCTGTCTTCTAGTCGCCATTGTTAATTATGAAGGTTA
4621 TTGTCTACAATTCTAATACCTTATGGATGATTGCCAGCTGGTTTCAAACCTCGTTACGC
4681 GCAAATGGTACGATTGAGGTATTATTCATTGTAAGTACCTCCGTACAGCGTCCCCAACTA
4741 TTTCCATTACGAGATGCCTCGCTTTTCGGTGCTTTTCGGAACAGGGCTGGCAGCGGATCA
4801 TGGCGCGATCAAAACATGGCGAGCAGCTGTCCAGGACGGAGGACAGGTTGGGGACTGATG
4861 CCTCCCGGACGCATTAAGGTGAGAAGATAGACACGTTTACACAGCGTTGAGACCGACAA
4921 GCCACATTAGGCAGCGCCGGTTGCACCACCGCCGTACGGGCAACGGTTCAATCAATCGA
4981 CAACAGTGGAAGACAAAGTACTGAAGATCAGGTATTAATAGTGTGAGAGAGAAACAGACG
5041 GTGGAAGTAGGGTGCTAATATTTCTCTTGATTTCGGTGCTCCATGGTAGTACAGAACACAA
5101 GAAAAAGAAGGAGGAGTGAGCGGAGAGGAGGAGGGGGAAGCCAGAAAAAAGAACATGAA
5161 AAAGCATACACATTGGAGTCGGTCAGTCGGTTGATTGGTTTGGTAGAGAGCGAAAAAGCA
5221 AGCGTCACCTGTAGGATTCGAACCTACGCTCCCGAAGGAACTGCCTAAGAACGCTAAGCA

```

FIGURA 13 (CONTINUADA)

SEQ ID NOs: 11 e 12 gene de bgl1 que codifica BGL

```
5281      AGGTTAGCAGGGCAGCGCGTTAACCCTCCGCCAAAGTGACTGTCGTTGATCATGGTCGA
5341      ATTCAGTAGCTTATAGGAGTTCAACCAGATCACAATGCATAGGTGCTCGTAGAACGGT
5401      CTAAGTATGAGTTGATTATAAGCAACCGAATGGCTCTCAGCGGCAACACCGTAGCTGAAG
5461      TAACAAAACGCACCTTTGGTTACTTTCTGACTATAAAAATGGGATATTTGGAAATGACCA
5521      CCCGATAAGGTGTCAAATTTCTAAATGACTGTCTGGGTGTGAAGATGTTACTGTGGTTCCA
5581      CCACGAACCAGTTTTAGTATCCGCATGCTTCAGTCTCTGCGCCTCGACAGGCGGAGGGTG
5641      TGTGTTAGATCAGAATCGATGTGACGCTGTGACCGCGAGGCTCTCGAGCCTAGGTGCGGT
5701      AGTTCTGTTCAAAAAGAAGTGTGTGGCCGGGTTTGGGCGCCCTTATAGCCTACCATCCTG
5761      GCTGTGGTTCCCGAGCGGGAGCCGTTCTCCGTTTTGGTTCCGATAAAGTGTCATATCTG
5821      CCTCCCGGTTTCGCATCTAATTTCTGACTTCGTTCCGGGACCTCTGGAGACGTAGGGATAG
5881      GTATGGGATATGCCCGGCATTTCTGTAATGTCCATAGTCTCTTTCGGGACGAGGCGGCAA
5941      GCTCTCAGAGCTATCTAAGCTTAACCAACCCCTGATCCTTAACCTCCAGACCACACCT
6001      CCTGGGAGAATAAACCGGGCTCCAAGATCGAAATCGAAATCAGTGCGGCAACTTGAAATC
```

FIGURA 14

SEQ ID NOs: 13 e 14 gene de eg5 que codifica EGV

atg cat ctc tcc gcc acc acc ggg ttc ctc gcc ctc ccg gcc ctg gcc	48
Met His Leu Ser Ala Thr Thr Gly Phe Leu Ala Leu Pro Ala Leu Ala	
1 5 10 15	
ctg gcc cag ctc tcc ggc agc ggc cag acg acc ccg tac tgg gac tgc	96
Leu Ala Gln Leu Ser Gly Ser Gly Gln Thr Thr Arg Tyr Trp Asp Cys	
20 25 30	
tgc aag ccg agc tgc gcc tgg ccc ggc aag ggc ccc tcc tct ccg gtg	144
Cys Lys Pro Ser Cys Ala Trp Pro Gly Lys Gly Pro Ser Ser Pro Val	
35 40 45	
cag gcc tgc gac aag aac gac aac ccg ctc aac gac ggc ggc tcc acc	192
Gln Ala Cys Asp Lys Asn Asp Asn Pro Leu Asn Asp Gly Gly Ser Thr	
50 55 60	
ccg tcc ggc tgc gac gcg ggc ggc agc gcc tac atg tgc tcc tcc cag	240
Arg Ser Gly Cys Asp Ala Gly Gly Ser Ala Tyr Met Cys Ser Ser Gln	
65 70 75 80	
agc ccc tgg gcc gtc agc gac gag ctg tcc tac ggc tgg gcg gcc gtc	288
Ser Pro Trp Ala Val Ser Asp Glu Leu Ser Tyr Gly Trp Ala Ala Val	
85 90 95	
aag ctc gcc ggc agc tcc gag tcc cag tgg tgc tgc gcc tgc tac gag	336
Lys Leu Ala Gly Ser Ser Glu Ser Gln Trp Cys Cys Ala Cys Tyr Glu	
100 105 110	
ctg acc ttc acc agc ggg ccg gtc gcg ggc aag aag atg att gtg cag	384
Leu Thr Phe Thr Ser Gly Pro Val Ala Gly Lys Lys Met Ile Val Gln	
115 120 125	

FIGURA 14 (CONTINUADA)

SEQ ID NOs: 13 e 14 gene de eg5 que codifica EGV

gcg acc aac acc ggt ggc gac ctg ggc gac aac cac ttt gac ctg gcc	432
Ala Thr Asn Thr Gly Gly Asp Leu Gly Asp Asn His Phe Asp Leu Ala	
130 135 140	
gtgagttgcc tccccctctc cccggaccgc tcagattaga tgagattaga ctttgcctgt	492
aaatcggtcc aagattccct tgactgacca acaaacatca tacgggcag atc ccc ggt	550
Ile Pro Gly	
145	
ggc ggt gtc ggt att ttc aac g gtaagctggt gccccggac ccctcccg	602
Gly Gly Val Gly Ile Phe Asn	
150	
acccctcccc ctttctctcc agcgagccga gttgggatcg ccgagatcga gaactcacac	662
aacttctctc tcgacag cc tgc acc gac cag tac ggc gct ccc ccg aac	711
Ala Cys Thr Asp Gln Tyr Gly Ala Pro Pro Asn	
160 165	
ggc tgg ggc gac cgc tac ggc ggc atc cat tcc aag gaa gag tgc gaa	759
Gly Trp Gly Asp Arg Tyr Gly Gly Ile His Ser Lys Glu Glu Cys Glu	
170 175 180	
tcc ttc ccg gag gcc ctg aag ccc ggc tgc aac tgg cgc ttc gac tg	806
Ser Phe Pro Glu Ala Leu Lys Pro Gly Cys Asn Trp Arg Phe Asp Trp	
185 190 195	
gtacgttgct ttgacatacc ggaacccaat tctccaacc ccccccttt tctccccaa	866
ctccgggggt agtcggaatg tcgcgactga ccctatttca g g ttc caa aac gcc	920
Phe Gln Asn Ala	
200	

FIGURA 14 (CONTINUADA)
SEQ ID NOs: 13 e 14 gene de eg5 que codifica EGV

gac aac ccg tgc gtc acc ttc cag gag gtg gcc tgc ccg tgc gag ctc	968
Asp Asn Pro Ser Val Thr Phe Gln Glu Val Ala Cys Pro Ser Glu Leu	
205 210 215	
acg tcc aag agc ggc tgc tcc cgt taa	995
Thr Ser Lys Ser Gly Cys Ser Arg	
220 225	

FIGURA 15

SEQ ID NOs: 15 e 16 eg7 que codifica EG VI

```

1      GCGCTTCCGGCCTGGGCGAGTAAATGACGGAAGCCgggccccgtccgactgcgtttgtc
61     ccaactcggaagcaggcatcgtttttggcgagggaagcgttgcaacacgcactatcg
121    ccaaggtggactcggcgcaatctggaggttcggcccgaggagacggaatccgggctgaa
181    tctgcgcaaaggctgaccctgcgatggtgggaaaatgtaaatatgtgaagttataggcat
241    ataggactcagcgatgacatgaaaattgcagaggcatgtgggatttcagcgtttggcatg
301    cattggtcggatctctcgccttgtctgatgtgatccgcgcggaggtgttccggtctctgg
361    ggaagggacccccctggccccacactgccccgcacatgcctcgccacgactcccgcg
421    cgccgaggaagaacttcgggtctttgtgacgggagattccactgagtgagcattggccaa
481    ccaagcacacaattactccgtacatacacagtacttctgactccgtaaagtaaaccgtgt
541    gtttcaaagatcgtaataccgtaacaggactccgtatctaaggtaaatttacccctgtgc
601    acggagcagaacctgaacttctccccctcttactcgagtagtcaccctactccaacca
661    gcggtctttcaactcgcaaagtctgtttataacagtgcataacctgcatttcgtatct
721    cgctagtgtaaagacgaccacacgcggacaagaaagaaaaatccaattgcccgatggct
781    cttagtttgaggacagcagcgaaggactacactgcgcgtagtgaccaggccaagaaacg
841    cgaatcgatatattaacggcaaatcaaatggattatatgccatttcgcttccgggttgcg
901    tgctcgtccgaagtctggtgccgatcgattgcgaacccccggaatcgcgggatgattcct
961    acagccgcgaaagggggggggggggggggggtctggacgggacgtgcataacttcgaa
1021   tttctagaatattgocgattgggttcccttcagccctgcgagcgcgcccccttctggaa
1081   cgcacccttcaccggttccacacacagaggacatgggtggaaatgtgtacctgacgggtg
1141   ccccttgggacagtggagaggcggatgttcggataaccatccggagccgcagtgctcgac
1201   caagatcttggcttaccatcgacaccaacatgcggactcgtccctcagtcattggagcctt
1261   ggctcgcggagcctccgttcgaagcggtatccogtccctgccagcggaggatctcgtacc
1321   gcttccgcgaactgtgaatgtcctgggtataagagcatggcgcgacctgtctcgtcagg
1381   aacggggaggaggagggttggttagggtcgcgttcggttggagattgctgagctctgag
1441   ccttcgggtccttggtacccctgcgggtccccgggtcctctctctctctctctctctc
1501   tctctctctctctctccacgctcgttcgacagacgcctccccctctctcgtctctcttccc
1561   tcgcacgtagcacactaatagtgacaccATGCGCGTCTCTAGTTTGGTCGCGGCCCTTGCT
                                     M R V S S L V A A L A

1621   ACCGGTGGTCTTTGTGCGCCGCCACGCCTAAGCCCAAGGGTCTGTCGCCCCCTGGGGCCGTG
      T G G L V A A T P K P K G S S P P G A V

1681   GACGCGAACCCCTTTCAAGGGCAAGACGCAGTTTCGTCAACCCGGCATGGGCGGCCAAGCTG
      D A N P F K G K T Q F V N P A W A A K L

```

FIGURA 15 (CONTINUADA)
SEQ ID NOs: 15 e 16 eg7 que codifica EG VI

1741 GAACAGACCAAAAAGGCGTTCCTGGCCAGGAACGACACCGTCAATGCCGCCAAGACGGAG
E Q T K K A F L A R N D T V N A A K T E

1801 AAGGTCCAGCAGACCAGCTCGTTCGTCTGGGTCTCGAGGATCGCCGAGCTCTCCAACATC
K V Q Q T S S F V W V S R I A E L S N I

1861 GACGACGCCATCGCGGCTGCCCCGAAGGCGCAGAAGAAGACGGGCAGGAGGCAGATCGTC
D D A I A A A R K A Q K K T G R R Q I V

1921 GGCTGGTGCTCTACAACCTTCCGACCGCGACTGCAGCGGGCGAGAGCGGGCGGAG
G L V L Y N L P D R D C S A G E S A G E

1981 CTCAGCAGCGACAAGAACGGGCTCGAGATCTACAAGACTGAGTTCGTCAAGCCCTTCGCC
L S S D K N G L E I Y K T E F V K P F A

2041 GACAAGGTGGCGGCCGCAAAGGACCTCGACTTCGCCATCGTCCTGGAGCCCGACTCGCTG
D K V A A A K D L D F A I V L E P D S L

2101 GCCAACCTGGTCAACCACTGGGCATCGAGTTCGCGCCAACGCCGCCCCCTCTACCGC
A N L V T N L G I E F C A N A A P V Y R

2161 GAGGGCATCGCCTATGCCATCTCCAGCCTTCAGCAGCCAAACGTGCACTTGTACATCGAT
E G I A Y A I S S L Q Q P N V H L Y I D

2221 GCTGCCCACGGCGGCTGGCTCGGCTGGGACGACAACCTGCCGCTGGCCGCCAAGGAGTTT
A A H G G W L G W D D N L P L A A K E F

2281 GCCGAGGTGGTCAAGCTTGCCGGCGAGGGCAAGAAGATCCGCGGCTTCGTACCAACGTG
A E V V K L A G E G K K I R G F V T N V

2341 TCCAACTACAACCCCTTCCACGCCGTCGTGCGCGAGAACTTTACCGAGTGGAGCAACTCG
S N Y N P F H A V V R E N F T E W S N S

2401 TGGGACGAGTCTCACTACGCCTCCTCGCTCACACCGTTCCTCGAGAAAGAGGGGCTGCCG
W D E S H Y A S S L T P F L E K E G L P

FIGURA 15 (CONTINUADA)
SEQ ID NOs: 15 e 16 eg7 que codifica EG VI

2461 GCACGCTTCATCGTCGACCAGGGTCGCGTTGCCCTCCCGGGAGCCCGCAAGGAGTGgtga
A R F I V D Q G R V A L P G A R K E W

2521 gtttcgaccagattgaccctcgacccatgcgaccgagattgctgacgattgaattgctg

2581 tcccggtccccagGGGTGAATGGTGGCAACGTGGCACCCGCCGATTGGGCCCCGCGCCCA
G E W C N V A P A G F G P A P

2641 CGACCAGGGTCAACAACACCGTCGTCGATGCTCTCGTCTGGGTCAAGCCTGGCGGCGAGA
T T R V N N T V V D A L V W V K P G G E

2701 GCGACGGCGAGTGTGGCTTGGCTGGCGCCCCCAAGGCCGCCAGTGGTTCGACGAGTACG
S D G E C G L A G A P K A G Q W F D E Y

2761 CCCAGATGCTGGTCGAGAATGCCACCCGTCGTCTCGTCCACAAGTGGTAGataaattttg
A Q M L V E N A H P S V V H K W *

2821 gagtccgagaaggggtcccagatagacttttggttttaaaacaaaatgcaaggtgtcgacag

2881 atactggcttaacattaaccaagcaccatgaacatgacttgtcaacatattgatacatte

2941 cgctgctttcccatagctgtctcaggtctcagggatcaaattggataggtcggtaatgca

3001 aaacgatccattggatatccagaagagagaaaaaaaaaaggacatgcatgccttgtctgt

3061 catcatgaggaaacaaaggaacaaacgatcgctcgtgttccaacaagctttccaagac

3121 cacaagaccatccaccaacacaacaaacgacaagcaatacgaatggaccgctgtgttc

3181 catctctcaagagctgactaaacgaacagtcgttgaaatcatcctacatgagtacgccgc

3241 accaocctgttatcgtgtaaaccaaatcgctgttaaagtgcacatctcttaggtatgat

3301 cgtaagttccgggtcacgggtcacggatcagggatgggtctcaattcgtgtgtcgcgtagcc

3361 gccgcggtatctggacaagacttcttgattgtctccgaaccgcttttgccgcctaata

3421 atctgtagccttcttacctgggtggtgccttgaaagacgcggcaggcaaaccttcgcagg

3481 ctgtggcgccaccagcaccaggctgtggtgatgccccggaaccggctcgtcgacttgctcgc

3541 ggtgtcctcggctggtggggatgggggtgatgagggcttgagggtgtgttgccgccgc

3601 aacatccgggtccgggtccggaccgtccacagacattggacctgcgagcatgactcgtgc

3661 cttcagccagaccaaagccatgccatcatcgctctgcccgcgctgttgagcgggaggct

3721 gatgttctcagccagaactcggggctgtacggccatgacctgggctgttcgggtctggcc

3781 gtcttgccggcggtttctccctgccagcttggtgtgcgcgggtgctgcgagattcgacttc

FIGURA 15 (CONTINUADA)
SEQ ID NOs: 15 e 16 eg7 que codifica EG VI

```
3841      gacctgggcgtggcagaggggtgacgagggacgttgacgccttgatctccttgctcccat
3901      gtcttccaccctacagggcgacgggtgccatacgcgtccacagcctgcacgagaacct
3961      cagggcgtcgtcaatgagttctgtcaacttgctctccagcctctctatgccgcgagcatc
4021      ctgatcctggagcagaaaaccgtgccgagcctccgaggaaacgctccttcagcttccgcgc
4081      gtagtttaggcgtgattcaacaaacgtccggcgggactcgttggtgcccgcagcagcgac
4141      gtccttgatgctgaagcccgctcggcgaacaggcgcatcatctgggccc
```

FIGURA 16

SEQ ID NOs: 17 e 18 gene de xyl2 que codifica Xyl II

```

1      cgcgccccgtctttgaacgcttgagaagcgacggtgaagaaccatcaactccgattcc
61     gctcctcatcctcccacgaagccgattgaaatagccacagcggctatgtacggattactc
121    tgctcgtttgcacatccatacacagcgctatttttaaagttcaggacggccaagcccg
181    gttcttggaacggacgacccggattccgaaagctccagcgctcaatgcggtcagtcgtgg
241    cgctgatcctgctgatctgctgatctcataaaccgcgaacttcaacttttctactttgaag
301    cgtatacacgcagcgctctttcacccggcgattcactactcgaaattaaccgctaatat
361    cctcgcaacttgataatgtgtagccgacacggaggagggggttgggggggggttggggg
421    gagacatgatggtctgcccacggatattattattttgttgtttgtataattactgcgg
481    caacattctcaaaggggcggtgcctcgccggcggaagcccatgacagagaattggacag
541    ctccaagctcgcgatatactctaacaacggcggtgactcggcaatgaaggcctgccgctcg
601    agtgatagggcgaagtaaaacggacggttacatgcggcacttagccggctgatgccggaga
661    atacgggattcaacgatacaatcacacgatgcgacacacctcggcgacttggcgctctat
721    ggaagaaggctgggttaaagctggcgtagattttgcgcgtcttggtttcttaaccgggtt
781    atttctatttctcatatgccgcgagcgaatgcggggtgcagagcgcccgaggatcgatgg
841    tcctatcagacaagagcctggccccggaaacctgggataatagaagccaaattaagccatg
901    ggagtatcgtccgggggtaggaaccgcacgggcaactagaggaggaagaatttggtataa
961    agggaggacggcggaacaggcttgatggacatgaatcagaagacgacactgggcaactaa
1021   acagcttgacagcagagttttgtgccttgcataggccctcgatatcATGGTCTCGTTCACT
                                           M V S F T
1081   CTCCTCCTCACGGTCATCGCCGCTGCGGTGACGACGGCCAGCCCTCTCGAGGTGGTCAAG
361    L L L T V I A A A V T T A S P L E V V K

1141   CGCGGCATCCAGCCGGGCACGGGCACCCACGAGGGGTACTTCTACTCGTTCTGGACCGAC
381    R G I Q P G T G T H E G Y F Y S F W T D
1201   GGCCGTGGCTCGGTGCACTTCAACCCCGGGCCCCGCGGCTCGTACAGCGTCACCTGGAAC
401    G R G S V D F N P G P R G S Y S V T W N
1261   AACGTCAACAACCTGGGTTGGCGGCAAGGGCTGGAACCCGGGCCCCGCGCGCAAGATTGCG
421    N V N N W V G G K G W N P G P P R K I A
1321   TACAACGGCACCTGGAACAACATAACGTGAACAGCTgtgcgttgctcctcctctttctcc
441    Y N G T W N N Y N V N S
1381   Ctttcgcttggttttcttgatgattgggatccatttttaaagagaaggaaaaaaaaaaca
      C1 xyl10 423 for
1441   aaggaaaatagaagataactaacgccaagctctggcagACCTCGCCCTGTACGGCTGGAC
      Y L A L Y G W T

```

FIGURA 16 (CONTINUADA)

SEQ ID NOs: 17 e 18 gene de xyl2 que codifica Xyl II

1501 TCGCAACCCGCTGGTCGAGTATTACATCGTGGAGGCATACGGCACGTACAACCCCTCGTC
 501 R N P L V E Y Y I V E A Y G T Y N P S S
 1561 GGGCACGGCGCGGCTGGGCACCATCGAGGACGACGGCGGCGTGTACGACATCTACAAGAC
 521 G T A R L G T I E D D G G V Y D I Y K T
 1621 GACGCGGTACAACCAGCCGTCATCGAGGGGACCTCCACCTTCGACCAGTACTGGTCCGT
 541 T R Y N Q P S I E G T S T F D Q Y W S V
 1681 CCGCCGCCAGAAGCGCGTCGGCGGCACTATCGACACGGGCAAGCACTTTGACGAGTGGAA
 561 R R Q K R V G G T I D T G K H F D E W K
 C1 Xyl10 722 rev
 1741 GCGCCAGGGCAACCTCCAGCTCGGCACCTGGAACTACATGATCATGGCCACCGAGGGCTA
 581 R Q G N L Q L G T W N Y M I M A T E G Y
 1801 CCAGAGCTCTGGTTCGGCCACTATCGAGGTCCGGGAGGCCTAAagaagccaggcgccctt
 601 Q S S G S A T I E V R E A *
 1861 cttttgttttgcaggaggggtagagggggggggagggaacgaaaagtagcagggg
 1921 ggttttatgccggcagccgtgggcatcgagtgcaacctgtatctctctctcccaag
 1981 tctccgggctccttctcagagaacttcaatatgtctggggacaaaccaccttgtgaaata
 2041 caacggttaattatctaagtttgagtgccctatcgatgcttctgaaaatttctgctcct
 2101 tgatacaagtcggtttgagccgagccaatgagactgtgtcgattgatagaggccctgaag
 2161 gatcaagcgcgatgcaacaattaagcatgactacgtgcctagctgcagataaatggaagc
 2221 cactaccaaggtcaaccccgcatactggcacgtaagaaccttccgtgtacaaggcccaa
 2281 ccgactcacatatctatctgcttgggttttgggatgcggttttttaccacaaaaaaat
 2341 ttgatacaaatgctctgctgtgccgggttgctgagaccaagccgtaaatcagcgggcaggg
 2401 aatcgagtaggtcacgcctgttgccttggtctagaacaaactaatattaaaaagccttgtg
 2461 ctcggcacacatacagaactcgacctgaggcatgttcttggaggcgctagccagtgcaa
 2521 gtctggcaccaggccttggctcgtcgaggataccgaggcgaggaggatgaggaagacc
 2581 tctttctcgctcagatctcttaggggacgaagaagacaacgccggagccacacaataat
 2641 taggtctcatatcagacgtttcggcctggccgagctaataatgtctaattatgcccatcag
 2701 ccgtatgtcgaggcaggttgacccgatacgctcgccgcgcgcctcattcatctccgact
 2761 gggcacaatgtcgccatctcgccgtcaaggtggtgcaagatacctattatgcaagcaga
 2821 ggatcagatggcgggcccatacagagcggtgctccggcttgogagaaagccgcttcgcag
 2881 caaggtatcggtggcaggccgcattttcggttgggtattctttgtcttgtttgcttcgta
 2941 attatgtcctggctggcattgtgggaagggcgaaacctcttgatttccgatgggggtcga

RESUMO

"CONSTRUÇÃO DE COMPOSIÇÕES DE CELULASE ALTAMENTE EFICAZES PARA A HIDRÓLISE ENZIMÁTICA DE CELULOSE"

- Esta invenção fornece novas composições enzimáticas usando enzimas de *C. lucknowense* recém identificadas e isoladas, incluindo CBH Ib, CBH IIb, EG II, EG VI, β -glicosidase, e xilanase II em combinação com as enzimas CBH Ia, CBH IIa (previamente descrita como Endo 43), e EG V previamente identificadas. Estas composições enzimáticas demonstram uma capacidade extremamente alta para converter biomassa lignocelulósica (por exemplo, Avicel, algodão, madeira da espécie Douglas fir pré-tratada por organosolvente) em glicose. CBH Ia e IIb, as quais têm um módulo de ligação de celulose (CBM) exibiram um sinergismo considerável com três endoglicanases principais (EG II, EG V, EG VI) provenientes do mesmo fungo na hidrólise de algodão, assim como uma forte sinergia entre si. As composições enzimáticas são eficazes na hidrólise da biomassa lignocelulósica.