



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

(11) PI 0718518-9 B1

(22) Data do Depósito: 21/09/2007

(45) Data de Concessão: 05/12/2017



(54) Título: POLINUCLEOTÍDEO CODIFICANDO MARCADOR SELECIONÁVEL DE ACETOLACTATO SINTASE (ALS) DE TRICHODERMA REESEI, MICROORGANISMO TRANSGÊNICO E MÉTODO DE SELEÇÃO DE UMA CÉLULA

(51) Int.Cl.: C12N 15/82; C12N 9/88

(30) Prioridade Unionista: 22/09/2006 US 60/846,656, 22/09/2006 US 60/846,804

(73) Titular(es): DANISCO US INC., GENENCOR DIVISION

(72) Inventor(es): BENJAMIN S. BOWER; NIGEL DUNN-COLEMAN; NICHOLAS LEIVA

Relatório Descritivo da Patente de Invenção para **"POLINUCLEOTÍDEO CODIFICANDO MARCADOR SELECIONÁVEL DE ACETOLACTATO SINTASE (ALS) DE *TRICHODERMA REESEI*, MICROORGANISMO TRANSGÊNICO E MÉTODO DE SELEÇÃO DE UMA CÉLULA".**

REFERÊNCIA CRUZADA A PEDIDOS RELACIONADOS

O presente pedido reivindica prioridade dos pedidos de patente provisórios dos Estados Unidos US 60/846.804, depositado em 22 de setembro de 2006, e US 60/846.656, depositado em 22 de setembro de 2006, ambos aqui incorporados a título de referência em sua totalidade.

Antecedentes da Invenção

Plantas, fungos e bactérias sintetizam os aminoácidos valina, leucina e isoleucina através de um curso comum. Uma das enzimas neste curso é acetolactato sintase (ALS) (de outro modo conhecida como acetohidroxíácido sintase ou AHAS), que converte piruvato em 2-acetolactato como a primeira etapa na síntese de valina e leucina, e também converte piruvato e 2-cetobutirato em 2-aceto-2-hidroxi-butirato, o precursor de isoleucina. A atividade de acetolactato sintase do tipo selvagem é sensível à ação de várias classes conhecidas de compostos tóxicos, incluindo compostos sulfonilureia e imidazolinona. Deste modo, tais compostos tóxicos podem ser empregados para matar células contendo proteínas acetolactato sintase que são sensíveis a esses compostos.

Essa descrição refere-se a enzimas acetolactato sintase que provêem resistência a compostos tóxicos, e então são marcadores selecionáveis úteis para células recombinantes.

Sumário da Invenção

Uma proteína acetolactato sintase que provê resistência a inibidores de ALS, por exemplo, compostos sulfonilureia e imidazolinona, é provida, bem como polinucleotídeo codificando a mesma. Em certas modalidades, a sequência de aminoácido da proteína acetolactato é pelo menos 80% idêntica à SEQ ID NO:1, a proteína acetolactato do tipo selvagem de *Trichoderma reesei*. Em modalidades particulares, a proteína ALS pode conter um

1a

aminoácido ácido na posição 190, um aminoácido ácido na posição 241 ou uma histidina na posição 372. Em algumas modalidades, o gene ou polipep-

tídeo de ALS pode ser empregado como um marcador selecionável em uma variedade ampla de espécies. Em certos casos, a proteína pode ser de ocorrência não-natural.

Em certas modalidades, um polinucleotídeo codificando ALS
 5 pode ser operavelmente ligado a um promotor e um terminador para prover expressão da proteína conferindo resistência a inibidor de ALS em uma célula hospedeira. O promotor e o terminador podem ser endógenos para a célula hospedeira em que o polinucleotídeo deve ser empregado e, em certos casos, o promotor e o terminador podem ser o promotor e o terminador de
 10 um gene de ALS da célula. O polinucleotídeo pode, em certas modalidades, conter uma estrutura de leitura aberta única codificando a proteína acetolactato sintase, caso onde o polinucleotídeo pode ser pelo menos 70% idêntico a ou pode hibridizar com SEQ ID NO.:2. Em outras modalidades, o polinucleotídeo pode compreender introns, caso onde o polinucleotídeo pode ter
 15 uma sequência de nucleotídeo que é pelo menos 70% idêntica a ou pode hibridizar com SEQ ID NO.:4. Em outras modalidades, o polinucleotídeo pode ser otimizado no códon para expressão da proteína acetolactato sintase em uma célula hospedeira particular.

Um vetor compreendendo o polinucleotídeo é também provido.
 20 Em adição ao polinucleotídeo, o vetor objeto pode conter um cassete de expressão para expressão de uma proteína recombinante, por exemplo, uma enzima ou proteína terapêutica, na célula.

Uma célula hospedeira compreendendo o polinucleotídeo objeto é também provida. Em certas modalidades, a célula hospedeira é resistente
 25 a um inibidor de ALS, por exemplo, um composto sulfonilureia ou imidazolinona tóxico. A célula hospedeira pode ser qualquer célula que seja sensível ao inibidor de ALS na ausência do polinucleotídeo. Em certas modalidades, a célula hospedeira pode ser uma célula de planta, por exemplo, uma célula de milho, soja ou *Arabidopsis*, uma célula fúngica, por exemplo, uma célula
 30 fúngica filamentosa tal como uma célula de *Trichoderma sp.* ou *Aspergillus sp.* ou uma célula bacteriana, por exemplo, um *Bacillus sp.* Em modalidades particulares, a célula hospedeira é um fungo filamentoso. A célula pode estar

presente *in vivo* ou em um organismo multicelular (por exemplo, uma planta). O polinucleotídeo pode estar presente em um genoma da célula hospedeira ou pode estar presente em um vetor que replica autonomamente na célula hospedeira.

5 É também provido um método de seleção de uma célula. Em certas modalidades, o método inclui: introdução de um polinucleotídeo objeto que codifica uma ALS em uma pluralidade de células, contato da pluralidade de células com um inibidor de ALS e cultura das células para prover seleção da célula. Em certas modalidades, a célula é uma célula fúngica. A célula
10 pode ser cultivada em um meio líquido contendo o inibidor de ALS ou um meio sólido contendo o inibidor de ALS. Esses métodos podem também incluir introdução de um segundo polinucleotídeo na célula hospedeira, onde o segundo polinucleotídeo codifica um polipeptídeo que deve ser produzido pela célula hospedeira. O polinucleotídeo objeto e o segundo polinucleotídeo
15 podem estar presentes no mesmo ácido nucleico ou um diferente, por exemplo, o mesmo vetor ou vetores diferentes. Se vetores diferentes forem empregados, eles podem ser cotransformados nas mesmas células.

São também providas modalidades relacionando-se aos promotores 1818A e 1818B.

20 Breve Descrição das Figuras

A figura 1 mostra a sequência de aminoácido da proteína acetolactato sintase do tipo selvagem de *T. reesei* (SEQ ID NO.:1).

A figura 2 mostra a sequência de nucleotídeo de um cDNA deduzido de *T. reesei* (SEQ ID NO.:1) codificando a proteína acetolactato do
25 tipo selvagem da figura 1.

A figura 3 mostra a sequência de nucleotídeo de um gene de *T. reesei* (SEQ ID NO.:3) codificando a proteína acetolactato sintase do tipo selvagem da figura 1, em que o promotor está em itálico, a sequência codificando ALS está sublinhada (SEQ ID NO.:4) e a sequência terminadora está
30 em itálico e sublinhada.

A figura 4 é um diagrama geral de um vetor que pode ser usado para testar o marcador de ALS, em que o vetor inclui dois genes. O primeiro

gene corresponde a um polinucleotídeo codificando uma proteína de interesse (por exemplo, uma enzima glicoamilase) e o segundo gene corresponde a um polinucleotídeo codificando o marcador de ALS.

5 As figuras 5A e 5B mostram a sequência de nucleotídeo dos promotores 1818A (SEQ ID NO.:7) e 1818B (SEQ ID NO.:8), respectivamente.

10 A figura 6 mostra a sequência de nucleotídeo (SEQ ID NO.:9) do vetor pTrex-glicoamilase usado para testar o marcador de ALS que compreende o promotor de *cbh1* de *Trichoderma reesei*, AttB1, um polinucleotídeo codificando uma glicoamilase de *Trichoderma reesei*, attB2, o terminador de *cbh1* de *Trichoderma reesei* e um marcador de ALS (A190D) no vetor de *E. coli* pSL1180.

15 A figura 7 ilustra um gel SDS-PAGE de amostras de sobrenadante de transformante cultivado em frasco de agitação. A faixa 1 mostra o controle não-transformado. Expressão de glicoamilase de *T. reesei* é mostrada nas faixas 3 e 4 como destacado pela seta.

A figura 8 é um diagrama do vetor usado para testar os promotores 1818A e 1818B.

20 A figura 9 mostra a sequência de nucleotídeo de um construto usado para testar os promotores 1818A e 1818B (SEQ ID NO:10).

Descrição Detalhada

Definições

25 A menos que de outro modo aqui definido, todos os termos técnicos e científicos usados aqui têm o mesmo significado que geralmente conhecido por uma pessoa comum versada na técnica à qual a presente invenção pertence. Singleton e outros, *Dictionary of Microbiology and Molecular Biology*, 2ª Ed., John Wiley and Sons, Nova York (1994) e Hale & Marham, *The Harper Collins Dictionary of Biology*, Harper Perennial, NY (1991) provê um versado na técnica com dicionários gerais de muitos dos termos
30 usados na presente invenção. Embora quaisquer métodos e materiais similares ou equivalentes àqueles descritos aqui possam ser usados na prática ou teste da presente invenção, os métodos e materiais preferidos são descritos.

Todas as patentes e publicações, incluindo todas as sequências descritas dentro de tais patentes e publicações, referidas aqui são expressamente incorporadas a título de referência.

Faixas numéricas são inclusivas dos números definindo a faixa.

- 5 A menos que de outro modo indicado, ácidos nucleicos são escritos da esquerda para a direita na orientação 5' a 3', sequências de aminoácido são escritas da esquerda para a direita em orientação amino para carbóxi, respectivamente.

- 10 Os cabeçalhos providos aqui não são limitações dos vários aspectos ou modalidades da invenção que possam ter tido por referência ao relatório como um todo. Deste modo, os termos definidos imediatamente abaixo são mais completamente definidos por referência ao relatório como um todo.

- 15 Conforme aqui usado, o termo "marcador selecionável" refere-se a um gene ou polinucleotídeo cuja expressão permite identificação de células que foram transformadas com um construto ou vetor de DNA contendo o gene ou polinucleotídeo. Marcadores selecionáveis podem prover resistência a compostos tóxicos tal como antibióticos, herbicidas e similares.

- 20 O termo "acetolactato sintase (ALS)" refere-se a uma enzima que tem uma atividade definida como EC 2.2.1.6 de acordo com *IUBMB Enzyme Nomenclature*. A enzima catalisa uma reação entre duas moléculas de piruvato para produzir 2-acetolactato e CO₂. A enzima requer tiamina difosfato e pode ser referida como acetoidroxiácido sintase (AHAS) em outro lugar.

- 25 O termo "inibidor de ALS" refere-se a um composto que inibe proteína ALS do tipo selvagem e é tóxico para células que contêm ALS do tipo selvagem. Tais compostos incluem herbicidas conhecidos e incluem um composto sulfonilureia, imidazolinona, triazolopirimidina, pirimidiniltiobenzoato ou sulfonilamino-carbonil-triazolinona discutidos abaixo.

- 30 O termo "promotor" é definido aqui como um ácido nucleico que direciona a transcrição de um polinucleotídeo a jusante em uma célula. Em certos casos, o polinucleotídeo pode conter uma sequência de codificação e

o promotor pode direcionar a transcrição da sequência de codificação em RNA traduzível.

O termo "isolado" conforme aqui definido significa um composto, uma proteína, célula, sequência de ácido nucleico ou aminoácido que é re-
5 movido de pelo menos um componente com o qual ele é naturalmente asso-
ciado.

O termo "sequência de codificação" é definido aqui como um á-
cido nucleico que, quando posto sob o controle de sequências-controle a-
propriadas incluindo um promotor, é transcrito em mRNA que pode ser tra-
10 duzido em um polipeptídeo. Uma sequência de codificação pode conter uma
estrutura de leitura aberta única, ou várias estruturas de leitura aberta sepa-
radas por íntrons, por exemplo. Uma sequência de codificação pode ser cD-
NA, DNA genômico, DNA sintético ou DNA recombinante, por exemplo. Uma
sequência de codificação geralmente inicia em um códon de início (por e-
15 xemplo, ATG) e termina em um códon de parada (por exemplo, UAA, UAG e
UGA).

O termo "recombinante" refere-se a um polinucleotídeo ou poli-
peptídeo que não acontece naturalmente em uma célula hospedeira. Uma
molécula recombinante pode conter duas ou mais sequências de ocorrência
20 natural que são ligadas de uma maneira que não acontece naturalmente.

O termo "heterólogo" refere-se a elementos que não estão nor-
malmente associados um com o outro. Por exemplo, uma proteína heterólo-
ga é uma proteína que não é produzida em uma célula hospedeira do tipo
selvagem, um promotor heterólogo é um promotor que não está presente em
25 ácido nucleico que é endógeno para uma célula hospedeira do tipo selva-
gem e um promotor operavelmente ligado a uma sequência de codificação
heteróloga é um promotor que é operavelmente ligado a uma sequência de
codificação que não é geralmente operavelmente ligada a uma célula hos-
pedeira do tipo selvagem.

30 O termo "operavelmente ligado" refere-se a uma disposição de
elementos que permite que eles sejam funcionalmente relacionados. Por
exemplo, um promotor é operavelmente ligado a uma sequência de codifica-

ção se ele controlar a transcrição da sequência, e uma sequência de sinal é operavelmente ligada a uma proteína se a sequência de sinal direcionar a proteína através do sistema de secreção de uma célula hospedeira.

5 O termo "ácido nucleico" compreende DNA, RNA, de filamento simples ou duplo e modificações dos mesmos. Os termos "ácido nucleico" e "polinucleotídeo" podem ser usados intercomutavelmente aqui.

O termo "construto de DNA" conforme aqui usado significa uma sequência de ácido nucleico que compreende pelo menos dois fragmentos de polinucleotídeo de DNA.

10 O termo "sequência de sinal" ou "peptídeo de sinal" refere-se a uma sequência de aminoácidos na porção N-terminal de uma proteína que facilita a secreção da forma madura da proteína fora da célula. A forma madura da proteína extracelular não tem a sequência de sinal que é clivada durante o processo de secreção.

15 O termo "vetor" é definido aqui como um polinucleotídeo feito para carregar sequências de ácido nucleico a serem introduzidas em um ou mais tipos de célula. Vetores incluem vetores de clonagem, vetores de expressão, vetores de embaralhamento, plasmídeos, fagos ou partículas de vírus, construtos de DNA, cassetes e similares. Vetores de expressão po-
20 dem incluir sequências reguladoras tal como promotores, sequências de sinal, uma sequência de codificação e terminadores de transcrição.

Um "vetor de expressão" conforme aqui usado significa um construto de DNA compreendendo uma sequência de codificação que é operavelmente ligada a sequências-controle adequadas capazes de realizar ex-
25 pressão de uma proteína em um hospedeiro adequado. Tais sequências-controle podem incluir um promotor para realizar a transcrição, uma sequência operadora adicional para controlar transcrição, uma sequência codifican- do sítios de ligação de ribossomo adequados no mRNA, realçadores e se-
quências que controlam término de transcrição e tradução.

30 Conforme aqui usado, os termos "polipeptídeo" e "proteína" são usados intercomutavelmente e incluem referência a um polímero de qual- quer número de resíduos de aminoácido. Os termos se aplicam a polímeros

de aminoácido onde um ou mais resíduos de aminoácido são um análogo químico artificial de um aminoácido de ocorrência natural correspondente, bem como a polímeros de aminoácido de ocorrência natural. Os termos também se aplicam a polímeros contendo substituições de aminoácido conservativas de modo que o polipeptídeo permanece funcional. "Peptídeos" são polipeptídeos tendo menos do que 50 resíduos de aminoácidos.

Uma "célula hospedeira" é uma célula que contém um ácido nucleico recombinante objeto, ou no genoma da célula hospedeira ou em um vetor extracromossomal que replica autonomamente a partir do genoma da célula hospedeira. Uma célula hospedeira pode ser qualquer tipo de célula.

O termo "fungos filamentosos" refere-se a todas as formas filamentosas da subdivisão *Eumycotina* (Vide Alexopoulos, C.J. (1962), *INTRODUCTORY MYCOLOGY*, Wiley, Nova York). Esses fungos são caracterizados por um micélio vegetativo com uma parede celular composta de quitina, glicanas e outros polissacarídeos complexos. Os fungos filamentosos da presente invenção são morfologicamente, fisiologicamente e geneticamente distintos de leveduras. Crescimento vegetativo por fungos filamentosos é através de alongamento hifal e catabolismo de carbono é aeróbico obrigatório.

Uma célula "não-patogênica" é uma linhagem que é não-patogênica para humanos.

"Transformação" significa introdução de DNA em uma célula de modo que o DNA é mantido na célula ou em um elemento extracromossomal ou integrante cromossomal.

A menos que de outro modo indicado todas as posições de aminoácido em uma proteína acetolato sintase são relativas à SEQ ID NO.:1, após alinhamento desta proteína com SEQ ID NO.:1 usando o programa BLASTP (Altschul, *Nucl. Acids Res.* 1997 25:3389-3402; Schäffer, *Bioinformatics* 1999 15:1000-1011) sob condições padrão, conforme disponível da *world wide website* do National Center of Biotechnology Information (NCBI).

Polinucleotídeos

É aqui provido um polinucleotídeo que codifica uma proteína a-

cetolactato sintase que provê resistência a inibidores de ALS. Em certas modalidades, a proteína acetolactato sintase é uma proteína de ocorrência não-natural. Em termos gerais, o polinucleotídeo codifica uma proteína que:

5 a) possui atividade de acetolactato sintase (isto é, pode catalisar uma reação entre duas moléculas de piruvato para produzir 2-acetolactato), b) confere resistência a inibidores de ALS, por exemplo, compostos sulfonilureia e imidazolinona e c) tem uma sequência de aminoácido que é pelo menos 85% idêntica (por exemplo, pelo menos 90% idêntica, pelo menos 93% idêntica, pelo menos 95% idêntica, pelo menos 96% idêntica, pelo menos 97% idêntica,

10 ca, pelo menos 98% idêntica, pelo menos 98,5% idêntica, pelo menos 99% idêntica ou pelo menos 99,5% idêntica) à SEQ ID NO.:1, onde SEQ ID NO.:1 mostra a sequência de aminoácido de uma acetolactato sintase do tipo selvagem de *Trichoderma reesei*.

Em certas modalidades, o polinucleotídeo pode codificar uma

15 proteína tendo um ou mais de: a) um aminoácido ácido (por exemplo, Asp ou Glu) na posição 190, b) um aminoácido ácido (por exemplo, Asp ou Glu) na posição 241 ou um His na posição 372. Posições 190, 241 e 372 em proteína ALS de *T. reesei* do tipo selvagem estão sublinhadas na figura 1. Conforme acima mencionado, posições 190, 241 e 372 em uma proteína ALS

20 diferente (por exemplo, uma proteína ALS que é, por exemplo, mais curta, mais longa ou contém deleções e/ou inserções relativas à proteína ALS de *T. reesei*), são definidas aqui como sendo as posições na proteína que correspondem às (isto é, alinham com ou estão através da) posições 190, 241 e 372 da proteína ALS de *T. reesei* do tipo selvagem quando a proteína ALS

25 de *T. reesei* do tipo selvagem e a outra proteína estão alinhadas usando métodos de alinhamento de sequência padrão, por exemplo, BLASTP (Altschul, *Nucl. Acids Res.* 1997 25:3389-3402; Schäfer, *Bioinformatics* 1999 15:1000-1011) usando parâmetros padrão. Proteínas ALS, em geral, são enzimas bem caracterizadas e foram investigadas em grande detalhe funcional e es-

30 truturalmente. Proteínas ALS foram revistas em várias publicações (vide, por exemplo, Chipman, *Biochim. Biophys. Acta* 1998 1385:401-19; Chipman, *Curr. Opin. Chem. Biol.* 2005 9:475-81) e foram cristalizadas (vide, por e-

xemplo, Pang, *J. Biol. Chem.* 2004 279:2242-53; Pang, *J. Mol. Biol.* 2002 317:249-62), bem como submetidas à mutagênese para identificar resíduos essenciais e não-essenciais (vide, por exemplo, Ibdah, *Biochemistry* 1996 35:16282-91; Mendel, *J. Mol. Biol.* 2001:465-771; Hill, *Biochem. J.* 1998 335:653-61). Outras mutações que conferem resistência à herbicida são também conhecidas (vide, por exemplo, Jung, *Biochem. J.* 2004 383:53-61; Duggelby, *Eur. J. Biochem.* 2003 207:2895-904). Ainda, as sequências de aminoácido de várias centenas de proteínas ALS são conhecidas e publicamente disponíveis através do banco de dados Genbank do NCBI. Deste modo, uma ampla variedade de mudanças de aminoácido que poderiam ser feitas a uma proteína ALS objeto, algumas das quais podem conferir resistência a inibidores de ALS, sem abolir sua atividade seriam prontamente aparentes.

Em certos casos, as mudanças de aminoácido descritas aqui podem ser transferidas para qualquer outra proteína ALS, de qualquer espécie, para tornar aquela proteína resistente a herbicida. Em outras palavras, um aminoácido na posição 190, 241 ou 372 (com relação à SEQ ID NO.:1) de qualquer outra proteína ALS pode ser substituído por um aminoácido ácido (por exemplo, Asp ou Glu), um aminoácido ácido (por exemplo, Asp ou Glu) ou um His, respectivamente, para prover uma proteína ALS que confere resistência a um inibidor de ALS.

Por exemplo, as sequências de aminoácido das proteínas ALS de várias espécies fúngicas são conhecidas e são depositadas no banco de dados Genbank da NCBI. Em certas modalidades, as alterações de aminoácido descritas acima podem ser transferidas para as proteínas ALS daqueles fungos a fim de prover outras proteínas resistentes a inibidor de ALS. Em outras modalidades, a sequência de aminoácido de outras proteínas ALS fúngicas pode ser empregada para fazer mudanças adicionais nas proteínas ALS baseadas em *T. reesei* objeto que não abolem a atividade de ALS dessas proteínas. Por exemplo, uma fusão entre duas proteínas ALS de espécies diferentes ou uma proteína contendo substituições, deleções ou inserções de aminoácido poderia ser feita. Sequências de aminoácido ALS e-

xemplares de outras espécies fúngicas, incluindo outras espécies fúngicas filamentosas, estão depositadas no banco de dados Genbank do NCBI como GIDS: 39977967 e 2547090 (*Magnaporthe grisea*), GID:85108881 (*Neurospora crassa*), GID:46108408 (*Gibberella zeae*), GID:90302929 (*Coccidioides immitis*), GID:67537572 (*Aspergillus nidulans*), GID:70999742 (*Aspergillus fumigatus*), GIDs:83767597 e 83771596 (*Aspergillus oryzae*); GID:111063308 (*Phaeosphaeria nodorum*), GID:50547615 (*Yarrowia lipolytica*), GID:49657303 (*Debaryomyces hansenii*), GID:68468265 (*Candida albicans*), GID:21615550 (*Daccharomycopsis fibuligera*), GID:49641223 (*Kluyveromyces lactis*); GID:49527687 (*Candida glabrata*) e GID:817866 (*Saccharomyces cerevisiae*). Os acessos ao Genbank acima referidos são incorporados a título de referência em sua totalidade, incluindo as sequências de ácido nucleico e proteína nele, e as anotações dessas sequências, como da primeira data de depósito deste pedido de patente.

Por causa da redundância do código genético, um polinucleotídeo objeto pode compreender qualquer uma de várias sequências de nucleotídeo. Em modalidades particulares, o polinucleotídeo objeto pode ter uma sequência de nucleotídeo que é: a) pelo menos 70% idêntica (por exemplo, pelo menos 80%, pelo menos 90%, pelo menos 95%, pelo menos 98% ou pelo menos 99% idêntica a) ou b) hibridiza sob condições de hibridização estritas para SEQ ID NO:2 ou SEQ ID NO:4, sequências que são mostradas na sequência de nucleotídeo de um cDNA do tipo selvagem e gene de *T. reesei*, respectivamente. Deste modo, o polipeptídeo de codificação pode conter íntrons ou pode conter uma estrutura de leitura aberta única codificando a proteína.

Como seria aparente, em certas modalidades, o polinucleotídeo pode ter uma sequência de nucleotídeo que codifica uma proteína tendo um ou mais de: a) um aminoácido ácido (por exemplo, Asp ou Glu) na posição 190, b) um aminoácido ácido (por exemplo, Asp ou Glu) na posição 241 ou um His na posição 372.

O termo "identidade" no contexto de duas sequências de ácido nucleico referem-se a resíduos de nucleotídeo nas duas sequências que são

iguais quando alinhados para correspondência máxima, conforme medido usando qualquer um dos algoritmos de comparação de sequência que se-
guem. Alinhamento ótimo de sequências para comparação pode ser realiza-
do, por exemplo, através do algoritmo de homologia local de Smith & Wa-
5 terman, *Adv. Appl. Math.* 2:482 (1981), através do algoritmo de alinhamento
de homologia de Needleman & Wunsch, *J. Mol. Biol.* 48:443 (1970), através
da pesquisa pelo método de similaridade de Pearson & Lipman, *Proc. Nat'l.
Acad. Sci. USA* 85:2444 (1988), através de implementações computadoriza-
das desses algoritmos (GAP, BESTFIT, FASTA e TFASTA no Wisconsin
10 Genetics Software Package, Genetics Computer Group, 575 Science Dr.,
Madison, WI) ou através de inspeção visual.

Um exemplo de um algoritmo que é adequado para determina-
ção de similaridade de sequência é o algoritmo BLAST, que é descrito em
Altus e outros, *J. Mol. Biol.* 215:403-410 (1990). *Software* para realizar
15 análises BLAST está publicamente disponível através do National Center for
Biotechnology Information disponível no *world wide web* (www) nc-
bi.nlm.nih.gov. O algoritmo BLAST realiza uma análise estatística da simila-
ridade entre duas sequências (vide, por exemplo, Karlin & Altschul, *Proc.
Nat'l Acad. Sci. USA* 90:5873-5787 (1993)). Uma sequência de nucleotídeo
20 de pelo menos 100, pelo menos 500, pelo menos 1000, até o comprimento
total das SEQ ID NO:2 ou 4 pode ser empregada em comparações de se-
quência.

Conforme acima mencionado, o polinucleotídeo pode incluir
uma sequência de nucleotídeo que hibridiza sob condições de hibridização
25 estridentes para um polinucleotídeo tendo a sequência de nucleotídeo da
SEQ ID NO:2 ou 4, onde condições de hibridização estridentes compreen-
dem condições de hibridização de estridência baixa, média, alta e muito
alta.

Condições de "baixa estridência" referem-se à lavagem com
30 uma solução de 1X SSC/SDS 0,1% a 20°C por 15 minutos. Condições de
"estridência média" referem-se à lavagem com uma solução de 1X
SSC/SDS 0,1% a 65°C por 60 minutos. Condições de "alta estridência" re-

fere-se à lavagem com uma solução de 0,2X SSC/SDS 0,1% a 65°C por 10 minutos. Condições de "estringência muito alta" referem-se à lavagem com uma solução de 0,2X SSC/SDS 0,1% a 65°C por 60 minutos.

Métodos de hibridização são descritos em mais detalhes em

5 Sambrook e outros, *Molecular Cloning: A Laboratory Manual* (2ª Ed., 1989 Cold Spring Harbor, NY). Em um ensaio de hibridização exemplar, uma amostra de DNA é eletroforada através de gel de agarose (por exemplo, agarose 0,8%) de modo que o fragmento de DNA pode ser visualizado através de tingimento com brometo de etídio. O gel é então rapidamente enxaguado

10 em H₂O destilada e subsequentemente despurinado em uma solução apropriada (tal como, por exemplo, HCl a 0,25M) com agitação suave seguido por desnaturação por 30 minutos (em, por exemplo, NaOH a 0,4M) com agitação suave. Uma etapa de renaturação pode ser incluída, no qual o gel é posto em NaCl a 1,5M, Tris a 1M, pH 7,0 com agitação suave por 30 minutos.

15 O DNA é então transferido para uma membrana positivamente carregada apropriada, por exemplo, membrana *Maximum Strength Nytran Plus* (Schleicher & Schuell, Keene, N.H.), usando uma solução de transferência (tal como, por exemplo, 6X SSC, isto é, NaCl a 900 mM, citrato de trissódio 90 a mM). Uma vez a transferência completa, geralmente após cerca de 2 horas,

20 a membrana é enxaguada em, por exemplo, 2X SSC (NaCl a 300 mM, citrato de trissódio a 30 mM) e secada ao ar em temperatura ambiente. A membrana pode ser pré-hibridizada (por aproximadamente 2 horas ou mais) em uma solução de pré-hibridização adequada (tal como, por exemplo, uma solução aquosa contendo por 100 mL: 20-50 mL de formamida, 25 mL de 20X

25 SSPE (1X SSPE = NaCl a 0,18 M, EDTA a 1 mM, NaH₂PO₄ a 10 mM, pH 7,7), 2,5 mL de SDS 20% e 1 mL de 10 mg/mL de DNA de esperma de arenque cisalhado). Como seria conhecido a um versado na técnica, a quantidade de formamida na solução de pré-hibridização pode ser variada dependendo da natureza da reação obtida de acordo com métodos de rotina.

30 Então, uma quantidade menor de formamida pode resultar em hibridização mais completa em termos de identificação de moléculas de hibridização do que o mesmo procedimento usando uma quantidade maior de formamida.

Por outro lado, uma faixa de hibridização forte pode ser mais fácil e visualmente identificada usando mais formamida.

Uma sonda de DNA de pelo menos 100, pelo menos 500, pelo menos 1000, até o comprimento total das SEQ ID NO:2 ou 4 pode ser empregada em ensaios de hibridização ou em comparações de sequência. A sonda de DNA pode ser isolada através de eletroforese em gel de agarose, o fragmento excisado do gel e recuperado da agarose excisada. Este fragmento purificado de DNA pode ser marcado (usando, por exemplo, o sistema de marcação *Megaprime* de acordo com as instruções do fabricante) para incorporar P^{32} no DNA. A sonda marcada é desnaturada através de aquecimento para 95°C por 5 minutos e imediatamente adicionada à membrana e solução de pré-hibridização. A reação de hibridização deve prosseguir por um tempo apropriado e sob condições apropriadas, por exemplo, por 18 horas a 37°C com agitação ou giro suave. A membrana é enxaguada (por exemplo, em 2X SSC/SDS 0,3%) e então lavada em uma solução de lavagem apropriada, conforme descrito acima, com agitação suave. Hibridização pode ser detectada através de autoradiografia.

Em modalidades particulares, o polinucleotídeo pode ser otimizado no códon para expressão em célula hospedeira particular. Em outras modalidades, o polinucleotídeo pode ter uma sequência de nucleotídeo que contém menos do que 10 (por exemplo, 9, 8, 7, 6, 5, 4, 3, 2 ou 1) diferenças comparado com a sequência de ALS do tipo selvagem (por exemplo, ou a sequência genômica ou de cDNA) da célula hospedeira particular em que o polinucleotídeo deve ser empregado.

Em certos casos, em adição a uma sequência de codificação, o polinucleotídeo pode conter ainda outros elementos que são necessários para expressão da proteína codificada em uma célula hospedeira. Por exemplo, em uma modalidade, o polinucleotídeo pode ser flanqueado por sequências para formar um cassete de expressão que provê expressão da proteína codificada em uma célula hospedeira. Em certas modalidades, o cassete de expressão pode conter um promotor para transcrição da sequência de codificação e um terminador transcripcional e uma sequência codificando

uma região não-traduzida 5' (UTR) que permite iniciação de tradução, cada um em ligação operável com a região de codificação. Promotores, realçadores, terminadores, UTRs, sinais de poliadenilação e outras sequências reguladoras para uma ampla variedade de células hospedeiras, particularmente
 5 aquelas de plantas, bactérias e fungos, são bem conhecidos na técnica (vide, por exemplo, Ausubel e outros, *Short Protocols in Molecular Biology*, 3ª ed., Wiley and Sons, 1995; Sambrook e outros, *Molecular Cloning: A Laboratory Manual*, Segunda Edição, 1989 Cold Spring Harbor, N.Y.).

Em modalidades particulares, o cassete de expressão objeto
 10 pode conter um promotor e um terminador para expressão da proteína objeto em uma célula fúngica filamentosa. Exemplos de promotores e terminadores adequados para direcionamento da transcrição de um ácido nucleico objeto em uma célula hospedeira fúngica filamentosa são promotores e terminados obtidos dos genes para amilase TAKA de *Aspergillus oryzae*, proteínase aspártica de *Rhizomucor miehei*, alfa-amilase neutra de *Aspergillus Niger*, alfa-amilase estável a ácido de *Aspergillus Niger*, glicoamilase de *Aspergillus Niger* ou *Aspergillus awamori* (glaA), lipase de *Rhizomucor miehei*,
 15 protease alcalina de *Aspergillus oryzae*, isomerase de triose fosfato de *Aspergillus oryzae*, acetamidase de *Aspergillus nudulans*, aminoglicosidase de
 20 *Fusarium venenatum*, protease do tipo tripsina de *Fusarium oxysporum* (WO 96/00787), celobioidrolase I de *Trichoderma reesei*, celobioidrolase II de *Trichoderma reesei*, endoglucanase I de *Trichoderma reesei*, endoglucanase II de *Trichoderma reesei*, endoglucanase III de *Trichoderma reesei*, endoglucanase IV de *Trichoderma reesei*, endoglucanase V de *Trichoderma reesei*,
 25 xilanase I de *Trichoderma reesei*, xilanase II de *Trichoderma reesei*, beta-xilosidase de *Trichoderma reesei*, bem como o promotor de NA2-tpi (um híbrido dos promotores dos genes para alfa-amilase neutra de *Aspergillus Niger* e triose fosfato isomerase de *Aspergillus oryzae*); e seus promotores mutantes, truncados e híbridos dos mesmos.

30 O promotor e/ou terminador pode ser nativo ou não-endógeno para a célula hospedeira na qual a proteína ALS objeto deve ser expressa e, em certas modalidades, o promotor e o terminador podem ser o promotor e

terminador do gene de ALS da célula hospedeira. Por exemplo, em uma modalidade, um cassete de expressão para uso em células de *Trichoderma reesei* pode compreender o promotor do gene de ALS de *Trichoderma reesei*, a sequência de codificação de ALS de *Trichoderma reesei* (contendo
5 uma ou mais das sequências de nucleotídeo mostradas acima) e terminador de ALS de *Trichoderma reesei*, em ligação operável (figura 3). O promotor e/ou terminador pode ser de um gene de ALS da célula hospedeira na qual o polinucleotídeo deve ser empregado.

O polinucleotídeo pode ser integrado no genoma da célula hospedeira ou pode estar presente em um vetor que replica autonomamente na
10 célula hospedeira.

Ácido Nucléico Recombinante

Um ácido nucleico recombinante compreendendo um polinucleotídeo objeto é também provido. Um ácido nucleico recombinante objeto
15 pode compreender um polinucleotídeo objeto, por exemplo, um cassete de expressão para produção de uma proteína ALS de conferência de resistência em uma célula hospedeira, bem como um segundo cassete de expressão para expressão de uma proteína de interesse na célula hospedeira. Em uma modalidade particular, o polinucleotídeo objeto é empregado como um
20 marcador para seleção de células hospedeiras que contêm o ácido nucleico recombinante sobre outras células que não contêm o ácido nucleico recombinante (isto é, o polinucleotídeo objeto é empregado como um "marcador selecionável" para células que contêm o polinucleotídeo objeto).

A proteína de interesse codificada pelo segundo cassete de expressão
25 pode ser, por exemplo, uma enzima, uma proteína terapêutica, uma proteína repórter, um aditivo alimentar ou um gênero alimentício ou similares.

Em uma modalidade, a proteína de interesse codificada pelo segundo cassete de expressão pode ser uma enzima tal como uma carbohidrase, tal como uma α -amilase, uma α -amilase alcalina, uma β -amilase, uma
30 celulase; uma dextranase, uma α -glicosidase, uma α -galactosidase, uma glicoamilase, uma hemicelulase, uma pentosanase, uma xilanase, uma in-

vertase, uma lactase, uma naringanase, uma pectinase ou uma pululanase; uma protease tal como uma protease ácida, uma protease alcalina, bromelaina, ficina, uma protease neutra, papaína, pepsina, uma peptidase, rennet, renina, quimosina, subtilisina, termolisina, uma proteinase aspártica ou tripsina; uma lipase ou esterase, tal como trigliceridase, uma fosfolipase, uma esterase pré-gástrica, uma fosfatase, uma fitase, uma amidase, uma iminoacilase, uma glutaminase, uma lisozima ou uma penicilina acilase; uma isomerase tal como glicose isomerase; uma oxidorreductase, por exemplo, uma aminoácido oxidase, uma catalase, uma cloroperoxidase, uma glicose oxidase, uma hidroxiesteroide desidrogenase ou uma peroxidase; uma liase tal como uma acetolactato descarboxilase, uma β -descarboxilase aspártica, uma fumarese ou uma histadese; uma transferase tal como ciclodextrina glicosiltransferase; ou uma ligase, por exemplo. Em modalidades particulares, a proteína pode ser uma aminopeptidase, uma carboxipeptidase, uma quitinase, uma cutinase, uma desoxirribonuclease, uma α -galactosidase, uma β -galactosidase, uma β -glicosidase, uma lacase, uma manosidase, uma mutanase, uma enzima pectinolítica, uma polifenoloxidase, ribonuclease ou transglutaminase, por exemplo. A enzima pode ser uma enzima do tipo selvagem ou uma variante de uma enzima do tipo selvagem. Em adição a enzima pode ser uma enzima híbrida que inclui fragmentos de enzimas diferentes.

Em outra modalidade, a proteína de interesse codificada pelo segundo cassete de expressão pode ser uma proteína terapêutica (isto é, uma proteína tendo uma atividade biológica terapêutica). Exemplos de proteínas terapêuticas adequadas incluem: eritropoietina, citocinas tal como interferon- α , interferon- β , interferon- γ , interferon- δ e granulócito-CSF, GM-CSF, fatores de coagulação tal como fator VII, fator IX e proteína C humana, anti-trombina III, trombina, receptor de cadeia α de IgE solúvel, IgG, fragmentos de IgG, fusões de IgG, IgM, IgA, e inibidor de interleucinas, urocinase, quimase ureia tripsina, proteínas de ligação de IGF, fator de crescimento epidérmico, fator de liberação de hormônio do crescimento, proteína de fusão de anexina V, angiostatina, fator-2 de crescimento endotelial vascular, fator-1

inibidor de progenitor mielóide, osteoprotegerina, α -1-antitripsina, α -feto proteínas, Dnase II, kringle 3 de plasminogênio humano, glicocerebrosidase, proteína de ligação de TNF, hormônio de estimulação de folículo, antígeno 4-Ig associado a linfócito T citotóxico, ativador de transmembrana e modula-
5 dor de cálcio e ligante de ciclofilina, fusão Fc de receptor de TNF solúvel, agonista de receptor de proteína 1 do tipo glucagon e IL-2. Proteínas de anticorpo, por exemplo, anticorpos monoclonais que podem ser humanizados, são de interesse particular.

Em uma modalidade adicional, a proteína codificada pelo se-
10 gundo cassete de expressão pode ser uma proteína repórter. Tais proteínas repórter podem ser opticamente detectáveis ou colorigênicas, por exemplo. Nesta modalidade, a proteína pode ser uma β -galactosidase (lacZ), β -glucuronidase (GUS), luciferase, fosfatase alcalina, nopalina sintase (NOS), cloranfenicol acetiltransferase (CAT), peroxidase de rábano silvestre (HRP)
15 ou uma proteína verde fluorescente, por exemplo, proteína verde fluorescente (GFP) ou um derivado da mesma.

Em certas modalidades, particularmente aquelas nas quais a célula hospedeira é uma célula hospedeira fúngica filamentosa, a sequência de codificação do segundo cassete de expressão pode codificar uma proteína
20 de fusão. Em algumas dessas modalidades, a proteína de fusão pode prover secreção da proteína a partir da célula hospedeira na qual ela é expressa e, desta maneira, pode conter uma sequência de sinal operavelmente ligada ao terminal N da proteína, onde a sequência de sinal contém uma sequência de aminoácidos que direciona a proteína para o sistema secretor da célula hos-
25 pedeira, resultando em secreção da proteína a partir da célula hospedeira no meio no qual a célula hospedeira está crescendo. A sequência de sinal é clivada a partir da proteína de fusão antes da secreção da proteína. A sequência de sinal empregada pode ser endógena ou não-endógena para a célula hospedeira e, em certas modalidades, pode ser sequência de sinal de
30 uma proteína que é conhecida ser altamente secretada a partir de uma célula hospedeira. Em modalidades particulares, a proteína de sequência de sinal pode ser qualquer sequência de sinal que facilite a secreção de proteí-

na a partir de uma célula hospedeira fúngica filamentosa (por exemplo, *Trichoderma* ou *Aspergillus*). Tal sequência de sinal inclui, mas não está limitada a: a sequência de sinal de celobioidrolase I, celobioidrolase II, endoglucanase I, endoglucanase II, endoglucanases III, α -amilase, aspartil proteases, glicoamilase, manase, glicosidase e endopeptidase B de cevada (vide Saarelainen, *Appl. Environ. Microbiol.*, 1997 63:4938-4940), por exemplo. Outras de sequências de sinal são aquelas originando-se do gene de amiloglicosidase (GA) fúngica (glA), o gene de fator α (leveduras, por exemplo, *Saccharomyces*, *Kluyveromyces* e *Hansenula*) ou o gene de α -amilase (*Bacillus*). Em certas modalidades, então, o ácido nucleico recombinante objeto pode compreender: um ácido nucleico codificando sequência de sinal operavelmente ligada a um ácido nucleico de codificação de proteína, onde a tradução do ácido nucleico em uma célula hospedeira produz uma proteína de fusão compreendendo uma proteína tendo uma sequência de sinal N-terminal para secreção da proteína a partir da célula hospedeira.

Em modalidades particulares, a proteína de fusão pode conter uma "proteína-veículo", que é uma porção de uma proteína que é endógena para e altamente secretada pela célula hospedeira. Proteínas-veículos adequadas incluem aquelas de mannanase I de *T. reesei* (Man5A ou MANI), celobioidrolase II de *T. reesei* (Cel6A ou CBHII) (vide, por exemplo, Paloheimo e outros, *Appl. Environ. Microbiol.* 2003, Dezembro; 69(12):7073-7082) ou celobioidrolase I de *T. reesei* (CHBI). Em uma modalidade, a proteína-veículo é uma proteína CBH1 de *T. reesei* truncada que inclui a região de núcleo de CBH1 e parte da região de ligação de CBH1. Uma proteína de fusão contendo, do terminal amino para o terminal carbóxi, uma sequência de sinal, uma proteína-veículo e uma proteína objeto em ligação operável é então provida, bem como um ácido nucleico codificando a mesma.

Em certas modalidades, o polinucleotídeo pode ser otimizado no códon para expressão da proteína em uma célula hospedeira particular. Uma vez que tabelas de uso de códon listando o uso de cada códon em muitas células são conhecidas na técnica (vide, por exemplo, Nakamura e outros, *Nucl. Acids Res.* 200 28:292) ou prontamente deriváveis, tais ácidos

nucleicos podem ser prontamente determinados dando a sequência de aminoácido de uma proteína a ser expressa.

Um ácido nucleico recombinante objeto pode estar presente, por exemplo, integrado, em um genoma (isto é, o genoma nuclear) de uma célula hospedeira, ou pode estar presente em um vetor, por exemplo, um vetor de fago, plasmídeo, viral ou retroviral, que replica autonomamente na célula hospedeira. Em certas modalidades, o vetor contém ainda o segundo cassete de expressão discutido acima. Em certas modalidades, o vetor pode ser um vetor de expressão para expressão de um polipeptídio recombinante em uma célula fúngica filamentosa.

Vetores para expressão de proteínas recombinantes são bem conhecidos na técnica (Ausubel e outros, *Short Protocols in Molecular Biology*, 3ª ed., Wiley & Sons, 1995; Sambrook e outros, *Molecular Cloning: A Laboratory Manual*, Segunda Edição, (1989) Cold Spring Harbor, N.Y.).

15 Métodos de Seleção

Um método para seleção de células hospedeiras que contêm um polinucleotídeo objeto é também provido. Em certas modalidades, o método inclui introdução de um polinucleotídeo objeto em uma pluralidade de células, contato da pluralidade de células com um inibidor de ALS, por exemplo, um composto sulfonilureia ou imidazolinona tóxico e cultura das células para prover a seleção da célula. A célula pode ser selecionada em meio sólido, por exemplo, plaqueando uma pluralidade de células em uma placa de ágar contendo o inibidor de ALS, ou em meio líquido, por exemplo, através de cultura de uma pluralidade de em um meio líquido contendo o inibidor de ALS.

Em certas modalidades, os métodos de seleção objeto podem ser empregados para selecionar células contendo um segundo cassete de expressão que codifica uma proteína de interesse. O segundo cassete de expressão pode estar presente em um ácido nucleico recombinante que também contém o presente polinucleotídeo de codificação de proteína ALS (isto é, em uma molécula de ácido nucleico recombinante única). Nessas modalidades, os presentes métodos de seleção podem ser empregados pa-

ra selecionar células que contêm o ácido nucleico recombinante. Uma vez que o ácido nucleico recombinante também contém o segundo cassete de expressão, células que contêm o segundo cassete de expressão são selecionadas usando o inibidor de ALS. Em modalidades alternativas, o segundo

5 cassete de expressão pode estar presente em um ácido nucleico recombinante que não contém o presente polinucleotídeo de codificação de proteína ALS (por exemplo, um vetor diferente). Nessas modalidades, um polinucleotídeo objeto pode ser cotransformado com (isto é, transformado ao mesmo tempo que) uma molécula de polinucleotídeo separada e distinta (por exemplo,

10 plo, uma molécula ou vetor de ácido nucleico diferente) que contém o cassete de expressão. Deste modo, os presentes métodos de seleção podem ser empregados para selecionar células que contêm o segundo cassete de expressão, mesmo que o segundo cassete de expressão esteja em uma molécula de ácido nucleico diferente ao polinucleotídeo.

15 Deste modo, os presentes métodos de seleção podem ser empregados para selecionar células hospedeiras que expressam a proteína de interesse. A proteína de interesse pode ser nativa ou não-nativa para as células hospedeiras usadas.

A concentração exata de inibidor de ALS empregada pode variar

20 de acordo com o inibidor de ALS particular usado e o tipo de célula hospedeira a ser selecionado. Em termos gerais, o inibidor de ALS é usado em uma concentração que provê seleção da célula hospedeira contendo o polinucleotídeo. Em certas modalidades, o inibidor de ALS pode ser empregado em uma concentração de 0,5 ppm a 10.000 ppm, por exemplo, 1 ppm a

25 10.000 ppm ou 10 ppm a 1.000 ppm. Por exemplo, em certos casos, o inibidor de ALS pode ser empregado em uma concentração na faixa de 25 ppm a 100 ppm, por exemplo, 50 ppm ou 100 ppm, ou em uma concentração na faixa de 100 ppm a 500 ppm, por exemplo, 200 ppm ou 500 ppm. Por exemplo, em uma modalidade, o inibidor de ALS pode ser empregado em uma

30 concentração na faixa de 1 µg/ml a 1 mg/ml, por exemplo, 10 µg/ml a 500 µg/ml.

Inibidores de ALS incluem quaisquer compostos que: a) matem

células que não têm resistência aos compostos através da inibição de uma enzima de ALS e b) não matem células tendo um polinucleotídeo objeto. Inibidores de ALS de interesse particular incluem sulfonilureia (SU), imidazolinona (IMI), triazolpirimidina (TP), pirimidiniltiobenzoato (PTB) e compostos sulfonilamino-carbonil-triazolinona (SCT) que são inibidores de ALS conhecidos e, em certos casos, podem ser geralmente empregados como herbicidas. Exemplos de compostos sulfonilureia que podem ser empregados nos métodos objetos incluem: I) fenilsulfonilureias, incluindo a) clorimuron etila (vide Agricultural Chemicals Book II "Herbicides" de W.T. Thompson, Thompson Publications, Fresno Calif., U.S.A. 1990 página 152); b) primisulfuron (CGA 136.872, vide Brighton Crop Prot. Conf. "Weeds" 1989, p. 41-48), c) 3-(4-etil-6-metóxi-1,3,5-triazin-2-il)-1-(2,3-di-idro-1,1-dioxo-2-metilbenzo[b]tiofen-7-sulfonil)-ureia (vide, por exemplo, EP-A-79.683), d) 3-(4-etóxi-6-etil-1,3,5-triazin-2-il)-1-(2,3-di-idro-1,1-dioxo-2-metilbenzo[b]tiofen-7-sulfonil)ureia (vide, por exemplo, EP-A-79.683), e) tribenuron-metila (vide "The Pesticide Manual", British Crop Protection Council 9th Edition (1990/91), página 840), f) metsulfuron-metila (vide, Proc. Int. Congr. Plant Prot., 10ª, 1983, vol. 1, 324), g) clorsulfuron (vide patente U.S. Nº 4.127.405; Weeds Weed Control, 1980, 21a, 24), h) triasulfuron (vide "The Pesticide Manual", 9ª Ed., p. 837) e i) sulfometuron-metila (vide "The Pesticide Manual", 9ª Ed., p. 774); II) tienilsulfonilureias, por exemplo, tifensulfuron-metila (vide Agricultural Chemicals Book II "Herbicides" de W.T. Thompson, Thompson Publications, Fresno Calif., U.S.A. 1990, página 155); III) pirazolisulfonilureias, por exemplo: a) pirazolsulfuron-etila (NC 311, vide "The Pesticide Manual", 9ª Ed., p. 735) e b) metil 3-cloro-5-(4,6-dimetoxipiridin-2-ilcarbamoilsulfonil)-1-metil-pirazol-4-carboxilato (vide EP 282.613); IV) derivados de sulfonodiamina, por exemplo, amidosulfuron e análogos estruturais (vide EP-A-0.131258 e Z. Pfl. Krankh, Pfl. Schutz, Special Issue XII, 489-497 (1990); V) piridilsulfonilureias, por exemplo: a) nicosulfuron (SL-950, vide Kimura e outros, Brighton Crop Protection Conference "Weeds", 1989, p. 29-34); b) DPX-E 9636 (vide Brighton Crop Prot. Conf. – Weeds – 1989, p. 23 et seq.) e c) piridilsulfonilureias como são descritos nos pedidos de patente A-

lemãs P 4000503.8 (WO-91/10660) e P 4030577.5 e VI) Fenoxisulfonilureias tal como aquelas descritas nas, por exemplo, EP-A-0.342.569, EP-A-4.163, EP-A-113.956, patente norte-americana Nº 4.678.500 e pat. norte-americana Nº 4.581.059. Exemplos de compostos imidazolinona que podem ser empregados nos métodos objeto incluem: a) imazetapir (vide "The Pesticide Manual" de Ch. R. Worthing, 8ª Edição, 1987, do British Crop Protection Council, página 473), b) imazaquin (vide "The Pesticide Manual" de Ch. R. Worthing, 8ª Edição, 1987, do British Crop Protection Council, página 474) e c) imazetapir (nome químico: ácido rac-2-[4,5-di-idro-4-metil-4-(1-metiletil)-5-oxo-1H-imidazol-2-il]-5-metil-3-piridino-carboxílico; vide "Weed Techn.", 1991 (5), 430-433 e 434-438) e outros compostos relacionados.

Células Hospedeiras

Uma célula hospedeira compreendendo um ácido nucleico recombinante objeto é também provida. A célula hospedeira pode ser qualquer tipo de célula, por exemplo, células hospedeiras bacterianas (tal como *E. coli*, *Bacillus* sp. ou *Streptomyces* sp.), fúngicas (tal como fúngica não-filamentosa ou filamentosa) ou de planta (tal como uma *Arabidopsis*, planta de milho ou soja). Em algumas modalidades, a célula hospedeira pode ser uma célula de uma espécie que tem uma história de uso para produção de proteínas que têm status GRAS, isto é, uma Geralmente Reconhecida como Segura, pelo FDA.

Em modalidades particulares, a célula hospedeira objeto pode ser uma célula fúngica da espécie que segue: *Trichoderma* (por exemplo, *Trichoderma reesei* (previamente classificada como *T. longibrachiatum* e atualmente também conhecida como *Hypocrea jecorina*), *Trichoderma viride*, *Trichoderma koningii* e *Trichoderma harzianum*)); *Penicillium* sp., *Humicola* sp. (por exemplo, *Humicola insolens* e *Humicola grisea*); *Chrysosporium* sp. (por exemplo, *C. lucknowense*), *Gliocladium* sp., *Aspergillus* sp. (por exemplo, *Aspergillus oryzae*, *Aspergillus Niger*, *Aspergillus nidulans*, *Aspergillus kawachi*, *Aspergillus aculeatus*, *Aspergillus japonicus*, *Aspergillus sojae* e *Aspergillus awamori*), *Fusarium* sp., *Mucor* sp., *Neurospora* sp., *Hypocrea* sp. ou *Emericella* sp. (Vide também Innis e outros (1985) *Sci.*, 228:21-26),

dentre outros.

Células hospedeiras bacterianas exemplares incluem *Bacillus* *sp.*, incluindo, mas não limitado a, *B. subtilis*, *B. licheniformis*, *B. lentus*, *B. brevis*, *B. stearothermophilus*, *B. alkalophilus*, *B. amyloliquefaciens*, *B. clausii*, *B. halodurans*, *B. megaterium*, *B. coagulans*, *B. circulans*, *B. lautus* e *B. thuringiensis* e *Streptomyces* *sp.*, incluindo, mas não limitado a: *S. lividans*, *S. carbophilus* e *S. helveticus*.

Células hospedeiras de planta exemplares incluem células de planta monocotiledôneas e dicotiledôneas, incluindo, mas não limitado a, células hospedeiras de milho (*Zea mays*), *Brassica* *sp.*, arroz (*Oryza sativa*), trigo (*Triticum aestivum*), soja (*Glycine Max*), tabaco (*Nicotiana tabacum*) e *Arabidopsis thaliana* e tomate (*Lycopersicon esculentum*). Uma célula hospedeira pode ser uma célula hospedeira cultivada *in vitro* ou uma célula hospedeira de um organismo multicelular, isto é, uma planta. Métodos de transferência de ácidos nucleicos exógenos para tais células hospedeiras são bem conhecidos na técnica.

Em modalidades particulares, uma célula fúngica objeto pode ser uma linhagem de *Trichoderma* e particularmente *T. reesei* que inclui equivalentes funcionais de RL-P37 (Sheir-Neiss e outros (1984) *Appl. Microbiol. Biotechnology*, 20:46-53). Outras linhagens hospedeiras úteis incluem: NRRL 15709, ATCC 13631, ATCC 26921 (QM 9414) ATCC 32098, ATCC 32086 e ATCC 56765 (RUT-30). Em outras modalidades, célula hospedeira fúngica pode ser de uma linhagem de *Aspergillus* *sp.*, incluindo ATCC 22342, ATCC 44733, ATCC 14331, ATCC 11490, NRRL 3112 e linhagens derivadas da mesma.

Em algumas modalidades, uma célula hospedeira pode ser uma em que os genes nativos foram deletados ou inativados. Por exemplo, genes correspondendo a genes de protease (por exemplo, aspartil protease) (Berka e outros (1990), *Gene* 86:153-162 e USP 6.509.171) ou genes correspondendo a genes de celulase podem ser deletados ou inativados (por exemplo, *cbh1*, *cbh2* e *egl1* e *egl2*) tal como a linhagem quadruplamente deletada de *T. reesei* descrita no WO 05/001036 e derivados da mesma.

Introdução de um ácido nucleico em uma célula hospedeira inclui técnicas tal como transformação; eletroporação; microinjeção nuclear; transdução; transfecção (por exemplo, transfecção mediada por lipofecção e mediada por DEAE-Dextrina); incubação com precipitado de fosfato de cálcio; bombardeamento de alta velocidade com microprojéteis revestidos com DNA; e fusão de protoplasto. Técnicas de transformação gerais são conhecidas na técnica (vide, por exemplo, Ausubel e outros (1987), *supra*, capítulo 9; e Sambrook (1989) *supra* e Campbell e outros (1989) *Curr. Genet.* 16:53-56). Referência é também feita aos WO 05/001036; USP 6.022.725; USP 6.103.490; USP 6.268.328; e pedidos de patente norte-americana publicados 20060041113, 200600403532, 20060040353 e 20050208623, cujas publicações são aqui incorporadas a título de referência.

Transformação e expressão de proteína em *Aspergillus* e *Trichoderma* são adicionalmente descritas na, por exemplo, Pat. norte-americana Nº 5.364.770; Pat. U.S. Nº 6.022.725; e Nevalainen e outros, 1992, *The Molecular Biology of Trichoderma and its Application to the Expression of Both Homologous and Heterologous Genes*, em MOLECULAR INDUSTRIAL MYCOLOGY, Eds. Leon e Berka, Marcel Dekker, Inc. pp. 129-148.

Conforme acima mencionado, uma célula hospedeira pode conter ainda um ácido nucleico recombinante para expressão de uma proteína de interesse em uma célula hospedeira, em adição ao ácido nucleico codificando ALS. O ácido nucleico recombinante objeto e o ácido nucleico codificando ALS podem ser ligados intimamente em *cis*, ou no genoma ou em um plasmídeo, de modo que o inibidor de ALS seleciona o ácido nucleico recombinante, e então seleciona as células que produzem a proteína.

Produção de Proteína

Métodos de uso da célula hospedeira descrita acima são também providos. Em certas modalidades, os métodos objetos incluem: cultura da célula compreendendo um ácido nucleico recombinante compreendendo um primeiro cassete de expressão para produção de uma enzima ALS objeto e um segundo cassete de expressão para produção de uma proteína, pa-

ra produzir a proteína. Em certas modalidades e conforme acima discutido, a proteína pode ser secretada no meio de cultura. Deste modo, certas modalidades do método incluem a etapa de recuperação da proteína a partir do meio de cultura.

- 5 Células podem ser cultivadas em um meio padrão contendo sais fisiológicos e nutrientes (vide, por exemplo, Pourquie, J. e outros, BIO-CHEMISTRY AND GENETICS OF CELLULOSE DEGRADATION, Eds. Aubert, J. P. e outros, Academic Press, pp. 71-86, 1988 e Ilmen, M. e outros (1997) *Appl. Environ. Microbiol.* 63:1298-1306). Meios preparados comercialmente comuns (por exemplo, caldo de Extrato de Malte de Levedura (YM),
10 caldo Luria Bertani (LB) e caldo de Dextrose Sabouraud (SD) também encontram uso na presente invenção. Condições de cultura preferidas para um dado fungo filamentoso são conhecidas na técnica e podem ser encontradas na literatura científica e/ou da fonte dos fungos tal como o *American Type*
15 *Culture Collection* (ATCC) e *Fungal Genetics Stock Center*.

- Em algumas modalidades, uma célula hospedeira objeto pode ser cultivada sob condições de fermentação em batelada ou contínua. Uma fermentação em batelada clássica é um sistema fechado, em que a composição do meio é ajustada no início da fermentação e não é submetida a alterações artificiais durante a fermentação. Então, no início da fermentação, o
20 meio é inoculado com o(s) organismo(s) desejado(s). Neste método, fermentação é permitida acontecer sem a adição de quaisquer componentes ao sistema. Tipicamente, uma fermentação em batelada qualifica como uma "batelada" com relação à adição da fonte de carbono e tentativas são frequentemente feitas em controlar fatores tal como pH e concentração de oxigênio. As composições de metabólito e biomassa do sistema de batelada mudam constantemente até o momento que a fermentação é parada. Dentro das culturas de batelada, as células progridem através de uma fase lag estática para uma fase log de crescimento alta e finalmente para uma fase estacionária onde a taxa de crescimento é diminuída ou parada. Se não tratadas,
30 células na fase estacionária eventualmente morrem. Em geral, células na fase log são responsáveis pelo volume da produção de produto final.

Uma variação do sistema de batelada padrão é o sistema "de fermentação em batelada alimentada", que também encontra uso com a presente invenção. Nesta variação de um sistema em batelada típico, o substrato é adicionado em incrementos conforme a fermentação progride. Sistemas *fed-batch* são úteis quando repressão de catabólito está apta a inibir o metabolismo das células e onde for desejável ter quantidades limitadas de substrato no meio. Medição da concentração de substrato real nos sistemas *fed-batch* é difícil e então estimada com base nas mudanças de fatores mensuráveis tal como pH, oxigênio dissolvido e a pressão parcial de gases residuais tal como CO₂. Fermentações em batelada e *fed-batch* são comuns e conhecidas na técnica.

Fermentação contínua é um sistema aberto onde um meio de fermentação definido é adicionado continuamente a um biorreator e uma quantidade igual de meio condicionado é removida simultaneamente para processamento. Fermentação contínua geralmente mantém as culturas em uma densidade alta constante onde as células estão principalmente em crescimento de fase log.

Fermentação contínua permite a modulação de um fator ou qualquer número de fatores que afetam crescimento celular e/ou concentração de produto final. Por exemplo, em uma modalidade, um nutriente limitante tal como a fonte de carbono ou fonte de nitrogênio é mantido em uma taxa fixa e todos os outros parâmetros são deixados moderar. Em outros sistemas, vários fatores que afetam o crescimento podem ser alterados continuamente enquanto a concentração celular, medida através de turbidez média, é mantida constante. Sistemas contínuos tentam manter condições de crescimento de estado uniforme. Então, perda celular devido a meio ser extraído deve ser equilibrada contra a taxa de crescimento celular na fermentação. Métodos de modulação de nutrientes e fatores de crescimento para processos de fermentação contínuos bem como técnicas para maximização da taxa de formação de produto são conhecidos.

Uma célula hospedeira fúngica pode ser cultivada em um meio padrão contendo sais e nutrientes fisiológicos (vide, por exemplo, Pourquie,

J. e outros, BIOCHEMISTRY AND GENETICS OF CELLULOSE DEGRADATION, Eds. Aubert, J. P. e outros, Academic Press, pp. 71-86, 1988 e Ilmen, M. e outros (1997) *Appl. Environ. Microbiol.* 63:1298-1306). Meios comercialmente preparados comuns (por exemplo, caldo de Extrato de Malte de Levedura (YM), caldo Luria Bertani (LB) e caldo de Dextrose Sabouraud (SD) também encontram uso nos presentes métodos. Condições de cultura preferidas para células hospedeiras fúngicas são conhecidas na técnica e podem ser encontradas na literatura científica e/ou da fonte dos fungos tal como a *American Type Culture Collection* (ATCC) e *Fungal Genetics Stock Center*.

Proteína pode ser recuperada dos meios de crescimento através de qualquer meio conveniente, por exemplo, através de precipitação, centrifugação, afinidade, filtragem ou qualquer outro método conhecido na técnica. Em outra modalidade, uma cultura de células é provida, onde a cultura de células compreende: a) meio de crescimento e b) a célula hospedeira acima descrito.

Promotores 1818A e 1818B

Um promotor que pode ser usado para expressar uma proteína em uma célula hospedeira é também provido. Em uma modalidade, o promotor compreende a sequência de nucleotídeo da SEQ ID NO:7 ou 8 (mostrada nas figuras 5A e 5B, respectivamente) ou uma subsequência ou equivalente funcional da mesma que tem atividade promotora em uma célula hospedeira. Também providos são ácidos nucleicos recombinantes e vetores contendo o promotor e células hospedeiras contendo o ácido nucleico ou vetor. Métodos de produção de uma proteína usando as células hospedeiras são também providos.

Em certas modalidades, o promotor pode compreender a sequência de nucleotídeo de: a) SEQ ID NO:7 ou 8; b) uma subsequência de SEQ ID NO:7 ou 8 que retém atividade de promotor; c) uma sequência funcionalmente equivalente de SEQ ID NO:7 ou 8 que retém a atividade de promotor ou d) uma sequência de ácido nucleico que hibridiza sob condições de hibridização estridentes com SEQ ID NO:7 ou 8 ou a subsequência da

mesma. Em modalidades particulares, a sequência de nucleotídeo pode ser pelo menos 80% idêntica à sequência de nucleotídeo da SEQ ID NO:7 ou 8.

Em modalidades particulares, uma subsequência de um promotor objeto pode conter pelo menos cerca de 100 nucleotídeos, pelo menos
5 cerca de 200 nucleotídeos; pelo menos cerca de 250 nucleotídeos; pelo menos cerca de 300 nucleotídeos; pelo menos cerca de 350 nucleotídeos; pelo menos cerca de 400 nucleotídeos; pelo menos cerca de 450 nucleotídeos; pelo menos cerca de 500 nucleotídeos; pelo menos cerca de 550 nucleotídeos; pelo menos cerca de 600 nucleotídeos; pelo menos cerca de 650 nucleotídeos; pelo menos cerca de 700 nucleotídeos; pelo menos cerca de 800
10 nucleotídeos; pelo menos cerca de 850 nucleotídeos que são contíguos na SEQ ID NO:7 ou 8, a sequência contínua completa da SEQ ID NO:7 ou 8 ou um equivalente funcional da mesma que retém atividade promotora.

Em certas modalidades, um promotor equivalente funcional pode
15 de incluir uma ou mais mudanças (por exemplo, 1, 2, 3, 4, 5, 6, 7, 8 ou 9, mais de 10, até 20 ou 30 ou mais mudanças) com relação à sequência de nucleotídeo da SEQ ID NO:7 ou 8, onde uma mudança pode ser uma deleção, substituição ou inserção, por exemplo. Em uma modalidade exemplar, a sequência de nucleotídeo do promotor equivalente funcional pode incluir
20 uma a cinco diferenças de nucleotídeo com relação à sequência de nucleotídeo do promotor de origem tal como SEQ ID NO:7 ou 8.

Em outras modalidades, o promotor pode incluir uma sequência de nucleotídeo que hibridiza sob condições de hibridização estridentes para um polinucleotídeo tendo a sequência de nucleotídeo da SEQ ID NO:7 ou 8,
25 onde condições de hibridização estridentes compreendem condições de hibridização de estridência baixa, média, alta e muito alta, onde tais condições são descritas acima.

Em outra modalidade, um promotor objeto pode conter uma sequência de nucleotídeo continua que é pelo menos 80%, pelo menos 85%,
30 pelo menos 90%, pelo menos 95%, pelo menos 96%, pelo menos 97%, pelo menos 98% ou pelo menos 99% idêntica à SEQ ID NO:7 ou 8 ou uma subsequência da mesma. Em uma modalidade, o promotor objeto pode conter

uma sequência de nucleotídeo contígua que é pelo menos 95% idêntica à SEQ ID NO:7 ou 8.

Conforme mencionado na seção Exemplos abaixo, o ácido nucleico das SEQ ID NOS:7 e 8 foi obtido de *Trichoderma reesei*, um fungo filamentoso. Como ficaria prontamente aparente, equivalentes funcionais de SEQ ID NOS:7 e 8 que retêm atividade promotora podem ser identificados através de sequências de identificação que são similares às SEQ ID NOS:7 e 8 em outros fungos filamentosos. Uma vez que a maioria ou todas das sequências de genoma de outros fungos filamentosos, por exemplo, *Aspergillus* (por exemplo, *Aspergillus fumigatus*, *Aspergillus oryzae* (vide, por exemplo, Machida e outros, *Nature*, 2005, 438, 1157-1161), *Aspergillus nidulans*, *Aspergillus fumigatus*, *Aspergillus Níger*, *Aspergillus flavus*, *Aspergillus terreus*), *Neurospora* (por exemplo, *Neurospora crassa*) e *Fusarium* (por exemplo, *Fusarium graminearum*) estão disponíveis, equivalentes funcionais de SEQ ID NOS:7 e 8 que têm atividade promotora são prontamente identificáveis.

Conforme acima mencionado, um promotor objeto pode ter atividade promotora em uma célula hospedeira. Atividade promotora pode ser detectada usando qualquer ensaio adequado. Em certas modalidades, um promotor objeto pode ser operavelmente ligado a um polinucleotídeo, e transcrição do polinucleotídeo pode ser detectada usando qualquer método adequado, por exemplo, *Northern blotting* ou RT-PCR, etc. Em outras modalidades, o promotor pode ser operavelmente ligado a um polinucleotídeo que codifica uma proteína, por exemplo, uma proteína repórter, e a atividade do promotor pode ser avaliada através da detecção da proteína. Nessas modalidades, se necessário, uma região 5' não-traduzida pode ser ligada ao promotor de modo que o transcrito resultante tem uma 5' UTR seguida por uma sequência de codificação. Como seria reconhecido, os resultados obtidos a partir de tal ensaio podem ser comparados com resultados comparados com um controle adequado, por exemplo, um controle negativo ou positivo, para determinar a significância de resultados obtidos. Qualquer célula hospedeira, por exemplo, uma célula bacteriana tal como uma célula hospedeira *E. coli*,

Bacillus ou *Streptomyces*, ou uma célula fúngica filamentosa, por exemplo, uma célula hospedeira *Aspergillus ssp.*, *Trichoderma SSP.* ou *Fusarium ssp.* pode ser empregada. Não há nenhuma necessidade que um promotor objeto esteja contido dentro de uma célula hospedeira particular. Em certos casos, o promotor pode ser testado quando à atividade promotora em uma célula hospedeira *Trichoderma reesei*.

Um ácido nucleico recombinante compreendendo o promotor objeto é também provido. Em certos casos, o ácido nucleico recombinante pode compreender um promotor objeto e um polinucleotídeo, onde o promotor e o polinucleotídeo são operavelmente ligados de modo que o promotor causa transcrição do polinucleotídeo em uma célula. Em certos casos, o promotor e o polinucleotídeo não estão normalmente ligados em natureza, isto é, são heterólogos um para o outro. Em certos casos, o polinucleotídeo pode conter uma sequência de codificação para uma proteína. A proteína pode ser uma enzima, uma proteína repórter ou terapêutica (por exemplo, uma proteína de anticorpo), conforme acima discutido, por exemplo. Em certas modalidades, a proteína pode ser uma proteína de fusão que pode, em certos casos, conter uma sequência de sinal ou porção-veículo para secreção da proteína.

Um vetor de ácido nucleico compreendendo o ácido nucleico recombinante objeto é também provido, bem como uma célula hospedeira contendo o mesmo. Em certas modalidades, o ácido nucleico recombinante pode estar presente no genoma da célula hospedeira. Em outras modalidades, o ácido nucleico recombinante pode estar presente em um vetor que replica na célula. A célula hospedeira pode ser qualquer uma de uma variedade de células hospedeiras diferentes, incluindo células bacterianas, fúngicas, de levedura, de planta e de mamífero. Em uma modalidade, a célula hospedeira pode ser uma célula hospedeira fúngica filamentosa, e, em outra modalidade, a célula hospedeira pode ser uma célula bacteriana.

Uma cultura de células compreendendo meio de cultura e uma célula hospedeira objeto é também provida.

Um método de produção de uma proteína é também provido.

Em termos gerais, este método inclui manutenção de uma cultura objeto de células sob condições adequadas para produzir a proteína. Este método pode incluir ainda recuperação da proteína a partir do meio de cultura.

A fim de ilustrar mais a presente invenção e vantagens da mesma, os exemplos específicos que seguem são dados com a compreensão de que eles estão sendo oferecidos para ilustrar a presente invenção e não devem ser considerados de modo algum como limitando seu escopo.

Exemplo 1

Isolamento de genes de ALS de células resistentes à clorimuron-etila

Clorimuron-etila (Chem Service Inc., West Chester, PA), recém-preparada e dissolvida em DMF para uma concentração de 20 mg/ml, foi adicionada em várias concentrações a ágar Vogels derretido imediatamente antes de verter as placas para produzir meios contendo 25, 50, 100, 200, 300, 400 ou 500 ppm de clorimuron etila.

Aproximadamente 25 milhões de esporos da linhagem *T. reesei* QM6a (ATCC13631) foram plaqueadas em cada placa. Após 10 dias de crescimento a 28°C, cinco colônias foram isoladas, duas de placas de 50 ppm e uma de cada uma das placas de 25, 200 e 300 ppm. As colônias foram mais isoladas fazendo novas listras delas sobre placas de ágar Vogels fresco contendo 200 ppm de clorimuron etila.

DNA genômico foi preparado a partir de cada uma de cinco linhagens resistentes à clorimuron etila a fim de amplificar e sequenciar o gene de acetolactato sintase (*als*) dessas linhagens. Polimerase de DNA Herculese (Stratagene, La Jolla, CA) foi usada para amplificação seguindo instruções do fabricante e os iniciadores que seguem: Avançado (5'-3') GGCGCGCCTGAGACAATGGCCGGCAATGGTAAAA (SEQ ID NO:5) e Reverso (5'-3') GCGATCGCCATCCCGTCGCGTCAAAAACACTGC (SEQ ID NO:6).

Sítios de restrição únicos foram adicionados às extremidades 5' dos *primers* para manipulação subsequente. Os fragmentos de 4,1 kb resultantes foram isolados, sequenciados e comparados com a sequência do gene de acetolactato sintase de *T. reesei* (do website JGI genome).

Três mutações por ponto únicas no gene *als* que confere resistência à clorimuron etila foram identificadas.

Comparado com a sequência da proteína acetolactato sintase de *T. reesei* do tipo selvagem (figura 1; SEQ ID NO:1), as proteínas acetolactato sintase resistentes à clorimuron etila tinham cada uma das substituições de aminoácido que seguem: A190D, K241E ou R372H. Os aminoácidos nas posições 190, 241 e 372 da proteína acetolactato sintase de *T. reesei* do tipo selvagem de SEQ ID NO:1 estão sublinhados.

Comparado com a sequência do cDNA deduzido da acetolactato sintase de *T. reesei* do tipo selvagem (figura 2; SEQ ID NO:2), os cDNAs de acetolactato sintase resistentes à clorimuron etila tinham cada uma das substituições de nucleotídeo que seguem: C596A (esta corresponde à substituição de aminoácido A190D), A721G (esta corresponde à substituição de aminoácido K241E) ou G1115A (esta corresponde à substituição de aminoácido R372H). Cada um dos códons alterados (GCC, que é alterado para GAC; AAG, que é alterado para GAG; e GCT, que é alterado para CAT) é indicado na figura 2.

Comparado com a sequência do gene de acetolactato sintase de *T. reesei* do tipo selvagem (figura 3; SEQ ID NO:3), os genes de acetolactato sintase resistentes à clorimuron etila tinham cada um uma das substituições de nucleotídeo que seguem: C1023A (esta corresponde à substituição de aminoácido A190D), A1175G (esta corresponde à substituição de aminoácido K241E) ou G1569A (esta corresponde à substituição de aminoácido R372H).

Por causa da degeneração do código genético, outras mutações na sequência de codificação de *als* podem codificar a substituição de aminoácido A190D, K241E ou R372H.

Exemplo 2

Transformação de *T. reesei* com gene de acetolactato sintase A190D

O vetor pTrex-glicoamilase foi criado para expressar uma glicoamilase derivada de *Trichoderma reesei* em *T. reesei*. A figura 4 mostra um diagrama geral do vetor. Três construtos de vetor diferentes foram feitos com

promotores diferentes: 1818A, 1818B e o promotor de *cbh1* de *T. reesei*. A figura 5 ilustra as sequências do promotor designado 1818A e a figura 5B ilustra as sequências do promotor 1818B.

A sequência de nucleotídeo inteira do pTrex-glicoamilase incluindo o promotor de *cbh1* é mostrada na figura 6. O vetor pTrex-glicoamilase é baseado no vetor pTrex3g conforme descrito em detalhes no WO 05/001036. Resumindo, o pTrex3g é baseado no vetor de *E. coli* pSL1180 (Pharmacia, Inc., Piscataway, NJ) que é um vetor baseado em fagemídeo pCU18 com um sítio de clonagem múltiplo estendido contendo sequências de reconhecimento de enzima de restrição de 64 hexâmeros. Ele foi chamado um vetor de destino Gateway (Hartley, J.L. e outros (2000) *Genome Research* 10:1788-1795) para permitir inserção usando Tecnologia Gateway (Invitrogen) de qualquer estrutura de leitura aberta desejada entre as regiões promotora e terminadora do gene de *cbh1* de *T. reesei*.

No vetor pTrex-glicoamilase, o gene de ALS A190D está sob controle de seu promotor e terminador nativo e é usado para substituir o marcador selecionável fúngico *amdS* que é usado em pTrex3g.

O vetor foi transformado na linhagem de *T. reesei* deletada de quad ($\Delta chb1$, $\Delta cbh2$, $\Delta eg1$ e $\Delta eg2$) (WO 05/001036) originalmente derivada de RL-P37 (Sheir-Neiss e outros (1984) *Appl. Microbiol. Biotechnol.* 20:46-53; USP 4.797.361) usando o procedimento mostrado abaixo.

Uma suspensão de esporos (aproximadamente 5×10^8 esporos/ml) da linhagem de *Trichoderma* foi preparada. 100 μ l-200 μ l de suspensão de esporo foram espalhados no centro de placas de meio Vogels modificado com 200 ppm de clorimuron etila e deixados secar.

Vogels modificado tinha a composição que segue: 2,5 g/L de Citrato de $\text{Na}_3 \cdot 2\text{H}_2\text{O}$, 5,0 g/L de KH_2PO_4 , 2,0 g/L de NH_4NO_3 , 0,2 g/L de $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$, 0,1 g/L de $\text{CaCl}_2 \cdot 2\text{H}_2\text{O}$, 5 mL/L de Solução de Elementos Traços Vogels Modificado, 2,5 mL/L de Solução de Biotina Vogels Modificado, 20 g/L de solução de elementos traços Vogels Modificado continha 50 g/L de Ácido Cítrico, 50 g/L de $\text{ZnSO}_4 \cdot 7\text{H}_2\text{O}$, 10 g/L $\text{Fe}(\text{NH}_4)_2\text{SO}_4 \cdot 6\text{H}_2\text{O}$, 2,5 g/L de $\text{CuSO}_4 \cdot 5\text{H}_2\text{O}$, 0,5 g/L de $\text{MnSO}_4 \cdot 4\text{H}_2\text{O}$, 0,5 g/L de H_3BO_3 , 0,5 g/L de NaMo-

$O_4 \cdot 2H_2O$.

Solução de Biotina de Vogels Modificado (Modified Vogels Biotin Solution) continha 0,1 g /L de d-Biotina. Após autoclave as adições que se-
guem são feitas antes de verter as placas: 20 mL/L de glicose 50%, 10 mL/L
5 de 20 mg/mL de clorimuron etila dissolvido em DMF.

Transformação da linhagem de *Trichoderma* através do método de transformação biolística foi realizada usando um Biolistic® PDS-1000/he Particle Delivery System da Bio-Rad (Hercules, CA) seguindo as instruções do fabricante (vide WO 05/001036 e US 2006/0003408).

10 Transformantes foram isolados após 3 a 4 dias de crescimento a 28°C. Transformantes foram serialmente passados duas vezes em placas de Vogels fresco com 200 ppm de clorimuron etila a fim de isolar monocarions estáveis. Transformantes foram cultivados e o sobrenadante cultivado foi testado através de eletroforese em SDS Page e usando um ensaio de enzi-
15 ma. Um gel de SDS PAGE é mostrado na figura 7 ilustrando a expressão da glicoamilase de *T. reesei*.

Exemplo 3

Sequenciamento e clonagem dos promotores 1818A e 1818B

Os promotores 1818A e 1818B (conforme mostrado nas figuras
20 5A e 5B, respectivamente) foram identificados através de exploração do banco de dados de sequência do genoma de *Trichoderma reesei* do *US Department of Energy Joint Genome Institute* (conforme encontrado no *world web site* de jgi.doe.gov) para sequências a montante de ESTs altamente re-
presentados.

25 Os promotores foram amplificados através de PCR e clonados por trás de uma sequência de codificação de glicoamilase de acordo com o mapa do vetor mostrado na figura 8. Toda a sequência de nucleotídeo deste construto é mostrada na figura 9. O vetor usado é baseado no vetor p-Trex3g, que é descrito em detalhes no Exemplo 6 do WO 05/001036. Resu-
30 mindo, o pTrex3g é baseado no vetor de *E. coli* pSL1180 (Pharmacia, Inc., Piscataway, NJ) que é um vetor baseado em fagemídeo pUC118 com um sítio de clonagem múltiplo estendido contendo sequências de reconheci-

to de enzima de restrição de 64 hexâmeros. Ele foi chamado um vetor de destino Gateway (Hartley, J.L. e outros (2000) *Genome Research* 10:1788-1795) para permitir inserção usando Tecnologia Gateway (Invitrogen) de qualquer estrutura de leitura aberta desejada entre as regiões de promotor e terminador do gene *cbh1* de *T. reesei*. No vetor usado neste exemplo, o promotor de *cbh1* foi substituído ou com o promotor 1818A ou o 1818B, o marcador selecionável *amdS* foi substituído com marcador de acetolactato sintase de *Trichoderma reesei*, dirigido pelos seus promotor e terminador nativos.

10 Exemplo 4

Transformação de células hospedeiras de *T. reesei* com vetores incluindo os promotores 1818A e 1818B

O vetor foi transformado na linhagem de *T. reesei* deletada de quad ($\Delta chb1$, $\Delta cbh2$, $\Delta egl1$ e $\Delta egl2$) (WO 05/001036) originalmente derivada de RL-P37 (Sheir-Neiss e outros (1984) *Appl. Microbiol. Biotechnol.* 20:46-53; USP 4.797.361) usando o procedimento mostrado abaixo.

Uma suspensão de esporos (aproximadamente 5×10^8 esporos/ml) da linhagem de *Trichoderma* foi preparada. 100 μ l-200 μ l de suspensão de esporo foram espalhados no centro de placas de meio de Vogels modificado com 200 ppm de clorimuron etila. Vogels modificado tinha a composição que segue: 2,5 g/L de Citrato de $\text{Na}_3^*2\text{H}_2\text{O}$, 5,0 g/L de KH_2PO_4 , 2,0 g/L de NH_4NO_3 , 0,2 g/L de $\text{MgSO}_4^*7\text{H}_2\text{O}$, 0,1 g/L de $\text{CaCl}_2^*2\text{H}_2\text{O}$, 5 mL/L de Solução de Elementos Traços de Vogels Modificado (Modified Vogels Trace Elements Solution, 2,5 mL/L de Solução de Biotina de Vogels Modificado, 20 g/L de solução de elementos traços de Vogels Modificado continha 50 g/L de Ácido Cítrico, 50 g/L de $\text{ZnSO}_4^*7\text{H}_2\text{O}$, 10 g/L $\text{Fe}(\text{NH}_4)_2\text{SO}_4^*6\text{H}_2\text{O}$, 2,5 g/L de $\text{CuSO}_4^*5\text{H}_2\text{O}$, 0,5 g/L de $\text{MnSO}_4^*4\text{H}_2\text{O}$, 0,5 g/L de H_3BO_3 , 0,5 g/L de $\text{NaMoO}_4^*2\text{H}_2\text{O}$. Solução de Biotina de Vogels Modificado continha 0,1 g /L de d-Biotina. Após autoclave as adições que seguem são feitas antes de verter as placas: 20 mL/L de glicose 50%, 10 mL/L de 20 mg/mL de clorimuron etila dissolvido em DMF. A suspensão foi deixada secar sobre a superfície das placas de Vogels Modificado.

Transformação da linhagem de *Trichoderma* através do método de transformação biolística foi realizada usando um Biolistic® PDS-1000/he Particle Delivery System da Bio-Rad (Hercules, CA) seguindo as instruções do fabricante (vide WO 05/001036 e US 2006/0003408).

5 Transformantes foram isolados.

Exemplo 5

Avaliação de transformantes quanto à expressão de atividade de glicoamilase dirigida por promotores 1818A e 1818B

Transformantes estáveis foram cultivados em placas de Ágar de Lactose de Vogels Modificado com amido (placas de Petri de 9 cm de diâmetro). Após cerca de 4 dias de crescimento a 28°C, 10 ml de uma solução de 1 mg/ml de 4-metil-umbeliferil- α -D-glicose foram vertidos nas colônias de crescimento. Após 30 minutos em temperatura ambiente, linhagens expressando glicoamilase foram visualizadas como colônias azuis fluorescentes quando vistas iluminadas por uma lâmpada UV de comprimento de onda longo. A linhagem controle de *T. reesei* de origem, não-transformada, não mostrou fluorescência azul.

Placas de Ágar de Lactose de Vogels Modificado com amido é a mesma receita de Ágar de Vogels Modificado exceto que 25 mL/L de solução de α -lactose a 20% (adicionada após autoclave) substitui solução de glicose e 20 g/L de amido de milho *Pure Food Powder* são adicionados antes do autoclave.

O substrato de glicoamilase 4-metil-umbeliferil- α -D-glicose foi preparado como segue: 200 mg de 4-metil-umbeliferil- α -D-glicose (Sigma-Aldrich Co.) são dissolvidos em 5 ml de DMSO e 195 ml de Tampão de Fosfato de Potássio 75 mM pH 6,3 são adicionados.

Exemplo 6

Experimentos de frasco agitado com transformantes de *Trichoderma reesei*

Transformantes fúngicos individuais serão cultivados em cultura de frasco agitado para determinar o nível de expressão de proteína glicoamilase. O experimento será conduzido essencialmente conforme descrito no Exemplo 1 da patente norte-americana 5.874.276 com a modificação que

segue: 16 g/L de alfa-lactose substitui celulose em meio TSF.

Em geral, o protocolo de fermentação conforme descrito em Foreman e outros (Foreman e outros (2003) *J. Biol. Chem.* 278:31988-31997) será seguido. Meio mínimo Vogels (Davis e outros (1970) *Methods in Enzymology*, 17^a, pg. 79-143 e Davis, Rowland, *NEUROSPORA, CONTRIBUTIONS OF A MODEL ORGANISM*, Oxford University Press (2000) contendo glicose 5% será inoculado com 1,5 ml de suspensão de esporo congelada. Após cerca de 48 horas, cada cultura será transferida para 6,2 L do mesmo meio em um fermentador Biolafitte de 14L. O fermentador será ativado a 25°C, 750 RPM e fluxo de ar de 8 litros padrão por minuto. Cerca de uma hora após a glicose inicial ser esgotada, uma alimentação de lactose a 25% (p/p) será iniciada e alimentada de uma maneira limitante de carbono para prevenir acúmulo de lactose. As concentrações de glicose e lactose serão monitoradas. Amostras serão obtidas em intervalos regulares para monitorar o progresso da fermentação. Amostras coletadas serão giradas em um tubo de centrífuga de 50 ml em uma velocidade $\frac{3}{4}$ em uma centrífuga clínica da International Equipment Company (Needham Heights, MA). Amostras de sobrenadante cultivadas em frasco de agitação serão administradas em géis BIS-TRIS SDS-PAGE (Invitrogen) sob condições de redução com tampão de administração de SDS de MOPS (ácido morfolinopropanossulfônico) e tampão de amostra de LDS.

REIVINDICAÇÕES

1. Polinucleotídeo, caracterizado pelo fato de que compreende:

(a) uma sequência de nucleotídeos selecionada a partir do grupo consistindo em SEQ ID NO: 2, 4 e sequências degeneradas das mesmas codificando uma proteína acetolactato sintase de SEQ ID NO:1 que provê resistência a compostos sulfonilureia tóxicos; ou

(b) uma sequência de nucleotídeos de (a) compreendendo ainda uma ou mais das seguintes mutações:

(i) o códon na posição de aminoácido 190 é GAT, GAC, GAA ou GAG;

(ii) o códon na posição de aminoácido 241 é GAT, GAC, GAA ou GAG;

(iii) o códon na posição de aminoácido 372 é CAT ou CAC;

em que o referido polinucleotídeo é operacionalmente ligado a um promotor heterólogo.

2. Polinucleotídeo de acordo com a reivindicação 1, caracterizado pelo fato de que é operavelmente ligado a um terminador.

3. Polinucleotídeo de acordo com a reivindicação 1, caracterizado pelo fato de que é códon-otimizado para expressão da dita proteína acetolactato sintase em uma célula hospedeira particular.

4. Polinucleotídeo de acordo com a reivindicação 1, caracterizado pelo fato de que é inserido no genoma de uma célula hospedeira.

5. Polinucleotídeo de acordo com a reivindicação 1, caracterizado pelo fato de que a referida célula hospedeira é uma célula de planta.

6. Vetor, caracterizado pelo fato de que compreende o polinucleotídeo como definido na reivindicação 1.

7. Vetor de acordo com a reivindicação 6, caracterizado pelo fato de que compreende ainda um cassete de expressão para expressão de uma proteína recombinante.

8. Micro-organismo transgênico, caracterizado pelo fato de que compreendendo o polinucleotídeo isolado como definido na reivindicação 1, em que o referido micro-organismo transgênico é resistente a compostos

sulfonilureia tóxicos.

9. Micro-organismo transgênico de acordo com a reivindicação 8, caracterizado pelo fato de que é um fungo.

5 10. Micro-organismo transgênico de acordo com a reivindicação 9, caracterizado pelo fato de que é um fungo filamentoso.

11. Micro-organismo transgênico de acordo com a reivindicação 10, caracterizado pelo fato de que é *Trichoderma reesei*.

10 12. Micro-organismo transgênico de acordo com a reivindicação 8, caracterizado pelo fato de que o referido polinucleotídeo está presente no genoma do referido micro-organismo transgênico.

13. Micro-organismo transgênico de acordo com a reivindicação 8, caracterizado pelo fato de que o referido polinucleotídeo está presente em um vetor que replica autonomamente no referido micro-organismo transgênico.

15 14. Método de seleção de uma célula, caracterizado pelo fato de que compreende:

introdução de um polinucleotídeo isolado como definido na reivindicação 1 em uma pluralidade de células;

contato da dita pluralidade de células com composto sulfonilureia ou imidazolinona tóxico; e

20 cultura das ditas células para prover seleção da dita célula.

15. Método de acordo com a reivindicação 14, caracterizado pelo fato de que a dita célula é uma célula fúngica.

16. Método de acordo com a reivindicação 15, caracterizado pelo fato de que a dita célula é uma célula de *Trichoderma reesei*.

25 17. Método de acordo com a reivindicação 14, caracterizado pelo fato de que a dita célula é uma célula de planta.

18. Método de acordo com a reivindicação 14, caracterizado pelo fato de que compreende ainda cultura da dita célula em meio líquido.

MLRSRQVTARAVRALGQARAFTSTTKPVMIQSSQRKQANASAAPQVRPVP
SPAFNAEDKDRSHVQPLVNPSKPDMDSEFIGKTGGEIFHEMMLRQGVKHI
FGYPGGAILPVFDAIYNSKHFDIFILPRHEQGAGHMAEGYARASGKPGVVL
VTSGPGATNVITPMQDALSDGTPLVVFCGQVPTTAIGSDAFQEADVVGIS
RACTKWNVMVKSVAELEPRRINEAFEIATSGRPGPVLVDLPKDVTAGILRR
AIPTETALPSLPSAASRAAMELSSKQLNASIKRAADLINIAKKPVIYAGQ
GVIQSEGGVELLKQLADKASIPVTTTLHGLGAFDELDEKSLHMLGMHGSA
YANMAMQQADLI IALGSRFDDRVTNLNVSKEFAPAARQAAAEGRGGIIHFEI
MPKNINKVIQATEAVEGDVATNLKHLIPQIAEKSMADRGWFGLINEWKK
KWPLSNYQRAERAGLIKQTVMEEISNLTANRKDKTYIATGVGQHQMWWA
QHFRWRHPRSMITSGGLGTMGYGLPAAIGAKVAQPDALVIDVDGDASFNM
TLTELSTAAQFNIGVKVVVLNNEEQGMVTQWQNLFYEDRYAHTHQKNPDF
MKLADAMGVQHQRVTEPEKLVDAITWLINTDGPALLEVVTDKKVPVLPV
PAGSALHEFLVFEPEKDKQRRELMKERTKGVHS

Fig. 1

ATGCTCCGAAGTCGCCAAGTCACAGCCAGGGCCGTCCGGGCTCTGGGCCAGGCGCG
 CGCCTTTACCTCGACGACCAAGCCTGTCATGATCCAGAGCAGCCAGAGGAAACAGG
 CCAACGCCAGCGCTGCTCCACAAGTCCGCCCTGTACCGAGCCCTGCTTTCAACGCT
 GAAGACAAAGACCGCAGCCATGTGCAGCCTCTGGTCAACCCGTCGAAGCCCCGACAT
 GGATGAATCATTCATTGGCAAAACCGGAGGCGAAATCTTCCACGAGATGATGCTGC
 GACAGGGTGTCAAGCACATTTTCGGATACCCTGGCGGCGCTATCCTGCCCGTCTTC
 GACGCCATCTACAACCTCAAAACACTTCGACTTCATCCTGCCCGTCATGAGCAGGG
 AGCTGGCCATATGGCCGAGGGCTATGCCCGTGCCTCGGGCAAACCCGGTGTCTGTC
 TGGTGAATTCGGCCCCGGTGCTACCAATGTTCATCACGCCCATGCAGGATGCCCTG
 TCGGACGGAACGCCCTTGGTCTCTTCTGCGGCCAGGTCCCCACCACGGCCATCGG
 CAGCGATGCCCTTCCAAGAGGCCGACGTCGTGGGCATCTCGCGGGCCTGCACCAAGT
 GGAACGTCATGGTCAAGAGCGTTGCTGAGCTGCCGCGGAGAATCAACGAGGCCTTT
 GAGATTGCCACCAGCGGCCGCCCTGGCCCCGTCTCGTCGACCTGCCCAAGGATGT
 CACGGCTGGTATCCTGAGGAGAGCCATCCCTACGGAGACTGCTCTGCCGTCTCTGC
 CCAGTGCCGCTCCCGCGCCGCCATGGAGCTGAGCTCCAAGCAGCTCAACGCCTCC
 ATCAAGCGTGCCGCGGACCTCATCAACATCGCCAAGAAGCCCGTCATCTACGCCGG
 TCAGGGTGTATCCAGTCCGAGGGCGGCGTTGAGCTCCTGAAGCAGCTGGCGGACA
 AGGCCTCCATCCCCGTCAACCACCACCTCCATGGCCTGGGTGCCTTTGATGAGCTG
 GACGAGAAGTCGCTGCACATGCTGGGCATGCACGGCTCGGCGTATGCCAACATGGC
 CATGCAGCAGGCCGACCTCATCATCGCCCTCGGCAGCCGATTTCGACGACCGTGTTA
 CTCTGAATGTCTCAAATTTGCGCCTGCAGCCAGGCAAGCTGCTGCCGAGGGCCGC
 GGCGGCATCATTCACCTTTGAGATCATGCCCAAGAACATCAACAAGGTCATCCAGGC
 GACCGAGGCCGTGAGGGCGACGTGCCACCAACCTGAAGCACCTCATTCCCCAGA
 TTGCCGAAAAGTCCATGGCGGACCGAGGAGAGTGGTTCGGCCTCATCAATGAGTGG
 AAGAAGAAGTGGCCCCCTGTCAAACCTACCAGCGCGCGGAGCGGGCTGGCCTCATCAA
 GCCGCAGACGGTCATGGAGGAGATTAGCAACCTGACGGCCAAACCGAAAGGACAAGA
 CGTACATTGCCACGGGTGTGCGCCAGCACCAGATGTGGGTGCCCAGCACTTCCGC
 TGGAGGCACCTTCGATCCATGATTACCTCTGGTGGTCTGGGCACCATGGGCTACGG
 TCTCCCCGCGGCCATTGGCGCCAAGGTGGCCAGCCGACGCTCTCGTAATTGACG
 TTGATGGCGATGCCTCGTTTAACATGACGCTGACGGAGCTGTGCGACTGCTGCACAG
 TTCAACATTGGCGTCAAGGTGGTTGTGCTCAACAACGAGGAGCAGGGCATGGTGAC
 GCAGTGGCAGAACCTCTTTTACGAGGACCGATATGCCCAACGACACCAAGAAGACC
 CCGACTTCATGAAGCTGGCCGACGCCATGGGCGTTCAGCACCAGCGCGTGACGGAG
 CCGGAGAAGCTGGTCGATGCCCTGACGTGGCTGATCAACACCGATGGCCCGGCCCT
 GTTGGAGGTTGTACGGACAAGAAGGTGCCTGTCCTGCCCATGGTGCCCGCCGGAT
 CGGCCCTGCACGAGTTCCTCGTCTTTGAACCTGAAAAGGATAAGCAGCGCCGTGAG
 CTGATGAAGGAGAGAACAAAGGGTGTGCACTCCTAA

Fig. 2

TGAGACAATGGCCGGCAATGGTAAAAAGGACCAAGATGTACTAGGTAGTTGCAATGTGGCTTAT
 TACCTACCTACTACCTGGTAGGCACCTACTAGGTACTTGGGTAGACGGACAATGAAATTTGAAG
 TCGGGGTTGCAGGAAAGCAGGGCGCTGGACACATTGTGCTTCAGGCGGTACCCGTCGTCATCGT
 CAGCCAATGTGAGGCGCCGGCAGCCGAGGAGCGAGACAACCTTGGCCGGAGGAGCCCGCAGGT
 ACCTGCCAAAGCGCGCTGGTACCTCTCAACCCTCTCAGGCCTGTTGGATGCCCTATGACATGC
 CCTGGGGGATGCAGCTGTTGCCCCGGCCCCGCACTTTCGGGTGACCGCGAGGCTGCTGATTGGC
 TGGTTGCCACGGGCTGGGCGGTCCCTGAAGTTGTTGCCATCTGAACTCTGTGCGCGCTGGCGTC
 GGCTGCGCCCAATGGGAGGCGAGACAACCTCAGGGTACTAGAATCACTGACAGAAGAAGAGAATC
 GAAAGTAGGTAGACAGCCAATTTCGTTGCATGGCAGGCAACCGCACAGGAGAAAAATTGACTACC
 CCACAATCAGGCACAGTAAGTAGGGCACAGTACGTATGTACAGACAAGGCGCAAGCGATACCTGC
 GCGACCCGCTACCTCGCCGGCTTGACACGTGCGACAGGCTACTTTACTAGTATTCGACGCGCG
 GGTGCGCGATTATTACATGTACTGTGCGCCATTGATGACTGGGCTGCTGCAGTATTAGTAGA
 TCTGCCCCGCATCGCCCTTCCATGGGCGCGACCCGGGACTGGACCCTCTGACTCTACCTACATG
 TACCTAGGCCGGGCGGGCTTGGTGACTTTTGTCCGATCAGGTCGTTGCGCTGGCTACCTATTA
 TTTCTCTTTCTTCTCTCCATCCTGCTTCTGGCCTTGCAATCTTCTTCTCGCCACTCTCTCCCTCT
 TCCCCCGCGATACCTTGAATTTCGTGAGAGGAAAAAGACGAGAAAAAAGGGCAGCAGAGA
 CGTCGGTCTGGCTCACGTGCTGCATCTCTGCGCACTCTCATTTTTTTTTATTGTCCGACCCCTCC
 CTCAACCTTCTCTTTCGTTGACAGGCTAAGCCTTGCTTCGACGCTCTCTCTTTGAATTTTTCTA
 CTCTACCTTCTTTCTTTCGCTGTTACCCACCATAGCTCGATTACGATGCTCCGAAAGTCGCCA
AGTCACAGCCAGGGCCGTCCGGGCTCTGGGCGAGGCGCGCGCCTTTACCTCGACGACCAAGCCT
GTCATAGTCAGAGCAGCCAGAGGAAACAGGCCAACGCCAGCGCTGCTCCGTAAGTCGCCCAT
GCCATTGCATCTTCTGTTTGATATATACTTCTGCTGCTTGGCTGGCGTCGCTCTCTCGGTTATG
CGTGTCAAGGACCAGGTGTGTTTCGCATCGTGGTTTTCCAGCGCCGATTACCGGGGACGAATTT
TTGGCTGCTCAACTCGCGCGCGCGCATTCTGATTCTTCGTTTTCAATCTTGAGCGACAACCTGGC
TAACATAATGGCCATTGGCAATTGCTTACACAGACAAGTCCGCCCTGTACCGAGCCCTGCTT
CAACGCTGAAGACAAAGACCGCAGCCATGTGCAGCCTCTGGTCAACCCGTCGAAGCCCGACATG
GATGAATCGTATGTCCACGTCCCCTCGTCCCGCCCTACAAAATGAACACGATTACACCAGAATT
TTTGCAACAATCGACACTTCTATAACAGACCAATTGAGCTTTGTTCTGACCAATCATGTTGCTC
TAGATTCAATTGGCAAAACCGGAGGCGAAATCTTCCACGAGATGATGCTGCGACAGGCTGTCAAG
CACATTTGTAGGTTCCGATGCGCGGCGCCACACGGGCTCCATCCTTGCTCCATCTCTCCAGCT
AGGCAAAATCTCGCTAACCTTGAGTCACCATCCAGTCGGATACCCCTGGCGGCGCTATCCTGCCCG
TCTTCGACGCCATCTACAACTCAAAACACTTCGACTTCATCCTGCCCCGTCATGAGCAGGGAGC
TGGCCATATGGCCGAGGGCTATGCCCGTGCTCGGGCAAACCGGTGTGCTCCTGGTGACTTCC
GGCCCCGGTGCTACCAATGTATCAGCCCATGCAGGATGCCCTGTGCGACGGAACGCCCTTGG
TCGTCTTCTGCGGCCAGGTCCCCACCACGGCCATCGGCAGCGATGCCCTTCCAAGAGGCCGACGT
CGTGGGCATCTCGCGGGCTGCACCAAGTGGAACGTATGGTCAAGAGCGTTGCTGAGCTGCCG
CGGAGAATCAACGAGGCCTTTGAGATTGCCACCAGCGGCGCCCTGGCCCCGTCTCTGTCGACC
TGCCCAAGGATGTACGGCTGGTATCCTGAGGAGAGCCATCCCTACGGAGACTGCTCTGCCGTC
TCTGCCCAGTGCCGCTTCCGCGCCGCCATGGAGCTGAGCTCCAAGCAGCTCAACGCCCTCCATC
AAGCGTGCCGCGACCTCATCAACATCGCCAAGAAGCCCGTCATCTACGCCGGTCAGGGTGTCA
TCCAGTCCGAGGGCGGCGTTGAGCTCCTGAAGCAGCTGGCGGACAAGGCCTCCATCCCCGTAC
CACCACCCTCCATGGCTGGGTGCCCTTGTATGAGCTGGACGAGAAGTCGCTGCACATGCTGGGC
ATGCACGGCTCGGCGTATGCCAATGCGCCATGCAGCAGGCCGACCTCATCATCGCCCTCGGCA

GCCGATTGACGACCGTGTTACTCTGAATGTCCTCAAATTTGCGCCTGCAGCCAGGCAAGCTGC
 TGCCGAGGGCCGCGCGGCATCATTCACTTTGAGATCATGCCCAGAACATCAACAAGGTCATC
 CAGGCGACCGAGGCCGTCGAGGGCGACGTCGCCACCAACCTGAAGCACCTCAATCCCCAGATTG
 CCGAAAAGTCCATGGCGGACCGAGGAGAGTGGTTCCGCCCTCATCAATGAGTGAAGAAGAAGTG
 GCCCCGTGTCAAACCTACGAGCGCGCGGAGCGGGCTGGCCTCATCAAGCCGCAGACGGTCATGGAG
 GAGATTAGCAACCTGACGGCCAACCGAAAGGACAAGACGTACATTGCCACGGGTGTCGGCCAGC
 ACCAGATGTGGGTTGCCCAGCACTTCCGCTGGAGGCACCCCTCGATCCATGATTACCTCTGGTGG
 TCTGGGCACCATGGGCTACGGTCTCCCCGCGGCCATTGGCGCCAAGGTGGCCAGCCCGACGCT
 CTCGTAATTGACGTTGATGGCGATGCCTCGTTTAACATGACGCTGACGGAGCTGTGACTGCTG
 CACAGTTCAACATTGGCGTCAAGGTGGTTGTGCTCAACAACGAGGAGCAGGGCATGGTGACGCA
 GTGGCAGAACCTCTTTTACGAGGACCGATATGCCACACGCACCAGAAGAACCCCGACTTCATG
 AAGCTGGCCGACGCCATGGGCGTTCAGCACCGAGCGCGTGACGGAGCCCGAGAAGCTGGTCCATG
 CCCTGACGTGGCTGATCAACACCGATGGCCCGGCCCTGTTGGAGGTTGTCACGGACAAGAAGGT
 GCCTGTCTGCCCATGGTGGCCGCGGATCGGCCCTGCACGAGTTCCTCGTCTTTGAACCTGGT
 GAGTCTACTTCAGACATATTGCTTGCGCATTGCAGATACTAACACTCTCACAGAAAAGGATAAG
 CAGCGCCGTGAGCTGATGAAGGAGAGAACAAAGGGTGTGCACTCCTAAAGCGATGATGTCTGCG
 AGGGGTCTTCGTTGAACCCTAGTTCAAGGCACCATCTTACCCTCTTATTTTTCCCGTGGGCTT
 TCATTTGTGTCATCCGAGCATGACGTTGTAGGGTTGGAGTTTCTTCCCTTTTATCTTGTCAAT
 TACTGGTACCCATAGGCGCGAGACTAGGCTTCCATGTTTTGTTTTGCGACTTTCAAAAAGTACT
 TTTAGTGGTTTTGGGGCAGCAGAGGGGGGGCAACCTCTTCTGTGAAAAAGGTGGCTGGATGGA
 TGAGATGAGATGAGATGAGGGTGAAGATAGATACCTGCAGTGTTTTTGACGCGACGGGATG

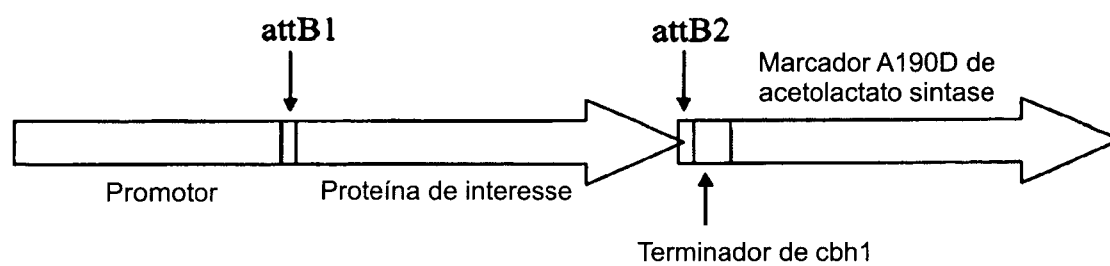


FIG. 4

Fig. 5A

contig1818 A Um promotor de versão (868 Pares de bases) :

TAGGCTGGCGGGTTGGTGCTTTTGCCCTTTGTCTGCCGCGACCTCTTGGTGACGCTGCATCGTGTGGCTGTAGCTGCAC
GCCCCGGCCTCTTTTCTGGCATCTCGCACAGACTGTACGCAGAACATGTGTAAGTGACTTTCTATTGTAGGACTTGGATGG
ATGGGAGATGGCTATAGACAACAAGAGCGACAACAAGGATTATAGACAGTTCCATTACCTGGGCGACCACCATCGCACTT
CACCTTGCTCGCCCCGGCAGGACTGGCAGGCGTTGGCAATCTTGAGTCGCTCCCTGTACAGCGACAAGGGCGCCTGAGCCT
GCACCTGAGCCTGGGCTGAGTCTGAGCCTGAAGCCGTTGATCCTCGTCTTGACGGGCGGACATGACGCTTTACGTGAGC
TGAGTCCAGGGCTGTCTGTCGGCAAGTGATTCCGCAGCCGGCAGGCACAGACGAAAGATAAGAAGCGGCTCGTTCACCAGA
TCCGAAGGCGGAATGCAAGAATGTCCGATGATGTCCGGGAAGAGAAGGCGATGCCTCGTCTATCTGGGAGGCGAGGGAAG
AAGGGAAAGCAACTCCCGCCTCGGCACAGCTACTCATTGGTCAGAGCGGACTTCGGAGTATAACCACACGGCTGACAAAC
CTGCATCGACCGTGATGGCGTTCTTTGGTCCAATCATGGCTGAATGCGACGTTGGTTGTGTGTCGGCCCCCTGCAGCCAATG
AGAAGCTGGTCCGATCCTGGGACGAGGGAGAGGCAAGCATATCGTATAAGTCCAAGCTCAAGCTTGGCGGCAGGTACCG
GTTGATGTGGTCGCTGTATATAAGTAAAAGAACTCGATTCCGCTTCGGAGTTTCTCTGTCAACCCTTC

Fig. 5B

contig1818 B Um promotor de versão (928 Pares de bases) :

TAGGCTGGCGGGTTGGTGCTTTTGCCCTTTGTCTGCCGCGACCTCTTGGTGACGCTGCATCGTGTGGCTGTAGCTGCAC
GCCCCGGCCTCTTTTCTGGCATCTCGCACAGACTGTACGCAGAACATGTGTAAGTGACTTTCTATTGTAGGACTTGGATGG
ATGGGAGATGGCTATAGACAACAAGAGCGACAACAAGGATTATAGACAGTTCCATTACCTGGGCGACCACCATCGCACTT
CACCTTGCTCGCCCCGGCAGGACTGGCAGGCGTTGGCAATCTTGAGTCGCTCCCTGTACAGCGACAAGGGCGCCTGAGCCT
GCACCTGAGCCTGGGCTGAGTCTGAGCCTGAAGCCGTTGATCCTCGTCTTGACGGGCGGACATGACGCTTTACGTGAGC
TGAGTCCAGGGCTGTCTGTCGGCAAGTGATTCCGCAGCCGGCAGGCACAGACGAAAGATAAGAAGCGGCTCGTTCACCAGA
TCCGAAGGCGGAATGCAAGAATGTCCGATGATGTCCGGGAAGAGAAGGCGATGCCTCGTCTATCTGGGAGGCGAGGGAAG
AAGGGAAAGCAACTCCCGCCTCGGCACAGCTACTCATTGGTCAGAGCGGACTTCGGAGTATAACCACACGGCTGACAAAC
CTGCATCGACCGTGATGGCGTTCTTTGGTCCAATCATGGCTGAATGCGACGTTGGTTGTGTGTCGGCCCCCTGCAGCCAATG
AGAAGCTGGTCCGATCCTGGGACGAGGGAGAGGCAAGCATATCGTATAAGTCCAAGCTCAAGCTTGGCGGCAGGTACCG
GTTGATGTGGTCGCTGTATATAAGTAAAAGAACTCGATTCCGCTTCGGAGTTTCTCTGTCAACCCTTCATGGACGAAGAG
CTATTGAGCCGCATCTTTGCTTCGTTGTCCAGCGTGAATTCCTACAC

FIGS. 5A E 5B

AAGCTTACTAGTACTTCTCGAGCTCTGTACATGTCCGGTCCGACGTACGCGTATCGATGGCGCCAGCT
 GCAGGCGGCCCGCTGCAGCCACTTGCAGTCCCCTGGAATTCTCACGGTGAATGTAGGCCTTTTGTAGGG
 TAGGAATTGTCACTCAAGCACCCCCAACCTCCATTACGCTCCCCATAGAGTTCCCAATCAGTGAGTC
 ATGGCACTGTTCTCAAATAGATTGGGGAGAAGTTGACTTCCGCCCAGAGCTGAAGGTGCGACAACCGCA
 TGATATAGGGTCCGGCAACGGCAAAAAAGCACGTGGCTCACGAAAAGCAAGATGTTTGCATCTAACAT
 CCAGGAACCTGGATACATCCATCATCACGCACGACCACTTTGATCTGCTGGTAAACTCGTAITCGCCCT
 AAACCGAAGTGGTGGTAAATCTACACGTGGGCCCCCTTTCCGTATACCTGCGTGTGTCTTCTCTAGGTGC
 CATCTTTTCCCTTCTCTAGTGTGAATTGTTTGTGTGGAGTCCGAGCTGTAACCTACCTCTGAATCT
 CTGGAGAATGGTGGACTAACGACTACCGTGCACCTGCATCATGTATATAATAGTGATCCTGAGAAGGGG
 GGTGTGGAGCAATGTGGGACTTTGATGGTCAATAACAAAGAACGAAGACGCCTCTTTTGCAAAGTTTT
 GTTTCGGCTACGGTGAAGAACTGGATACTTGTGTGTCTTCTGTGTATTTTTGTGGCAACAAGAGGCCA
 GAGACAATCTATTCAAACACCAAGCTTGCTCTTTTGAGCTACAAGAACCTGTGGGGTATATATCTAGAG
 TTGTGAAGTCGGTAATCCCCTGTATAGTAATACGAGTCGCATCTAAATACTCCGAAGCTGCTGCGAAC
 CCGGAGAATCGAGATGTGCTGGAAAGCTTCTAGCGAGCGGCTAAATTAGCATGAAAGGCTATGAGAAAT
 TCTGGAGACGGCTTGTGAATCATGGCGTTCCATTCTTCGACAAGCAAAGCGTTCGGTCCGAGTAGCAG
 GCACTCATCTCCGAAAAAATCTCGGAGATTCTAAGTAGCGATGGAACCGGAATAATATAATAGGCAATA
 CATTGAGTTGCCCTCGACGGTTGCAATGCAGGGGTACTGAGCTTGGACATAACTGTTCCGTACCCACCT
 CTCTCAACCTTTTGGCGTTTCCCTGATTACGGTACCCGTACAAGTCGTAATCACTATTAACCCAGACT
 GACCGGACGTGTTTTTGGCCCTTCATTTGGAGAAATAATGTCAATTGCGATGTGTAATTTGCCCTGCTTGACC
 GACTGGGGCTGTTTGAAGCCCGAATGTAGGATTGTTATCCGAACCTCTGCTCGTAGAGGCATGTTGTGAA
 TCTGTGTCCGGCAGGACACGCCCTCGAAGGTTACGGCAAGGGAAACCACCGATAGCAGTGTCTAGTAGC
 AACCTGTAAAGCCGCAATGCAGCATCACTGGAATAACAAACCAATGGCTAAAAGTACATAAGTTAATG
 CCTAAAGAAGTCATATACCAGCGGCTAATAATTGTACAATCAAGTGGCTAAACGTACCGTAATTTGCCA
 ACGGCTTGTGGGGTTGCAGAAGCAACGGCAAGCCCCACTTCCCCACGTTTGTCTTCACTCAGTCCA
 ATCTCAGCTGGTGTATCCCCAATTGGGTCCGCTTGTGTTCGGTGAAGTGAAAGAAGACAGAGGTAAG
 AATGTCTGACTCGGAGCGTTTTCATACAACCAAGGGCAGTGATGGAAGACAGTGAAATGTTGACATTC
 AAGGAGTATTTAGCCAGGGATGCTTGAGTGTATCGTGTAAGGAGGTTTGTCTGCCGATACGACGAATAC
 TGTATAGTCACTTCTGATGAAGTGGTCCATATTGAAATGTAAAGTCCGCACTGAACAGGCAAAAGATTG
 AGTTGAAACTGCCTAAGATCTCGGGCCCTCGGGCCTTTCGGCTTTCGGGTGTACATGTTTGTGCTCCGGG
 CAAATGCAAAGTGTGGTAGGATCGAACACACTGCTGCCCTTACCAAGCAGCTGAGGGTATGTGATAGGC
 AAATGTTTCAGGGGCCACTGCGATGGTTTCGAATAGAAAGAGAAGCTTAGCCAAGAAACAATAGCCGATAAA
 GATAGCCTCATTAAACGGAATGAGCTAGTAGGCAAGTCAGCGAATGTGTATATATAAAGGTTTCGAGGT
 CCGTGCCCTCCCTCATGCTCTCCCCATCTACTCATCAACTCAGATCCTCCAGGAGACTTGTACACCATCT
 TTTGAGGCACAGAAACCCAATAGTCAACCATCACAAGTTTGTACAAAAAGCAGGCTATGCACGTCCTG
 TCGACTGCGGTGCTGCTCGGCTCCGTTGCCGTTCAAAAGGTCCTGGGAAGACCAGGATCAAGCGGTCTG
 TCCGACGTACCAAGAGGTCTGTTGACGACTTCATCAGCACCGAGACGCCTATTGCACTGAACAATCTT
 CTTTGCAATGTTGGTCCGTGATGGATGCCGTGCATTCCGGCACATCAGCTGGTGCGGTGATTGCATCTCCC
 AGCACAATTGACCCGGAAGTGAAGTTGGCCTTGATGAACCATATCATATATCGCCGAGAAGTGGACCGC
 GTGCTGAGACTGAGACAGACTATTACATGTGGACGCGAGATAGCGCTCTTGTCTTCAAGAACCTCATCG
 ACCGCTTCACCGAAACGTACGATGCGGGCCTGCAGCGCCGATCGAGCAGTACATTACTGCCAGGTCA
 CTCTCCAGGGCCTCTCTAACCCCTCGGGCTCCCTCGCGGACGGCTCTGGTCTCGGCGAGCCCAAGTTTG
 AGTTGACCCTGAAGCCTTTACCGGCAACTGGGTCCGACCGCAGCGGGATGGCCAGCTCTGCGAGCCA
 TTGCCCTTGATTGGATACTCAAAGTGGCTCATCAACAACAATATCAGTCGACTGTGTCCAACGTCTCT
 GGCCTATTGTGCGCAACGACCTCAACTATGTTGCCCAGTACTGGTCAGTGCTTGCTTGTGAATTA
 CGTCTTTGCTTGTGTGTCTAATGCCCTCCACCACAGGAACCAACCGGCTTTGACCTCTGGGAAGAAGTC
 AATGGGAGCTCATCTTTACTGTTGCCAACCAGCACCGAGGTATGAAGCAAATCCTCGACATTGCTGCTG
 TACTGCACATG

FIG. 6

AGCATTGTTACTGACCAGCTCTACAGCACTTGTGCGAGGGCGCCACTCTTGCTGCCACTCTTGCCAGTC
 GGGAAAGCGCTTATTCATCTGTTGCTCCCCAGGTTTGTGCTTTCTCCAACGATTCTGGGTGTCGTCTGG
 TGGATACGTCGACTCCAACAGTATGTCTTTTCACTGTTTATATGAGATTGGCCAATACTGATAGCTCGC
 CTCTAGTCAACACCAACGAGGGCAGGACTGGCAAGGATGTCAACTCCGTCTGACTTCCATCCACACCT
 TCGATCCCAACCTTGGCTGTGACGCAGGCACCTTCCAGCCATGCAGTGACAAAGCGCTCTCCAACCTCA
 AGGTTGTTGTGCGACTCCTTCCGCTCCATCTACGGCGTGAACAAGGGCATTCTGCCGGTGCTGCCGTGCG
 CCATTGGCCGGTATGCAGAGGATGTGTACTACAACGGCAACCCTTGGTATCTTGCTACATTTGCTGCTG
 CCGAGCAGCTGTACGATGCCATCTACGTCTGGAAGAAGACGGGCTCCATCACGGTGACCGCCACCTCCC
 TGGCCTTCTTCCAGGAGCTTGTTCCGTGGCGTGACGGCCGGGACCTACTCCAGCAGCTCTTCGACCTTTA
 CCAACATCATCAACGCCGTCTCGACATACGCCGATGGCTTCCCTCAGCGAGGCTGCCAAGTACGTCCCCG
 CCGACGGTTCTGCTGGCCGAGCAGTTTGACCGCAACAGCGGCACTCCGCTGTCTGCGCTTACCTGACGT
 GGTCTGACGCCCTCGTTCTTGACAGCCACGGCCCGTCCGGCTGGCATCGTGCCCCCTCGTGGGCCAACA
 GCAGCGCTAGCAGATCCCCTCGACGTGCTCCGGCGCGTCCGTGGTCCGATCCTACTCGCGTCCACCG
 CCACGTCAATCCCTCCGTCCGACAGCGCCCAAGCCTGGCGTGCTTCCGGTACTCCCTACACGCCCTGCG
 CCTGCGCGACCCCAACCTCCGTGGCGCTCACCTTCCACGAGCTCGTGTGACACAGTTTGGCCAGACGG
 TCAAGGTGGCGGGCAACGCCGCGGCCCTGGGCAACTGGAGCACGAGCGCCCGTGGCTCTGGACGCCG
 TCAACTATGCCGATAACCACCCCTGTGGATTGGGACGGTCAACCTCGAGGCTGGAGACGTCTGGAGT
 ACAAGTACATCAATGTGGGCCAAGATGGCTCCGTGACCTGGGAGAGTGATCCCAACCACACTTACACGG
 TTCTGCGGTGGCTTGTGTGACGCAGGTTGTCAAGGAGGACACCTGGCAGTCGTAAACCCAGCTTTCTT
 GTACAAAGTGGTGATCGCGCCAGCTCCGTGCGAAAGCCTGACGCACCGGTAGATTCTTGGTGAGCCCGT
 ATCATGACGGCGGCGGGAGCTACATGGCCCCGGGTGATTATTTTTTTTGTATCTACTTCTGACCTTT
 TCAAATATACGGTCAACTCATCTTTCACTGGAGATCGCGCCTGCTTGGTATTGCGATGTTGTGAGCTTG
 GCAAATTGTGGCTTTTCGAAAACACAAAACGATTCTTTAGTAGCCATGATTTTAAGATAACGGAATAGA
 AGAAAGAGGAAATTAAAAAACAACATCCCGTTCATAACCCGTAGAATCGCCGCTGGCG
 CGCCTGAGACAATGGCCGGCAATGGTAAAAGGACCAAGATGTACTAGGTAGTTGCAATGTGGCTTATT
 ACCTACCTACTACCTGGTAGGCACCTACTAGGTACTTGGGTAGACGGAACAATGAAATTTGAAGTCGGGG
 TTGCAGGAAAGCAGGGCGCTGGACACATTGTGCTTCAGGCGGTACCCGTCGTCATCGTCAGCCAATGTC
 GAGGCCCGGCAGCCCGAGGAGCGAGACAACCTTGGCCGGAGGAGCCCGCAGGTACCTGCCAAAGCGCGG
 CTGGTACCTCTCAACCCTCTCAGGCCGTGTTGGATGCCCTATGACATGCCCTGGGGGATGCAGCTGTTGC
 CCGGCCCGGCACCTTTCCGGTGACCGCGAGGCTGCTGATTGGCTGGTTGCCACGGGCTGGGCGGTCCCT
 GAAGTTGTTGCCATCTGAACTCTGTGCGCGCTGGCGTCCGCTGCGCCCAATGGGAGGCGAGACAACCTCA
 GGGTACTAGAATCACTGACAGAAGAAGAGAATCGAAAAGTAGGTAGACAGCCAATTCGTTGTCATGGCAGG
 CAACCGCACAGGAGAAAAATTGACTACCCACAATCAGGCACAGTAAGTAGGGCACAGTACGTATGTAC
 AGACAAGGCGCAAGCGATACTGCGCGACCCGGTACCTCGCCGGCTTGACACGTGCGACAGGCTACTTTA
 CTAGTATTGCGAGCGGCGGGTTCGCGCATTATTACATGTACTGTGCCGCCATTGATGACTGGGCTGCTG
 CAGTATTAGTAGATCTGCCCAGCATCGCCCTTCCATGGGCGCGACCCGGGACTGGACCCTCTGACTCTA
 CCTACATGTACCTAGGCCGGGCGGGCTTGGTGACTTTTGTCCGATCAGGTGCTTCGCTGGCTACCTA
 TTATTTCTCTTTCTTCTTCTCCATCCTGCTTCTGGCCTTGCAATTCTTCTTCGCCACTCCTCCCTCTTC
 CCCCCGCGATACCTTGAATTCGTGAGAGAGGAAAAGACGAGAAAAAAGGGCAGCAGAGACGTCCGT
 CTGGCTCACGTGCTGCATCTCTGCGCACTCTCATTTTTTTTATTGTCCGACCCCTCCCTCAACCTTCTC
 CTTCTGTTGACAGGCTAAGCCTTGCTTCGACGCTCTCTCTTTGAATTTTTCTACTTCTACCTTCTTTCT
 TGCGTGTTACCCACCATAGCTCGATTACGATGCTCCGAAGTCGCCAAGTCACAGCCAGGGCCGTCCGG
 GCTCTGGGCCAGGCGCGCGCTTTACCTCGACGACCAAGCCTGTATGATCCAGAGCAGCCAGAGGAAA
 CAGGCCAACGCCAGCGCTGCTCCGTAAGTCGCCCATTGCCATTGCATCTTCTGTTTGATATATACTTCC
 TGCTGCTTGCGTGGCGTCG

FIG. 6

TCTCTCGGTTATGCGTGTCAAGGACCAGGTGTGTTTCGCATCGTGGTTTTCCAGCGCCGATTACCGGG
 GGACGAATTTTTGGCTGCTCAACTCGCGCGCGCGCATTCTGATTCTTCGTTTTCAATCTTGAGCGAC
 AACTGGCTAACATAATGGCCATTGGCAATTGCTTCACACAGACAAGTCCGCCCTGTACCGAGCCCTG
 CTTTCAACGCTGAAGACAAAGACCGCAGCCATGTGCAGCCTCTGGTCAACCCGTCGAAGCCCGACAT
 GGATGAATCGTATGTCCACGTCCCCTCGTCCCGCCCTACAAAATGAACACGATTACACCAGAATTTT
 TGCAACAATCGACACTTCTATAACAGACCAATTGAGCTTTGTTCTGACCAATCATGTTGCTCTAGAT
 TCATTGGCAAAACCGGAGGCGAAATCTTCCACGAGATGATGCTGCGACAGGGTGTCAAGCACATTTG
 TAGGTTCCGATGCCGGCCGCCACACGGGCTCCATCCTTGCTCCATCTCTCCAGCTAGGCAAATCTC
 GCTAACCTTGAGTCAACATCCAGTCGGATACCTTGGCGGCGCTATCCTGCCCGTCTTCGACGCCATC
 TACAACCTCAAAACACTTCGACTTCATCCTGCCCGTCATGAGCAGGGAGCTGGCCATATGGCCGAGG
 GCTATGCCCGTGCTCGGGCAAACCCGGTGTCTGCTCCTGGTGACTTCCGGCCCCGGTGCTACCAATGT
 CATCACGCCCATGCGAGGATGCCCTGTGCGACGGAACGCCCTTGGTGCTCTTCTGCGGCCAGGTCCCC
 ACCACGGCCATCGGCAGCGATGACTTCCAAGAGGCCGACGTCGTGGGCATCTCGCGGCCCTGCACCA
 AGTGGAACGTCTGTTCAAGAGCGTTGCTGAGCTGCCCGGAGAAATCAACGAGGCCCTTGAGATTGC
 CACCAGCGGCCCGCTGGCCCCGTCTCTGCTGACCTGCCCAAGGATGTACGGCTGGTATCCTGAGG
 AGAGCCATCCCTACGGAGACTGCTCTGCCGTCTCTGCCAGTGCCCGCTCCCGCGCCCGCATGGAGC
 TGAGCTCCAAGCAGCTCAACGCCTCCATCAAGCGTGCCGCCGACCTCATCAACATCGCCAAGAAGCC
 CGTCATCTACGCCGGTCAGGGTGTCTCAGTCCGAGGGCGGCGTTGAGCTCCTGAAGCAGCTGGCG
 GACAAGGCCTCCATCCCCGTCAACACCACCTCCATGGCCTGGGTGCCTTTGATGAGCTGGACGAGA
 AGTCGCTGCACATGCTGGGCATGCACGGCTCGGCGTATGCCAACATGGCCATGCAGCAGGCCGACCT
 CATCATCGCCCTCGGCAGCCGATTGACGACCGTGTACTCTGAATGTCTCCAAATTTGCGCCTGCA
 GCCAGGCAAGCTGTGCGCGAGGGCCGCGCGGCATCATTCACTTTGAGATCATGCCCAAGAACATCA
 ACAAGGTCTCAGGCGACCGAGGCCGTGAGGGCGACGTCGCCACCAACCTGAAGCACCTCATTCC
 CCAGATTGCCGAAAAGTCCATGGCGGACCGAGGAGAGTGGTTCCGGCTCATCAATGAGTGGAAGAAG
 AAGTGGCCCCGTGCAAACTACCAGCGCGCGGAGCGGGCTGGCCTCATCAAGCCGACAGCGGTCTG
 AGGAGATTAGCAACCTGACGGCCAACCGAAAGGACAAGACGTACATTGCCACGGGTGTGCGCCAGCA
 CCAGATGTGGGTTGCCAGCACTTCCGCTGGAGGCACCTCTGATCCATGATTACCTCTGGTGGTCTG
 GGCACCATGGGCTACGGTCTCCCCGCGGCCATTGGCGCCAAGGTGGCCCGAGCCGACGCTCTCGTAA
 TTGACGTTGATGGCGATGCCTCGTTTAACATGACGCTGACGGAGCTGTGACTGTCTGCACAGTTCAA
 CATTGGCGTCAAGGTGGTTGTGCTCAACAACGAGGAGCAGGGCATGGTGACGCAGTGGCAGAACCTC
 TTTTACGAGGACCGATATGCCCACACGCACCAGAAGAACCCGACTTCATGAAGCTGGCCGACGCCA
 TGGGCGTTGACGACCAGCGCGTGACGGAGCCGAGAAGCTGGTGCATGCCCTGACGTGGCTGATCAA
 CACCGATGGCCCGGCCCTGTGGAGGTTGTACGGACAAGAAGGTGCCTGTCTGCCCCATGGTGGCC
 GCCGGATCGGCCCTGCACGAGTTCTCTGCTTTTGAACCTGGTGAGTCTACTTCAGACATATTGCTTG
 CGCATTGCAGATACTAACTCTCACAGAAAAGGATAAGCAGCGCCGTGAGCTGATGAAGGAGAGAA
 CAAAGGGTGTGCACTCCTAAAGCGATGATGTCTGCGAGGGGTTCTTCGTTGAACCCCTAGTTCAGGCA
 CCATCTTACCCTCTTATTTTTTCCCGTGGGCTTTCAATTTGTGTCTATCCGAGCATGACGTTGTAGGG
 TTGGAGTTTCTTCTTTTATCTTGTCTTACTGGTACCCATAGGCGCGAGACTAGGCTTCCATGT
 TTTGTTTTGCGACTTTCAAAAAGTACTTTTAGTGGTTTGGGGCAGCAGAGGGGGGGCAACCTCTTC
 TGTGAAAAAAGGTGGCTGGATGGATGAGATGAGATGAGATGAGGGTGAAGATAGATACCTGCAGTGT
 TTTTGACGCGACGGGATGGCGATCGCTCAGGGTTGCGTTTCCAGTCTAGACACGTATAACGGCACAA
 GTGTCTCTCACCAAATGGGTTATATCTCAAATGTGATCTAAGGATGGAAGGCCAGAATATCGATCG
 CGCGCAGATCCATATATAGGGCCCGGTTATAATTACCTCAGGTGACGTCCCATGGCCATTCGAAT
 TCGTAATCATGGTCATAGCTGTTCCTGTGTGAAATTGTTATCCGCTCACAATTCACACAACATAC
 GAGCCGGAAGCATAAAGTGTAAAGCCTGGGGTGCCTAATGAG

TGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCAG
 CTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTTCGTATTGGGCGCTCTTCCGCTTCCTC
 GCTCACTGACTCGCTGCGCTCGGTGCTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAA
 TACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAATGTGAGCAAAAGGCCAGCAAAAGGCC
 AGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTTCATAGGCTCCGCCCCCTGACGAGCATCACAA
 AAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTG
 GAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCCTGCCGCTTACCGGATACCTGTCCGCTTTCTCCCT
 TCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTGCTTCGCTC
 CAAGCTGGGCTGTGTGCACGAACCCCCCGTTTCCAGCCGACCGCTGCGCTTATCCGGTAACTATCGTC
 TTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGA
 GCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGAAC
 AGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGAAAAAGAGTTGGTAGCTCTTGATCCG
 GCAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTGTTGCAAGCAGCAGATTACGCGCAGAAAAAA
 GGATCTCAAGAAGATCCTTTGATCTTTTTCTACGGGCTGTGACGCTCAGTGGAAACGAAACTCACGTTA
 AGGGATTTTGGTCATGAGATTATCAAAAAGGATCTTACCTAGATCCTTTTAAATTAATAATGAAGTT
 TTAATCAATCTAAAGTATATATGAGTAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCA
 CCTATCTCAGCGATCTGTCTATTTTCGTTTATCCATAGTTGCCTGACTCCCCGTCGTGTAGATAACTAC
 GATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCGCGAGACCCACGCTCACCGGCTC
 CAGATTTATCAGCAATAAACAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAACTTTATCC
 GCCTCCATCCAGTCTATTAATTGTTGCCGGAAGCTAGAGTAAGTAGTTCCGAGTTAATAGTTTGCG
 CAACGTTGTTGCCATTGCTACAGGCATCGTGGTGTACGCTCGTCGTTTGGTATGGCTTCATTACGCT
 CCGGTTCCCAACGATCAAGGCGAGTTACATGATCCCCCATGTTGTGCAAAAAGCGGTTAGCTCCTTC
 GGTCCTCCGATCGTTGTGAGAAGTAAGTTGGCCGAGTGTTATCACTCATGGTTATGGCAGCACTGCA
 TAATTCCTTACTGTGATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCAT
 TCTGAGAATAGTGATGCGGCGACCGAGTTGCTCTTGGCCGCGTCAATACGGGATAATACCGCGCCA
 CATAGCAGAACTTTTAAAGTGCTCATCATTGGAACGTTCTTCGGGGCGAAACTCTCAAGGATCTT
 ACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGACCCCAACTGATCTTCAGCATCTTTTACTT
 TCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAGGGAATAAGGGCGACA
 CGGAAATGTTGAATACTCATACTCTTCTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCT
 CATGAGCGGATACATATTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCC
 GAAAAGTGCCACCTGACGCTAAGAAACCATATTATCATGACATTAACCTATAAAAATAGGCGTATC
 ACGAGGCCCTTTCGCTCTCGCGCGTTTCGGTGATGACGGTGAACCTCTGACACATGCAGCTCCCGGA
 GACGGTCACAGCTTGTCTGTAAGCGGATGCCGGGAGCAGACAAGCCCGTCAGGGCGCGTCAGCGGGTG
 TTGGCGGGTGTCGGGGCTGGCTTAACTATGCGGCATCAGAGCAGATTGTACTGAGAGTGACCATATAA
 ATTGTAAACGTTAATATTTTGTAAAAATTTCGCGTTAAATTTTGTAAATCAGCTCATTTTTTAAACCA
 ATAGGCCGAAATCGGCAAAATCCCTTATAAATCAAAAGAATAGCCCGAGATAGGGTTGAGTGTGTTTC
 CAGTTTGGAAACAAGAGTCCACTATTAAAGAACGTGGACTCCAACGTCAAAGGGCGAAAAACCGTCTAT
 CAGGGCGATGGCCCACTACGTGAACCATCACCCAAATCAAGTTTTTGGGGTCGAGGTGCCGTAAAGC
 ACTAAATCGGAACCTAAAGGGAGCCCCGATTTAGAGCTTGACGGGGAAGCCGGCGAACGTGGCGA
 GAAAGGAAGGGAAGAAAGCGAAAGGAGCGGGCGCTAGGGCGCTGGCAAGTGTAGCGGTACGCTGCGC
 GTAACCACCACACCCCGCGCTTAATGCGCCGCTACAGGGCGCGTACTATGGTTGCTTTGACGTATG
 CGGTGTGAAATACCGCACAGATGCGTAAGGAGAAAAATACCGCATCAGGCGCCATTGCGCATTACGGCT
 GCGCAACTGTTTGGGAAGGGCGATCGGTGCGGGCTCTTCGCTATTACGCCAGCTGGCGAAAGGGGAT
 GTGCTGCAAGGCGATTAAAGTTGGGTAACGCCAGGGTTTTCCAGTCACGACGTTGTAAACGACGGCC
 AGTGCCC

FIG. 6



Fig. 7

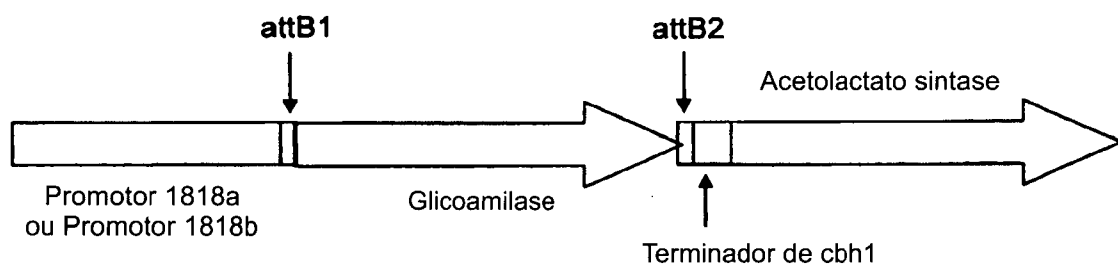


Fig. 8

CATTACGCCTCCCCATAGATAGGTCTAGGCTGGCGGGTTGGTGCTTTTGCCCTTTGTCTGCCGCGACC
 TCTTGGTGACGCTGCATCGTGTTGGCTGTAGCTGCACGCCCGGCCTCTTTCTGGCATCTCGCACAGAC
 TGTACGCAGAACATGTGTAAGTGACTTTCTATTGTAGGACTTGGATGGATGGGAGATGGCTATAGACAA
 CAAGAGCGACAACAAGGATTATAGACAGTTCCATTACCTGGGCGACCACCATCGCACTTCACCTTGCTC
 GCCC GG CAGGACTGGCAGGCGTTGGCAATCTTGAGTCGCTCCCTGTACAGCGACAAGGGCGCCTGAGCC
 TGCACCTGAGCCTGGGCTGAGTCCAGGCTGAGCCTGAAGCCGTTGATCCTCGTCTTGACGGGCGGACATGACG
 CTTTACGTGAGCTGAGTCCAGGCTGTCGTGGCAAGTGATTCCGGCAGCCGGCAGGCACAGACGAAAGA
 TAAGAAGCGGCTCGTTTACCAGATCCGAAGGCGGAATGCAAGAATGTCCGATGATGTCCGGGAAGAGAA
 GGCGATGCCTCGTCTATCTGGGAGGCGAGGGAAGAAGGGAAGCAACTCCCGCTCGGCACAGCTACTC
 ATTTGGTCAGAGCGGACTTCGGAGTATAACCACACGGCTGACAAACCTGCATCGACCGTGATGGCGTTCT
 TTTGGTCCAATCATGGCTGAATGCGACGTTGGTTGTGTGGCCCTTCAGCCAATGAGAAGCTGGTCCG
 ATCCTGGGGACGAGGGAGAGGCAAGCATATCGTATAAGTCCAAGCTCAAGCTTGGCGGCAGGTACCGGT
 TGATGTGGTTCGCTGTATATAAGTAAAGAACTCGATTCCGCTTCGGAGTTTCTCTGTCAACCTTCACA
 AGTTTGTACAAAAAGCAGGCTATGCACGTCCTGTCCGACTGCGGTGCTGCTCGGCTCGGTGACCGTTCA
 AAAGGTCCTGGGAAGACCAGGATCAAGCGGTCTGTCCGACGTCACCAAGAGGTCTGTTGACGACTTCAT
 CAGCACCGAGACGCCTATTGCACTGAACAATCTTCTTTGCAATGTTGGTCTGATGGATGCCGTGCATT
 CGGCACATCAGCTGGTGGGTGATTGCATCTCCAGCACAAITGACCCGGACTGTAAGTTGGCTTGAT
 GAACCATATCATATATCGCCGAGAAGTGGACCGCGTGCTGAGACTGAGACAGACTATTACATGTGGACG
 CGAGATAGCGCTCTTGCTTTCAAGAACCTCATCGACCGCTTCACCGAAACGTACGATGCGGGCTGCAG
 CGCCGCATCGAGCAGTACATTACTGCCAGGTCACTCTCCAGGGCCTCTCTAACCCCTCGGGCTCCCTC
 GCGGACGGCTCTGGTCTCGGCGAGCCCAAGTTTGAGTTGACCCTGAAGCCTTTCACCGGCAACTGGGGT
 CGACCGCAGCGGGATGGCCAGCTCTGCGAGCCATTGCCTTGATTGGATACTCAAAGTGGCTCATCAAC
 AACAACTATCAGTCGACTGTGTCCAACGTCACTGCGCCTATTGTGCGCAACGACCTCAACTATGTGACC
 CAGTACTGGTCACTGCTTGCTCTTGAATTACGTCTTTGCTTGTGTGTCTAATGCCCTCCACCACAG
 GAACCAAAACCGGCTTTGACCTCTGGGAAGAAGTCAATGGGAGCTCATTCTTTACTGTTGCCAACCAGCA
 CCGAGGTATGAAGCAAATCCTCGACATTGCTGCTACTGCACATGAGCATTGTTACTGACCAGCTCTAC
 AGCACTTGTGCGAGGGCGCCACTCTTGCTGCCACTCTTGCCAGTCGGGAAGCGCTTATTCATCTGTTGC
 TCCCCAGGTTTGTGCTTTCTCCAACGATTCTGGGTGTGCTCTGGTGGATACGTCGACTCCAACAGTAT
 GTCTTTTCACTGTTTATATGAGATTGGCCAATACTGATAGCTCGCCTCTAGTCAACACCAACGAGGGCA
 GGACTGGCAAGGATGTCAACTCCGCTCTGACTTCCATCCACACCTTCGATCCCAACCTTGGCTGTGACG
 CAGGCACCTTCCAGCCATGCAGTGACAAAGCGCTCTCCAACCTCAAGGTTGTTGTGCACTCCTTCCGCT
 CCATCTACGGCGTGAAACAAGGGCATTCTGCGGGTGCTGCGCTCGCCATTGGCCGGTATGCAGAGGATG
 TGTACTACAACGGCAACCCTTGGTATCTTGCTACATTTGCTGCTGCGGAGCAGCTGTACGATGCCATCT
 ACGTCTGGAAGAAGACGGGCTCCATCACGGTGACCGCCACCTCCCTGGCCTTCTTCCAGGAGCTTGTTT
 CTGGCGTGACGGCCGGGACCTACTCCAGCAGCTCTTCGACCTTTACCAACATCATCAACGCCGTCTCGA
 CATAACGCCGATGGCTTCTCAGCGAGGCTGCCAAGTACGTCCCCGCCGACGGTTGCTGGCCGAGCAGT
 TTGACCGCAACAGCGGCACTCCGCTGTCTGCGCTTCACCTGACGTGGTTCGTACGCCTCGTTCTTGACAG
 CCACGGCCCGTGGGCTGGCATCGTGCCCCCTCGTGGGCCAACAGCAGCGCTAGCACGATCCCCCTCGA
 CGTGCTCCGGCGCGTCCGTGGTGGGATCCTACTCGCGTCCCACCGCCACGTCAITCCCTCCGTGCGCAGA
 CGCCCAAGCCTGGCGTGCTTCCGGTACTCCCTACACGCCCTGCCCCTGCGCGACCCCAACCTCCGTGG
 CCGTCACCTTCCACGAGCTCGTGTGACACAGTTTGGCCAGACGGTCAAGGTGGCGGGCAACGCCGCGG
 CCTGGGCAACTGGAGCAGGAGCGCCGCGGTGGCTCTGGACGCCGTCAACTATGCCGATAACCAACCCCG
 TGTGGATTGGGACGGTCAACCTCGAGGCTGGAGACGTG

Fig. 9

TGGAGTACAAGTACATCAATGTGGGCCAAGATGGCTCCGTGACCTGGGAGAGTGATCCCAACCACACT
 TACACGGTTCCCTGCGGTGGCTTGTGTGACGACAGGTTGTCAAGGAGGACACCTGGCAGTCGTAAACCCA
 GCTTTCTTGTACAAAGTGGTGATCGCGCCAGCTCCGTGCGAAAGCCTGACGCACCGGTAGATTCTTGG
 TGAGCCCGTATCATGACGGCGGCGGGAGCTACATGGCCCCGGGTGATTTATTTTTTTTGTATCTACTT
 CTGACCCCTTTTCAAATATACGGTCAACTCATCTTTCACTGGAGATGCGGCCTGCTTGGTATTGCGATG
 TTGTGAGCTTGGCAAATTGTGGCTTTGAAAAACAAAAACGATTCTTAGTAGCCATGCATTTTAAGA
 TAACGGAATAGAAAGAGGAAATTAAAAAACAACATCCCGTTCATAACCCGTAGA
 ATCGCCGCTGGCGCGCTGAGACAATGGCCGGCAATGGTAAAAAGGACCAAGATGTACTAGGTAGTTG
 CAATGTGGCTTATTACCTACCTACTACCTGGTAGGCACCTACTAGGTACTTGGGTAGACGGACAATGA
 AATTTGAAGTCGGGGTTGACAGAAAGCAGGGCGCTGGACACATTGTGCTTCAGGCGGTACCCGTCGTC
 ATCGTCAGCCAATGTGAGGCCCGGCAGCCCGAGGAGCGAGACAACCTTGGCCGGAGGAGCCCGCAGG
 TACCTGCCAAAGCGCGCTGGTACCTCTCAACCTCTCAGGCCTGTTGGATGCCCTATGACATGCCCT
 GGGGGATGCAGCTGTTGCCCCGGCCCCGCACTTTCGGGTGACCGCGAGGCTGCTGATTGGCTGGTTGC
 CACGGGCTGGGCGGTCCCTGAAGTTGTTGCCATCTGAACTCTGTGCGCGCTGGCGTGGCTGCGCCCA
 ATGGGAGGCGAGACAACCTCAGGGTACTAGAATCACTGACAGAAGAAGAGAATCGAAAGTAGGTAGACA
 GCCAATTCTTGCATGGCAGGCAACCGCACAGGAGAAAAATTGACTACCCACAATCAGGCACAGTAA
 GTAGGGCACAGTACGTATGTACAGACAAGGCGCAAGCGATACTGCGCGACCCGGTACCTCGCCGGCTT
 GACACGTGCGACAGGCTACTTTACTAGTATTGCGAGCGGCGGGTCGCGCATTATTACATGTACTGTGC
 CGCCATTTGATGACTGGGCTGCTGACGTATTAGTAGATCTGCCCCGGCATCGCCCTTCCATGGGCGCGA
 CCCGGGACTGGACCTCTGACTCTACCTACATGTACCTAGGCCGGGCGGGCTTGGTGACTTTTGTCC
 GATCAGGTGCTTCGCTGGCTACCTATTATTCTCTTTCTTCTTCTCCATCCTGCTTCTGGCCTTGCA
 ATTCTTCTTCGCCACTCCTCCCTCTTCCCCCGCGATACCCCTTGAATTCGTGAGAGGAAAAAGACGA
 GAAAAAAGGGCAGCAGAGACGTGGTCTGGCTCACGTGCTGCATCTCTGCGCACTCTCATTTTTTT
 TATTGTCCGACCCCTCCCTCAACCTTCTCCTTCTGTTGACAGGCTAAGCCTTGCTTCGACGCTCTCTCT
 TTGAATTTTTCTACTTCTACCTTCTTTTCTTGGCTGTTACCCACCATAGCTCGATTACGATGCTCCG
 AAGTCGCCAAGTCACAGCCAGGGCCGTCCGGGCTCTGGGCCAGGCGCGCGCCTTTACCTCGACGACCA
 AGCCTGTGATGATCCAGAGCAGCCAGAGGAAACAGGCCAACGCCAGCGCTGCTCCGTAAGTCGCCCCAT
 TGCCATTGCATCTTCTGTTTGATATATACTTCTGCTGCTTGGCTGGCGTCTCTCTCGGTTATGCGT
 GTCAAGGACCAGGTGTGTTGCGATCGTGGTTTTCCAGCGCCGATTACCGGGGACGAATTTTTGGCTG
 CTCAACTCGCGCGCGCGCATCTGATTCTTCTGTTTTCAATCTTGAGCGACAACCTGGCTAACATAATGG
 CCATTGGCAATTGCTTCACACAGACAAGTCCGCCCTGTACCGAGCCCTGCTTTCAACGCTGAAGACAA
 AGACCGCAGCCATGTGCAGCCTCTGGTCAACCCGTGAAAGCCGACATGGATGAATCGTATGTCCACG
 TCCCTCTGTCGCCCTACAAAATGAACACGATTACACCAGAATTTTTGCAACAATCGACACTTCTAT
 AACAGACCAATTGAGCTT

Fig. 9

TGTTCTGACCAATCATGTTGCTCTAGATTCAATTGGCAAAACCGGAGGCGAAATCTTCCACGAGATGA
 TGCTGCGACAGGGTGTCAAGCACATTTGTAGGTTCCGATGCCGGCCGCCCACACGGGCTCCATCCTT
 GCTCCATCTCTCCAGCTAGGCAAATCTCGCTAACCTTGAGTCACCATCCAGTCGGATACCCCTGGCGG
 CGCTATCCTGCCCCGTCTTCGACGCCATCTACAACTCAAAACACTTCGACTTCATCCTGCCCCGTCTCAT
 GAGCAGGGAGCTGGCCATATGGCCGAGGGCTATGCCCGTGCCCTCGGGCAAACCCGGTGTCTGCTCTGG
 TGACTTCCGGCCCCGGTGCTACCAATGTCATCACGCCCATGCAGGATGCCCTGTCTGGACGGAACGCC
 CTTGGTCTGCTCTCTGCGGCCAGGTCCCCACCACGGCCATCGGCAGCGATGACTTCCAAGAGGCCGAC
 GTCGTGGGCATCTCGCGGGCCTGCACCAAGTGAACGTCATGGTCAAGAGCGTTGCTGAGCTGCCGC
 GGAGAATCAACGAGGCCTTTGAGATTGCCACCAGCGGCCGCTCGGCCCGCTCTCGTCGACCTGCC
 CAAGGATGTCACGGCTGGTATCCTGAGGAGAGCCATCCCTACGGAGACTGCTCTGCCGTCTCTGCCC
 AGTGCCGCCTCCCGCGCCGCCATGGAGCTGAGCTCCAAGCAGCTCAACGCCTCCATCAAGCGTGCCG
 CCGACCTCATCAACATCGCCAAGAAGCCCGTCATCTACGCCGGTCAGGGTGTCTCCAGTCCGAGGG
 CGGCGTTGAGCTCCTGAAGCAGCTGGCGGACAAGGCCCTCCATCCCGTCACCACCACCCTCCATGGC
 CTGGGTGCCTTTGATGAGCTGGACGAGAAGTCTGCTGCACATGCTGGGCATGCACGGCTCGGCGTATG
 CCAACATGGCCATGCAGCAGGCCGACCTCATCATCGCCCTCGGCAGCCGATTTCGACGACCGTGTTAC
 TCTGAATGTCTCCAAATTTGCGCCTGCAGCCAGGCAAGTGTCTGCCGAGGGCCGCGCGGCATCATT
 CACTTTGAGATCATGCCCAAGAACATCAACAAGGTCATCCAGGCGACCGAGGCCGTGAGGGCGACG
 TCGCCACCAACCTGAAGCACCTCATTCCCCAGATTGCCGAAAAGTCCATGGCGGACCGAGGAGAGTG
 GTTCCGCCTCATCAATGAGTGGAAGAAGAAGTGGCCCCCTGTCAAACCTACCAGCGCGCGGAGCGGGCT
 GGCCTCATCAAGCCGACAGCGGTATGGAGGAGATTAGCAACCTGACGGCCAACCGAAAGGACAAGA
 CGTACATTGCCACGGGTGTGCGCCAGCACCATGATGTGGGTTGCCAGCACTTCCGCTGGAGGCACCC
 TCGATCCATGATTACCTCTGGTGGTCTGGGCACCATGGGCTACGGTCTCCCCGCGGCCATTGGCGCC
 AAGGTGGCCCGACCCGACGCTCTCGTAATTGACGTTGATGGCGATGCCCTCGTTTAACATGACGCTGA
 CGGAGCTGTGCACTGCTGCACAGTTCAACATTGGCGTCAAGGTGGTTGTGCTCAACAACGAGGAGCA
 GGGCATGGTGACGCAGTGGCAGAACCTCTTTTACGAGGACCGATATGCCACACGCACCAGAAGAAC
 CCGACTTCATGAAGCTGGCCGACGCCATGGGCGTTTACGACCAGCGCGTGACGGAGCCGAGAAGC
 TGGTTCGATGCCCTGACGTGGCTGATCAACACCGATGGCCCGGCCCTGTTGGAGGTTGTACCGACAA
 GAAGGTGCCTGTCTGCCCATGGTGGCCCGCGGATCGGCCCTGCACGAGTTCCTCGTCTTTGAACCT
 GGTGAGTCTACTTCAGACATATTGCTTGCGCATTCAGATACTAACACTCTCACAGAAAAGGATAAG
 CAGCGCCGTGAGCTGATGAAGGAGAGAACAAGGGTGTGCACTCCTAAAGCGATGATGTCTGCGAGG
 GGTTCTTCGTTGAACCCTAGTTGAGGCACCATCTTACCCTCTTATTTTTTCCCGTGGGCTTTCATTT
 TGTGTATCCGAGCATGACGTTGTAGGGTTGGAGTTTCTTCCCTTTTATCTTGTCTATTTACTGGTAC
 CCATAGGCGCGAGACTAGGCTTCCATGTTTTGTTTTGCGACTTTCAAAAAGTACTTTTAGTGGTTTG
 GGGCACGACGAGGGGGGGCAACCTCTTCTGTGAAAAAGGTGGCTGGATGGATGAGATGAGATGAGA
 TGAGGGTGAAGATAGATACCTGCAGTGTTTTTGACGCGACGGGATGGCGATCGC

Fig. 9