



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

(21) PI 0807834-3 A2



* B R P I 0 8 0 7 8 3 4 A 2 *

(22) Data de Depósito: 30/01/2008

(43) Data da Publicação: 22/07/2014
(RPI 2272)

(51) Int.Cl.:

C12N 15/62

(54) Título: " SEQUÊNCIAS-LÍDER BACTERIANAS
PARA EXPRESSÃO AUMENTADA ".

(57) Resumo:

(30) Prioridade Unionista: 31/01/2007 US 60/887,476,
31/01/2007 US 60/887,486

(73) Titular(es): Dow Global Technologies Inc.

(72) Inventor(es): Charles Douglas Hershberger, Diane M.
Retallack, Jane C. Schneider, Russell John Coleman, Sol Resnick,
Stacey L. Lee, Thomas M. Ramseier

(74) Procurador(es): Dannemann ,Siemsen, Bigler &
Ipanema Moreira

(86) Pedido Internacional: PCT US2008052434 de
30/01/2008

(87) Publicação Internacional: WO 2008/094986de
07/08/2008

Relatório Descritivo da Patente de Invenção para **"SEQUÊNCIAS-LÍDER BACTERIANAS PARA EXPRESSÃO AUMENTADA"**.

CAMPO DA INVENÇÃO

5 A presente invenção refere-se ao campo de produção de proteína, particularmente ao uso de polipeptídeos direcionados para a produção de proteínas heterólogas apropriadamente processadas.

ANTECEDENTES DA INVENÇÃO

10 Mais do que 150 proteínas e polipeptídeos produzidos recombinantemente foram aprovados pelo U.S. Food and Drug Administration (FDA) para uso como fármacos e vacinas em biotecnologia, com outras 370 em ensaios clínicos. Diferentes de terapêuticos de moléculas pequenas que são produzidos através de síntese química, as proteínas e os polipeptídeos são mais eficientemente produzidos em células vivas. Entretanto, os métodos de produção atuais de proteínas recombinantes em bactérias geralmente produzem proteínas impropriamente enoveladas, agregadas ou inativas e vários

15 tipos de proteínas requerem modificações secundárias que são obtidas de maneira ineficaz usando métodos conhecidos.

Um problema primário com os métodos conhecidos está situado na formação de corpos de inclusão feitos de agregados de proteínas no citoplasma, que ocorre quando uma quantidade excessiva de proteína se acumula na célula. Outro problema na produção de proteína recombinante é o estabelecimento da conformação secundária e terciária apropriada para as proteínas expressas. Uma barreira é o citoplasma bacteriano que resiste

20 ativamente a formação de ligações de dissulfeto, o que geralmente dá suporte ao enovelamento apropriado da proteína (Derman *et al.* (1993) Science 262:1744-7). Como resultado, várias proteínas recombinantes, particularmente aquelas de origem eucariótica, são impropriamente enoveladas e inativas quando produzidas em bactérias.

Numerosas tentativas têm sido desenvolvidas para aumentar a

30 produção de proteínas apropriadamente enoveladas em sistemas recombinantes. Por exemplo, pesquisadores têm alterado as condições de fermentação (Schein (1989) Bio/Technology, 7:1141-1149), variado a força do promotor ou

usado proteínas chaperonas superexpressas (Hockney (1994) Trends Biotechnol. 12:456-463), que podem ajudar a prevenir a formação de corpos de inclusão.

Uma abordagem alternativa para aumentar a coleta de proteínas
5 apropriadamente enoveladas é secretar a proteína a partir do ambiente intracelular. A forma mais comum de secreção de polipeptídeos com uma sequência de sinal envolve o sistema Sec. O sistema Sec é responsável pela exportação de proteínas com polipeptídeos de sinal N-terminais através das
10 membranas citoplasmáticas (veja Agarraberes e Dice (2001) Biochim Biophys Acta. 1513:1-24; Muller *et al* (2001) Prog Nucleic Acid Res Mol Biol. 66:107-157).

Estratégias têm sido desenvolvidas para excretar proteínas a partir da célula no sobrenadante. Por exemplo, Pat. U.S. Nº 5.348.867; Pat. U.S. Nº 6.329.172; Publicação PCT Nº WO 96/17943; Publicação PCT Nº
15 WO 02/40696 e Pedido de Publicação U.S. 2003/0013150. Outras estratégias para expressão aumentada são dirigidas para o direcionamento da proteína para o periplasma. Algumas pesquisas concentram-se na secreção tipo não Sec (veja, por exemplo, Publicação PCT Nº WO 03/079007; Publicação U.S. Nº 2003/0180937; Publicação U.S. Nº 2003/0064435 e Publicação PCT
20 Nº WO 00/59537). Entretanto, a maioria das pesquisas tem se concentrado na secreção de proteínas exógenas com um sistema de secreção tipo Sec.

Vários sinais de secreção têm sido descritos para uso em expressão de proteínas ou polipeptídeos recombinantes. Veja, por exemplo, Pat. U.S Nº 5.914.254; Pat. U.S Nº 4. 963.495; Patente Européia Nº 0 177
25 343; Pat. U.S Nº 5.082.783; Publicação PCT Nº WO 89/10971; Pat. U.S Nº 6.156.552; Pat. U.S Nº 6.495.357; 6.509.181; 6.524.827; 6.528.298; 6.558.939; 6.608.018; 6.617.143; Pat. U.S Nº 5.595.898; 5.698.435; e 6.204.023; Pat. U.S Nº 6.258.560; Publicação PCT Nº WO 01/21662, WO 02/068660 e Publicação de Pedido U.S. 2003/0044906; Pat. U.S Nº
30 5.641.671 e Patente Européia Nº EP 0 121 352.

Estratégias que dependem de sequências de sinal para proteínas direcionadas para fora do citoplasma geralmente produzem proteínas

impropriamente processadas. Isso é particularmente verdadeiro para sinais de secreção amino-terminais tais como aqueles que levam à secreção através do sistema Sec. As proteínas que são processadas através desse sistema geralmente tanto retêm uma porção do sinal de secreção, o que requer
5 um elemento ligante que é, geralmente, impropriamente clivado, quanto são truncadas na terminação.

Como está esclarecido a partir da técnica acima descrita, várias estratégias têm sido desenvolvidas para direcionar proteínas para o periplasma de uma célula hospedeira. Entretanto, as estratégias conhecidas não
10 têm resultado em um rendimento consistentemente elevado de proteína recombinante ativa, apropriadamente processada, que possa ser purificada para uso terapêutico. Uma limitação principal das estratégias prévias tem sido a expressão de proteínas com sequências de sinal de secreção pobres em sistemas celulares inadequados.

15 Como resultado, permanece uma necessidade na técnica por sistemas de expressão em larga escala aperfeiçoados capazes de secretar e processar apropriadamente polipeptídeos recombinantes para produzir proteínas transgênicas em uma forma apropriadamente processada.

SUMÁRIO DA INVENÇÃO

20 A presente invenção fornece composições e processos aperfeiçoados para produzir altos níveis de proteína ou polipeptídeo de interesse apropriadamente processados em um sistema de expressão celular. Em particular, a invenção fornece novas sequências de aminoácidos e de nucleotídeos para sinais de secreção derivadas de um organismo bacteriano. Em
25 uma modalidade, os sinais de secreção da invenção incluem um polipeptídeo isolado com uma sequência que é, ou é substancialmente homóloga ao polipeptídeo de secreção de *Pseudomonas fluorescens* (*P. fluorescens*) selecionado a partir de uma proteína dissulfeto isomerase C (dsbC), uma proteína mutante que se ligue a fosfato (pbp*), uma proteína dissulfeto isomerase A (dsbA), Bce, CupA2, CupB2, CupC2, NikA, Flgl, uma proteína da família de tetratricopeptídeo repetido (ORF5550), uma proteína de tolerância ao
30 tolueno (Tgt2C) e um sinal de secreção de proteína quimiotática que aceita

metilação (ORF8124), assim como suas variantes ativas, fragmentos e derivados. Em outra modalidade, os sinais de secreção da invenção incluem um polipeptídeo isolado com uma sequência que é, ou que é substancialmente homólogo à sequência de sinal de secreção Bce de *Bacillus coagulans*. As

5 sequências de nucleotídeos que codificam as sequências de sinal da invenção são úteis em vetores e sistemas de expressão para promover o direcionamento de uma proteína ou polipeptídeo de interesse expressos para o periplasma de bactérias Gram-negativas ou no ambiente extracelular.

Construtos de DNA compreendendo as sequências de sinal de

10 secreção são úteis em células hospedeiras para expressar proteínas recombinantes. Sequências de nucleotídeos para as proteínas de interesse estão operativamente ligadas a um sinal de secreção como aqui descrito. A célula pode expressar a proteína no compartimento periplasmático. Em certas modalidades, a célula também pode secretar a proteína recombinante expressa

15 extracelularmente através da parede celular externa. As células hospedeiras incluem células eucarióticas, incluindo células de levedura, células de inseto, células de mamíferos, células de planta, etc., e células procarióticas, incluindo células de bactérias tais como *P. fluorescens*, *E. coli* e semelhantes. Qualquer proteína de interesse pode ser expressa usando as sequências-

20 líder de polipeptídeo de secreção da invenção, incluindo proteínas terapêuticas, hormônios, fatores de crescimento, receptores ou ligantes extracelulares, proteases, quinases, proteínas do sangue, quimiocinas, citocinas, anticorpos e semelhantes.

BREVE DESCRIÇÃO DAS FIGURAS

25 Figura 1 descreve o construto de expressão para a proteína de fusão dsbC SS-skp.

Figura 2 mostra a expressão da proteína Skp em torno de 17 kDa (setas). As bandas marcadas 2 e 3 foram consistentes com a proteína Skp. A banda 1 parece ter ambas, a proteína de ligação ao DNA (3691) e Skp.

30 Figura 3 é uma análise após a expressão de dsbC SS-skp em *Pseudomonas fluorescens* após 0 e 24 horas em frações solúvel (S) e insolúvel (I) para amostras marcadas 2B-2 (Figura 3A) e 2B-4 (Figura 3B). Na

figura 3A, as bandas 5, 7 e 9 eram a proteína dsbC-skp não processada na fração insolúvel. As bandas 6, 8 e 10 eram de dsbC-skp processada na fração insolúvel. As bandas 1 e 3 eram de dsbC-skp processada na fração solúvel. As bandas 2 e 4 eram de uma proteína desconhecida. Na figura 3B, as

5 bandas 15,17 e 19 eram de proteína dsbC-skp não processada na fração insolúvel. As bandas 16, 18 e 20 eram de dsbC-skp processada na fração insolúvel. As bandas 11 e 13 eram de dsbC-skp processada na fração solúvel. As bandas 12 e 14 eram de uma proteína desconhecida.

Figura 4 mostra uma análise Western de acumulação de proteína após a expressão de DC694 (dsbA-PA83). Acumulação de solúvel (S), insolúvel (I) e em caldo acelular (B) em 0 e 24 horas foi avaliada pela análise Western.

10

Figura 5 mostra uma análise Western de acumulação de proteína após a expressão de EP468-002.2 (dsbA). A acumulação de proteína solúvel (S) e insolúvel (I) em 0 e 24 horas após a indução foi avaliada pela análise Western.

15

Figura 6 demonstra a atividade de fosfatase alcalina do mutante pINS-008-3 (pbp*) comparada com o sinal de secreção de pINS-008-5 (pbp tipo selvagem). As culturas celulares foram ajustadas para 1 unidade OD₆₀₀, então a atividade de PhoA foi medida pela adição de 4-metilumbeliferona (MUP) e pela medida de formação de produto fluorescente aos 10 min. Os

20 controles negativos continham MUP, mas não células.

Figura 7 mostra a análise Western de acumulação de proteína após a expressão de pINS-008-3 (pbp*) e pINS-008-5 (pbp tipo selvagem). A acumulação de proinsulina-PhoA na fração solúvel (S), insolúvel (I) e extra-

25 celular (Bro) em 10, 116 e 140 horas foi avaliada pela análise Western. Aliquotas da cultura foram ajustadas para 20 unidades OD₆₀₀, separadas por SDS-PAGE, transferidas para um filtro e visualizadas com um anticorpo para insulina (policlonal de galinha, Abcam cat#ab14042).

Figura 8 mostra a análise SDS-PAGE das frações EP484-003 e EP484-004. Os resultados representativos da análise de SDS-PAGE são mostrados. Os marcadores de peso molecular (L) são mostrados no centro.

30

Os padrões de BSA (BSA Stds) estão indicados. A seta indica a banda induzida. Abaixo de cada banda está a fração do tipo: solúvel (Sol), insolúvel (Ins) e de caldo celular (CFB). Acima de cada banda está a amostra do tempo de indução (10) ou 24 horas após indução (I24). O número da cepa está mostrado abaixo de cada agrupamento de amostras. A banda larga de proteína na fração insolúvel I24 de EP484-004 corresponde à expressão de gene intensificada facilitada pela sequência-líder Bce.

Figura 9 demonstra as análises SDS-PAGE e Western da expressão de Gal2 scFv. As frações solúvel (S) e insolúvel (I) foram analisadas. Acima de cada par de bandas está indicada a líder de secreção fundida a Gal2. Marcadores de peso molecular estão descritos à esquerda de cada gel de SDS-PAGE (alto) ou Western blot (embaixo). As setas indicam a migração de Gal2.

Figura 10 representa uma análise por SDS-PAGE da expressão de Tiorredoxina (TrxA). As frações solúveis foram analisadas. Acima de cada par de bandas está indicada a líder de secreção fundida com TrxA. Marcadores de peso molecular estão descritos à esquerda do gel de SDS-PAGE. As setas indicam a migração de TrxA não processada (seta superior) e processada (seta inferior).

20 DESCRIÇÃO DETALHADA

I. Visão geral

São fornecidos composições e métodos para produção de níveis elevados de polipeptídeos apropriadamente processados em uma célula hospedeira. Em particular, são fornecidos novos sinais de secreção que promovem o direcionamento de um polipeptídeo de interesse operativamente ligado para o periplasma de bactérias Gram-negativas ou para o ambiente extracelular. Para os propósitos da presente invenção, um "sinal de secreção", um "polipeptídeo de sinal de secreção", "peptídeo de sinal" ou "sequência-líder" é compreendido como uma sequência de peptídeo (ou a sequência de peptídeo que codifica um polinucleotídeo) que é útil para direcionar uma proteína ou polipeptídeo de interesse operativamente ligados para o periplasma de bactérias Gram-negativas ou para o espaço extracelular. As

sequências de sinal de secreção da invenção incluem os polipeptídeos de secreção selecionados entre os sinais de secreção de pbp*, dsbA, dsbC, Bce, CupA2, CupB2, CupC2, NikA, F1gl, ORF5550, Tgt2C e ORF8124, e seus fragmentos e variantes. As sequências de aminoácidos para os sinais de secreção estão descritas nas SEQ ID Nº 2, 4,6, 8, 10, 12,14, 16, 18, 20, 22 e 24. As sequências de nucleotídeos correspondentes são fornecidas pelas SEQ ID Nº 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21 e 23, respectivamente. A invenção compreende essas sequências assim como seus fragmentos e variantes.

Os métodos da invenção fornecem aperfeiçoamentos de métodos atuais de produção de proteínas recombinantes em bactérias que geralmente produzem proteínas inapropriadamente enoveladas, agregadas ou inativas. Além disso, vários tipos de proteína requerem modificações secundárias que são obtidas de modo ineficaz usando os métodos conhecidos. Os métodos desse aumentam a coleta de proteínas apropriadamente enoveladas pela secreção da proteína do ambiente intracelular. Em bactérias Gram-negativas, uma proteína secretada do citoplasma pode terminar no espaço periplasmático, acoplada à membrana externa ou no líquido extracelular. Os métodos também evitam corpos de inclusão, que são feitos de agregados de proteínas. A secreção para o espaço periplasmático tem o efeito bem conhecido de facilitar a formação de ligação de dissulfeto apropriada (Bardwell *et al.*, (1994) *Phosphate Microorg.* 270-5; Manoil (2000) *Methods in Enzymol.* 326:35-47). Outros benefícios da secreção de proteína recombinante incluem o isolamento mais eficiente da proteína; enovelamento e formação de ligação de dissulfeto da proteína transgênica apropriados, levando a um aumento no percentual da proteína na forma ativa; formação reduzida de corpos de inclusão e toxicidade reduzida para a célula hospedeira; e percentual aumentado da proteína recombinante na forma solúvel. O potencial para excreção da proteína de interesse para o meio de cultura também pode promover a produção de proteína potencialmente contínua, mais do que a cultura em lotes.

As bactérias Gram-negativas desenvolveram numerosos siste-

mas para a exportação ativa de proteínas através de suas membranas duplas. Essas vias de secreção incluem, por exemplo, a via ABC (Tipo I), a via Path/Fla (Tipo III) e a via Path/Vir (Tipo IV) para a translocação em uma etapa através de ambas as membranas plasmática e externa; as vias Sec (Tipo II), Tat, MscL e Holins para translocação através da membrana plasmática; e as vias Sec mais fímbria que conduz porina (FUP), Sec mais autotransportador (AT), Sec mais dois parceiros de secreção (TPS), Sec mais ramo terminal principal (MTB) e Tat mais MTB para translocação em duas etapas através das membranas plasmática e externa. Nem todas as bactérias têm todas essas vias de secreção.

Três sistemas de proteína (tipos I, II e III) secretam proteínas através de ambas as membranas em uma etapa única ligada a energia. Quatro sistemas (Sec, Tat, MscL e Holins) secretam apenas através da membrana interna e quatro sistemas (MTB, FUP, AT e TPS) secretam apenas através da membrana externa.

Em uma modalidade, as sequências de sinal da invenção utilizam o sistema de secreção Sec. O sistema Sec é responsável pela exportação de proteínas com os polipeptídeos de sinal N-terminal através das membranas citoplasmáticas (veja Agarraberes e Dice (2001) *Biochim Biophys Acta*. 66:107-157). Os complexos de proteína da família Sec são encontrados universalmente em procarióticos e eucarióticos. O sistema Sec bacteriano consiste em proteínas de transporte, uma proteína chaperona (SecB) ou uma partícula de reconhecimento de sinal (SRP) e peptidases de sinal (SPase I e SPase II). O complexo de transporte Sec em *E. coli* consiste em três proteínas de membrana interna integrais, SecY, SecE e SecG, e a ATPase citoplasmática, SecA. SecA recruta os complexos SecY/E/G para formar o canal de translocação ativa. A proteína chaperona SecB se liga à cadeia do polipeptídeo nascente para protegê-lo do enovelamento e o direciona para SecA. A cadeia linear de polipeptídeo é subsequentemente transportada através do canal SecYEG e, após a clivagem do polipeptídeo de sinal, a proteína é enovelada no periplasma. Três proteínas auxiliares (SecD, SecF e YajC) formam um complexo que não é essencial para a se-

creção, mas que estimula a secreção em até dez vezes sob várias condições, particularmente em temperaturas baixas.

As proteínas que são transportadas para o periplasma, isto é, através de um sistema de secreção tipo II, também podem ser exportadas através de um autotransportador, um sistema de dois parceiros de secreção, um sistema de ramo terminal principal ou de fímbria que conduz porina.

Dos doze sistemas de secreção conhecidos em bactérias Gram-negativas, oito são conhecidos por utilizar polipeptídeos de sinal direcionais encontrados com parte da proteína expressa. Esses polipeptídeos de sinal interagem com as proteínas dos sistemas de secreção de maneira que a célula direciona apropriadamente a proteína para a sua destinação apropriada. Cinco desses oito sistemas de secreção baseados em polipeptídeo de sinal são aqueles que envolvem o sistema Sec. Esses cinco são referidos como envolvidos na translocação através da membrana citoplasmática dependentes de Sec e seus polipeptídeos de sinal operativos nesse local podem ser referidos como sinais de secreção dependentes de Sec. Um dos problemas no desenvolvimento de um sinal de secreção apropriado é garantir que o sinal seja expresso e clivado da proteína expressa apropriadamente.

Polipeptídeos de sinal para a via sec geralmente consistem em três domínios: (i) uma região n positivamente carregada, (ii) uma região hidrofóbica h e (iii) uma região c não carregada, mas polar. O sítio de clivagem para a peptidase de sinal está localizado na região c. Entretanto, o grau de conservação da sequência de sinal e a extensão, assim como a posição do sítio de clivagem, podem variar entre proteínas diferentes.

Uma indicação de exportação de proteína dependente de Sec é a presença de uma sequência de sinal curta (cerca de 30 aminoácidos), amino-terminal predominantemente hidrofóbica na proteína exportada. A sequência de sinal auxilia na exportação de proteína e é clivada por uma peptidase de sinal periplasmática quando a proteína exportada alcança o periplasma. Um polipeptídeo de sinal N-terminal típico de Sec contém um domínio N com pelo menos um resíduo de arginina ou lisina, seguido por um domínio que contém uma fita de resíduos hidrofóbicos e um domínio C conten-

do o sítio de clivagem para peptidases de sinal.

Sistema bacterianos de produção de proteína têm sido desenvolvidos nos quais construtos de proteína transgênica são manipulados como proteínas de fusão que contêm ambos, uma proteína de interesse e um
5 sinal de secreção, em uma tentativa de direcionar a proteína para fora do citoplasma.

P. fluorescens tem demonstrado ser uma plataforma aperfeiçoada para a produção de uma variedade de proteínas e vários sinais de secreção eficientes que foram identificados a partir desse organismo (veja, Publicação de Pedido U.S. Número 20060008877, incorporado aqui por referên-
10 cia em sua totalidade). *P. fluorescens* produz proteínas exógenas de uma forma corretamente processada em um nível mais elevado do que o tipicamente visto em outros sistemas de expressão bacterianos e o transporte dessas proteínas em um nível mais elevado para o periplasma da célula,
15 levando a recuperação aumentada de proteína recombinante totalmente processada. Portanto, em uma modalidade, a invenção fornece um método para a produção de proteína exógena em uma célula de *P. fluorescens* pela expressão da proteína-alvo ligada a um sinal de secreção.

As sequências de sinal de secreção da invenção são úteis em
20 *Pseudomonas*. O sistema em *Pseudomonas* oferece vantagens para a expressão comercial de polipeptídeos e enzimas, em comparação com outros sistemas de expressão bacterianos. Em particular, *P. fluorescens* tem sido identificada como um vantajosos sistema de expressão. *P. fluorescens* abrange um grupo de saprófitas comuns, não patogênicas, que colonizam o
25 solo, a água e ambientes superficiais de planta. Enzimas comerciais derivadas de *P. fluorescens* têm sido usadas para reduzir a contaminação ambiental, como aditivos para detergentes e para hidrólise estereoseletiva. *P. fluorescens* TAM é usada na agricultura para controlar patógenos. A Patente U.S. Número 4.695.462 descreve a expressão de proteínas bacterianas re-
30 combinantes em *P. fluorescens*. Entre 1985 e 2004, várias empresas capitalizaram o uso agrícola de *P. fluorescens* para a produção de toxinas pesticidas, inseticidas e nematocidas, assim como sequências tóxicas específicas

e manipulação genética para aumentar a expressão dessas. Veja, por exemplo, Pedido PCT Nº WO 03/068926 e WO 03/068948; publicação PCT Nº WO 03/089455; Pedido PCT Nº WO 04/005221; e Publicação de Patente U.S. Número 20060008877.

5 II. Composições

A. Polipeptídeos isolados

Em uma modalidade da presente invenção, um polipeptídeo isolado é fornecido, em que o polipeptídeo isolado é um novo sinal de secreção útil para o direcionamento de uma proteína operativamente ligada ou um
10 polipeptídeo de interesse para o periplasma de bactérias Gram-negativas para o espaço extracelular. Em uma modalidade, o polipeptídeo tem uma sequência de aminoácidos que é, ou é substancialmente, homóloga a um sinal de secreção de pbp*, dsbA, dsbC, Bce, CupA2, CupB2, CupC2, NikA, F1gl, ORF5550, Ttg2C ou ORF8124, ou seus fragmentos ou variantes. Em
15 outra modalidade, esse polipeptídeo isolado é uma proteína de fusão do sinal de secreção e uma proteína ou polipeptídeo de interesse.

Em outra modalidade, a sequência do polipeptídeo é, ou é substancialmente, homóloga ao polipeptídeo de sinal de secreção descrito nas SEQ ID Nº 2,4,6, 8, 10, 12,14, 16, 18, 20, 22 ou 24 ou é codificada pela se-
20 quência de nucleotídeos descrita nas SEQ ID Nº 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21 ou 23. Em outra modalidade, a sequência do polipeptídeo compreende pelo menos 2 a 24 aminoácidos de SEQ ID Nº 2, pelo menos 2 a 22 aminoácidos de SEQ ID Nº 4, pelo menos 2 a 21 aminoácidos de SEQ ID Nº 6, pelo menos 2 a 33 aminoácidos de SEQ ID Nº 8, pelo menos 2 a 25 aminoá-
25 cidos de SEQ ID Nº 10, pelo menos 2 a 24 aminoácidos de SEQ ID Nº 12, pelo menos 2 a 23 aminoácidos de SEQ ID Nº 14, pelo menos 2 a 21 aminoácidos de SEQ ID Nº 16, pelo menos 2 a 21 aminoácidos de SEQ ID Nº 18, pelo menos 2 a 21 aminoácidos de SEQ ID Nº 20, pelo menos 2 a 33 aminoácidos de SEQ ID Nº 22 ou pelo menos 2 a 39 aminoácidos de SEQ
30 ID Nº 24. Em outra modalidade, a sequência de polipeptídeo compreende um fragmento das SEQ ID Nº 2, 4, 6, 8, 10, 12,14, 16, 18, 20, 22 ou 24, que é truncada por 1, 2, 3, 4, 5, 6, 7, 8, 9 ou 10 aminoácidos da terminação ami-

no, mas retém a atividade biológica, isto é, atividade de sinal de secreção.

Em uma modalidade, a sequência de aminoácidos do polipeptídeo homólogo é uma variante de um dado polipeptídeo original, em que a sequência da variante é obtível pela substituição de até ou cerca de 30% dos resíduos de aminoácidos do polipeptídeo original por outros resíduos de aminoácidos, incluindo até cerca de 1%, 2%, 3%, 4%, 5%, 6%, 7%, 8%, 9%, 10%, 11%, 12%, 13%, 14%, 15%, 16%, 17%, 18%, 19%, 21%, 21%, 22%, 23%, 24%, 25%, 26%, 27%, 28%, 29% ou 30%, desde que a variante retenha a função desejada do polipeptídeo original. Uma variante de aminoácido com homologia substancial será pelo menos cerca de 70%, pelo menos cerca de 75%, pelo menos cerca de 80%, pelo menos cerca de 85%, pelo menos cerca de 90%, pelo menos cerca de 95%, pelo menos cerca de 96%, pelo menos cerca de 97%, pelo menos cerca de 98% ou pelo menos cerca de 99% homóloga ao dado polipeptídeo. Uma variante de aminoácido pode ser obtida de várias maneiras incluindo substituições, deleções, truncamentos e inserções de aminoácidos de um ou mais aminoácidos das SEQ ID N^o 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22 ou 24, incluindo até cerca de 1, cerca de 2, cerca de 3, cerca de 4, cerca de 5, cerca de 6, cerca de 7, cerca de 8, cerca de 9, cerca de 10, cerca de 15, cerca de 20, cerca de 25 ou mais substituições, deleções ou inserções de aminoácidos.

Por "substancialmente homóloga" ou "substancialmente similar" é entendida uma sequência de aminoácidos ou sequência de nucleotídeos que tem pelo menos cerca de 60% ou 65% de identidade de sequência, cerca de 70% ou 75% de identidade de sequência, cerca de 80% ou 85% de identidade de sequência, cerca de 90%, cerca de 91%, cerca de 92%, cerca de 93%, cerca de 94%, cerca de 95%, cerca de 96%, cerca de 97%, cerca de 98% ou cerca de 99% ou mais de identidade de sequência comparada com uma sequência de referência que usa um dos programas de alinhamento aqui descritos usando parâmetros padronizados. Aquele versado na técnica reconhecerá que esses valores podem ser apropriadamente ajustados para determinar a identidade correspondente de proteínas codificadas pelas duas sequências de nucleotídeos levando em consideração a degeneração

de códon, a similaridade de aminoácido, o posicionamento da fase de leitura e semelhantes.

Por exemplo, preferivelmente, as substituições conservativas de aminoácidos podem ser feitas em um ou mais resíduos de aminoácidos pre-
 5 ditos, preferivelmente não essenciais. Um resíduo de aminoácido "não es-
 sencial" é um resíduo que pode ser alterado a partir da sequência de tipo
 selvagem de um polipeptídeo de sinal de secreção sem alterar a atividade
 biológica, enquanto que um resíduo de aminoácido "essencial" é necessário
 para a atividade biológica. Uma "substituição conservativa de aminoácido" é
 10 aquela na qual o resíduo de aminoácido é substituído por outro resíduo de
 aminoácido que tem uma cadeia lateral similar. As famílias de resíduos de
 aminoácidos conservativos e não conservativos estão listadas na Tabela 1.

Tabela 1. Grupos de Substituição de Aminoácido Similar	
Grupos Conservativos (8)	Grupos Semiconservativos (7)
Arg, Lys	Arg, Lys, His
Asp, Gln	Asn, Asp, Gly, Gln
Asn, Glu	
Ile, Leu, Val	Ile, Leu, Val, Met, Phe
Ala, Gly	Ala, Gly, Pro, Ser, Thr
Ser, Thr	Ser, Thr, Tyr
Phe, Thr	Phe, Trp, Tyr
Cys (não cisteína), Ser	Cys (não cisteína), Ser, Thr

Proteínas variantes abrangidas pela presente invenção são bio-
 logicamente ativas, ou seja, elas continuam a possuir a atividade biológica
 15 desejada da proteína nativa, ou seja, retêm a atividade de sinal de secreção.
 Por "reter a atividade" é pretendido que a variante tenha pelo menos cerca
 de 30%, pelo menos cerca de 50%, pelo menos cerca de 70%, pelo menos
 cerca de 80%, cerca de 90%, cerca de 95%, cerca de 100%, cerca de 110%,
 cerca de 125%, cerca de 150%, pelo menos cerca de 200% ou mais da ati-
 20 vidade de sinal de secreção da proteína nativa.

B. Polinucleotídeos isolados

A invenção também inclui um ácido nucleico isolado com uma

sequência que codifique um novo sinal de secreção útil para o direcionamento de uma proteína ou polipeptídeo de interesse operativamente ligados para o periplasma de bactérias Gram-negativas ou para o espaço extracelular. Em uma modalidade, o polinucleotídeo isolado codifica uma sequência de polipeptídeos substancialmente homóloga ao polipeptídeo de sinal de secreção de pbp*, dsbA, dsbC, Bce, CupA2, CupB2, CupC2, NikA, F1gl, ORF5550, Ttg2C ou ORF8124. Em outra modalidade, a presente invenção fornece um ácido nucleico que codifica uma sequência de polipeptídeos substancialmente homóloga a pelo menos 2 a 24 aminoácidos de SEQ ID Nº 2, pelo menos 2 a 22 aminoácidos de SEQ ID Nº 4, pelo menos 2 a 21 aminoácidos de SEQ ID Nº 6, pelo menos 2 a 33 aminoácidos de SEQ ID Nº 8, pelo menos 2 a 25 aminoácidos de SEQ ID Nº 10, pelo menos 2 a 24 aminoácidos de SEQ ID Nº 12, pelo menos 2 a 23 aminoácidos de SEQ ID Nº 14, pelo menos 2 a 21 aminoácidos de SEQ ID Nº 16, pelo menos 2 a 21 aminoácidos de SEQ ID Nº 18, pelo menos 2 a 21 aminoácidos de SEQ ID Nº 20, pelo menos 2 a 33 aminoácidos de SEQ ID Nº 22 ou pelo menos 2 a 39 aminoácidos de SEQ ID Nº 24 ou fornece um ácido nucleico substancialmente homólogo a SEQ ID Nº 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21 ou 23, incluindo variantes e fragmentos biologicamente ativos dessas.

Em outra modalidade, a sequência de ácido nucleico é pelo menos cerca de 60%, pelo menos cerca de 65%, pelo menos cerca de 70%, pelo menos cerca 75%, cerca de 80%, cerca de 85%,cerca de 90%, cerca de 95%, cerca de 96%, cerca de 97%, cerca de 98% ou pelo menos cerca de 99% idêntica a sequência de SEQ ID Nº 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21 ou 23. Em outra modalidade, o ácido nucleico codifica um polipeptídeo que é pelo menos cerca de 70%, cerca 75%, pelo menos cerca de 80%, cerca de 85%,cerca de 90%, cerca de 95%, cerca de 96%, cerca de 97%, cerca de 98% ou pelo menos cerca de 99% idêntica a sequência de aminoácidos de SEQ ID Nº 6, 2, 4, 8, 10, 12, 14, 16, 18, 20, 22 ou 24.

Os polipeptídeos de sinal de secreção preferidos da presente invenção são codificados por uma sequência de nucleotídeos substancialmente homóloga as sequências de nucleotídeos de SEQ ID Nº 1 ou 3. U-

sando métodos tais como PCR, hibridização e semelhantes, as sequências de polipeptídeo de sinal de secreção correspondentes podem ser identificadas, tais sequências tendo identidade substancial com as sequências da invenção. Veja, por exemplo, Sambrook J., e Russell, D.W. (2001) *Molecular Cloning: A Laboratory Manual*. (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY) e Innis, et al (1990) *PCR Protocols: A Guide to Methods and Applications* (Academic Press, NY). Sequências de nucleotídeos variantes também incluem sequências de nucleotídeos derivadas sinteticamente que foram geradas, por exemplo, pelo uso de mutagênese direcionada ao sítio mas que ainda codificam os polipeptídeos de sinal de secreção descritos na presente invenção como descrito *infra*. Polipeptídeos de sinal de secreção variantes abrangidos pela presente invenção são biologicamente ativos, ou seja, eles continuam possuindo a atividade biológica desejada da proteína nativa, ou seja, retêm a atividade de sinalização de secreção. Por "reter a atividade" é pretendido que a variante tenha pelo menos cerca de 30%, pelo menos cerca de 50%, pelo menos cerca de 70%, pelo menos cerca de 80%, pelo menos cerca de 85%, pelo menos cerca de 90%, pelo menos cerca de 95%, cerca de 96%, cerca de 97%, cerca de 98%, pelo menos cerca de 99% ou mais da atividade do polipeptídeo de sinal de secreção nativo. Métodos para medir a atividade do polipeptídeo de sinal de secreção serão discutidos aqui em outra parte.

O técnico versado apreciará ainda que podem ser introduzidas alterações pela mutação nas sequências de nucleotídeos da invenção levando, dessa maneira, a alterações na sequência de aminoácidos dos polipeptídeos de sinal de secreção codificados, sem alterar a atividade biológica dos polipeptídeos de sinal de secreção. Assim moléculas de ácido nucleico isolado variante podem ser criadas pela introdução de uma ou mais substituições, adições ou deleções de nucleotídeo na sequência de nucleotídeos correspondente descritas aqui, tal que uma ou mais substituições, adições ou deleções são introduzidas na proteína codificada. As mutações podem ser introduzidas por técnicas padronizadas, tais como mutagênese direcionada ao sítio e mutagênese mediada por PCR. Tais sequências variantes de

nucleotídeos também são abrangidas pela presente invenção.

C. Homologia de ácido nucleico e aminoácido

5 A homologia de sequência de ácido nucleico e aminoácido é determinada de acordo com qualquer um dos vários métodos bem conhecidos na técnica. Exemplos de alinhamento de sequência útil e de metodologias para determinação de homologia incluem aqueles descritos abaixo.

Alinhamentos e pesquisas por sequências similares podem ser realizados usando o programa de U.S. National Center for Biotechnology Information (NCBI), MegaBLAST (disponível atualmente em <http://www.ncbi.nlm.nih.gov/BLAST/>). O uso desse programa com as opções para percentual de identidade começando, por exemplo, com 70% para sequências de aminoácidos ou começando, por exemplo, com 90% para sequências de nucleotídeos, identificará aquelas sequências com 70% ou 90% ou mais de identidade de sequência com a sequência investigada. Outro programa conhecido na técnica também está disponível para alinhamento e/ou pesquisa de sequências similares, por exemplo, sequências pelo menos 70% ou 90% idênticas a uma série de informação que contém uma sequência de sinal de secreção de acordo com a presente invenção. Por exemplo, alinhamentos de sequência para comparação para identificar sequências pelo menos 70% ou 20 90% idênticas a uma sequência investigada podem ser realizados pelo uso, por exemplo, dos programas GAP, BESTFIT, BLAST, FASTA, e TFASTA disponíveis no GCG Sequence Analysis Software Package (disponibilizado por Genetics Computer Group, University of Wisconsin Biotechnology Center, 1710 University Avenue, Madison, Wis. 53705), com os parâmetros pre- 25 determinados como especificado nele, mais um parâmetro para a extensão de identidade de sequência começando no percentual desejado. Por exemplo, o programa CLUSTAL (disponível no pacote de programas PC/Gene de Intelligenetics, Mountain View, Cal.) pode ser usado.

30 Esses e outros métodos de alinhamento de sequência são bem conhecidos na técnica e podem ser conduzidos por alinhamento manual, por inspeção visual ou pela aplicação manual ou automática de um algoritmo de alinhamento de sequência, tal como qualquer um daqueles incorporados

pelos programas acima descritos. Vários algoritmos úteis incluem, por exemplo: o método de busca de similaridade descrito em W.R. Pearson & D.J. Lipman, Proc. Natl. Acad. Sci. USA 85:2444-48 (Abril de 1988); o método de homologia local descrito em T.F. Smith & M.S. Waterman, em Adv. Appl. Math. 2:482-89 (1981) e no J. Molec. Biol. 147:195-97 (1981); o método de alinhamento por homologia descrito por S.B. Needleman & C.D. Wunsch, J. Molec. Biol. 48(3):443-53 (Março de 1970); e os vários métodos descritos, por exemplo, por W.R. Pearson, em Genomics 11(3):635-50 (Novembro de 1991); por W.R. Pearson, em Methods Molec. Biol. 24:307-31 e 25:365-89 (1994); e por D.G. Higgins & P.M. Sharp, em Comp. Appl'ns in Biosci. 5:151-53 (1989) e em Gene 73(1):237-44 (15 de Dezembro de 1988).

A menos que estabelecido de outra maneira, GAP versão 10, que usa o algoritmo de Needleman e Wunsch (1970) *supra*, será usado para determinar a identidade de sequência ou a similaridade usando os seguintes parâmetros: % de identidade e % de similaridade para uma sequência de nucleotídeos usando "GAP Weight" de 50 e "Lenght Weight" de 3 e o escore de matriz nwsgapdna.cmp; % de identidade ou % de similaridade para uma sequência de aminoácidos usando "GAP weight" de 8 e "lenght weight" de 2 e o programa de escore BLOSUM62. Programas equivalentes também podem ser usados. Por "programa equivalente" é pretendido qualquer programa de comparação de sequência que, para quaisquer duas sequências em questão, gera um alinhamento que tem pareamentos de resíduos de nucleotídeos idênticos e um percentual idêntico de identidade de sequência quando comparado com o alinhamento correspondente gerado por GAP Versão 10. Em várias modalidades, a comparação de sequência é realizada através da totalidade da sequência em questão ou do indivíduo ou de ambas.

D. Condições de hibridização

Em outro aspecto da invenção, é fornecido um ácido nucleico que hibridiza com um ácido nucleico isolado com uma sequência que codifica um polipeptídeo com uma sequência substancialmente similar ao polipeptídeo de sinal de secreção de pbp*, dsbA, dsbC, Bce, CupA2, CupB2, CupC2, Nika, F1gl, ORF5550, Ttg2C ou ORF8124. Em certas modalidades,

o ácido nucleico hibridizante se ligará sob condições de alta estringência. Em várias modalidades, a hibridização ocorrerá, substancialmente, através da extensão completa da sequência de nucleotídeos que codifica o polipeptídeo de sinal de secreção, por exemplo, substancialmente através da extensão completa de uma ou mais de SEQ ID Nº 1, 3,5, 7, 9, 11, 13, 15, 17, 19, 21 ou 23. Uma molécula de ácido nucleico hibridiza com "substancialmente a extensão completa" de uma sequência de nucleotídeos que codifica um sinal de secreção aqui descrito quando a molécula de ácido nucleico hibridiza com pelo menos 80% da extensão completa de uma ou mais das SEQ ID Nº 1, 3,5, 7, 9, 11, 13, 15, 17, 19, 21 ou 23, pelo menos 85%, pelo menos 90% ou pelo menos 95% da extensão completa. A menos que especificado de outra maneira, "substancialmente a extensão completa" refere-se à pelo menos 80% da extensão completa da sequência de nucleotídeos que codifica um sinal de secreção onde a extensão é medida em nucleotídeos contíguos (por exemplo, hibridiza a pelo menos 53 nucleotídeos contíguos da SEQ ID Nº 3, a pelo menos 51 nucleotídeos contíguos da SEQ ID Nº 5, a pelo menos 80 nucleotídeos contíguos da SEQ ID Nº 7, etc.).

Em um método de hibridização, toda ou parte da sequência de nucleotídeos que codifica o polipeptídeo de sinal de secreção pode ser usada para rastrear cDNA ou bibliotecas genômicas. Métodos para construção de tais cDNA e bibliotecas genômicas são comumente conhecidos na técnica e estão descritos em Sambrook e Russell, 2001. As assim chamadas sondas de hibridização podem ser fragmentos de DNA genômico, fragmentos de cDNA, fragmentos de RNA ou outros oligonucleotídeos e podem ser marcadas com um grupo detectável tal como ^{32}P ou qualquer outro marcador detectável, tal como outros radioisótopos, um composto fluorescente, uma enzima ou um cofator de enzima. Sondas para hibridização podem ser feitas pela marcação de oligonucleotídeos sintéticos baseados na sequência de nucleotídeos que codifica o polipeptídeo de sinal de secreção aqui descrita. Iniciadores degenerados determinados com base em nucleotídeos conservados ou resíduos de aminoácidos da sequência de nucleotídeos ou sequência codificada de aminoácidos podem ser adicionalmente usados. A

sonda compreende tipicamente uma região da sequência de nucleotídeos que hibridiza sob condições estridentes com pelo menos cerca de 10, pelo menos cerca de 15, pelo menos cerca de 16, 17, 18, 19, 20 ou mais nucleotídeos consecutivos de uma sequência de nucleotídeos que codifica o polipeptídeo de sinal de secreção da invenção ou seu fragmento ou variante. Métodos para a preparação de sondas para hibridização são comumente conhecidos na técnica e estão descritos em Sambrook e Russell, 2001, incorporado aqui por referência.

Em técnicas de hibridização toda ou parte da sequência de nucleotídeos é usada como uma sonda que hibridiza, seletivamente, com outras sequências de nucleotídeos correspondentes presentes em uma população de fragmentos de DNA genômico ou fragmentos de cDNA clonados (isto é, bibliotecas genômicas ou de cDNA) de um organismo escolhido. As sondas de hibridização podem ser fragmentos de DNA genômico, fragmentos de cDNA, fragmentos de RNA ou outros oligonucleotídeos e podem ser marcadas com um grupo detectável tal como ^{32}P ou qualquer outro marcador detectável. Assim, por exemplo, as sondas para hibridização podem ser feitas pela marcação de oligonucleotídeos sintéticos baseados na sequência de nucleotídeos que codifica o polipeptídeo de sinal de secreção da invenção. Métodos para preparação de sondas para hibridização e para a construção de bibliotecas de cDNA e genômicas são comumente conhecidos na técnica e estão descritos em Sambrook *et al.* (1989) *Molecular Cloning: A Laboratory Manual* (2ª ed., Cold Spring Harbor Laboratory Press, Plainview, New York).

Por exemplo, a sequência de nucleotídeos que codifica o polipeptídeo de sinal de secreção completa aqui descrita ou uma ou mais de suas porções, pode ser usada como uma sonda capaz de hibridizar, especificamente, com sequências de nucleotídeos correspondentes e RNAs mensageiros que codificam polipeptídeos de sinal de secreção. Para obter a hibridização específica sob uma variedade de condições, tais sondas incluem sequências que são únicas e têm, preferivelmente, pelo menos cerca de 10 nucleotídeos de extensão ou pelo menos cerca de 15 nucleotídeos de ex-

tensão. Tais sondas podem ser usadas para amplificar sequências de nucleotídeos que codificam o polipeptídeo de sinal de secreção correspondentes a partir de um organismo escolhido por PCR. Essa técnica pode ser usada para isolar sequências codificadoras adicionais de um organismo desejado ou como um ensaio diagnóstico para determinar a presença de sequências codificadoras em um organismo. As técnicas de hibridização incluem rastreamento de hibridização de bibliotecas de DNA plaqueado (tantos placas quanto colônias; veja, por exemplo, Sambrook *et al.* (1989) *Molecular Cloning: A Laboratory Manual* (2ª ed., Cold Spring Harbor Laboratory Press, Plainview, New York).

A hibridização de tais sequências pode ser realizada sob condições estringentes. Por "condições estringentes" ou "condições estringentes de hibridização" são pretendidas condições sob as quais uma sonda hibridizará à sua sequência-alvo em um grau detectavelmente maior do que a outras sequências (por exemplo, pelo menos 2 vezes o antecedente). Condições estringentes são dependentes da sequência e serão diferentes em circunstâncias diferentes. Pelo controle da estringência de hibridização e/ou condições de lavagem, sequências-alvo que são 100% complementares a sonda podem ser identificadas (sondagem homóloga). Alternativamente, as condições de estringência podem ser ajustadas para permitir alguns não pareamentos nas sequências tal que graus inferiores de similaridade são detectados (sondagem heteróloga). Geralmente, um sonda tem menos do que 1000 nucleotídeos de extensão, preferivelmente menos do que 500 nucleotídeos de extensão.

Tipicamente, condições estringentes serão aquelas nas quais a concentração de sal é menor do que 1,5 M de íon de Na, tipicamente cerca de 0,01 a 1,0 M de concentração de íon de Na (ou outros sais) em pH 7.0 a 8.3 e a temperatura é pelo menos cerca de 60°C, preferivelmente cerca de 68°C. Condições estringentes também podem ser obtidas com a adição de agentes desestabilizadores tais como formamida. Condições de baixa estringência exemplares incluem hibridização com uma solução tampão de formamida 30 a 35%, NaCl 1 M, SDS 1% (dodecil sulfato de sódio) a 37°C e

uma lavagem em 1X a 2X SSC (20X SSC = NaCl 3,0 M/citrato trissódico 0,3 M) a 50 a 55°C. Condições de estringência moderada exemplares incluem hibridização em formamida 40 a 45%, NaCl 1,0 M, SDS 1% a 37°C e uma lavagem em 0,5X a 1X SSC a 55 a 60°C. Condições de alta estringência exemplares incluem hibridização em formamida 50%, NaCl 1,0 M, SDS 1% a 37°C e uma lavagem em 0,1X SSC a 60 a 68°C. Opcionalmente, os tampões de lavagem podem compreender cerca de 0,1% a cerca de 1% de SDS. A duração da hibridização é geralmente de menos do que cerca de 24 horas, geralmente cerca de 4 a cerca de 12 horas.

Especificidade é tipicamente a função de lavagens pós-hibridização, os fatores críticos sendo a força iônica e a temperatura da solução de lavagem final. Para híbridos de DNA-DNA, a T_m pode ser aproximada a partir da equação de Meinkoth e Wahl (1984) *Anal. Biochem.* 138:267-284: $T_m = 81,5^\circ\text{C} + 16,6 (\log M) + 0,41 (\%GC) - 0,61 (\%form) - 500/L$; onde M é a molaridade de cátions monovalentes, %GC é o percentual de nucleotídeos de guanosina e citosina no DNA, %forma é o percentual de formamida na solução de hibridização e L é a extensão do híbrido em pares de base. A T_m é a temperatura (sob força iônica e pH definidos) na qual 50% de uma sequência-alvo complementar hibridiza com uma sonda perfeitamente pareada. A T_m é reduzida em cerca de 1°C para cada 1% de não pareamento; assim, T_m , hibridização e/ou condições de lavagem podem ser ajustadas para hibridizar sequências da identidade desejada. Por exemplo, se são procuradas sequências com identidade $\geq 90\%$, a T_m pode ser diminuída 10°C. Geralmente, as condições estringentes são selecionadas para serem cerca de 5°C menores do que o ponto de fusão térmica (T_m) para a sequência específica e seu complemento em uma força iônica e pH definidos. Entretanto, condições severamente estringentes podem utilizar uma hibridização e/ou lavagem em 1, 2, 3 ou 4°C abaixo do ponto de fusão térmica (T_m); condições moderadamente estringentes podem utilizar uma hibridização e/ou lavagem em 6, 7, 8, 9 ou 10°C abaixo do ponto de fusão térmica (T_m); condições de baixa estringência podem utilizar uma hibridização e/ou lavagem em 11, 12, 13, 14, 15 ou 20°C abaixo do ponto de fusão térmica (T_m).

Usando a equação, a hibridização e as composições de lavagem e a T_m desejada aqueles versados na técnica entenderão que variações na estringência de hibridização e/ou soluções de lavagem estão inerentemente descritas. Se o grau de não pareamento desejado resulta em uma T_m menor do que 45°C (solução aquosa) ou 32°C (solução de formamida), é preferível aumentar a concentração de SSC tal que uma temperatura mais elevada possa ser usada. Um guia extenso para hibridização de ácidos nucleicos é encontrado em Tijssen (1993) *Laboratory Techniques in Biochemistry and Molecular Biology – Hybridization with Nucleic Acids Probes*, Parte I, Capítulo 2 (Elsevier, New York); e Ausubel *et al.*, Eds. (1995) *Current Protocols in Molecular Biology*, Capítulo 2 (Greene Publishing and Wiley-Interscience, New York). Veja Sambrook *et al.* (1989) *Molecular Cloning: A Laboratory Manual* (2ª ed., Cold Spring Harbor Laboratory Press, Plainview, New York).

E. Utilização de códon

As sequências de ácido nucleico aqui descritas podem ser ajustadas baseado na utilização de códon de um organismo hospedeiro. A utilização de códon ou preferência de códon é bem conhecida na técnica. A sequência codificadora selecionada pode ser modificada pela alteração do seu código genético para combinar com aquela empregada pela célula hospedeira bacteriana e a sua sequência de códon pode ser intensificada para melhor se aproximar daquela empregada pelo hospedeiro. A seleção de código genético e a intensificação da frequência do códon podem ser realizadas de acordo com qualquer um dos vários métodos conhecidos por aqueles versados na técnica, por exemplo, mutagênese direcionada ao sítio. Recursos úteis da Internet para auxiliar nesse processo incluem, por exemplo: (1) o Banco de Dados de Utilização de Códon do Kazusa DNA Research Institute (2-6-7 Kazusa-kamatari, Kisarazu, Chiba 292-0818 Japão) e disponível em www.kazusa.or.jp/codon; e (2) as Tabelas de Código Genético disponibilizadas pelo banco de dados de NCBI Taxonomy em www.ncbi.nlm.nih.gov/Taxonomy/Utils/wprintgc.cgi?mode=c. Por exemplo, espécies de *Pseudomonas* são relatadas como utilizando a Tabela 11 da Tradução do Código Genético do site NCBI Taxonomy e, no site da Kazusa como exibindo a frequência de utiliza-

ção de códon da tabela mostrada em www.kazusa.or.jp/codon/cgibin. É conhecido que a sequência codificadora tanto para o polipeptídeo de sinal de secreção, para o polipeptídeo de interesse descrito aqui em outra parte ou para ambos pode ser ajustada por utilização de códon.

5 F. Vetores de expressão

Outra modalidade da presente invenção inclui um vetor de expressão que inclui um ácido nucleico que codifica um novo polipeptídeo de secreção útil para direcionar uma proteína ou polipeptídeo de interesse operativamente ligados para o periplasma de bactérias Gram-negativas ou para
10 o espaço extracelular. Em uma modalidade, o vetor compreende uma sequência de polinucleotídeos que codifica um polipeptídeo que é substancialmente similar ao polipeptídeo de sinal de secreção aqui descrito, operativamente ligada a um promotor. Sequências codificadoras expressáveis serão operativamente acopladas a um promotor da transcrição capaz de funcionar na célula hospedeira escolhida, assim como todos os outros elementos regulatórios de transcrição e tradução necessários.
15

O termo "operativamente ligado" refere-se a qualquer configuração na qual os elementos regulatórios transcricional e traducional estão covalentemente acoplados a sequência codificadora em tal(tais) disposição(ões), em relação à sequência codificadora, que na e pela ação da célula hospedeira, os elementos regulatórios podem dirigir a expressão da sequência codificadora.
20

O vetor compreenderá tipicamente um ou mais marcadores fenotípicos selecionáveis e uma origem de replicação para assegurar a manutenção do vetor e, se desejável, fornecer a amplificação dentro do hospedeiro. Os hospedeiros adequados para transformação de acordo com a presente descrição incluem várias espécies dentro do gênero *Pseudomonas* e é, particularmente, preferida a célula hospedeira da cepa *P. fluorescens*.
25

Em uma modalidade, o vetor compreende ainda uma sequência codificadora para a expressão de uma proteína ou polipeptídeo de interesse, operativamente ligada a um sinal de secreção aqui descrito. As proteínas e polipeptídeos recombinantes podem ser expressos a partir de polinucleotí-
30

deos nos quais a sequência codificadora do polipeptídeo-alvo está operativamente ligada a uma sequência líder e a elementos regulatórios da transcrição e da tradução para formar um gene funcional a partir do qual a célula hospedeira pode expressar a proteína ou o polipeptídeo. A sequência codificadora pode ser uma sequência nativa para o polipeptídeo-alvo, se disponível, mas preferencialmente será uma sequência codificadora que foi selecionada, aperfeiçoada ou otimizada para uso na célula hospedeira de expressão selecionada: por exemplo, por sintetizar um gene para refletir o uso de códon preferencial de uma espécie hospedeira. Em uma modalidade da invenção, a espécie hospedeira é *P. fluorescens* e o códon preferencial de *P. fluorescens* é levado em consideração quando se planeja ambas a sequência de sinal e ou a sequência de proteína ou polipeptídeo. O(s) gene(s) é(são) construído(s) dentro ou inserido(s) em um ou mais vetores, que podem ser transformados na célula hospedeira de expressão.

Outros elementos regulatórios podem ser incluídos em um vetor (também denominado "construto de expressão"). Tais elementos incluem, mas não são limitados a, por exemplo, sequências intensificadoras transcricionais, sequências intensificadoras traducionais, outros promotores, ativadores, sinais de início e término da tradução, terminadores da transcrição, reguladores cistronicos, reguladores policistronicos, sequências tag, tais como sequência de nucleotídeos "tags" e sequências que codificam polipeptídeo "tag", que facilitam a identificação, separação, purificação e/ou isolamento de um polipeptídeo expresso.

Em outra modalidade, o vetor de expressão compreende ainda uma sequência tag adjacente a sequência codificadora para o sinal de secreção ou a sequência codificadora da proteína ou polipeptídeo de interesse. Em uma modalidade, essa sequência tag permite a purificação da proteína. A sequência tag pode ter um tag de afinidade, tal como um tag de afinidade de hexa-histidina. Em outra modalidade, o tag de afinidade pode ser uma molécula de glutathione-S-transferase. O tag também pode ser uma molécula fluorescente, tal como YFP ou GFP, ou análogos de tais proteínas fluorescentes. O tag também pode ser uma porção de uma molécula de um anti-

corpo ou um antígeno conhecido ou ligante para um parceiro de ligação conhecido útil para a purificação.

Um gene que codifica proteína de acordo com a presente invenção pode incluir, além da sequência codificadora da proteína, os seguintes

5 elementos regulatórios operativamente ligados a ele: um promotor, um sítio de ligação a ribossomo (RBS), um terminador da transcrição, sinais de início e término da tradução. RBSs úteis podem ser obtidos a partir de qualquer uma das espécies úteis como células hospedeiras em sistemas de expressão de acordo com a presente invenção, preferivelmente a partir da célula

10 hospedeira selecionada. Vários RBSs específicos e uma variedade de RBSs consenso são conhecidos, por exemplo, aqueles descritos e referenciados por D. Frishman *et al.*, Starts of bacterial genes: estimating the reliability of computer predictions, *Gene* 234(2):257-65 (8 de Julho de 1999); e B.E. Suzek *et al.*, A probabilistic method for identifying start códons in bacterial genomes, *Bioinformatics* 17(12):1123-30 (Dezembro de 2001). Além disso, tanto RBSs nativos quanto sintéticos podem ser usados, por exemplo, aqueles descritos em: EP 0207459 (RBSs sintéticos); O. Ikehata *et al.*, Primary structure of nitrile hydratase deduced from the nucleotide sequence of a *Rhodococcus* species and its expression in *Escherichia coli*, *Eur. J. Biochem.*

20 181(3):563-70 (1989) (sequência nativa AAGGAAG de RBS). Exemplos de métodos adicionais, vetores e elementos de tradução e transcrição e outros elementos úteis na presente invenção estão descritos, por exemplo, em: Pat. U.S. Nº 5.055.294 de Gilroy e Pat. U.S. Nº 5.128.130 de Gilroy *et al.*; Pat. U.S. Nº 5.281.532 de Rammner *et al.*; Pat. U.S. Nº 4.695.455 e 4.861.595 de

25 Barnes *et al.*; Pat. U.S. Nº 4.755.465 de Gray *et al.*; e Pat. U.S. Nº 5.169.760 de Wilcox.

A transcrição do DNA que codifica as proteínas da presente invenção é aumentada pela inserção de uma sequência intensificadora no vetor ou no plasmídeo. Intensificadores típicos são elementos cis-atuantes de

30 DNA, geralmente com cerca de 10 a 300 bp de tamanho que atuam sobre o promotor para aumentar sua transcrição. Exemplos incluem vários intensificadores de *Pseudomonas*.

Geralmente, os vetores de expressão recombinantes incluirão origens de replicação e marcadores selecionáveis que permitem a transformação da célula hospedeira e um promotor derivado de um gene altamente expresso para dirigir a transcrição de uma sequência estrutural à jusante.

- 5 Tais promotores podem ser derivados de óperons que codificam enzimas tais como 3-fosfoglicerato quinase (PGK), fosfatase ácida ou proteínas de choque térmico, entre outras. A sequência estrutural heteróloga é acoplada na fase apropriada com as sequências de iniciação e término da tradução e, preferivelmente, a sequência de secreção capaz de direcionar a secreção do
- 10 polipeptídeo traduzido. Opcionalmente, a sequência heteróloga pode codificar um polipeptídeo de fusão que inclui um polipeptídeo de identificação N-terminal transmitindo as características desejadas, por exemplo, estabilização ou purificação simplificada do produto recombinante expresso.

- Vetores são reconhecidos na técnica por expressar proteínas
- 15 recombinantes em células hospedeiras e qualquer um desses pode ser usado para expressar os genes de acordo com a presente invenção. Tais vetores incluem, por exemplo, plasmídeos, cosmídeos e vetores de expressão em fago. Exemplos de vetores de plasmídeo úteis incluem, mas não são limitados aos plasmídeos de expressão PBBR1MCS, pDSK519, pKT240, p-
- 20 ML122, pPS10, RK2, RK6, pRO1600 e RSF1010. Outros exemplos de tais vetores úteis incluem aqueles descritos em, por exemplo: N. Hayase, em *Appl. Envir. Microbiol.* 60(9):3336-42 (Setembro de 1994); A.A. Lushnikov *et al.*, em *Basic Life Sci.* 30:657-62 (1985); S. Graupner & W. Wackemagel, em *Biomolec. Eng.* 17(1):11-16 (outubro de 2000); H.P. Schweizer, em *Curr.*
- 25 *Opin. Biotech.* 12(5):439-45 (Outubro de 2001); M. Bagdasarian & K.N. Timmis, em *Curr. Topic Microbiol. Immunol.* 96:47-67 (1982); T. Ishii *et al.*, em *FEMS Microbiol. Lett.* 116(3):307-13 (1 de Março de 1994); I.N. Oleknovich & Y.K. Fomichev, em *Gene* 140(1):63-65 (11 de Março de 1994); M. Tsuda & T. Nakazawa, em *Gene* 136(1-2):257-62 (22 de Dezembro de 1993); C. Nieto *et al.*, em *Gene* 87(1):145-49 (1 de Março de 1990); J. D. Jones & N. Guttererson, em *Gene* 61(3):299-306 (1987); M. Bagdasarian *et al.*, em *Gene*
- 30 16(1-3):237-47 (dezembro de 1981); H.P. Schweizer *et al.*, em *Genet. Eng.*

(NY) 23:69-81 (2001); P. Mukhopadhyay *et al.*, em J. Bact. 172(1):477-80 (Janeiro de 1990); D.O. Wood *et al.*, em J. Bact. 145(3):1448-51 (Março de 1981); e R. Holtwick *et al.*, em Microbiology 147(Pt 2):337-44 (Fevereiro de 2001).

- 5 Exemplos adicionais de vetores de expressão que podem ser úteis em uma célula hospedeira que compreende os construtos de sinal de secreção da invenção incluem aqueles listados na Tabela 2 como derivados a partir de réplicons indicados.

Tabela 2. Exemplos de Vetores de Expressão Úteis	
Réplicon	Vetor(es)
PPS10	PCN39, PCN51
RSF1010	PKT261-3
	PMMB66EH
	PEB8
	PPLGNI
	PMYC1050
RK2/RP1	PRK415
	PJB653
PRO1600	PUCP
	PBSP

- 10 O plasmídeo de expressão, RSF1010, está descrito, por exemplo, em F. Heffron *et al.*, em Proc. Natl. Acad. Sci. USA 72(9):3632-27 (Setembro de 1975) e por K. Nagahari & K. Sakaguchi, em J. Bact. 133(3):1527-29 (Março de 1978). O plasmídeo RSF1010 e seus derivados são vetores particularmente úteis na presente invenção. Derivados exemplares, úteis de RSF1010, que são conhecidos na técnica, incluem, por exemplo, pKT212,
- 15 pKT214, pKT231 e plasmídeos relacionados e pMYC1050 e plasmídeos relacionados (veja, por exemplo, Pat. U.S. Nº 5.527.883 e 5.840.554 de Thompson *et al.*), tal como, por exemplo, pMYC1803. O plasmídeo pMYC1803 é derivado do plasmídeo pTJ260 baseado em RSF1010 (veja Pat. U.S. Nº 5.169.760 de Wilcox), que carrega um marcador de resistência a tetraciclina regulado e os
- 20 loci de replicação e mobilização do plasmídeo RSF1010. Outros vetores e-

xemplares úteis incluem aqueles descritos na Pat. U.S. Nº 4.680.264 de Puhler *et al.*

Em uma modalidade, um plasmídeo de expressão é usado como o vetor de expressão. Em outra modalidade, RSF1010 ou seu derivado é
5 usado como o vetor de expressão. Em outra modalidade ainda, pMYC1050 ou seu derivado ou pMYC4803 ou seu derivado são usados como vetor de expressão.

O plasmídeo pode ser mantido na célula hospedeira pela inclusão de um gene marcador de seleção no plasmídeo. Esse pode ser um gene
10 de resistência a antibiótico, onde o(s) antibiótico(s) correspondente(s) é(são) adicionado(s) ao meio de fermentação ou qualquer outro tipo de gene marcador de seleção conhecido na técnica, por exemplo, um gene que restaura a prototrofia onde um plasmídeo é usado em uma célula hospedeira que é auxotrófica para a característica correspondente, por exemplo, uma caracte-
15 rística biocatalítica tal como uma biosíntese de aminoácido ou uma característica de biosíntese de nucleotídeo, ou uma característica de utilização de fonte de carbono.

Os promotores usados de acordo com a presente invenção podem ser promotores constitutivos ou promotores regulados. Exemplos co-
20 muns de promotores regulados úteis incluem aqueles da família derivado do promotor lac (isto é, o promotor lacZ), especialmente os promotores tac e trc descritos na Pat. U.S. Nº 4.551.433 de DeBoer, assim como os promotores Ptac16, Ptac17, PtacII, PlacUV5 e T7lac. Em uma modalidade, o promotor não é derivado do organismo da célula hospedeira. Em certas modalidades,
25 o promotor é derivado de um organismo de *E.coli*.

Exemplos comuns de promotores tipo não-lac úteis em sistemas de expressão de acordo com a presente invenção incluem, por exemplo, aqueles listados na Tabela 3.

Tabela 3. Exemplos de Promotores não-lac	
Promotor	Indutor
P _R	Alta temperatura
P _L	Alta temperatura

Tabela 3. Exemplos de Promotores não- <i>lac</i>	
Pm	Alquila ou halo-benzoatos
Pu	Alquila ou halo-toluenos
Psal	Salicilatos

Veja, por exemplo: J. Sanchez-Romero & V. De Lorenzo (1999) Genetic Engineering of Nonpathogenic *Pseudomonas* strains as Biocatalysts for Industrial and Environmental Processes, em *Manual of Industrial Microbiology and Biotechnology* (A. Demain & J. Davies, eds) pp. 460-74 (ASM Press, Washington, D.C.); H. Schweizer (2001) Vectors to express foreign genes and techniques to monitor gene expression for *Pseudomonads*, *Current Opinion in Biotechnology*, 12:439-445; e R. Slater e J. Williams (2000) The Expression of Foreign DNA in Bactéria, em *Molecular Biology and Biotechnology* (J. Walker & R. Rapley, eds.) pp.125-54 (The Royal Society of Chemistry, Cambridge, UK)). Um promotor que tem a sequência de nucleotídeos de um promotor nativo da célula hospedeira bacteriana selecionada também pode ser usado para controlar a expressão do transgene que codifica o polipeptídeo-alvo, por exemplo, um promotor de operon de antranilato ou benzoato de *Pseudomonas* (Pant, Pben). Promotores em tandem também podem ser usados nos quais mais do que um promotor é covalentemente acoplado ao outro, ou igual ou diferente em sequência, por exemplo, um promotor em tandem Pant-Pben (híbrido interpromotor) ou um promotor em tandem Plac-Plac, ou derivados do mesmo ou de organismos diferentes.

Promotores regulados utilizam proteínas regulatórias de promotor a fim de controlar a transcrição de um gene do qual o promotor é uma parte. Onde um promotor regulado é usado aqui, uma proteína regulatória de promotor correspondente também será parte de um sistema de expressão de acordo com a presente invenção. Exemplos de proteínas regulatórias de promotor incluem: proteínas ativadoras, por exemplo, proteína ativadora de catabolito de *E. coli*, proteína MalT; ativadoras transcricionais da família AraC; proteínas repressoras, por exemplo, proteínas LacI de *E. coli*; e proteínas regulatórias com função dupla, por exemplo, proteína NagC de *E. coli*. Vários pares de promotor regulado/proteína regulatória de promotor são co-

nhecidos na técnica.

As proteínas regulatórias de promotor interagem com um composto efetor, isto é, um composto que se associa, reversível ou irreversivelmente, com a proteína regulatória de maneira a possibilitar que a proteína
 5 tanto libere quanto se ligue a pelo menos uma região regulatória da transcrição de DNA do gene que está sob o controle do promotor, permitindo ou bloqueando desse modo a ação de uma enzima transcriptase de iniciar a transcrição do gene. Compostos efetores são classificados tanto como indutores quanto como correpressores e esses compostos incluem compostos
 10 efetores nativos e compostos indutores gratuitos. Vários tríos de promotor regulado/proteína regulatória de promotor/composto efetor são conhecidos na técnica. Embora um composto efetor possa ser usado durante a cultura celular ou a fermentação, em uma modalidade preferida na qual um promotor regulado é usado, após o crescimento de uma quantidade ou densidade
 15 desejadas de biomassa de célula hospedeira, um composto efetor apropriado é adicionado à cultura para resultar, direta ou indiretamente, na expressão do(s) gene(s) desejado(s) que codifica(m) a proteína ou polipeptídeo de interesse.

Como exemplo, onde um promotor da família *lac* é utilizado, um
 20 gene *lacI* também está presente no sistema. O gene *lacI*, que é (normalmente) um gene expresso constitutivamente, codifica a proteína repressora *Lac* (proteína *LacD*) que se liga ao operador *lac* desses promotores. Assim onde um promotor da família *lac* é utilizado, o gene *lacI* também pode estar incluído e expresso no sistema de expressão. No caso de membros da família do promotor *lac*, por exemplo, o promotor *tac*, o composto efetor é um indutor, pre-
 25 ferivelmente um indutor gratuito tal como IPTG (isopropil-D-1-tiogalactopiranosídeo, também chamado de "isopropiltiogalactosídeo").

Para expressão de uma proteína ou polipeptídeo de interesse, qualquer promotor de planta também pode ser usado. Um promotor pode ser
 30 um promotor de RNA polimerase II de planta. Os elementos incluídos em promotores de planta podem ser um TATA Box ou Goldberg-Hogness box, tipicamente posicionado a aproximadamente 25 a 35 pares de bases à mon-

tante (5') do sítio de iniciação da transcrição e o CCAAT box, localizado entre 70 e 100 pares de bases à montante. Em plantas, CCAAT box pode ter uma sequência consenso diferente daquela sequência funcionalmente análoga de promotores de mamífero (Messing *et al.*, (1983) *Em: Genetic Engineering of Plants*, Kosuge *et al.*, Eds., pp.211-227). Além disso, virtualmente todos os promotores incluem sequências ativadoras à montante adicionais ou intensificadores (Benoist e Chambon (1981) *Nature* 290:304-310; Gruss *et al.* (1981) *Proc. Nat. Acad. Sci.* 78:943-947; e Khoury e Gruss (1983) *Cell* 27:313-314) que se estendem em torno de 100 bp a 1000 bp ou mais à montante do sítio de iniciação da transcrição.

G. Sistemas de Expressão

A presente invenção fornece ainda um sistema de expressão aperfeiçoado útil para direcionar uma proteína ou polipeptídeo de interesse operativamente ligados para o periplasma de bactérias Gram-negativas ou para o espaço extracelular. Em uma modalidade, o sistema inclui uma célula hospedeira e um vetor descritos acima que compreendem uma sequência de nucleotídeo que codifica uma proteína ou polipeptídeo de interesse operativamente ligados a um sinal de secreção selecionado do grupo consistindo em sequência de sinal de secreção de pbp*, dsbA, dsbC, Bce, CupA2, CupB2, CupC2, Nika, F1gl, ORF5550, Ttg2C ou ORF8124 ou uma sequência que é substancialmente homóloga a sequência de sinal de secreção descrita aqui como SEQ ID N° 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21 ou 23 ou uma sequência de nucleotídeos que codifica as SEQ ID N° 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22 ou 24. Em algumas modalidades, não são feitas modificações entre a sequência de sinal e a proteína ou polipeptídeo de interesse. Entretanto, em certas modalidades, sinais de clivagem adicionais são incorporados para promover o processamento correto da terminação amino do polipeptídeo.

O sistema de secreção também pode incluir um meio de fermentação, tal como descrito abaixo. Em uma modalidade, o sistema inclui um meio com sais minerais. Em outra modalidade, o sistema inclui um indutor químico no meio.

O sistema de expressão CHAMPION®pET fornece um nível alto de produção de proteína. A expressão é induzida a partir do promotor forte T7lac. Esse sistema tira vantagem da alta atividade e especificidade da RNA polimerase do bacteriófago T7 para um alto nível de transcrição do gene de interesse. O operador lac localizado na região promotora fornece uma regulação mais estreita do que os vetores baseados em T7 tradicionais, melhorando a estabilidade do plasmídeo e a viabilidade celular (Studier e Moffatt (1986) *J Molecular Biology* 189(1):113-30; Rosenberg *et al.* (1987) *Gene* 56(1):125-35). O sistema de expressão T7 usa o promotor T7 e T7 RNA polimerase (T7 RNAP) para um alto nível de expressão por que T7 RNAP é mais processável do que RNAP nativo de *E. coli* e é dedicado a transcrição do gene de interesse. A expressão do gene identificado é induzida pelo fornecimento de uma fonte de T7 RNAP na célula hospedeira. Isso é obtido pelo uso de *E. coli* BL21 hospedeira que contém uma cópia cromossômica do gene de T7 RNAP. O gene de T7 RNAP está sob o controle do promotor lacUV5 que pode ser induzido por IPTG. T7 RNAP é expresso após a indução e transcreve o gene de interesse.

O sistema de expressão pBAD permite uma expressão estreitamente controlada, titulável da proteína ou polipeptídeo de interesse através da presença de fontes de carbono específicas tais como glicose, glicerol e arabinose (Guzman *et al.*(1995) *J. Bacteriology* 177(14):4121-30). Os vetores pBAD são designados exclusivamente para dar o controle preciso sobre os níveis de expressão. A expressão de gene heterólogo a partir de vetores de pBAD é iniciada no promotor araBAD. O promotor é regulado positiva e negativamente pelo produto do gene de araC. AraC é um regulador transcricional que forma um complexo com L-arabinose. Na ausência de L-arabinose, o dímero AraC bloqueia a transcrição. Para a ativação transcricional máxima dois eventos são necessários: (i) L-arabinose se liga a AraC permitindo o início da transcrição; (ii) o complexo cAMP ativador de proteína (CAP)-cAMP se liga ao DNA e estimula a ligação de AraC na localização correta da região promotora.

O sistema de expressão *trc* permite a expressão em *E. coli* em

alto nível, regulada a partir do promotor *trc*. Os vetores de expressão *trc* foram otimizados para a expressão de genes eucarióticos em *E. coli*. O promotor *trc* é um promotor híbrido forte derivado dos promotores de triptofano (*trp*) e de lactose (*lac*). Ele é regulado pelo operador *lacO* e é produto do gene *lacI*Q (Brosius, J. (1984) *Gene* 27(2):161-72).

A transformação das células hospedeiras com o(s) vetor(es) descrito(s) aqui pode ser realizada usando qualquer metodologia de transformação conhecida na técnica e as células hospedeiras bacterianas podem ser transformadas como células intactas ou como protoplastos (isto é, incluindo citoplastos). Metodologias de transformação exemplares incluem metodologias de poração, por exemplo, eletroporação, fusão de protoplasto, conjugação bacteriana e tratamento com cátion divalente, por exemplo tratamento com cloreto de cálcio ou com $\text{CaCl}_2/\text{Mg}^{2+}$ ou outros métodos bem conhecidos na técnica. Veja, por exemplo, Morrison, *J. Bact.*, 132:349-351 (1977); Clark-Curtiss & Curtiss, *Methods in Enzimology*, 101:347-362 (Wu *et al.*, eds, 1983), Sambrook *et al.*, *Molecular Cloning, A Laboratory Manual* (2ª Ed. 1989); Kriegler, *Gene Transfer and Expression: A Laboratory Manual* (1990); e *Currents Protocols in Molecular Biology* (Ausubel *et al.*, eds., 1994)).

20 H. Célula hospedeira

Em uma modalidade a invenção fornece um sistema de expressão útil para direcionar uma proteína ou polipeptídeo de interesse operativamente ligados para o periplasma de bactérias Gram-negativas ou para o espaço extracelular. Em uma modalidade, esse sistema utiliza um peptídeo de sinal de secreção. Em outra modalidade, o sistema de expressão é um sistema de expressão em *P. fluorescens* para expressão de uma proteína que compreende um sinal de secreção aqui descrito. Esse aspecto da invenção é fundamentado na surpreendente descoberta de que *P. fluorescens* é capaz de processar apropriadamente e direcionar sinais de secreção para ambos os sistemas em *P. fluorescens* e não *P. fluorescens*.

Nessa modalidade, a célula hospedeira pode ser selecionada do "Subgrupo 18 de Proteobactérias Gram-negativas". O "Subgrupo 18 de Pro-

teobactérias Gram-negativas" é definido como o grupo de todas as espécies, variedades, cepas e outras unidades subespeciais da espécie de *Pseudomonas fluorescens*, incluindo aquelas que pertencem, por exemplo, aos seguintes (com os números de depósito na ATCC ou em outros das cepas e-
 5 xemplares mostrados entre parênteses):

Pseudomonas fluorescens biotipo A, também chamada de biovar 1 ou biovar I (ATCC 13525);

Pseudomonas fluorescens biotipo B, também chamada de biovar 2 ou biovar II (ATCC 17816);

10 *Pseudomonas fluorescens* biotipo C, também chamada de biovar 3 ou biovar III (ATCC 17400);

Pseudomonas fluorescens biotipo F, também chamada de biovar 4 ou biovar IV (ATCC 12983);

15 *Pseudomonas fluorescens* biotipo G, também chamada de biovar 5 ou biovar V (ATCC 17518);

Pseudomonas fluorescens biovar VI; *Pseudomonas fluorescens* Pf0-1; *Pseudomonas fluorescens* Pf-5 (ATCC BAA-477); *Pseudomonas fluorescens* SBW25; e *Pseudomonas fluorescens* subespécie cellulosa (NCIMB 10462).

20 A célula hospedeira pode ser selecionada do "Subgrupo 19 de Proteobactérias Gram-negativas". O "Subgrupo 19 de Proteobactérias Gram-negativas" é definido como o grupo de todas as cepas de *Pseudomonas fluorescens* biótipo A. Uma cepa desse tipo particularmente preferida é a cepa MB101 de *P. fluorescens* (veja Pat. U.S. Nº 5.169.760 de Wilcox) e seus de-
 25 rivados. Um exemplo de um derivado preferido dessa é a cepa MB214 de *P. fluorescens*, construído pela inserção no locus *asd* (gene aspartato desidrogenase) cromossômico de MB101, um construto nativo de *E. coli* PlacI-lacI-lacZYA (isto é, no qual PlacZ foi deletado).

Cepas adicionais de *P. fluorescens* que podem ser usadas na
 30 presente invenção incluem *Pseudomonas fluorescens* Migula e *Pseudomonas fluorescens* Loitokitok, que têm as seguintes designações da ATCC: [NCIB 8286]; NRRL B-1244; NCIB 8865 cepa CO1; NCIB 8866 cepa CO₂;

1291 [ATCC 17458; IFO 15837; NCIB 8917; LA; NRRL B-1864; pirrolidina; PW2 [ICMP 3966; NCPPB 967; NRRL b-899]; 13475; NCTC 10038; NRRL B-1603 [6; IFO 15840]; 52-1C; CCEB 488-A [BU 140]; CCEB 553 [EM 15/47]; IAM 1008 [AHH-27]; IAM 1055 [AHH-23]; 1 [IFO 15842]; 12 [ATCC 5 25323; NIH 11; den Dooren de Jong 216]; 18 [IFO 15833; WRRL P-7]; 93 [TR-10]; 108 [52-22; IFO 15832]; 143 [IFO 15836; PL]; 149 [2-40-40; IFO 15838]; 182 [IFO 3081; PJ 73]; 184 [IFO 15830]; 185 [W2 L-1]; 186 [IFO 15829; PJ79]; 187 [NCPPB 263]; 188 [NCPPB 316]; 189 [PJ227; 1208]; 191 [IFO 15834; PJ 236; 22/1]; 194 [Klinge R-60; PJ 253]; 196 [PJ 288]; 197 [PJ 10 290]; 198 [PJ 302]; 201 [PJ 368]; 202 [PJ 372]; 203 [PJ 376]; 204 [IFO 15835; PJ 682]; 205 [PJ 686]; 206 [PJ 692]; 207 [PJ 693]; 208 [PJ 722]; 212 [PJ 832]; 215 [PJ 849]; 216 [PJ 885]; 267 [B-9]; 271 [B-1612]; 401 [C71A; IFO 15831; PJ 187]; NRRL B-3178 [4; IFO. 15841]; KY 8521; 3081; 30-21; [IFO 3081]; N; PYR; PW; D946-B83 [BU 2183; FERM-P 3328]; P-2563 15 [FERM-P 2894; IFO 13658]; IAM-1126 [43F]; M-1; A506 [A5-06]; A505 [A5-05-1]; A526 [A5-26]; B69; 72; NRRL B-4290; PMW6 [NCIB 11615]; SC 12936; A1 [IFO 15839]; F 1847 [CDC-EB]; F 1848 [CDC 93]; NCIB 10586; P17; F-12; AmMS 257; PRA25; 6133D02; 6519E01; Ni; SC15208; BNL-WVC; NCTC 2583 [NCIB 8194]; H13; 1013 [ATCC 11251; CCEB 295]; IFO 20 3903; 1062; ou Pf-5.

Em uma modalidade, a célula hospedeira pode ser qualquer célula capaz de produzir uma proteína ou polipeptídeo de interesse, incluindo uma célula de *P. fluorescens* como descrita acima. Os sistemas mais comumente usados para produzir proteínas ou polipeptídeos de interesse incluem certas células bacterianas; particularmente *E. coli*, devido as suas necessidades de cultivo relativamente econômicas e a capacidade potencial de produzir proteína em grandes lotes de cultura. Leveduras também podem ser usadas para expressar proteínas e polipeptídeos biologicamente relevantes, particularmente para propósitos de pesquisa. Os sistemas incluem *Saccharomyces cerevisiae* ou *Pichia pastoris*. Esses sistemas são bem caracterizados, fornecem geralmente níveis aceitáveis de expressão de proteína total e são comparativamente rápidos e econômicos. Sistemas de expressão 30

em célula de inseto também têm surgido como uma alternativa para expressar proteínas recombinantes em forma biologicamente ativa. Em alguns casos, proteínas corretamente enoveladas que são modificadas após a tradução podem ser produzidas. Sistemas de expressão em célula de mamífero, tais como células de ovário de hamster chinês, também têm sido usados para expressão de proteínas ou polipeptídeos de interesse. Em pequena escala, esses sistemas de expressão são geralmente eficazes. Certos biológicos podem ser derivados de proteínas, particularmente em aplicações para saúde animal e humana. Em outra modalidade, a célula hospedeira é uma célula de planta incluindo, mas não limitada a uma célula de tabaco, milho, uma célula de uma espécie de *Arabidopsis*, célula de batata ou arroz. Em outra modalidade, um organismo multicelular é analisado ou é modificado no processo, incluindo mas não limitado a um organismo transgênico. As técnicas para analisar e/ou modificar um organismo multicelular são geralmente baseadas em técnicas descritas para modificar células descritas abaixo.

Em outra modalidade, a célula hospedeira pode ser um procarioto tal como uma célula bacteriana incluindo, mas não limitado a uma espécie de *Escherichia* ou uma *Pseudomonas*. Células bacterianas típicas são descritas, por exemplo, em "Biological Diversity: Bactéria and Archaeans", um capítulo do On-Line Biology Book, fornecido pelo Dr M J Farabee do Estrella Mountain Community College, Arizona, USA no site da web www.emc.maricotpa.edu/faculty/farabee/BIOBK/BioBookDiversity. Em certas modalidades, a célula hospedeira pode ser uma célula de *Pseudomonad* e pode tipicamente ser uma célula de *P. fluorescens*. Em outras modalidades, a célula hospedeira também pode ser uma célula de *E.coli*. Em outra modalidade, a célula hospedeira pode ser uma célula eucariótica, por exemplo uma célula de inseto, incluindo mas não limitada a uma célula de *Spodoptera*, *Trichoplusia*, *Drosophila* ou uma espécie de *Estigmene*, ou uma célula de mamífero, incluindo mas não limitada a uma célula de murino, célula de hamster, uma célula de macaco, primata ou humano.

Em uma modalidade, a célula hospedeira pode ser um membro de qualquer táxon bacteriano. A célula pode, por exemplo, ser um membro

de qualquer espécie de eubactérias. O hospedeiro pode ser qualquer um do táxon de: Acidobactéria, Actinobactéria, Aquificae, Bacteroidetes, Chlorobi, Clamídia, Chloroflexi, Chrysiogenetes, Cyanobactéria, Deferribacteres, Deinococcus, Dictyoglomi, Fibrobacteres, Firmicutes, Fusobactéria, Gemmatimonadetes, Lentisphaerae, Nitrospirae, Planctomycetes, Proteobactéria, Spirochaetes, Thermodesulfobactéria, Thermomicrobia, Thermotogae, Thermus (Thermales) ou Verrucomicrobia. Em uma modalidade de uma célula hospedeira eubacteriana, a célula pode ser um membro de qualquer espécie de eubactérias, excluindo Cianobactéria.

10 O hospedeiro bacteriano também pode ser um membro de qualquer espécie de Proteobactéria. Uma célula hospedeira proteobacteriana pode ser um membro de qualquer um dos táxons de Alfaproteobactérias, Betaproteobactérias, Gamaproteobactérias, Deltaproteobactérias ou Epsilonproteobactérias. Além disso, o hospedeiro pode ser um membro de qualquer um dos táxons de Alfaproteobactérias, Betaproteobactérias ou Gamaproteobactérias e um membro de qualquer espécie de Gamaproteobactérias.

15 Em uma modalidade de um hospedeiro de Gamaproteobactérias, o hospedeiro será um membro de qualquer um dos táxons de Aeromonadales, Alteromonadales, Enterobacteriales, Pseudomonadales ou Xanthomonadales; ou um membro de qualquer espécie de Enterobacteriales ou Pseudomonadales. Em uma modalidade, a célula hospedeira pode ser da ordem das Enterobacteriales, a célula hospedeira pode ser da família Enterobacteriaceae ou pode ser um membro do gênero Erwinia, Escherichia, ou Serratia; ou um membro do gênero Escherichia. Onde a célula hospedeira é da ordem Pseudomonadales, a célula hospedeira pode ser um membro da família Pseudomonadaceae, incluindo o gênero Pseudomonas. Hospedeiros de Gamaproteobactérias incluem membros da espécie Escherichia coli e membros da espécie *Pseudomonas fluorescens*.

20 Outros organismos de Pseudomonas também podem ser úteis. 30 Pseudomonads e espécies intimamente relacionadas incluem o as Proteobactérias Gram-negativas do Subgrupo 1, que inclui o grupo de Proteobactérias que pertence às famílias e/ou gêneros descritos como "Gram-Negative

Aerobic Rods and Cocci" por R.E. Buchanan e N.E. Gibbons (Eds.), Bergey's Manual of Determinative Bacteriology, pp 217-289 (8ª Ed., 1974) (The Williams & Wilkins Co., Baltimore, Md., USA) (daqui por diante "Bergey (1974)").

A Tabela 4 apresenta essas famílias e gêneros de organismos.

Tabela 4. Famílias e Gêneros Listados na Parte, "Gram-Negative Aerobic Rods and Cocci" (em Bergey (1974))	
Família I. Pseudomonaceae	Gliconobacter Pseudomonas Xanthomonas Zoogloea
Família II. Azotobacteraceae	Azomonas Azotobacter Beijerinckia Derxia
Família III. Rhizobiaceae	Agrobacterium Rhizobium
Família IV. Methylomonadaceae	Methylococcus Methylomonas
Família V. Halobacteriaceae	Halobacterium Halococcus
Outros Generos	Acetobacter Alcaligenes Bordetella Brucella Francisella Thermus

5 As "Proteobactérias Gram-Negativas do Subgrupo 1" também incluem Proteobactérias que poderiam ser classificadas tópico de acordo com os critérios usados na classificação. O tópico também inclui grupos que foram previamente classificados nessa seção mas que não são mais, tais como os gêneros Acidovorax, Brevundimonas, Burkholderia, Hydrogenophaga, Oceanimonas, Ralstonia e Stenotrophomonas, o gênero Spingomonas

10

(e o gênero *Blastomonas*, derivado dele), que foram criados pelo reagrupamento de organismos que pertencem (e previamente chamados espécies de) ao gênero *Xantomonas*, ao gênero *Acidomonas*, que foi criado pelo reagrupamento de organismos que pertencem ao gênero *Acetobacter* como definido em Bergey (1974). Além disso, hospedeiros também podem incluir células do gênero *Pseudomonas*, *Pseudomonas enalia* (ATCC 14393), *Pseudomonas nigrifaciens* (ATCC 19375) e *Pseudomonas putrefaciens* (ATCC 8071), que foram reclassificadas respectivamente como *Alteromonas haloplanktis*, *Alteromonas nigrifaciens* e *Alteromonas putrefaciens*. Similarmente, por exemplo, *Pseudomonas acidovorans* (ATCC 15668) e *Pseudomonas testosteroni* (ATCC 11996) foram reclassificadas como *Comamonas acidovorans* e *Comamonas testosteroni*, respectivamente; e *Pseudomonas nigrifaciens* (ATCC 19375) e *Pseudomonas piscicida* (ATCC 15057) foram reclassificadas como *Pseudoalteromonas nigrifaciens* e *Pseudoalteromonas piscicida*. "Proteobactérias Gram-Negativas do Subgrupo 1" também incluem as Proteobactérias classificadas como pertencendo a qualquer uma das famílias: Pseudomonadaceae, Azotobacteraceae (comumente chamada agora pelo sinônimo, o "grupo Azotobacter" de Pseudomonadaceae), Rhizobiaceae e Methylomonadaceae (comumente chamada agora pelo sinônimo, "Methylococcaceae"). Consequentemente, além daqueles gêneros descritos aqui de outra maneira, os gêneros Proteobacterianos que estão dentro de "Proteobactérias Gram-Negativas do Subgrupo 1" incluem: 1) Bactérias do grupo Azotobacter do gênero *Azorhizobium*; 2) bactérias da família Pseudomonadaceae dos gêneros *Cellvibrio*, *Oligella* e *Teredinibacter*; 3) bactérias da família Rhizobiaceae dos gêneros *Chelatobacter*, *Ensifer*, *Liberibacter* (também chamada de "Candidatus *Liberibacter*") e *Sinorhizobium*; e 4) bactérias da família Methylococcaceae dos gêneros *Methylobacter*, *Methylocaldum*, *Methylobacterium*, *Methylosarcina* e *Methylosphaera*.

Em outra modalidade, a célula hospedeira é selecionada do "Subgrupo 2 de Proteobactérias Gram-Negativas". O "Subgrupo 2 de Proteobactérias Gram-Negativas" é definido como o grupo de Proteobactérias dos seguintes gêneros (com os números de catálogo listados, disponíveis publi-

camente, suas cepas depositadas indicadas entre parênteses, todas depositadas na ATCC, exceto como indicado de outra maneira): *Acidomonas* (2); *Acetobacter* (93); *Gliconobacter* (37); *Brevundimonas* (23); *Beyerinckia* (13); *Derxia* (2); *Brucella* (4); *Agrobacterium* (79); *Chelatobacter* (2); *Ensifer* (3);
 5 *Rhizobium* (144); *Sinorhizobium* (24); *Blastomonas* (1); *Sphingomonas* (27); *Alcaligenes* (88); *Bordetella* (43); *Burkholderia* (73); *Ralstonia* (33); *Acidovorax* (20); *Hydrogenophaga* (9); *Zoogloea* (9); *Methylobacter* (2); *Methylocaldum* (1 no NCIMB); *Methylococcus* (2); *Methylomicrobium* (2); *Methylomonas* (9); *Methylosarcina* (1); *Methylosphaera*; *Azomonas* (9); *Azorhizophilus* (5);
 10 *Azotobacter* (64); *Cellvibrio* (3); *Oligella* (5); *Pseudomonas* (1139); *Francisella* (4); *Xanthomonas* (229); *Stenotrophomonas* (50) e *Oceanimonas* (4).

Exemplares de espécies de células hospedeiras do " Subgrupo 2 de Proteobactérias Gram-Negativas" incluem, mas não estão limitados as seguintes bactérias (com os números de depósito da ATCC ou de outros das suas cepas exemplares mostrados entre parênteses): *Acidomonas methanolic*
 15 *lica* (ATCC 43581); *Acetobacter acetii* (ATCC 15973); *Gluconobacter oxydans* (ATCC 19357); *Brevundimonas diminuta* (ATCC 11568); *Beijerinckia indica* (ATCC 9039 e ATCC 19361); *Derxia gummosa* (ATCC 15994); *Brucella melitensis* (ATCC 23456), *Brucella abortus* (ATCC 23448); *Agrobacterium tumefaciens* (ATCC 23308), *Agrobacterium radiobacter* (ATCC 19358), *Agrobacterium rhizogenes* (ATCC 11325); *Chelatobacter heintzii* (ATCC 29600); *Ensifer adhaerens* (ATCC 33212); *Rhizobium leguminosarum* (ATCC 10004); *Sinorhizobium fredii* (ATCC 35423); *Blastomonas natatória* (ATCC 35951); *Sphingomonas paucimobilis* (ATCC 29837); *Alcaligenes faecalis*
 20 (ATCC 8750); *Bordetella pertussis* (ATCC 9797); *Burkholderia cepacia* (ATCC 25416); *Ralstonia pickettii* (ATCC 27511); *Acidovorax facilis* (ATCC 11228); *Hydrogenophagaflava* (ATCC 33667); *Zoogloea ramigera* (ATCC 19544); *Methylobacter luteus* (ATCC 49878); *Methylocaldum gracile* (NCIMB 11912); *Methylococcus capsulatus* (ATCC 19069); *Methylomicrobium agile*
 25 (ATCC 35068); *Methylomonas methanica* (ATCC 35067); *Methylosarcina fibrata* (ATCC 700909); *Methylosphaera hansonii* (ACAM 549); *Azomonas agilis* (ATCC 7494); *Azorhizophilus paspali* (ATCC 23833); *Azotobacter c-*
 30

hroococcum (ATCC 9043); *Cellvibrio mixtus* (UQM 2601); *Oligella urethralis* (ATCC 17960); *Pseudomonas aeruginosa* (ATCC10145), *Pseudomonas fluorescens* (ATCC 35858); *Francisella tularensis* (ATCC 6223); *Stenotrophomonas maltophilia* (ATCC 13637); *Xanthomonas campestris* (ATCC 33913) e

5 *Oceanimonas doudoroffii* (ATCC 21123).

Em outra modalidade, a célula hospedeira é selecionada do "Subgrupo 3 de Proteobactérias Gram-Negativas". O "Subgrupo 3 de Proteobactérias Gram-Negativas" é definido como o grupo de Proteobactérias dos seguintes gêneros: *Brevundimonas*; *Agrobacterium*; *Rhizobium*; *Sinorhizobium*;

10 *Blastomonas*; *Sphingomonas*; *Alcaligenes*; *Burkholderia*; *Ralstonia*; *Acidovorax*; *Hydrogenophaga*; *Methylobacter*; *Methylocaldum*; *Methylococcus*; *Methylomicrobium*; *Methylomonas*; *Methylosarcina*; *Methylosphaera*; *Azomonas*; *Azorhizophilus*; *Azotobacter*; *Cellvibrio*; *Oligella*; *Pseudomonas*; *Teredinibacter*; *Francisella*; *Stenotrophomonas*; *Xanthomonas* e *Oceanimonas*.

15 Em outra modalidade, a célula hospedeira é selecionada do "Subgrupo 4 de Proteobactérias Gram-Negativas". O "Subgrupo 4 de Proteobactérias Gram-Negativas" é definido como o grupo de Proteobactérias dos seguintes gêneros: *Brevundimonas*; *Blastomonas*; *Sphingomonas*; *Burkholderia*; *Ralstonia*; *Hydrogenophaga*; *Methylobacter*; *Methylocaldum*; *Methylococcus*;

20 *Methylomicrobium*; *Methylomonas*; *Methylosarcina*; *Methylosphaera*; *Azomonas*; *Azorhizophilus*; *Azotobacter*; *Cellvibrio*; *Oligella*; *Pseudomonas*; *Teredinibacter*; *Francisella*; *Stenotrophomonas*; *Xanthomonas* e *Oceanimonas*.

Em outra modalidade, a célula hospedeira é selecionada do

25 "Subgrupo 5 de Proteobactérias Gram-Negativas". O "Subgrupo 5 de Proteobactérias Gram-Negativas" é definido como o grupo de Proteobactérias dos seguintes gêneros: *Methylobacter*; *Methylocaldum*; *Methylococcus*; *Methylomicrobium*; *Methylomonas*; *Methylosarcina*; *Methylosphaera*; *Azomonas*; *Azorhizophilus*; *Azotobacter*; *Cellvibrio*; *Oligella*; *Pseudomonas*; *Teredinibacter*;

30 *Francisella*; *Stenotrophomonas*; *Xanthomonas* e *Oceanimonas*.

A célula hospedeira pode ser selecionada do "Subgrupo 6 de Proteobactérias Gram-Negativas". O "Subgrupo 6 de Proteobactérias Gram-

Negativas" é definido como o grupo de Proteobactérias dos seguintes gêneros: Brevundimonas; Blastomonas; Sphingomonas; Burkholderia; Ralstonia; Acidovorax; Hydrogenophaga; Azomonas; Azorhizophilus; Azotobacter; Cellvibrio; Oligella; Pseudomonas; Teredinibacter; Francisella; Stenotrophomonas; Xanthomonas e Oceanimonas.

A célula hospedeira pode ser selecionada do "Subgrupo 7 de Proteobactérias Gram-Negativas". O "Subgrupo 7 de Proteobactérias Gram-Negativas" é definido como o grupo de Proteobactérias dos seguintes gêneros: Azomonas; Azorhizophilus; Azotobacter; Cellvibrio; Oligella; Pseudomonas; Teredinibacter; Francisella; Stenotrophomonas; Xanthomonas e Oceanimonas.

A célula hospedeira pode ser selecionada do "Subgrupo 8 de Proteobactérias Gram-Negativas". O "Subgrupo 8 de Proteobactérias Gram-Negativas" é definido como o grupo de Proteobactérias dos seguintes gêneros: Brevundimonas; Blastomonas; Sphingomonas; Burkholderia; Ralstonia; Acidovorax; Hydrogenophaga; Pseudomonas; Stenotrophomonas; Xanthomonas e Oceanimonas.

A célula hospedeira pode ser selecionada do "Subgrupo 9 de Proteobactérias Gram-Negativas". O "Subgrupo 9 de Proteobactérias Gram-Negativas" é definido como o grupo de Proteobactérias dos seguintes gêneros: Brevundimonas; Burkholderia; Ralstonia; Acidovorax; Hydrogenophaga; Pseudomonas; Stenotrophomonas; e Oceanimonas.

A célula hospedeira pode ser selecionada do "Subgrupo 10 de Proteobactérias Gram-Negativas". O "Subgrupo 10 de Proteobactérias Gram-Negativas" é definido como o grupo de Proteobactérias dos seguintes gêneros: Burkholderia; Ralstonia; Pseudomonas; Stenotrophomonas; e Xanthomonas.

A célula hospedeira pode ser selecionada do "Subgrupo 11 de Proteobactérias Gram-Negativas". O "Subgrupo 11 de Proteobactérias Gram-Negativas" é definido como o grupo de Proteobactérias dos seguintes gêneros: Pseudomonas; Stenotrophomonas; e Xanthomonas. A célula hospedeira pode ser selecionada do "Subgrupo 12 de Proteobactérias Gram-

Negativas". O "Subgrupo 12 de Proteobactérias Gram-Negativas" é definido como o grupo de Proteobactérias dos seguintes gêneros: Burkholderia; Ralstonia; Pseudomonas. A célula hospedeira pode ser selecionada do "Subgrupo 13 de Proteobactérias Gram-Negativas". O "Subgrupo 13 de Proteobactérias Gram-Negativas" é definido como o grupo de Proteobactérias dos seguintes gêneros: Burkholderia; Ralstonia; Pseudomonas; e Xanthomonas. A célula hospedeira pode ser selecionada do "Subgrupo 14 de Proteobactérias Gram-Negativas". O "Subgrupo 14 de Proteobactérias Gram-Negativas" é definido como o grupo de Proteobactérias dos seguintes gêneros: Pseudomonas e Xanthomonas. A célula hospedeira pode ser selecionada do "Subgrupo 15 de Proteobactérias Gram-Negativas". O "Subgrupo 15 de Proteobactérias Gram-Negativas" é definido como o grupo de Proteobactérias do seguinte gênero: Pseudomonas.

A célula hospedeira pode ser selecionada do "Subgrupo 16 de Proteobactérias Gram-Negativas". O "Subgrupo 16 de Proteobactérias Gram-Negativas" é definido como o grupo de Proteobactérias das seguintes espécies de Pseudomonas (com os números de depósito da ATCC e outros das cepas exemplares mostrados entre parênteses): *Pseudomonas abietaniphila* (ATCC 700689); *Pseudomonas aeruginosa* (ATCC 10145); *Pseudomonas alcaligenes* (ATCC 14909); *Pseudomonas anguilliseptica* (ATCC 33660); *Pseudomonas citronellolis* (ATCC 13674); *Pseudomonas flavescens* (ATCC 51555); *Pseudomonas mendoncina* (ATCC 25411); *Pseudomonas nitroreducens* (ATCC 33634); *Pseudomonas oleovorans* (ATCC 8062); *Pseudomonas pseudoalcaligenes* (ATCC 17440); *Pseudomonas resinovorans* (ATCC 14235); *Pseudomonas straminea* (ATCC 33636); *Pseudomonas agarici* (ATCC 25941); *Pseudomonas alcaliphila*; *Pseudomonas alginovora*; *Pseudomonas andersonii*; *Pseudomonas asplenii* (ATCC 23835); *Pseudomonas azelaica* (ATCC 27162); *Pseudomonas beyerinckii* (ATCC 19372); *Pseudomonas borealis*; *Pseudomonas boreopolis* (ATCC 33662); *Pseudomonas brassicacearum*; *Pseudomonas butanovora* (ATCC 43655); *Pseudomonas cellulosa* (ATCC 55703); *Pseudomonas aurantiaca* (ATCC 33663); *Pseudomonas chlororaphis* (ATCC 9946, ATCC 13985, ATCC 17418, ATCC

17461); *Pseudomonas fragi* (ATCC 4973); *Pseudomonas lundensis* (ATCC 49968); *Pseudomonas taetrolens* (ATCC 4683); *Pseudomonas cissicola* (ATCC 33616); *Pseudomonas coronafaciens*; *Pseudomonas diterpeniphila*; *Pseudomonas elongata* (ATCC 10144); *Pseudomonas flectens* (ATCC 12775); *Pseudomonas azotoformans*; *Pseudomonas brenneri*; *Pseudomonas cedrella*; *Pseudomonas corrugata* (ATCC 29736); *Pseudomonas extremorientalis*; *Pseudomonas fluorescens* (ATCC 35858); *Pseudomonas gessardi*; *Pseudomonas libanensis*; *Pseudomonas mandelii* (ATCC 700871); *Pseudomonas marginalis* (ATCC 10844); *Pseudomonas migulae*; *Pseudomonas mucidolens* (ATCC 4685); *Pseudomonas orientalis*; *Pseudomonas rhodesiae*; *Pseudomonas synxantha* (ATCC 9890); *Pseudomonas tolaasii* (ATCC 33618); *Pseudomonas veronii* (ATCC 700474); *Pseudomonas frederiksbergensis*; *Pseudomonas geniculata* (ATCC 19374); *Pseudomonas gingeri*; *Pseudomonas graminis*; *Pseudomonas grimontii*; *Pseudomonas halodendrificans*; *Pseudomonas halophila*; *Pseudomonas hibiscicola* (ATCC 19867); *Pseudomonas huttiensis* (ATCC 14670); *Pseudomonas hydrogenovora*; *Pseudomonas jessenii* (ATCC 700870); *Pseudomonas kilonensis*; *Pseudomonas lanceolata* (ATCC 14669); *Pseudomonas lini*; *Pseudomonas marginata* (ATCC 25417); *Pseudomonas mephitica* (ATCC 33665); *Pseudomonas denitrificans* (ATCC 19244); *Pseudomonas pertucinogena* (ATCC 190); *Pseudomonas pictorum* (ATCC 23328); *Pseudomonas psychrophila*; *Pseudomonas filva* (ATCC 31418); *Pseudomonas monteillii* (ATCC 700476); *Pseudomonas mosselii*; *Pseudomonas otyzihabitans* (ATCC 43272); *Pseudomonas plecoglossicida* (ATCC 700383); *Pseudomonas putida* (ATCC 12633); *Pseudomonas reactans*; *Pseudomonas spinosa* (ATCC 14606); *Pseudomonas balearica*; *Pseudomonas luteola* (ATCC 43273); *Pseudomonas atutzeri* (ATCC 17588); *Pseudomonas amygdali* (ATCC 33614); *Pseudomonas avellanae* (ATCC 700331); *Pseudomonas caricapapayae* (ATCC 33615); *Pseudomonas cichorii* (ATCC 10857); *Pseudomonas ficuserectae* (ATCC 35104); *Pseudomonas fuscovaginae*; *Pseudomonas meliae* (ATCC 33050); *Pseudomonas syringae* (ATCC 19310); *Pseudomonas viridiflava* (ATCC 13223); *Pseudomonas thermocarboxydovorans* (ATCC 35961);

Pseudomonas thermotolerans; *Pseudomonas thivervalensis*; *Pseudomonas vancouverensis* (ATCC 700688); *Pseudomonas wisconsinensis*; e *Pseudomonas xiamenensis*.

A célula hospedeira pode ser selecionada do "Subgrupo 17 de Proteobactérias Gram-Negativas". O "Subgrupo 17 de Proteobactérias Gram-Negativas" é definido como o grupo de Proteobactérias conhecidas na técnica como as "Pseudomonads fluorescentes" incluindo aquelas que pertencem, por exemplo, as seguintes espécies de *Pseudomonas*: *Pseudomonas azotoformans*, *Pseudomonas brenneri*; *Pseudomonas cedrella*; *Pseudomonas corrugata*; *Pseudomonas extremorientalis*; *Pseudomonas fluorescens*; *Pseudomonas gessardii*; *Pseudomonas libanensis*; *Pseudomonas mandelii*; *Pseudomonas marginalis*; *Pseudomonas migulae*; *Pseudomonas mucidolens*; *Pseudomonas orientalis*; *Pseudomonas rhodesiae*; *Pseudomonas synxantha*; *Pseudomonas tolaasi*; *Pseudomonas veronii*.

Outros hospedeiros adequados incluem aqueles classificados em outras partes da referência, tais como Proteobactérias Gram(+). Em uma modalidade, a célula hospedeira é uma *E. coli*. A sequência do genoma de *E. coli* foi estabelecida para a *E. coli* MG1655 (Blattner *et al.* (1997). A sequência genômica completa da *Escherichia coli* K-12, (*Science* 277(5331):1453-74) e microarranjos de DNA estão comercialmente disponíveis para *E. coli* K12 (MWG Inc, High Point, N.C.). *E. coli* pode ser cultivada tanto em um meio rico tal como Luria-Bertani (LB) (triptona 10 g/l, NaCl 5g/l, extrato de levedura 5g/l) quanto em um meio mínimo definido tal como M9 (Na₂PO₄ 6g/l, KH₂PO₄ 3g/l, NH₄Cl 1g/l, NaCl 0,5 g/l, pH 7.4) com uma fonte de carbono apropriada tal como glicose 1%. Rotineiramente, uma cultura no decorrer de uma noite de célula de *E. coli* é diluída e inoculada em meio rico fresco ou meio mínimo tanto em um frasco de agitação quanto em um fermentador e cultivada a 37°C.

Um hospedeiro também pode ser de origem mamífera, tal como uma célula derivada de um mamífero incluindo qualquer mamífero humano e não humano. Mamíferos podem incluir MS não estão limitados a primatas, macacos, porcos, ovinos, roedores, ungulados, porcos, suínos, ovelhas,

carneiros, cabras, gado, cervos, mulas, cavalos, macacos, chimpanzés, cães, gatos, ratos e camundongos.

Uma célula hospedeira também pode ter origem em uma planta. Qualquer planta pode ser selecionada para a identificação de genes e sequências regulatórias. Exemplos de plantas-alvo adequadas para o isolamento de genes e sequências regulatórias podem incluir mas não são limitados a alfafa, maçã, damasco, *Arabidopsis*, alcachofra, rúcula, aspargo, abacate, banana, cevada, feijões, beterraba, amora, mirtilo, brócolis, couve de Bruxelas, repolho, canola, melão amarelo, cenoura, mandioca, semente de mamona, couve-flor, aipo, cereja, chicória, coentro, cítricos, laranja mandarina, cravo, coco, café, milho, algodão, amora, pepino, abeto de Douglas, berinjela, endívia, escarola, eucalipto, erva-doce, figos, alho, abóbora, uva, grapefruit, melão, jicama, kiwi, alface, alho-poró, limão, lima, pinheiro Loblolly, semente de linho, manga, melão, cogumelo, nectarina, noz, aveia, óleo de palma, óleo de semente de colza, quiabo, azeitona, cebola, laranja, uma planta ornamental, palma, papaia, salsa, pastinaca, ervilha, pêssego, amendoim, pêra, pimenta, caqui, pinha, abacaxi, banana da terra, ameixa, romã, álamo, batata, abóbora, marmelo, pinheiro de Monterey, radichio, semente de colza, framboesa, arroz, centeio, sorgo, pinheiro do sul, soja, espinafre, melão squash, morango, beterraba de açúcar, cana de açúcar, girassol, batata doce, liquidambar, tangerina, chá, tabaco, tomate, tritcale, turfa, nabo, videira, melancia, inhame e abobrinha. Em algumas modalidades, as plantas úteis no método são *Arabidopsis*, milho, trigo, soja e algodão.

III. Métodos

Os métodos da invenção fornecem a expressão de proteínas de fusão que compreendem um polipeptídeo de sinal de secreção selecionado a partir de um sinal de secreção de pbp*, dsbA, dsbC, Bce, CupA2, CupB2, CupC2, NikA, F1gl, ORF5550, Tgt2C ou ORF812. Em uma modalidade, os métodos incluem uma célula hospedeira que expressa uma proteína de interesse ligada a um sinal de secreção da invenção. Os métodos incluem o fornecimento de uma célula hospedeira, preferivelmente uma célula hospedeira de *P. fluorescens*, compreendendo um vetor que codifica uma proteína re-

combinante que compreende a proteína ou polipeptídeo de interesse operativamente ligado a uma sequência de sinal de secreção aqui descrita e cultivar a célula sob condições que resultam na expressão da proteína ou polipeptídeo. Alternativamente, o método para expressar proteínas ou polipeptídeos usando os sinais de secreção identificados pode ser usado em qualquer sistema hospedeiro dado, incluindo células hospedeiras de origem eucariótica ou procariótica. O vetor pode ter qualquer uma das características descritas acima. Em uma modalidade, o vetor compreende uma sequência de nucleotídeos que codifica os polipeptídeos de sinal descritos aqui como SEQ ID Nº 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22 ou 24 ou seus fragmentos ou variantes. Em outra modalidade, o vetor compreende uma sequência de nucleotídeos que compreende as SEQ ID Nº 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21 ou 23.

Em outra modalidade, a célula hospedeira tem um periplasma e a expressão do polipeptídeo de sinal de secreção resulta no direcionamento de substancialmente quase toda proteína ou polipeptídeo de interesse para o periplasma da célula. É reconhecido que uma pequena fração da proteína expressa no periplasma pode realmente escapar através da membrana celular para o espaço extracelular; entretanto, a maioria da proteína direcionada deve permanecer dentro do espaço periplasmático.

A expressão pode levar ainda a produção de proteína extracelular. O método também pode incluir a etapa de purificação da proteína ou polipeptídeo de interesse do periplasma ou médio extracelular. O sinal de secreção pode ser expresso de uma maneira na qual ele é ligado à proteína e a proteína ligada ao sinal pode ser purificada a partir da célula. Portanto, em uma modalidade, esse polipeptídeo isolado é uma proteína de fusão do sinal de secreção e da proteína ou polipeptídeo de interesse. Entretanto, o sinal de secreção pode ser clivado da proteína quando a proteína é direcionada para o periplasma. Em uma modalidade, a ligação entre o sinal de secreção e a proteína ou polipeptídeo é modificada para aumentar a clivagem do sinal de secreção.

Os métodos da invenção também podem levar a uma produção

aumentada da proteína ou polipeptídeo de interesse dentro da célula hospedeira. A produção aumentada pode ser, alternativamente, um nível aumentado de proteína ou polipeptídeo apropriadamente processados por grama de proteína produzida ou por grama de proteína hospedeira. A produção
5 aumentada pode ser também um nível aumentado de proteína ou polipeptídeo recuperáveis produzidos por grama de proteína recombinante ou por grama de célula hospedeira. A produção aumentada também pode ser qualquer combinação de um nível aumentado de proteína total, nível aumentada de proteína apropriadamente processada ou nível aumentado de proteína
10 ativa ou solúvel. Nessa modalidade, a expressão "aumentado" é relativa ao nível de proteína ou polipeptídeo que é produzido, apropriadamente processado, solúvel e/ou recuperável quando a proteína ou polipeptídeo de interesse são expressos em uma célula sem o polipeptídeo de sinal de secreção da invenção.

15 Uma expressão aperfeiçoada de uma proteína ou polipeptídeo de interesse também pode referir-se a um aumento da solubilidade da proteína. A proteína ou polipeptídeo de interesse podem ser produzidos e recuperados do citoplasma, periplasma ou meio extracelular da célula hospedeira. A proteína ou polipeptídeo podem ser insolúveis ou solúveis. A proteína ou
20 polipeptídeo podem incluir uma ou mais sequências de direcionamento ou sequências para auxiliar a purificação, como discutido acima.

A expressão "solúvel" como usada aqui significa que a proteína não é precipitada pela centrifugação em aproximadamente 5,000 e 20,000 x a gravidade quando centrifugada por 10 a 30 minutos em um tampão sob
25 condições fisiológicas. Proteínas solúveis não são parte de um corpo de inclusão ou outra massa precipitada. Similarmente, "insolúvel" significa que a proteína ou polipeptídeo podem ser precipitados pela centrifugação entre 5,000 e 20,000 x a gravidade quando centrifugados por 10 a 30 minutos em um tampão sob condições fisiológicas. Proteínas ou polipeptídeos insolúveis
30 podem ser parte de um corpo de inclusão ou de outra massa precipitada. A expressão "corpo de inclusão" pretende incluir qualquer corpo intracelular contido dentro de uma célula no qual um agregado de proteínas ou polipep-

tídeos pode estar seqüestrado.

Os métodos da invenção podem produzir proteína localizada no periplasma da célula hospedeira. Em uma modalidade, o método produz proteínas ou polipeptídeos de interesse apropriadamente processados na célula. Em outra modalidade, a expressão do polipeptídeo de sinal de secreção pode produzir proteínas ou polipeptídeos de interesse ativos na célula. O método da invenção também pode levar a um rendimento aumentado de proteínas ou polipeptídeos de interesse quando comparado com a proteína que é expressa sem o sinal de secreção da invenção.

Em uma modalidade, o método produz pelo menos 0,1 g/l de proteína no compartimento periplasmático. Em outra modalidade, o método produz 0,1 a 10 g/l de proteína periplasmática na célula, ou pelo menos cerca de 0,2, cerca de 0,3, cerca de 0,4, cerca de 0,5, cerca de 0,6, cerca de 0,7, cerca de 0,8, cerca de 0,9 ou pelo menos cerca de 1,0 g/l de proteína periplasmática. Em uma modalidade, o total de proteína ou polipeptídeo de interesse produzidos é de pelo menos 1,0 g/l, pelo menos cerca de 2 g/l, pelo menos cerca de 3 g/l, cerca de 4 g/l, cerca de 5 g/l, cerca de 6 g/l, cerca de 7 g/l, cerca de 8 g/l, cerca de 10 g/l, cerca de 15g/l, cerca de 20g/l, pelo menos cerca de 25g/l ou mais. Em algumas modalidades, a quantidade de proteína periplasmática produzida é de pelo menos cerca de 5%, cerca de 10%, cerca de 15%, cerca de 20%, cerca de 25%, cerca de 30%, cerca de 40%, cerca de 50%, cerca de 60%, cerca de 70%, cerca de 80%, cerca de 90%, cerca de 95%, cerca de 96%, cerca de 97%, cerca de 98%, cerca de 99% ou mais do total de proteína ou polipeptídeo de interesse produzidos.

Em uma modalidade, o método produz pelo menos 0,1 g/l de proteína corretamente processada. Uma proteína corretamente processada tem uma terminação amino da proteína nativa. Em algumas modalidades, pelo menos 50% da proteína ou polipeptídeo de interesse compreendem uma terminação amino nativa. Em outra modalidade, pelo menos 60%, pelo menos 70%, pelo menos 80%, pelo menos 90% ou mais da proteína tem uma terminação amino da proteína nativa. Em várias modalidades, o método produz 0,1 a 10 g/l de proteína corretamente processada na célula, incluindo

pelo menos cerca de 0,2, cerca de 0,3, cerca de 0,4, cerca de 0,5, cerca de 0,6, cerca de 0,7, cerca de 0,8, cerca de 0,9 ou pelo menos cerca de 1,0 g/l de proteína corretamente processada. Em outra modalidade, o total de proteína ou polipeptídeo de interesse produzidos é de pelo menos 1,0 g/l, pelo menos cerca de 2g/l, pelo menos cerca de 3g/l, cerca de 4 g/l, cerca de 5 g/l, cerca de 6 g/l, cerca de 7g/l, cerca de 8 g/l, cerca de 10 g/l, cerca de 15 g/l, cerca de 20 g/l, cerca de 25 g/l, cerca de 30 g/l, cerca de 35 g/l, cerca de 40 g/l, cerca de 45 g/l, pelo menos cerca de 50 g/l ou mais. Em algumas modalidades, a quantidade produzida de proteína corretamente processada é de pelo menos cerca de 5%, cerca de 10%, cerca de 15%, cerca de 20%, cerca de 25%, cerca de 30%, cerca de 40%, cerca de 50%, cerca de 60%, cerca de 70%, cerca de 80%, cerca de 90%, cerca de 95%, cerca de 96%, cerca de 97%, cerca de 98%, cerca de 99% ou mais do total de proteína recombinante em uma forma corretamente processada.

Os métodos da invenção também podem levar a um rendimento aumentado da proteína ou polipeptídeo de interesse. Em uma modalidade, o método produz um proteína ou polipeptídeo de interesse como pelo menos cerca de 5%, pelo menos cerca de 10%, cerca de 15%, cerca de 20%, cerca de 25%, cerca de 30%, cerca de 40%, cerca de 45%, cerca de 50%, cerca de 55%, cerca de 60%, cerca de 65%, cerca de 70%, cerca de 75% ou mais do total de proteína da célula (tcp). "Percentual de proteína total da célula" é a quantidade de proteína ou polipeptídeo na célula hospedeira como um percentual de agregado de proteína celular. A determinação do percentual de proteína total da célula é bem conhecida na técnica.

Em uma modalidade particular, a célula hospedeira pode ter um nível de expressão do polipeptídeo recombinante, polipeptídeo, proteína ou seus fragmentos de pelo menos 1% de tcp e uma densidade celular de pelo menos 40 g/l, quando cultivada (isto é, dentro de uma faixa de temperatura de cerca de 4°C a cerca de 55°C, incluindo cerca de 10°C, cerca de 15°C, cerca de 20°C, cerca de 25°C, cerca de 30°C, cerca de 35°C, cerca de 40°C, cerca de 45°C e cerca de 50°C) em um meio com sais minerais. Em uma modalidade particularmente preferida, o sistema de expressão terá um nível

de expressão de proteína ou polipeptídeo de pelo menos 5 de tcp e uma densidade celular de pelo menos 40 g/l, quando cultivada (isto é, dentro de uma faixa de temperatura de cerca de 4°C a cerca de 55°C, inclusive) em um meio com sais minerais em uma escala de fermentação de pelo menos

5 cerca de 10 litros.

Na prática, proteínas heterólogas direcionadas para o periplasma são geralmente encontradas em caldo de cultura (veja Patente Europeia N° EP 0 288 451), possivelmente por causa de dano ou por um aumento da fluidez da membrana celular externa. A taxa dessa secreção "passiva" pode

10 ser aumentada pelo uso de uma variedade de mecanismos que permeabilizam a membrana celular externa: colicina (Miksch *et al.* (1997) *Arch. Microbiol.* 167:143-150); taxa de crescimento (Shokri *et al.*, (2002) *App Microbiol Biotechnol* 58:386-392); superexpressão de TolIII (Wan e Baneyx (1998) *Protein Expression Purif.* 14:13-22); proteína de liberação de bacteriocina

15 (Hsiung *et al.*, (1989) *Bio/Technology* 7:267-71); proteína de lise de colicina A (Lloubes *et al.* (1993) *Biochimie* 75:451-8); mutantes que extravasam proteínas periplasmáticas (Furlong e Sundstrom (1989) *Developments in Indus. Microbio.* 30:141-8); ligantes de fusão (Jeong e Lee (2002) *Appl. Environ. Microbio.* 68:4979-4985); recuperação por choque osmótico (Taguchi *et al.*

20 (1990) *Biochimica Biophysica Acta* 1049:278-85). O transporte de proteínas manipuladas para o espaço periplasmático com localização subsequente no caldo de cultura tem sido usado para produzir proteínas apropriadamente enoveladas e ativas em *E. coli* (Wan e Baneyx (1998) *Protein Expression Purif.* 14:13-22; Simmons *et al* (2002) *J. Immun. Meth.* 263:133-147; Lundell

25 *et al.* (1990) *J. Indust. Microbio.* 5:215-27).

A. Produção de proteína ativa

Em algumas modalidades, a proteína também pode ser produzida em uma forma ativa. A expressão "ativa" significa a presença de atividade biológica, em que a atividade biológica é comparável ou corresponde substancialmente à atividade biológica de uma proteína ou polipeptídeo nativos

30 correspondentes. No contexto de proteínas, isso significa tipicamente que um polinucleotídeo ou polipeptídeo compreende uma função ou efeito bioló-

gicos que têm pelo menos cerca de 20%, cerca de 50%, preferivelmente pelo menos cerca de 60 a 80% e o mais preferivelmente pelo menos cerca de 90 a 95% da atividade comparada com a proteína ou polipeptídeo nativos correspondentes usando parâmetros padronizados. A determinação da atividade de proteína ou polipeptídeo pode ser realizada utilizando ensaios biológicos padronizados, comparativos marcados correspondentes para proteínas ou polipeptídeos em particular. Uma indicação de que uma proteína ou polipeptídeo de interesse mantém a atividade biológica é que o polipeptídeo é imunologicamente interreativo com o polipeptídeo nativo.

10 A invenção pode também aperfeiçoar a recuperação da proteína ou polipeptídeo ativos de interesse. Proteínas ativas podem ter uma atividade específica de pelo menos cerca de 20%, pelo menos cerca de 30%, pelo menos cerca de 40%, cerca de 50%, cerca de 60%, pelo menos cerca de 70%, cerca de 80%, cerca de 90% ou pelo menos cerca de 95% daquela da
15 proteína ou polipeptídeo nativos dos quais a sequência é derivada. Adicionalmente, a especificidade do substrato (k_{cat}/K_m) é opcional e substancialmente similar a da proteína ou polipeptídeo nativos. Tipicamente, k_{cat}/K_m será pelo menos cerca de 30%, cerca de 40%, cerca de 50%, cerca de 60%, cerca de 70%, cerca de 80%, pelo menos cerca de 90%, pelo menos cerca
20 de 95% ou mais. Métodos para ensaiar e quantificar medidas de atividade de polipeptídeo e especificidade de substrato (k_{cat}/K_m) são bem conhecidos por aqueles versados na técnica.

A atividade da proteína ou do polipeptídeo de interesse também pode ser comparada com a atividade padrão previamente estabelecida de
25 uma proteína ou polipeptídeo nativos. Alternativamente, a atividade da proteína ou do polipeptídeo de interesse pode ser determinada em um ensaio comparativo simultâneo, ou substancialmente simultâneo, com a proteína ou polipeptídeo nativos. Por exemplo, ensaios in vitro podem ser usados para determinar qualquer interação detectável entre uma proteína ou polipeptídeo
30 de interesse e um alvo, por exemplo, entre uma enzima expressa e o substrato, entre um hormônio expresso e o receptor do hormônio, entre um anticorpo expresso e o antígeno, etc. Tal detecção pode incluir a medida de alte-

rações calorimétricas, alterações de proliferação, morte celular, repulsão celular, alterações na radioatividade, alterações na solubilidade, alterações no peso molecular como medidas por métodos de eletroforese em gel e/ou exclusão em gel, habilidades de fosforilação, ensaios de especificidade de anticorpo tais como ensaios ELISA, etc. Além disso, ensaios in vivo incluem, mas não são limitados a ensaios para detectar efeitos fisiológicos da proteína ou polipeptídeo produzidos em *Pseudomonas* em comparação com efeitos fisiológicos da proteína ou polipeptídeo nativos, por exemplo, ganho de peso, alteração no balanço eletrolítico, alteração no tempo de coagulação, alterações na dissolução do coágulo e a indução de resposta antigênica. Geralmente, qualquer ensaio in vitro ou in vivo pode ser usado para determinar a natureza ativa da proteína ou polipeptídeo de interesse que permita uma análise comparativa com a proteína ou polipeptídeo nativos, desde que tal atividade possa ser ensaiada. Alternativamente, as proteínas ou polipeptídeos produzidos pela presente invenção podem ser ensaiados quanto à habilidade de estimular ou inibir a interação entre a proteína ou o polipeptídeo e uma molécula que interage normalmente com a proteína ou o polipeptídeo, por exemplo, um substrato ou um componente da via de sinalização com o qual a proteína nativa normalmente interage. Tais ensaios podem incluir, tipicamente, as etapas de combinar a proteína com um substrato molecular sob condições que permitam a proteína ou o polipeptídeo interagir com a molécula alvo e detectar a consequência bioquímica da interação da proteína com a molécula alvo.

Ensaios que podem ser utilizados para determinar a atividade de proteína ou de polipeptídeo estão descritos, por exemplo, em Ralph, P.J. *et al.*, (1984) *J. Immunol.* 132:1858 ou Saiki *et al.*, (1981) *J. Immunol.* 127:1044, Steward, W. E. II (1980) *The Interferon Systems*. Springer-Verlag, Viena e New York, Broxmeyer, H.E. *et al.*, (1982) *Blood* 60:595, *Molecular Cloning: A Laboratory Manual*, 2ª ed., Cold Spring Harbor Laboratory Press, Sambrook, J., E.F.Fritsch e T. Maniatis eds., 1989 e *Methods in Enzymology: Guide to Molecular Cloning Techniques*, Academic Press, Berger, S.L. e A.R. Kimmel eds., 1987, A.K. Patra *et al.*, *Protein Expr Purif*, 18(2):p/182-92

(2000), Kodama *et al.*, *J. Biochem.* 99:1465-1472 (1986); Stewart *et al.*, *Proc. Nat'l Acad. Sci. USA* 90:5209-5213 (1993); (Lombillo *et al.*, *J. Cell Biol.* 128:107-115 (1995); (Vale *et al.*, *Cell* 42:39-50 (1985).

B. Condições de cultura celular

5 As condições de cultura celular para as células hospedeiras aqui descritas podem incluir aquelas que facilitam a expressão da proteína de interesse e/ou que facilitam a fermentação da proteína de interessa expressa. Como usada aqui, a expressão "fermentação" inclui ambas as modalidades nas quais a fermentação literal é empregada e modalidades nas quais
10 outros modos de cultura não fermentativos são empregados. A fermentação pode ser realizada em qualquer escala. Em uma modalidade, o meio de fermentação pode ser selecionado entre meios ricos, meios mínimos e meios com sais minerais; um meio rico pode ser usado, mas é preferivelmente evitado. Em outra modalidade, tanto um meio mínimo quanto um meio com sais
15 minerais são selecionados. Em outra modalidade, um meio mínimo é selecionado. Em outra modalidade ainda, um meio com sais minerais é selecionado. Meios com sais minerais são particularmente preferidos.

Meios com sais minerais consistem em sais minerais e uma fonte de carbono tal como, por exemplo, glicose, sucrose ou glicerol. Exemplos
20 de meios com sais minerais incluem, por exemplo, meio M9, meio para *Pseudomonas* (ATCC 179), meio de Davis e Mingioli (veja, BD Davie & ES Mingioli (1950) em *J. Bact.* 60:17-28). Os sais minerais usados para fazer meios com sais minerais incluem aqueles entre, por exemplo, fosfatos de potássio, cloreto ou sulfato de amônia, cloreto ou sulfato de magnésio e re-
25 síduos minerais tais como cloreto de cálcio, borato e sulfatos de ferro, cobre, manganês e zinco. Nenhuma fonte de nitrogênio orgânico, tais como peptona, triptona, aminoácidos ou extrato de levedura, é incluída em um meio com sais minerais. Ao contrário, é usada uma fonte de nitrogênio inorgânico e essa pode ser selecionada entre, por exemplo, sais de amônia, amônia a-
30 quosa e amônia gasosa. Um meio com sais minerais preferido conterá glicose como fonte de carbono. Em comparação com meios com sais minerais, meios mínimos também podem conter sais minerais e uma fonte de carbono,

mas podem ser suplementados com, por exemplo, níveis baixos de aminoácidos, vitaminas, peptonas ou outros ingredientes, embora esses sejam adicionados em níveis mínimos.

- Em uma modalidade, os meios podem ser preparados usando os componentes listados na Tabela 5 abaixo. Os componentes podem ser adicionados na seguinte ordem: primeiro, $(\text{NH}_4)\text{HPO}_4$, KH_2PO_4 e ácido cítrico podem ser dissolvidos em aproximadamente 30 litros de água destilada; então uma solução de elementos residuais pode ser adicionada, seguida pela adição de um agente antiespumante, tal como Ucolub N 115. Depois, após esterilização por aquecimento (tal como em aproximadamente 121°C), soluções estéreis de glicose, MgSO_4 e timina-HCl podem ser adicionadas. O controle do pH em aproximadamente 6,8 pode ser obtido usando amônia aquosa. Água destilada estéril pode então ser adicionada para ajustar o volume inicial para 371 menos o estoque de glicerol (123 ml). Os químicos estão comercialmente disponíveis a partir de vários fornecedores, tal como Merck. Esse meio pode permitir uma cultura com alta densidade celular (HCDC) para o crescimento de espécies de *Pseudomonas* e bactérias relacionadas. HCDC pode começar como um processo em lote que é seguido por um cultivo de duas fases de alimentação. Após o crescimento ilimitado na parte de lote, o crescimento pode ser controlado em uma taxa reduzida de crescimento específico durante um período de 3 tempos de duplicação nos quais a concentração da biomassa pode aumentar várias vezes. Mais detalhes de tais procedimentos de cultura são descritos por Riesenber, D.; Schulz, V; Knorre, W.A; Pohl, H.D.; Korz, D.; Sanders, E.A.; Ross, A.; Deckwer, W.D. (1991), "High cell density cultivation of *Escherichia coli* at controlled specific growth rate" *J. Biotechnol.* 20(1)17-27.

Tabela 5: Composição do meio	
Componente	Concentração inicial
KH_2PO_4	$13,3 \text{ g l}^{-1}$
$(\text{NH}_4)_2\text{HPO}_4$	$4,0 \text{ g l}^{-1}$
Ácido cítrico	$1,7 \text{ g l}^{-1}$
$\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$	$1,2 \text{ g l}^{-1}$

Tabela 5: Composição do meio	
Solução de resíduo de metal	10 ml 1 ⁻¹
Tiamina HCL	4,4 mg 1 ⁻¹
Glicose-H ₂ O	27,3 g 1 ⁻¹
Antiespumante Uncolub N115	0,1 ml 1 ⁻¹
Solução de alimentação	
MgSO ₄ -7H ₂ O	19,7 g 1 ⁻¹
Glicose-H ₂ O	770 g 1 ⁻¹
NH ₃	23 g
Solução de resíduo de metal	
citrato de Fe (III) 6 g 1 ⁻¹	
MnCl ₂ -4H ₂ O 1,5 g 1 ⁻¹	
ZmCH ₂ COOI ₂ -2H ₂ O 0,8 g 1 ⁻¹	
H ₃ BO ₃ 0,3 g 1 ⁻¹	
Na ₂ MoO ₄ -2H ₂ O 0,25 g 1 ⁻¹	
CoCl ₂ -6H ₂ O 0,3 g 1 ⁻¹	
CuCl ₂ 2H ₂ O 0,15 g 1 ⁻¹	
etileno 0,84 g 1 ⁻¹	
ácido dinitrilo-tetracético Na ₂ sah 2H ₂ O	
(Tritriplex III, Merck)	

O sistema de expressão de acordo com a presente invenção pode ser cultivado em qualquer forma de fermentação. Por exemplo, modos de fermentação em lote, com alimentação, semicontínua e contínua podem ser empregados aqui. Onde a proteína é excretada para o meio extracelular, a fermentação contínua é preferida.

Os sistemas de expressão de acordo com a presente invenção são úteis para a expressão de transgene em qualquer escala (isto é, volume) de fermentação. Assim, por exemplo, volumes de fermentação na escala de microlitro, escala de centilitro e na escala de decilitro podem ser usados; e volumes de fermentação na escala de 1 litro ou mais podem ser usados. Em

uma modalidade, o volume da fermentação será de ou acima de 1 litro. Em outra modalidade, o volume da fermentação será de ou acima de 5 litros, 10 litros, 15 litros, 20 litros, 25 litros, 50 litros, 75 litros, 100 litros, 200 litros, 500 litros, 1000 litros, 2000 litros, 5000 litros, 10.000 litros ou 50.000 litros.

5 Na presente invenção, o crescimento, cultivo e/ou fermentação das células hospedeiras transformadas é realizado dentro de uma faixa de temperatura que permite a sobrevivência das células hospedeiras, preferi-
velmente dentro de uma faixa de cerca de 4°C a cerca de 55°C, inclusive. Assim, por exemplo as expressões "crescimento" (e "crescer", "crescendo"),
10 "cultivo" (e "cultura") e "fermentação" (e "fermento", "fermentar"), com usa-
das aqui em relação às células hospedeiras da presente invenção, signifi-
cam inerentemente "crescimento", "cultivo" e "fermentação" dentro de uma
faixa de temperatura de cerca de 4°C a cerca de 55°C, inclusive. Além disso,
"crescimento" é usado para indicar ambos os estados biológicos de divisão
15 celular ativa ou expansão, assim com estados biológicos nos quais uma cé-
lula não dividida/não expandida está sendo metabolicamente sustentada, o
último significado da expressão "crescimento" sendo sinônimo da expressão
"manutenção".

Uma vantagem adicional em usar *Pseudomonas fluorescens* na
20 expressão de proteínas secretadas inclui a habilidade que a *Pseudomonas fluorescens* tem de crescer em altas densidades celulares comparada a al-
guns outros sistemas de expressão bacterianos. Para essa finalidade, siste-
ma de expressão em *Pseudomonas fluorescens* de acordo com a presente
invenção podem fornecer uma densidade celular de cerca de 20 g/l ou mais.
25 Os sistemas de expressão em *Pseudomonas fluorescens* de acordo com a
presente invenção podem, igualmente, fornecer uma densidade celular de
pelo menos cerca de 70 g/l, como determinado em termos de biomassa por
volume, a biomassa sendo medida como o peso celular seco.

Em uma modalidade, a densidade celular será de pelo menos
30 cerca de 20 g/l. Em outra modalidade, a densidade celular será de pelo me-
nos cerca de 25 g/l, cerca de 30 g/l, cerca de 35 g/l, cerca de 40 g/l, cerca de
45 g/l, cerca de 50 g/l, cerca de 60 g/l, cerca de 70 g/l, cerca de 80 g/l, cerca

de 90 g/l, cerca de 100 g/l, cerca de 110 g/l, cerca de 120 g/l, cerca de 130 g/l, cerca de 140 g/l, cerca de ou pelo menos cerca de 150 g/l.

Em outras modalidades, a densidade celular na indução estará entre cerca de 20 g/l a cerca de 150 g/l; entre cerca de 20 g/l a cerca de 120 g/l; entre cerca de 20 g/l a cerca de 80 g/l; entre cerca de 25 g/l a cerca de 80 g/l; entre cerca de 30 g/l a cerca de 80 g/l; entre cerca de 35 g/l a cerca de 80 g/l; entre cerca de 40 g/l a cerca de 80 g/l; entre cerca de 45 g/l a cerca de 80 g/l; entre cerca de 50 g/l a cerca de 80 g/l; entre cerca de 50 g/l a cerca de 75 g/l; entre cerca de 50 g/l a cerca de 70 g/l; entre cerca de 40 g/l a cerca de 80 g/l.

C. Isolamento da Proteína ou Polipeptídeo de Interesse

Para medir o rendimento, solubilidade, conformação e/ou atividade da proteína de interesse, pode ser desejável isolar a proteína da célula hospedeira e/ou do meio extracelular. O isolamento pode ser um isolamento bruto, semi-bruto ou puro, dependendo das necessidades do ensaio usado para fazer as medidas apropriadas. A proteína pode ser produzida no citoplasma, direcionada para o periplasma ou pode ser secretada no meio de cultura ou fermentação. Para liberar as proteínas direcionadas para o periplasma, os tratamentos que envolvem químicos tais como o clorofórmio (Ames *et al.*, (1984) *J. Bacteriol.*, 160:1181-1183), guanidina-HCl e Triton X-100 (Naglak e Wang (1990) *Enzyme Microb. Technol.*, 12:603-611) foram usados. Entretanto, esses químicos não são inertes e podem ter efeitos prejudiciais sobre vários produtos de proteína recombinante ou procedimentos de purificação subsequentes. O tratamento com glicina de células de *E.coli*, causando a permeabilização da membrana externa, também foi descrito liberar os conteúdos periplasmáticos (Ariga *et al.* (1989) *J. Ferm. Bioeng.*, 68:243-246). Os métodos mais amplamente usados de liberação periplasmática de proteínas recombinantes são o choque osmótico (Nosal e Heppel (1966) *J. Biol. Chem.*, 241:3055-3062; Neu e Heppel (1965) *J. Biol. Chem.*, 240:3685-3692), tratamento com clara de ovo de galinha (HEW)-lisozima/ácido etilenodiamino tetracético (EDTA) (Neu e Heppel (1964) *J. Biol. Chem.*, 239:3893-3900; Witholt *et al.* (1976) *Biochim. Biophys. Acta*,

443:534-544; Pierce *et al* (1995) I Cheme Research Event, 2:995-997) e o tratamento combinado de HEW-lisozima/choque osmótico (French *et al.* (1996) *Enzyme and Microb. Tech.*, 19:332-338). O método de French envolve a ressuspensão das células em um tampão de fracionamento seguido
5 pela recuperação da fração periplasmática, onde o choque osmótico segue imediatamente o tratamento com lisozima.

Tipicamente, esses procedimentos incluem um ruptura inicial no meio osmoticamente estabilizado seguido pela liberação seletiva em meio não estabilizado. A composição desses meios (pH, agente protetor) e os métodos de rompimento usados (clorofórmio, HEW-lisozima, EDTA, sonicação)
10 variam entre os procedimentos específicos relatados. Uma variação do tratamento com HEW-lisozima/EDTA usa um detergente iônico dipolar no lugar de DETA como discutido por Stabel *et al.* (1994) *Veterinary Microbiol.*, 38:307-314. Para uma revisão geral do uso de sistemas enzimáticos líticos
15 intracelulares para romper *E. coli*, veja Dabora e Cooney (1990) em *Advances in Biochemical Engineering/Biotechnology*, Vol. 43, A. Fiechter, Ed. (Springer-Verlag: Berlim), pp 11-30.

Métodos convencionais para a recuperação de proteínas e polipeptídeos de interesse a partir do citoplasma, como proteínas solúveis ou
20 partículas refrativas, envolve a desintegração da célula bacteriana por meio de ruptura mecânica. A ruptura mecânica envolve tipicamente a geração de cavitação local em uma suspensão líquida, agitação rápida com esferas rígidas, sonicação ou trituração de suspensão celular (*Bacterial Cell Surface Techniques*, Hancock e Poxton (John Wiley & Sons Ltd. 1988), Capítulo 3 p. 55).

25 HEW-lisozima atua bioquimicamente para hidrolisar o arcabouço de peptidoglicano da parede celular. O método foi desenvolvido primeiro por Zinder e Arndt (1956) *Proc. Natl. Acad. Sci. USA*, 42:586-590, que trataram a *E.coli* com albumina de ovo (que contém HEW-lisozima) para produzir esferas redondas celulares mais tarde conhecidas com esferoplastos. Essas estruturas retinham alguns componentes da parede celular mas tinham grandes áreas de superfície nas quais a membrana citoplasmática estava exposta.
30 A Pat. U.S. Nº 5.169.772 descreve um método para purificar heparinase

de bactérias que compreende romper o envelope das bactérias em um meio osmoticamente estabilizado, por exemplo, solução de sucrose 20% usando, por exemplo, EDTA, lisozima ou um composto orgânico, liberando proteínas não semelhantes à heparinase a partir do espaço periplasmático da bactéria rompida pela exposição da bactéria a um tampão de força iônica baixa e liberando as proteínas semelhantes à heparinase pela exposição de bactérias lavadas com o tampão de força iônica baixa a uma solução de sal tamponada.

Várias modificações diferentes desses métodos foram usadas em uma ampla faixa de sistemas de expressão com vários graus de sucesso (Joseph-Liazun *et al.* (1990) *Gene*, 86:291-295; Carter *et al.* (1992) *Bio/Technology*, 10:163-167). Esforços para induzir uma cultura de célula recombinante para produzir lisozima têm sido relatados. EP 0 155 189 descreve um meio para induzir uma cultura de célula recombinante a produzir lisozimas, que comumente seria esperado matar tais células hospedeiras por meio da destruição ou lise da estrutura da parede celular.

A Pat. U.S. Nº 4.595.658 descreve um método para facilitar a externalização de proteínas transportadas para o espaço periplasmático de *E. coli*. Esse método permite o isolamento seletivo de proteínas que se localizam no periplasma sem a necessidade de tratamento com lisozima, trituração mecânica ou tratamento por choque osmótico de células. A Pat. U.S. Nº 4.637.980 descreve a produção de um produto bacteriano pela transformação de um lisógeno sensível a temperatura com uma molécula de DNA que codifica, direta ou indiretamente, o produto, cultivando o transformante sob condições que permitem a expressão do produto de gene intracelularmente e externalizando o produto pela elevação da temperatura para induzir funções codificadas por fago. Asami *et al.* (1997) *J. Ferment. and Bioeng.*, 83:511-516, descrevem o rompimento sincronizado de células de *E. coli* pela infecção com fago T4 e Tanji *et al* (1998) *J. Ferment. and Bioeng.*, 85:74-78 descrevem a expressão controlada de genes de lise codificados no fago T4 para o rompimento moderado de células de *E. coli*.

Após a lise da célula, o DNA genômico extravasa do citoplasma para o meio e resulta em aumento significativo da viscosidade do fluido que

pode impedir a sedimentação de sólidos em um campo centrífugo. Na ausência de forças de cisalhamento, tais como aquelas exercidas durante o rompimento mecânico para quebrar os polímeros de DNA, a taxa de sedimentação mais lenta de sólidos através de um fluido viscoso resulta em uma

5 separação deficiente de sólidos e líquido durante a centrifugação. Diferentes da força de cisalhamento mecânica, existem enzimas nucleolíticas que degradam o polímero de DNA. Em *E. coli*, o gene endógeno *endA* codifica uma endonuclease (o peso molecular da proteína madura é de aproximadamente 24.5 kD) que é normalmente secretada para o periplasma e cliva o DNA em

10 oligodesoxiribonucleotídeos de uma maneira endonucleolítica. Tem sido sugerido que *endA* é fracamente expresso, relativamente, por *E. coli* (Wackemagel *et al.* (1995) *Gene* 154:55-59).

Em uma modalidade, nenhuma condição ou agentes que promovam ligações de dissulfeto adicionais são necessárias a fim de recuperar

15 o polipeptídeo identificado que contém ligação de dissulfeto em forma ativa, solúvel, da célula hospedeira. Em uma modalidade, o polipeptídeo transgênico, polipeptídeo, proteína ou seus fragmentos têm uma conformação intramolecular enovelada em seu estado ativo. Em uma modalidade, o polipeptídeo transgênico, polipeptídeo, proteína ou seus fragmentos contêm pelo menos uma ligação dissulfeto em seu estado ativo; e, talvez, até 2, 4, 6, 8,

20 10, 12, 14, 16, 18 ou 20 ou mais ligações dissulfeto.

A proteína dessa invenção pode ser isolada e purificada até uma pureza substancial por técnicas padronizadas bem conhecidas, que incluem mas não são limitadas a precipitação com sulfato de amônia ou etanol, ex-

25 tração ácida, cromatografia com troca de anion ou cátion, cromatografia em fosfocelulose, cromatografia de interação hidrofóbica, cromatografia de afinidade, cromatografia com níquel, cromatografia com hidroxapatita, cromatografia de fase reversa, cromatografia com lectina, eletroforese preparativa, solubilização com detergente, precipitação seletiva com substâncias tais

30 como cromatografia de coluna, métodos de imunopurificação e outros. Por exemplo, proteínas que têm propriedades de adesão molecular estabelecidas podem ser fundidas reversivelmente com um ligante. Com o ligante a-

apropriado, a proteína pode ser adsorvida seletivamente a uma coluna de purificação e então liberada da coluna em uma forma relativamente pura. A proteína fundida é removida, então, pela atividade enzimática. Além disso, a proteína pode ser purificada usando colunas de imunoafinidade ou colunas de Ni-NTA. Técnicas gerais são descritas em, por exemplo, R Scopes, *Protein Purification: Principles and Practice*, Springer-Verlag:N.Y. (1982); Deutscher, *Guide to Protein Purification*, Academic Press (1990); Pat. U.S. Nº 4.511.503; S. Roe, *Protein Purification Techniques: A Practical Approach* (Practical Approach Series), Oxford Press (2001); D. Bollag *et al.*, *Protein Methods*, Wiley-Lisa, Inc. (1996); AK Patra *et al.*, *Protein Expr Purif.*, 18(2): p/182-92 (2000); e R. Mukhija *et al.*, *Gene* 165(2):p. 303-6 (1995). Veja também, por exemplo, Ausubel *et al.*, (1987 e suplementos periódicos); Deutscher (1990) "Guide to Protein Purification", *Methods in Enzymology* vol 182 e outros volumes dessa série; Coligan *et al.*, (1996 e Suplementos periódicos) *Currents Protocols in Protein Science* Wiley/Greene, NY; e a literatura de fabricantes sobre o uso de produtos para purificação de proteína, por exemplo, Pharmacia, Piscataway, N.J. ou Bio-Rad, Richmond, Calif. A combinação com técnicas recombinantes permite a fusão de segmentos apropriados, por exemplo, de uma sequência FLAG ou um equivalente que pode ser fundido através de uma sequência removível com protease. Veja, por exemplo, Hochuli (1989) *Chemische Industrie* 12:69-70; Hochuli (1990) "Purification of Recombinant Proteins with Metal Chelate Absorbent" em Setlow (ed.) *Genetic Engineering, Principle and Methods* 12:87-98, Plenum Press, NY; e Crowe *et al.*, (1992) *QIA express: The High Level Expression & Protein Purification System* QIAGEN, Inc., Chatsworth, Calif.

A detecção da proteína expressa é obtida por métodos conhecidos na técnica e incluem, por exemplo, radioimunoensaios, técnicas de Western blotting ou imunoprecipitação.

Certas proteínas expressas nessa invenção podem formar agregados insolúveis ("corpos de inclusão"). Vários protocolos são adequados para purificação de proteínas a partir de corpos de inclusão. Por exemplo, a purificação de corpos de inclusão envolve tipicamente a extração, separação

e/ou purificação de corpos de inclusão pelo rompimento das células hospedeiras, por exemplo, pela incubação em um tampão de TRIS/HCl 50 mM pH 7.5, NaCl 50 mM, MgCl₂ 5 mM, DTT 1 mM, ATP 0,1 mM e PMSF 1 mM. A suspensão celular é tipicamente lisada usando 2-3 passagens através de
5 uma Prensa de French. A suspensão celular também pode ser homogeneizada usando um Polytron (Brinkman Instruments) ou sonicadas sobre gelo. Métodos alternativos de lisar bactérias são claros para aqueles versados na técnica (veja, por exemplo, Sambrook *et al.* supra; Ausubel *et al.* supra).

Se necessário, os corpos de inclusão podem ser solubilizados e
10 a suspensão de lisado celular pode ser, tipicamente, centrifugado para remover material insolúvel indesejado. As proteínas que formavam os corpos de inclusão podem ser renaturadas por diluição ou diálise com um tampão compatível. Solventes adequados incluem, mas não são limitados a ureia (entre cerca de 4 M a cerca de 8 M), formamida (pelo menos cerca de 80%
15 volume/volume) e hidrocloreto de guanidina (entre cerca de 4 M a cerca de 8 M). Embora o hidrocloreto de guanidina e agentes similares sejam desnaturantes, essas desnaturação não é irreversível e a renaturação pode ocorrer após a remoção (por diálise, por exemplo) ou diluição do desnaturante, permitindo a reformação da proteína imunológica e/ou biologicamente ativa. Outros tampões adequados são conhecidos por aqueles versados na técnica.
20

As proteínas heterologicamente expressas presentes no sobrenadante podem ser separadas das proteínas do hospedeiro por técnicas de separação padronizadas bem conhecidas por aqueles versados na técnica. Por exemplo, um fracionamento inicial com sal pode separar várias das proteínas indesejadas da célula hospedeira (ou proteínas derivadas dos meios
25 de cultura celular) da proteína ou polipeptídeo de interesse. Um exemplo pode ser o sulfato de amônia. O sulfato de amônia precipita proteínas pela redução efetiva da quantidade de água na mistura de proteína. A proteína, então, precipita baseado na sua solubilidade. Quanto mais hidrofóbica for a
30 proteína, mais provavelmente ela precipitará em concentrações menores de sulfato de amônia. Um protocolo típico inclui adicionar sulfato de amônia saturado a uma solução de proteína, tal que a concentração de sulfato de a-

mônia resultante está entre 20-30%. Essa concentração precipitará a maioria das proteínas hidrofóbicas. O precipitado é então descartado (a menos que a proteína de interesse seja hidrofóbica) e o sulfato de amônia é adicionado ao sobrenadante em uma concentração conhecida para precipitar a proteína de interesse. O precipitado é então solubilizado em tampão e o excesso de sal removido se necessário, tanto através de diálise quanto de diafiltração. Outros métodos que contam com a solubilidade de proteínas, tais como a precipitação com etanol, são bem conhecidos por aqueles versados na técnica e podem ser usados para fracionar misturas de complexos de proteína.

O peso molecular de uma proteína ou polipeptídeo de interesse pode ser usado para isolá-los de proteínas de tamanho maior ou menor usando a ultrafiltração através de membranas de diferentes tamanhos de poro (por exemplo, membranas Amicon ou Millipore). Como uma primeira etapa, a mistura de proteína pode ser ultrafiltrada através de uma membrana com um tamanho de poro que tem um limite de corte de peso molecular menor do que o peso molecular da proteína de interesse. O retentado da ultrafiltração pode então ser ultrafiltrado novamente contra uma membrana com um limite de corte molecular maior do que o peso molecular da proteína de interesse. A proteína ou polipeptídeo de interesse passarão através da membrana para o filtrado. O filtrado pode então ser cromatografado como descrito abaixo.

As proteínas ou polipeptídeos de interesse secretados também podem ser separados a partir de outras proteínas com base em seu tamanho, carga líquida da superfície, hidrofobicidade e afinidade por ligantes. Além disso, anticorpos surgidos contra as proteínas podem ser conjugadas a matrizes de coluna e as proteínas imunopurificadas. Todos esses métodos são bem conhecidos na técnica. Ficará claro para o versado que as técnicas cromatográficas podem ser realizadas em qualquer escala e usando equipamentos de vários fabricantes diferentes (por exemplo, Pharmacia Biotech).

D. Renaturação e Re-enovelamento

Em algumas modalidades da presente invenção, mais do que 50% do polipeptídeo transgênico, polipeptídeo, proteína ou seus fragmentos expressos produzidos podem ser produzidos em uma forma renaturável em

uma célula hospedeira. Em outra modalidade, cerca de 60%, 70%, 75%, 80%, 85%, 90%, 95% da proteína expressa é obtida ou pode ser renaturada para a forma ativa.

5 A proteína insolúvel pode ser renaturada ou re-enovelada para gerar uma conformação estrutural secundária e terciária da proteína. As etapas de re-enovelamento de proteína podem ser usadas, como necessário, na complementação da configuração do produto recombinante. O re-enovelamento e a renaturação podem ser obtido usando um agente que é conhecido na técnica por promover a dissociação/associação de proteínas. Por
10 exemplo, a proteína pode ser incubada com ditioneito seguido pela incubação com sal oxidado de glutathione dissódica seguido pela incubação com um tampão que contém um agente de re-enovelamento tal como a ureia.

A proteína ou polipeptídeo de interesse também podem ser renaturados, por exemplo, dialisando-os contra salina tamponada com fosfato
15 (PBS) ou acetato de Na 50 mM, tampão com pH 6 mais NaCl 200 mM. Alternativamente, a proteína pode ser re-enovelada enquanto imobilizada sobre uma coluna, tal como a coluna de Ni-NTA usando um gradiente linear de ureia 6 M-1 M em NaCl 500 mM, glicerol 20%, Tris/HCl 20 mM pH 7.4, contendo inibidores de protease. A renaturação pode ser realizada durante um
20 período de 1,5 horas ou mais. Após a renaturação, as proteínas podem ser eluídas pela adição de imidazol 250 mM. O imidazol pode ser removido por uma etapa final de diálise contra PBS ou acetato de sódio 50 mM tampão com pH 6 mais NaCl 200 mM. A proteína purificada pode ser armazenada a 4°C ou congelada a -80°C.

25 Outros métodos incluem, por exemplo, aqueles que são descritos em MH Lee *et al.*, *Protein Expr. Purif.*, 25(1): p. 166-73 (2002), W. K. Cho *et al.*, *J. Biotechnology*, 77(3): p.169-78 (2000), Ausubel *et al* (1987 e suplementos periódicos), Deutscher (1990) "Guide to Protein Purification" *Methods in Enzymology* vol 182 e outros volumes dessa série, Coligan *et al.*,
30 (1996 e suplementos periódicos) *Current Protocols in Protein Science* Wiley/Greene, N.Y., S. Roe, *Protein Purification Techniques: A Practical Approach* (Practical Approach Series), Oxford Press (2001); D. Bollag *et al.*, Prote-

in Methods, Wiley-Lisa, Inc. (1996).

E. Proteínas de interesse

Os métodos e composições da presente invenção são úteis para produzir níveis elevados de proteína ou polipeptídeo de interesse apropria-

5 damente processados em um sistema de expressão celular. A proteína ou polipeptídeo de interesse (também referidos aqui como "proteína-alvo" ou "polipeptídeo-alvo") podem ser de qualquer espécie e de qualquer tamanho. Entretanto, em certas modalidades, a proteína ou polipeptídeo de interesse são uma proteína ou polipeptídeo terapeuticamente úteis. Em algumas mo-

10 dalidades, a proteína pode ser uma proteína de mamífero, por exemplo, uma proteína humana, e pode ser, por exemplo, um fator de crescimento, uma citocina, uma quimiocina ou uma proteína sanguínea. A proteína ou polipeptídeo de interesse podem ser processados de uma maneira similar a da proteína ou polipeptídeo nativos. Em certas modalidades, a proteína ou polipeptídeo não incluem um sinal de secreção na sequência codificadora. Em cer-

15 tas modalidades, a proteína ou polipeptídeo de interesse tem menos do que 100 kD, menos do que 50 kD ou menos do que 30kD de tamanho. Em certas modalidades, proteína ou polipeptídeo de interesse é um polipeptídeo com pelo menos cerca de 5, 10, 15, 20, 30, 40, 50 ou 100 aminoácidos.

20 Informação abrangente sobre sequência necessária para técnicas de genética molecular e engenharia genética está publicamente disponível. O acesso a sequências de nucleotídeos completas de mamíferos, assim como humanos, genes, sequências de cDNA, sequências de aminoácidos e genomas pode ser obtido no GenBank no website [//www.ncbi.nlm.nih.gov/Entrez](http://www.ncbi.nlm.nih.gov/Entrez). In-

25 formações adicionais também podem ser obtidas de GeneCards, uma enciclopédia eletrônica que integra informações sobre genes e seus produtos e aplicações biomédicas do Weizmann Institute of Science Genome e Bioinformatics (bioinformatics.weizmann.ac.il/cards), informação sobre sequência de nucleotídeos também pode ser obtida do EMBL Nucleotide Sequence Data-

30 base (www.ebi.ac.uk/embl) ou do DNA Databank ou Japan (DDBJ, www.ddbi.nig.ac.jp/; sites adicionais de informação sobre sequências de aminoácidos incluem o website de recurso de informação sobre proteína de Ge-

orgetown (www-nbrf.georgetown.edu/pirl) e Swiss-Prot (au.expasy.org/sprot/sprot-top.html).

Exemplos de proteínas que podem ser expressas nessa invenção incluem moléculas tais como, por exemplo, renina, um hormônio do crescimento, incluindo o hormônio do crescimento humano; hormônio do crescimento bovino; fator de liberação de hormônio do crescimento; hormônio paratireoidiano; hormônio estimulante da tireóide; lipoproteínas; α -1-antitripsina; cadeia A da insulina; cadeia B da insulina; pró-insulina; trombo-poietina; hormônio folículo estimulante; calciotonina; hormônio luteinizante; glucagon; fatores da coagulação tais como fator VIIIc, fator IX, fator tecidual e fator de Von Willebrand; fatores anticoagulantes tais como Proteína C; fator natriurético atrial; surfactante pulmonar; um ativador de plasminogênio, tal como uroquinase ou ativador de plasminogênio do tipo da urina humana ou tecidual (t-PA); bombesina; trombina; fator de crescimento hematopoiético; fator alfa e beta de necrose de tumor; encefalinase; uma albumina sérica tal como albumina sérica humana; substância de inibição mulleriana; cadeia A de relaxina; cadeia B de relaxina; pró-relaxina; polipeptídeo associado a gonadotropina de camundongo; uma proteína microbiana, tal como beta-lactamase; Dnase; inibina; ativina; fator de crescimento endotelial vascular (VEGF); receptores para hormônios ou fatores de crescimento; integrina; proteína A ou D; fatores reumatóides; um fator neurotrópico tal como o fator neurotrópico derivado de cérebro (BDNF), neurotrofina-3, -4, -5, ou -6 (NT-3, NT-4, NT-5 ou NT-6) ou um fator de crescimento de nervo tal como NGF- β ; cardiotrofinas (fator de hipertrofia cardíaca) tal como cardiotrofina-1 (CT-1); fator de crescimento derivado de plaqueta (PDGF); fator de crescimento de fibroblasto tais como FGF e bFGF; fator de crescimento epidérmico (EGF); fator de transformação de crescimento (TGF) tais como TGF-alfa e TGF- β , incluindo TGF- β 1, TGF- β 2, TGF- β 3, TGF- β 4 ou TGF- β 5; fator de crescimento I e II semelhantes a insulina (IGF-I e IGF-II); des(1-3)-IGF-I (IGF-I cerebral), proteínas que se ligam ao fator de crescimento semelhante a insulina; proteínas CD tais como CD-3, CD-4, CD-8 e CD-19; eritropoietina; fatores ósteo-indutivos; imunotoxinas; uma proteína morfogenética óssea (BMP); um

interferon tais como interferon-alfa, -beta e -gama; fatores que estimulam colônia (CSFs), por exemplo, M-CSF, GM-CSF e G-CSF; interleucinas (ILs), por exemplo IL-1 a IL-10; anticorpo anti-HER-2; superóxido dismutase; antígeno viral tal como, por exemplo, uma porção do envelope da AIDS; proteínas de transporte; receptores residentes; endereçadores; proteínas regulatórias; anticorpos; e fragmentos de qualquer um dos polipeptídeos listados acima.

Em certas modalidades, a proteína ou polipeptídeo pode ser selecionado entre IL-1, IL-1a, IL-1b, IL-2, IL-3, IL-4, IL-5, IL-6, IL-7, IL-8, IL-9, IL-10, IL-11, IL-12, IL-12elasti, IL-13, IL-15, IL-16, IL-18, IL-18BP_a, IL-23, IL-24, VIP, eritropoietina, GM-CSF, G-CSF, M-CSF, fator de crescimento derivado de plaqueta (PDGF), MSF, ligante de FLT-3, EGF, fator de crescimento de fibroblasto (FGF; por exemplo, α -FGF (FGF-1), β -FGF (FGF-2), FGF-3, FGF-4, FGF-5, FGF-6 ou FGF-7), fatores de crescimento semelhantes a insulina (por exemplo, IGF-1, IGF-2), fatores de necrose de tumor (por exemplo, TNF, linfotoxina), fatores de crescimento de nervo (por exemplo, NGF), fator de crescimento endotelial vascular (VEGF), interferons (por exemplo, IFN- α , IFN- β , IFN- γ); fator inibitório de leucemia (LIF); fator neurotrófico ciliar (CNTF); oncostatina M; fator de célula tronco (SCF); fatores de transformação do crescimento (por exemplo, TGF- α , TGF- β 1, TGF- β 2, TGF- β 3), superfamília TNF (por exemplo, LIGHT/TNFSF14, STALL-1/TNFSF13B (BLY5, BAFF, THANK), TNF α /TNFSF2 e TWEAK/TNFSF12); ou quimiocinas (BCA-1/BLC-1, BRAK/Kec, CXCL16, ENA-78/LIX, Eotaxina-1, Eotaxina-2/MPIF-2, Exodus-2/SLC, Fratalcina/Neurotactina, GRO α /MGSA, HCC-1, I-TAC, Linfotactina/ATAC/SCM, MCP-1/MCAF, MCP-3, MCP-4, MDC/STCP-1/ABCD-1, MIP-1.quadratura., MIP-2.quadratura./GRO.quadratura., MIP-3.quadratura./Exodus/LARC, MIP-3/Exodus-3/ELC, MIP-4/PARC/DC-CK1, PF-4, RANTES, SDF1, TARC ou TECK).

Em uma modalidade da presente invenção, a proteína de interesse pode ser uma multissubunidade de proteína ou polipeptídeo. Multissubunidades de proteína podem ser expressas incluindo proteínas homoméricas e heteroméricas. As multissubunidades de proteínas podem incluir duas

ou mais subunidades, que podem ser iguais ou diferentes. Por exemplo, a proteína pode ser uma proteína homomérica que compreende 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 ou mais subunidades. A proteína também pode ser uma proteína heteromérica que inclui 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 ou mais subunidades. Subunidades de proteínas exemplares incluem: receptores que incluem receptores de canal de íon; proteínas da matriz extracelular incluindo condroitina; colágeno; imunomoduladores incluindo proteínas MHC, anticorpos de cadeia completa e fragmentos de anticorpo; enzimas que incluem RNA polimerases e DNA polimerases; e proteínas de membrana.

Em outra modalidade, a proteína de interesse pode ser uma proteína do sangue. As proteínas sanguíneas expressas nessa modalidade incluem mas não estão limitadas a proteínas de transporte, tais como albumina, incluindo albumina humana e albumina bovina, transferrina, meias moléculas de transferrina recombinante, haptoglobina, fibrinogênio e outros fatores de coagulação, componentes de complemento, imunoglobulinas, inibidores de enzima, precursores de substâncias tais como angiotensina e bradicina, insulina, endotelina e globulina, incluindo alfa, beta e gama-globulina e outros tipos de proteínas, polipeptídeos e seus fragmentos encontrados primariamente no sangue de mamíferos. As sequências de aminoácidos das numerosas proteínas sanguíneas foram descritas (veja, S.S. Baldwin (1993) *Comp. Biochem. Physiol.* 106b:203-218), incluindo a sequência de aminoácidos para a albumina sérica humana (Lawn, L., *et al.* (1981) *Nucleic Acids Research*, 9:6103-6114) e a transferrina sérica humana (Yang, F. *et al.* (1984) *Proc. Natl. Acad. Sci USA* 81:2752-2756).

Em outra modalidade, a proteína de interesse pode ser uma enzima ou cofator recombinantes. As enzimas e co-fatores expressos nessa modalidade incluem mas não são limitadas a aldolases, amina oxidases, aspartases, enzimas dependentes de B12, carboxipeptidases, carboxiesterases, carboxilases, quimiotripsina, enzimas que requerem CoA, cianidrina sintetases, cistationa sintetases, decarboxilases, desidrogenases, álcool desidrogenases, desidratases, diaforases, dioxigenases, enoato redutases, epóxido hidrases, fumarases, galacto oxidases, glicose isomerases, glicose

oxidases, glicosiltransferases, metiltransferases, nitrilo hidrases, nucleosídeo fosforilases, oxidoredutases, oxinitilases, peptidases, glicosiltransferases, peroxidases, enzimas fundidas a um polipeptídeo terapeuticamente ativo, ativador de plasminogênio tecidual, uroquinase, reptilase, estreptoquinase, catalase, superóxido dismutase, DNase, aminoácido hidrolases (por exemplo, asparaginase, amido-hidrolases), carboxipeptidases, proteases, tripsina, pepsina, quimotripsina, papaína, bromelaina, colagenase, neuraminidase, lactase, maltase, sucrase e arabinofuranosidases.

Em outra modalidade, a proteína de interesse pode ser um anticorpo de cadeia simples, fragmento Fab e/ou anticorpo de cadeia completa ou seus fragmentos ou porções. Um anticorpo de cadeia simples pode incluir regiões que se ligam ao antígeno de anticorpos sobre uma cadeia simples de polipeptídeo estavelmente enovelada. Fragmentos Fab podem ser um pedaço de um anticorpo em particular. O fragmento Fab pode conter o sítio de ligação ao antígeno. O fragmento Fab pode conter 2 cadeias: um fragmento de cadeia leve e um de cadeia pesada. Esses fragmentos podem ser ligados através de um ligante ou de uma ligação dissulfeto.

A sequência codificadora para a proteína ou polipeptídeo de interesse pode ser uma sequência codificadora nativa para o polipeptídeo-alvo, se disponível, mas muito preferivelmente será uma sequência codificadora que foi selecionada, melhorada ou otimizada para uso na célula hospedeira de expressão selecionada: por exemplo, pela síntese do gene para refletir a frequência do uso do códon de uma espécie de *Pseudomonas* tal como *Pseudomonas fluorescens* ou outro organismo adequado. O(s) gene(s) resultante(s) será (serão) construído(s) dentro ou será (serão) inserido(s) em um ou mais vetores, que serão então transformados na célula hospedeira de expressão. Um ácido nucleico ou um polipeptídeo ditos serem fornecidos em uma "forma expressível" significam um ácido nucleico ou um polipeptídeo que contêm pelo menos um gene que pode ser expresso pela célula hospedeira de expressão selecionada.

Em certas modalidades, a proteína de interesse é, ou é substancialmente, homóloga a proteína nativa, tal como uma proteína nativa de

mamífero ou proteína humana. Nessas modalidades, a proteína não é encontrada em uma forma concatamérica, mas está ligada apenas a um sinal de secreção e opcionalmente a uma sequência tag para purificação e/ou reconhecimento.

5 Em outras modalidades, a proteína de interesse é uma proteína que é ativa em uma temperatura entre cerca de 20 a cerca de 42°C. Em uma modalidade, a proteína é ativa em temperaturas fisiológicas e é inativada quando aquecida a temperaturas altas ou extremas, tais como temperaturas superiores a 65°C.

10 Em uma modalidade, a proteína de interesse é uma proteína que é ativa em uma temperatura entre cerca de 20 a cerca de 42°C e/ou é inativada quando aquecida a temperaturas altas ou extremas, tais como temperaturas superiores a 65°C; é, ou é substancialmente, homóloga a proteína nativa, tal como uma proteína nativa de mamífero ou proteína humana e não
15 é expressa a partir de ácidos nucleicos em forma concatamérica; e o promotor não é um promotor nativo em *P. fluorescens*, mas é derivado de outro organismo, tal como *E. coli*.

 Em outras modalidades, a proteína quando produzida também inclui uma sequência direcionadora adicional, por exemplo uma sequência
20 que direciona a proteína para o meio extracelular. Em uma modalidade, a sequência direcionadora adicional está operativamente ligada a terminação carboxi da proteína. Em outra modalidade a proteína inclui um sinal de secreção para um autotransportador, um sistema de secreção de dois parceiros, um sistema de ramo terminal principal ou uma fímbria que conduz porina.

25 Os seguintes exemplos são oferecidos como meio de ilustração e não como meio de limitação.

EXEMPLOS EXPERIMENTAIS

Exemplo 1. Identificação de sequência líder *dsbC*

1. MATERIAIS E MÉTODOS

30 A. Construção do Plasmídeo de Expressão pDOW2258

Técnicas de DNA recombinante padronizadas foram usadas na construção do plasmídeo pDOW2258, usado para a expressão do peptídeo

líder DsbC-proteína de fusão Skp (Figura 1).

Uma reação de amplificação por PCR foi realizada usando Hercules Master Mix (Stratagene, #600610-51), iniciadores RC-322 (5'- AAT-TACTAGTAGGAGGTACATTATGCGCTT-3', SEQ ID Nº 25) e RC-323 (5'-
 5 TATACTCGAGTTATTTAACCTGTTTCAGTA-3', SEQ ID Nº26) e o modelo de plasmídeo pDOW3001 (já contendo a fusão da sequência líder clonada *dsbC*- sequência codificadora *skf* gerada por SOE PCR) para amplificar a sequência codificadora com 521 bp de *dsbC-skp* usando o protocolo do fabricante. O fragmento de PCR foi purificado usando o QIAQUICK® Gel Ex-
 10 traction Kit (Qiagen, #28704), digerido com as nucleases de restrição *SpeI* e *XhoI* (New England Biolabs, M0202) de acordo com o protocolo do fabricante. A reação de ligação foi transformada em *P. fluorescens* DC454 (*lsc::lacI^{Q1} ΔpyrF*) por eletroporação, recuperada em meio SOC-com-Soja e plaqueada em um meio seletivo (M9 glicose agar). As colônias foram analisadas pela
 15 digestão de restrição de plasmídeo de DNA (Qiagen, cat. #27106). Dez clones contendo os insertos foram seqüenciados para confirmar a presença da sequência codificadora *dsbC-skp* correta. O plasmídeo de isolados da sequência confirmada foram designados como pDOW2258.

B. Crescimento e Análise de Expressão em Frascos de Agitação

20 Isolados da cepa DC454 de *P. fluorescens* (*lsc::lacI^{Q1} ΔpyrF*) contendo pDOW2258 foram analisados pelo protocolo padronizado de expressão em frasco de agitação Dow 1 L-scale. Resumidamente, cultura de sementes cultivadas em meio M9 suplementado com glicose 1% e elementos residuais foram usadas para inocular 200 ml de meio mínimo definido
 25 com sais com glicerol 5% como fonte de carbono. Depois de uma fase de crescimento inicial de 24 horas, a expressão através do promotor *P_{tac}* foi induzida com isopropil-β-D-1-tiogalactopiranosídeo (ITPG) 0,3 mM.

As culturas foram amostradas no momento da indução (I0), 24 horas após a indução (I24) e 40 horas após a indução (I48). A densidade celular foi medida por densidade óptica em 600 nm (OD₆₀₀). A densidade
 30 celular foi ajustada para OD₆₀₀ = 20 e alíquotas de 100 µl foram centrifugadas a 14.000 x RPM por 5 minutos e o sobrenadante foi removido.

As frações solúvel e insolúvel das amostras dos frascos de agitação foram geradas usando EASYLYSE® (Epicentre Technologies). O pélete celular foi ressuspenso e diluído 1:4 em tampão de lise e incubado com agitação em temperatura ambiente por 30 minutos. O lisado foi centrifugado a 14.000 rpm por 20 minutos (4°C) e o sobrenadante removido. O sobrenadante foi resgatado como a fração solúvel. As amostras foram misturadas 1:1 com 2 X tampão de amostra de Laemmli que contém β -mercaptoetanol (BioRad, cat. #161-0737) e fervido por 5 minutos antes de carregar 20 μ l sobre um gel Bio-Rad Criterion 12% Bis-Tris (BioRad, cat. #45-0112 Lote # cx090505(C2) e a eletroforese em 1 X tampão MES (cat # 161-0788 Lote #210001188). Os géis foram corados com SIMPLYBLUE® SafeStain (Invitrogen cat #LC6060) de acordo com o protocolo do fabricante e visualizado usando o sistema Alpha Innotech Imaging.

C. Análise de Sequenciamento N-terminal

As frações solúvel e insolúvel separadas por SDS-PAGE foram transferidas para uma membrana de Sequenciamento série PVDF (BioRad, cat.#162-0236) por 1,5 horas em 40 V usando CAPS 10 mM (2,21 g/l), pH 11 (com NaOH) e metanol 10% como tampão de transferência. O "blot" foi corado com solução corante (Comassie Brilliant Blue R-250 0,2%, metanol 40%, ácido acético 10%) por 10 segundos e então imediatamente descorado três vezes, 10 segundos cada. As bandas das proteínas de interesse foram cortadas do "blot" e seqüenciadas usando a degradação de Edman realizada em um PROCISE® Protein Sequencer (modelo 494) da Applied Biosystems (Foster City, CA).

II. RESULTADOS

A análise por SDS-PAGE confirmou acumulação significativa de proteína no peso molecular predito para Skp (~17 kDa) com 24 horas (I24) e 48 horas (I48) pós-indução em ambas as frações solúvel e insolúvel (Figura 2).

A análise de sequenciamento N-terminal confirmou que a banda solúvel induzida do tamanho esperado para a proteína Skp em I24 produziu os primeiros 5 aminoácidos da sequência predita da proteína para a forma processada de DsbC-Skp (ADKIA, SEQ ID Nº 27). A análise N-terminal tam-

bém mostrou que as duas bandas que se acumularam na fração insolúvel em I24 produziu ambas as formas, processada e não processada, de DsbC-Skp. A banda de maior peso molecular produziu os 10 primeiros aminoácidos da sequência predita da proteína para a forma não processada de DsbC-Skp (MRLTQIIAAA, SEQ ID Nº 28), enquanto que a banda de menor peso molecular produziu os 10 primeiros aminoácidos da sequência predita da proteína para a forma processada da proteína DsbC-Skp (ADKIAIVNMG, SEQ ID Nº 29). Veja as figuras 3A e 3B.

Exemplo 2. Identificação do sinal de secreção de pbp*

10 I. MATERIAIS E MÉTODOS

A. Cepas

DC206 ($\Delta pyrF$, $lsc::lacI^{Q1}$) foi construída por amplificação por PCR do gene *lacI* de pCN51/*lacI* (Scheneider *et al.* 2005) usando iniciadores que incorporam a mutação do promotor de *lacI*^{Q1} (Calos *et al.* 1980) e re-combinando o gene no locus *lsc* (levansucrase) de MB101 $\Delta pyrF$ (Schneider, Jenings *et al.* 2005b) por troca alélica.

B. Construção de transposoma para rastrear sequências de sinal de *P. fluorescens*

Um vetor de transposoma foi manipulada pela inserção do gene *kanR* (que codifica a resistência a canamicina) e um gene repórter *phoA*, que não possui o códon de partida e a sequência de sinal N-terminal, entre os sítios de ligação de transposase codificada pelo vetor ("terminações em mosaico) no vetor de transposoma pMOD-2<MCS> (Epicentre Technologies). O gene *kanR* de 1,6 kB foi purificado de pUC4-KIXX (Pharmacia) por digestão por restrição com *XhoI*, então ligado a pMOD2<MCS> digerido por *Sall* para formar pDOW1245. A sequência de sinal menos o gene *phoA* foi amplificada por PCR a partir de *E. coli* K12 (ATCC) com os sítios de *BamHI* e *XbaI* adicionados aos iniciadores. Após a digestão por restrição, o gene foi ligado em pDOW 1245 digerido por *BamHI* e *XbaI* para fazer pDOW1208. O transposoma linear foi preparado por digestão por restrição de pDOW1208 com *PshA1* e purificação em gel do fragmento de 3,3 kb flanqueado pela extremidade em mosaico usando Ultrafree DA (Amicon). Após a passagem

por uma coluna MicroBioSpin 6 (Biorad), 30 ng foram misturados com 4 unidades de transposase (Epicentre) e as alíquotas foram eletroporadas em *P. fluorescens* MB101.

C. Identificação da sequência de sinal aperfeiçoada de *pbp*

5 Um plasmídeo de expressão *pbp-pró-insulina-phoA* foi designado para fundir a proteína *pbp-pró-insulina* a uma enzima PhoA madura, tal que a acumulação de *pró-insulina-phoA* no periplasma pode ser medida e as cepas com acumulação melhorada puderam ser ensaiadas quanto a atividade de PhoA. A fusão entre a sequência de sinal de *pbp*, *pró-insulina* humana
10 e *phoA* em pINS-008 foi construída por SOE-PCR (Horton *et al.* 1990), usando iniciadores que superpõem a sequência codificadora para a sequência líder e *pró-insulina* e para *pró-insulina* e a forma madura de PhoA (isto é sem a sequência líder nativa). A fusão foi clonada sob o controle do promotor *tac* em pDOW1169 (Schneider *et al.* 2005a; Schneider, Jennings *et al.*
15 2005b) que foi digerido por restrição com *SpeI* e *XhoI* e tratada com fosfatase alcalina de camarão, e então ligada e eletroporada em DC206, para formar pINS-008. O gene padrão de *pró-insulina* foi códon-otimizado para expressão em *P. fluorescens* e sintetizado (DNA 2.0) O gene *phoA* foi amplificado a partir do DNA genômico de *E.coli* MG1655. As colônias foram rastreadas sobre placas de agar contendo BCIP, um indicador colorimétrico da
20 atividade de fosfatase alcalina, com ITPG para induzir a expressão do gene de *pbp-pró-insulina-phoA*. Das colônias que exibiram hidrólise de BCIP, uma cresceu muito maior do que as outras. Esse isolado foi encontrado ter uma única mutação de C para T na região que codifica o peptídeo de secreção, causando uma alteração de alanina para valina no aminoácido 20 (A20V, SEQ ID N° 2; veja Tabela 6).

A expressão das duas cepas foi avaliada pelo protocolo padrão de frasco de agitação. O crescimento de ambas estabilizou-se pouco depois da adição do indutor ITPG. A atividade de fosfatase alcalina na cepa mutante pINS-008-3 foi 3-4 vezes maior (Figura 6) e a acumulação da proteína
30 (solúvel) foi maior (Figura 7).

Tabela 6. Sinais de secreção Sec identificados em *P. fluorescens*

Função da Proteína Madura	Abreviação	Sequência de sinal predita (SinalP-HMM)	SEQ ID NO:	Sequência de DNA	SEQ ID NO:
pbp (sequência de sinal mutante)	pbp*	MKLKRLMAAMTFV AAGVATVNAVA	2	atgaaactgaacggttgatggcg gcaatgactttgtcgtcgtggcg ttcgaccgtcaacgcgtggcc	1
precursor E1 de porina	PO	MKKSTLAVAVTLGA IAQQAGA	31	atgaagaagtcaccttggctgtg gctgaacgttggcgcaatcgc ccagcaagcagggcgt	30
Porina F da membrana externa	OP	MKLKNTLGLAIGSLI AATSFGVLA	33	atgaaactgaacacacttggcg cttggcacttggctcttattggcg ctacttcttggcggttctggca	32
Proteína periplasmática de ligação ao fosfato (pbp)	PB	MKLKRLMAAMTFV AAGVATANAVA	35	atgaaactgaacggttgatggcg gcaatgactttgtcgtcgtggcg ttcgaccgtcaacgcgttggcc	34
azurina	AZ	MFAKLVAVSLTLA SGQLLA	37	atgttgcacaaactcgttgcgtttc cttgcgtactctggcgagcgcc agtgtgtgtct	36
precursor da lipoproteína B raia	L	MIKRNLVMGLAVL LSA	39	atgacaaacgcaatctgctgttta tggcgcttgcgtgtgtgagcg ct	38
Proteína de ligação a lisina-arginina-ornitina	LAO	MQNYKKFLAAAV SMAFSATAMA	41	atgcagaactataaaaattcttctc tggcgcgcggtctcgtatggcg ttcagcgccacgccaatggca	40
Proteína de ligação a Ferro(III)	IB	MIRDNRLKTSLLRG LTLTLLSLTLLSPAA HS	43	atgatccgtgacaaccgactcaa gacatcccttctcgtcggttgac cctcaccctactcagcctgacct gtctctcgcgcgcgcgcattct	42

D. Sequenciamento Genômico

O DNA genômico foi purificado pelo kit DNA Easy (Invitrogen) e 10 µg foram usados como modelo para o sequenciamento com um iniciador específico para transposon usando 2X ABI PRISM BigDye Terminators v3.0 Cycle Kit (Applied Biosystems). As reações foram purificadas e carregadas no ABI PRISM 3100 Genetic Analyzer (Applied Biosystems) de acordo com as orientações do fabricante.

E. Clonagem de regiões codificadoras de sequência de sinal

As sequências de sinal foram determinadas pelo programa SPS-can ou como em (De *et al* 1995). Os resultados desses experimentos foram descritos no Pedido de Patente U.S. Número 20060008877 copendente, depositado em 22 de Novembro de 2004. As líderes de secreção de porina F da membrana externa (*oprF*), proteína que se liga a fosfato (*pbp*), porina E1 (*porE*), azurina, lipoproteína B e proteína que se liga ao ferro foram amplifi-

cadas por reação em cadeia da polimerase (PCR). Os produtos de PCR re-
 sultantes foram clonados no vetor pCRII Blunt TOPO e transformados em
 células Top 10 de *E. coli* (Invitrogen) de acordo com o protocolo do fabrican-
 te. Os transformantes resultantes foram rastreados quanto ao inserto correto
 5 pelo sequenciamento com os iniciadores à frente M13 e reverso M13. Os
 clones positivos foram denominados como se segue: oprF (pDOW1112), pbp
 (pDOW1113), porinaE1 (pDOW1183), azurina (pDOW1180), lipoproteína B
 (pDOW1182), proteína que se liga ao ferro (pDOW1181).

F. Construção de clones de gal2 scFv para secreção em *P. fluorescens*

10 As sequências de sinal de oprF e pbp foram amplificadas para fun-
 dir com a sequência codificadora de Gal2 na posição +2 usando pDOW1112
 ou pDOW1113 como modelo. A sequência codificadora de gal2 foi amplifi-
 cada usando pGal2 (Martineau *et al.* 1998) como modelo. O produto de 837
 bp de SOE-PCR foi clonado no vetor PCR BLUNT II TOPO e a sequência
 15 confirmada. O gene de scFv fundido à sequência de sinal foi removido do
 vetor TOPO com as enzimas de restrição XbaI e SalI e clonado nos sítios de
 SpeI e XhoI de pMYC1803 para produzir pDOW1122 (oprF:gal2) e
 pDOW1123 (pbp:gal2) usando técnicas de clonagem padronizadas (Sam-
 brook *et al.* 2001). Os plasmídeos resultantes foram transformados em
 20 DC191 selecionada sobre LB agar suplementado com tetraciclina 30 µg/ml e
 canamicina 50 µg/ml.

A sequência de sinal de porE de pDOW1183 e fundida por SOE-
 PCR a gal2 amplificada a partir de pDOW1123 e os produtos de PCR foram
 purificadas por extração em gel. O produto de PCR resultante foi clonado em
 25 PCRII Blunt TOPO e subsequentemente transformados em células Top10 de
E. coli de acordo com as instruções do fabricante (INVITROGEN). Os clones
 resultantes foram seqüenciados e um clone positivo (pDOW1185) seleciona-
 do para subclonagem. pDOW1185 foi digerido por restrição com SpeI e SalI
 e o fragmento porE-gal2 foi purificado em gel. O fragmento purificado foi li-
 30 gado a pDOW1169 digerido por SpeI-SalI usando T4 DNA ligase (NEB). A
 mistura de ligação foi transformada em DC454 eletrocompetentes e selecio-
 nada sobre placas M9 glicose 1% agar. Os transformantes foram rastreados

por digestão de restrição de plasmídeo de DNA usando *SpeI* e *Sall*. Um clone positivo foi isolado e armazenado como pDOW1186.

As sequências de sinal de azurina, proteína que se liga ao ferro e lipB foram amplificadas a partir dos clones pDOW1180, pDOW1181 e pDOW1182, respectivamente. O gene *gal2* foi amplificado a partir de pDOW1123 usando os iniciadores apropriados para fundir cada líder de secreção, e os produtos de PCR resultantes foram isolados e fundidos por SOE-PCR como descrito acima. Os produtos de SOE-PCR foram clonados em PCR-BLUNT II TOPO, os clones resultantes foram seqüenciados e os clones positivos para cada fusão foram subclonados em pDOW1169 como descrito acima.

G. Construção de um vetor de secreção em *P. fluorescens* com tag de histidina C-terminal

Um clone contendo um inserto com a líder de secreção de pbp, MCS com His tag C-terminal e terminadores transcrpcionais *rrnT1T2* foi sintetizado por DNA 2.0 (pJ5:G03478). O cassete de secreção com 450 bp foi isolado por digestão com restrição com *SpeI* e *NdeI* e purificado em gel. O fragmento foi ligado a pDOW1219 (derivado de pMYC1803 (Shao *et al.*2006)) digerido com as mesmas enzimas. Os produtos de ligação foram transformados em *E. coli* JM109 quimicamente competentes. O plasmídeo de DNA foi preparado e rastreado quanto ao inserto por PCR usando iniciadores específicos para o vetor. O plasmídeo resultante teve a sequência confirmada e foi denominado pDOW3718. DC454 eletrocompetente de *P. fluorescens* foi então transformada com o plasmídeo e selecionada sobre LB agar suplementado com uracil 250 µg/ml e tetraciclina 30 µg/ml.

Fases de leitura aberta que codificam proteínas humanas foram amplificadas usando modelos da coleção de ORFeome humano (Invitrogen). Produtos de PCR foram digeridos por restrição com *NheI* e *XhoI* e então ligados a pDOW3718 digerido por *NheI-XhoI*. Os produtos de ligação foram subsequentemente transformados em DC454 eletrocompetente de *E. coli* e os transformantes selecionados em LB agar suplementado com uracil 250 µg/ml e tetraciclina 30 µg/ml. Os clones positivos foram seqüenciados para confirmar a sequência do inserto.

H. Construção de clones de secreção de *E. coli*

ORFs humanas foram amplificadas como acima, exceto que os iniciadores foram estruturados com o sítio de *Nco*I no iniciador 5' e *Xho*I no iniciador 3'. Os produtos de PCR foram digeridos por restrição com *Nco*I e *Xho*I (NEB), então purificados usando o kit Qiaquick Extraction (Qiagen). Os produtos digeridos foram ligados em pET22b (Novagen) digerido com *Nco*I – *Xho*I usando T4 DNA ligase (NEB) e os produtos de ligação foram transformados em células Top10 quimicamente competentes de *E. coli*. Os transformantes foram selecionados em placas com LB agar ampicilina (Tecknova). O plasmídeo de DNA foi preparado (Qiagen) e os clones positivos foram sequenciados para confirmar a sequência do inserto. Um plasmídeo clonado confirmado para cada um foi subsequentemente transformado em BL21(DE3) (Invitrogen) para análise de expressão.

I. Sequenciamento de DNA

As reações de sequenciamento (Big dye versão 3.1 (Applied Biosystems)) foram purificadas usando G-50 (Sigma) e carregadas no seqüenciador ABI3100.

J. Análise de expressão de alto rendimento (HTP)

As cepas de *P. fluorescens* foram analisadas usando o protocolo de expressão padronizado Dow HTP. Resumidamente, culturas de sementes cultivadas em M9 suplementado com glicose 1% e elementos residuais foram usadas para inocular 0,5 ml meio mínimo com sais definido com glicerol 5% como uma fonte de carbono em uma placa de 96 poços funda de 2,0 ml. Após uma fase de crescimento inicial a 30°C, a expressão através do promotor *P_{tac}* foi induzida com isopropil- β -D-1-tiogalactopiranosídeo (IPTG) 0,3 mM. A densidade celular foi medida por densidade óptica em 600nm (OD₆₀₀).

K. preparação de amostra de HTP para análise por SDS-PAGE

As frações solúveis e insolúveis de amostras de cultura foram geradas usando EASY LYSE[®] (Epicentre Technologies, cat#RP03750). Um total de 25 μ l de amostra de caldo de cultura foram lisados pela adição de 175 μ l de tampão EASY LYSE[®], incubados com agitação moderada em temperatura ambiente por 30 minutos. O lisado foi centrifugado a 14.000 rpm

por 20 minutos (4°C) e o sobrenadante removido. O sobrenadante foi salvo como a fração solúvel. O pélete (fração insolúvel) foi então ressuspense em um volume igual de tampão de lise e ressuspense por pipetamento para cima e para baixo. Para os clones selecionados, amostras de caldo de cultura sem célula foram descongeladas e analisadas sem diluição.

L. Expressão e análise de secreção de proteínas ou polipeptídeos de interesse

As culturas de semente, cultivadas em 1X M9 suplementado com glicose 1% (Tecknova), suplementado com solução de elementos residuais foram usadas para inocular 50 ml de meio mínimo definido com sais DOW com inóculo 2% e incubadas a 30°C com agitação. As células foram induzidas com ITPG 0,3 mM (isopropil- β -D-1-tiogalactopiranosídeo)~24 horas decorridas do tempo de fermentação (EFT). As amostras foram retiradas no momento da indução (I0) e 16 (I16), 24 (I24) ou 40 (I40) horas após a indução para as análise. A densidade celular foi medida por densidade óptica em 600nm (OD₆₀₀). A densidade celular foi ajustada para OD₆₀₀ =20 e 1 ml foi centrifugado a 14.000 x g por cinco minutos. Os sobrenadantes (caldo CE cultura sem célula) foram pipetados para um novo tubo de microcentrifugação, então os pellets celulares e as amostras de caldo de cultura sem célula foram congelados a -80°C para processamento posterior.

M. Análise SDS-PAGE

As frações solúvel e insolúvel das amostras do frasco de agitação foram geradas usando tampão EASY LYSE™ (Epicentre Technologies). O pélete congelado foi ressuspense em 1 ml de tampão de lise. Cinquenta microlitros foram adicionados a um adicional de 150 μ l de tampão EASY LYSE® e incubados com agitação em temperatura ambiente por 30 minutos. O lisado foi centrifugado a 14.000 rpm por 20 minutos (4°C) e o sobrenadante removido. O sobrenadante foi salvo como a fração solúvel. O pélete foi então ressuspense em um volume igual (200 μ l) de tampão de lise e ressuspense por pipetagem para cima e para baixo; isso foi salvo como a fração insolúvel. Amostras de caldo de cultura sem célula foram descongeladas e usadas sem diluição.

N. Análise Western

As frações solúveis e insolúveis preparadas e separadas por SDS-PAGE foram transferidas para nitrocelulose (BioRad) usando 1X tampão de transferência (Invitrogen) preparado de acordo com o protocolo do fabricante, por 1 hora a 100 V. Após a transferência, o blot foi bloqueado com o diluente POLY-HRP (Research Diagnostics, Inc.) e sondado com uma diluição de 1:5000 de anticorpo anti-His tag (Sigma ou US Biologicals). O blot foi lavado com 1 X PBS-Tween e subsequentemente desenvolvido usando o Kit Immunopure Metal Enhanced DAB Substrate (Pierce).

10 O. Fermentação de 20 l

Os inóculos para as culturas do fermentador foram gerados, cada um, pela inoculação de um frasco de agitação contendo 600 ml de um meio definido quimicamente suplementado com extrato de levedura e glicérol com uma cultura de estoque congelada. Após 16 a 24 h de incubação com agitação a 32°C, a cultura do frasco de agitação foi então transferida assepticamente para um fermentador de 20 l contendo um meio designado para suportar uma grande biomassa. O oxigênio dissolvido foi mantido em um nível positivo no líquido da cultura pela regulação do fluxo de ar aspergido e das taxas de agitação. O pH foi controlado no ponto desejado através da adição de amônia aquosa. O processo de fermentação com alimentação de alta densidade foi dividido em uma fase inicial de crescimento de aproximadamente 24 h e uma fase de expressão de gene na qual foi adicionado ITPG para iniciar a expressão de gene recombinante. A fase de expressão da fermentação foi então deixada processar por 24 horas.

25 P. Análise da Sequência de Aminoácido N-terminal

As amostras foram corridas como descrito na análise SDS-PAGE acima e transferidas para uma membrana Criterion Sequi-Blot PVDF (Biorad). A membrana foi corada com o reagente corante GelCode Blue (Pierce) e subsequentemente descoradas com metanol 50%, ácido acético 1%, enxaguadas com metanol 10% seguido por água deionizada e então secadas. As bandas de interesse foram cortadas da membrana, extraídas e submetidas a 8 ciclos de degradação de Edman em um sistema de sequencia-

mento de proteína Procise, modelo 494 (Applied Biosystems, Foster City, CA). P Edman, *Acta Chem. Scand.* 4, 283 (1950); revisado por R. A. Laursen *et al.*, *Methods Biochem. Anal.* 26, 201-284 (1980).

II. RESULTADOS

5 A. Identificação de sequências de sinal de secreção nativas por mutagenese de transposon

Para identificar sequências de sinal de *P. fluorescens* que poderiam secretar uma proteína heteróloga para o periplasma ou caldo de cultura, um gene repórter de secreção foi clonado em um transposoma. O gene repórter de secreção usado é um gene de fosfatase alcalina de *E. coli* (*phoA*) 10 sem um códon de partida ou sequência de sinal N-terminal. PhoA é ativo no periplasma (mas não no citosol) devido a formação de ligações de dissulfeto intramoleculares que permitem a dimerização para a forma ativa (Derman *et al.* 1991). Um método similar referido como "rastreamento genômico" 15 foi usado para encontrar proteínas secretadas em *E. coli* (Bailey *et al.* 2002). O gene *phoA* também foi usado para analisar sinais de secreção no periplasma, membrana e as proteínas exportadas em *E. coli* (Manoil *et al.* 1985) e em outras bactérias (Gicquel *et al.* 1996). Após a eletroporação e plaqueamento sobre um meio indicador, foram isoladas oito colônias azuis. O sítio 20 de inserção do transposoma no genoma foi seqüenciado e usado para pesquisar uma banco de dados específico de genoma de *P. fluorescens* MB101. Oito fusões de gene identificadas como capazes de expressar PhoA ativa são mostrados na Tabela 6.

B. Clonagem de fusões de sequência de sinal-gal2

25 As sequências de sinal das proteínas secretadas identificadas acima, porina F da membrana externa (OP), proteína *porE* que se liga ao fosfato (PB), proteína que se liga ao ferro (IB), azurina (AZ), lipoproteína B (L) e proteína que se liga a lisina-ornitina-arginina (LOA) foram preditas usando o programa SignalP (J.D. Bendtsen 2004). As sequências de sinal 30 para OP, PE e AZ foram previamente identificadas em outros sistemas [Arvidsson, 1989 #25; De, 1995 #24; Yamano, 1993 #23]. A atividade de uma líder de secreção adicional identificada em outro estudo, *pbpA20V* (Schenei-

der *et al.*2006), também foi analisada em paralelo. Nesse estudo, a região codificadora de seis sequências de sinal nativas de *P. fluorescens* e um mutante da sequência de sinal da proteína que se liga ao fosfato de *P. fluorescens* (veja Tabela 6) foram fundidas, cada uma, ao gene de *gal2* scFv usando "splicing" de superposição de extensão PCR (SOE-PCR) como descrito em Materiais e Métodos tal que os 4 aminoácidos N-terminais de Gal2 após a clivagem do peptídeo de sinal poderiam ser AQVQ. Tentativas repetidas para amplificar a sequência de sinal de LAO falharam e essa sequência de sinal foi excluída da análise posterior. As fusões de gene foram clonadas no vetor de expressão de *P. fluorescens* pDOW1169 e transformadas na cepa hospedeira DC454 ($\Delta pyrF$ *Isc::laI^{Q1}*). As cepas resultantes foram subsequentemente avaliadas quanto a expressão de Gal2 scFv e o processamento apropriado de líderes de secreção.

C. Expressão de Gal2 scFv secretada

Na escala de frasco de agitação, as fusões de PB, OP, PO, Az, IB e L a *gal2* scFv atingiram a OD₆₀₀ esperada, exceto L-*gal2* scFv, que falhou em crescer após a subcultura em meio de produção (dados não mostrados). Análises por Western blot confirmaram que as sequências de sinal de PB, OP, PO, AZ e IB foram clivadas da fusão com Gal2 scFv. Entretanto, a análise por Western blot mostrou a presença de PB-Gal2 e OP-Gal2 não processadas. Algumas Gal2 scFv solúveis expressas a partir de fusões com AZ e IB foram encontradas em meio de cultura sem célula, indicando que a proteína solúvel foi expressa e extravasada para o espaço periplasmático. A análise da sequência amino terminal foi realizada para confirmar a clivagem da sequência de sinal. A proteína insolúvel Gal2 expressa a partir de fusões com azurina (pDOW1191) mostra uma mistura de proteína com sinal de secreção processado e não processado. Entretanto, a sequência de sinal foi observada estar totalmente processada a partir da fusão de IB-Gal2.

A expressão de Gal2 scFv fundida com cada uma das sete líderes foi avaliada na escala de fermentação de 20 l usando condições de fermentação padronizadas. Todas as cepas cresceram como esperado, alcançando indução em OD₆₀₀ (~180 unidades) em 18 a 24 horas. A cepa lipB-

Gal2 cresceu discretamente mais lentamente do que as outras cepas. Isso não era totalmente inesperado já que a cepa lipB-Gal2 não cresceu após a inoculação do meio do frasco de agitação em fermentação em pequena escala. A expressão e o processamento de Gal2 scFv foi avaliada por SDS-PAGE e Western blot. A análise por SDS-PAGE mostrou que altos níveis de Gal2 eram expressos quando fundida aos sinais de secreção de OP ou PB. Entretanto, apenas uma porção (~ 50%) da proteína de fusão OP-Gal2 pareceu ser secretada para o periplasma com a sequência de sinal clivada. Como observado em pequena escala, Gal2 foi predominantemente expressa na fração insolúvel, embora a proteína solúvel fosse detectada por Western blot. Uma pequena quantidade de proteína também foi detectada no sobrenadante da cultura, indicando extravasamento para o periplasma (Figura 7). A análise de sequência N-terminal confirmou que as líderes de ibp e azurina foram processadas como esperado, resultando na sequência de aminoácido N-terminal AQVQL (SEQ ID Nº 44). Igualmente, a líder de secreção de PorE pareceu ser processada pela análise Western e foi confirmada pela análise N-terminal. O nível de expressão de PorE-Gal2 foi discretamente menor do que o de ibp-Gal2 e azurina-Gal2 insolúveis. LipB-Gal2 mostrou a expressão de Gal2 processada em níveis similares àqueles de PorE-Gal2. A maior quantidade de proteína foi observada a partir de cepas que expressam pbp-Gal2 e pbpA20V-Gal2. A quantidade de Gal2 expressa a partir da cepa pbpA20V-Gal2 pareceu ser até maior do que aquela produzida pelas cepas pbp-Gal2 (Figura 6). Gal2 solúvel processada foi detectada pela análise Western, como se fosse uma mistura de proteína insolúvel processada e não processada (Figura 7). A análise da sequência N-terminal da proteína insolúvel confirmou uma mistura de Gal2 não processada e corretamente processada.

Exemplo 3. Identificação de sequência líder de bce

I. MATERIAIS E MÉTODOS

BceL é uma líder de secreção que foi identificada ser codificada pela parte do inserto de DNA que contém o gene para uma hidrolase de *Bacillus coagulans* CMC 104017. Essa cepa de *Bacillus coagulans* também é

conhecida como NCIMB 8041, ATCC 10545 e DSMZ 2311 em várias coleções de cultura comerciais e tem sua origem como NRS784. NRS784 é da coleção NR Smith de bactérias que formam esporo (Smith *et al* Aerobic spore forming bactéria US. Dep. Agr. Monogra. 16:1-148(1952)). A outra referência original para essa cepa citada por NCIMB é Cambell, L.L. e Sniff E.E. (1959, J. Bacteriol. 78:267. An investigation on Folic acid requirements of *Bacillus coagulans*).

Análise de Sequência e Bioinformática

Um inserto de BNA de 4.127 bp de *Bacillus coagulans* CMC 104017 foi seqüenciado e analisado para localizar sequências codificadoras que codificam potencialmente uma enzima hidrolase. Uma sequência codificadora de 1314 bp, designada CDS1, foi identificada atrás do promotor *lac* na terminação 5'. O DNA e as sequências de proteína preditas para CDS1 estão descritos na SEQ ID Nº 45 e 46, respectivamente. CDS1 foi determinada codificar mais provavelmente uma hidrolase baseada na análise BLAST da sequência predita da proteína. A sequência de CDS1 mostrou homologia (valor de E: 2e-36) com beta-lactamase de *Rhodopseudomonas palustris* HaA2. A análise SignalP 3.0 pelo modelo de Markov oculto (Bendtsen JD, Nielson G. Von Heijne G, Brunak S: Improved prediction of signal peptides: Signal 3.0. *J. Mol. Biol.* 2004, 340:783) de CDS1 previu a presença de uma sequência de sinal para o organismo da classe de bactérias Gram-positivas com um sítio de clivagem de peptidase de sinal entre os resíduos 33/34 de SEQ ID Nº 46.

Construção de Plasmídeos de Expressão de Proteína

Métodos de clonagem padronizados foram usados na construção de plasmídeos de expressão (Sambrook J, Russell D: Molecular Cloning a Laboratory Manual, terceira edição. Cold Spring Harbor: Cold Spring Harbor Press; 2001). As fusões de sequência de DNA foram realizadas usando o método SOE-PCR (Horton, R.M., Z. Cai, S.N. Ho e L.R.Pease (1990). "Gene splicing by overlap extension: tailor-made genes using the polymerase chain reaction". *BioTechniques* 8(5):528-30, 532, 534-5)). Phusion DNA polymerase (New England Biolabs cat#F531S) foi usada para todas as rea-

ções de PCR.

Os plasmídeos foram designados para expressar e localizar uma proteína esterase de *Bacillus coagulans* CMC104017 tanto no citoplasma quanto no espaço periplasmático de *P. fluorescens*. Os produtos finais de PCR foram digeridos com as endonucleases de restrição *SpeI* e *XhoI* (New England Biolabs cat. #R0133 e #R0146) e então ligados no vetor de expressão pDOW1169, também digerido com *SpeI* e *XhoI*, usando T4 DNA ligase (New England Biolabs cat. #M0202S) para produzir o vetor de expressão p484-001 de CMC104641 CDS-1 citoplasmática e o vetor de expressão p484-002 da sequência líder de Bce nativa CMC104641 CDS-1. As misturas de reação de ligação foram então transformadas na cepa DC454 ($\Delta pyrF$, lac^{Q1}) de *P. fluorescens* por eletroporação, recuperadas em meio SOC com soja (Tecknova cat. #2S2699) e plaqueadas em meio seletivo (M9 glicose agar, Tecknova cat #2M1200). As colônias foram analisadas com digestão por restrição de plasmídeo de DNA miniprep (Qiagen, cat. #27106). Dez clones de cada transformação foram seqüenciados para confirmar o inserto correto.

Análise de Expressão

A cepa DC454 de *P. fluorescens* carregando cada clone foi examinada em frascos de agitação contendo 200 ml de meio mínimo definido com sais com glicerol 5% como a fonte de carbono ("Dow Medium"). Após uma fase de crescimento inicial, a expressão através do promotor *tac* foi induzida com isopropil- β -D-1-tiogalactopiranosídeo 0,3 mM (ITPG). As culturas foram amostradas no momento da indução (I0) e 24 horas após a indução (I24). A densidade celular foi medida por densidade óptica em 600 nm (OD_{600}). A tabela que mostra o esquema de numeração do frasco de agitação é mostrada na Tabela 7.

Cepa Hospedeira	Número do plasmídeo (gene-líder)	Número do Frasco		
DC 454	P484-001 (484 citoplasmático)	EP484-001	EP484-002	EP484-003
DC454	P484-002 (484 líder nativa)	EP484-004	EP484-005	EP484-006

Em cada período de amostragem, a densidade celular das a-

mostras foi ajustada para $OD_{600} = 20$ e alíquotas de 1 ml foram centrifugadas a 14.000 x g por cinco minutos. Os sobrenadantes (caldo de cultura sem célula) foram pipetados para novos tubos de microcentrifugação e então os pellets celulares e as amostras de caldo de cultura sem célula foram congelados a -20°C .

Lise Celular e Análise por SDS-PAGE

As frações solúveis e insolúveis das amostras dos frascos de agitação foram geradas usando Easy Lyse (Epicentre Technologies). O pélete congelado foi ressuspensão e diluído 1:4 em tampão de lise e incubado com agitação em temperatura ambiente por 30 minutos. O lisado foi centrifugado a 14.000 rpm por 20 minutos (4°C) e o sobrenadante removido. O sobrenadante foi salvo como a fração solúvel. O pélete (fração insolúvel) foi então ressuspensão em igual volume de tampão de lise e ressuspensão por pipetagem para cima e para baixo. As amostras de caldo de cultura sem célula foram descongeladas e usadas sem diluição. As amostras foram misturadas 1:1 com 2X tampão de amostra Laemmli contendo β -mercaptoetanol (BioRad, cat. #161-0737) e fervidas por 5 minutos antes do carregamento de 20 μl sobre um gel Bio-Rad Criterion 10% Criterion XT (BioRad, cat. #450112) e separadas por eletroforese no tampão recomendado 1X MOPS (cat. #161-0788 Lote #210001188). Os géis foram corados com SIMPLY-BLUE™ SafeStain (Invitrogen cat. #LC6060) de acordo com o protocolo do fabricante e visualizados usando o sistema Alpha Innotech Imaging. A quantidade de proteína das bandas de gel de interesse foi estimada por comparação com modelos de proteína BSA carregadas no mesmo gel.

II-RESULTADOS

Um total de seis frascos de agitação (3 frascos por cepa) foi usado para avaliar a expressão de hidrolase. O crescimento das cepas periplasmáticas e citoplasmáticas designadas foi consistente com o crescimento normal das cepas de *P. fluorescens*, atingindo uma OD_{600} de aproximadamente 15 em vinte e quatro horas após a indução. A análise por SDS-PAGE foi realizada para avaliar a expressão da proteína hidrolase (proteína CDS1) no momento da indução e 24 horas após a indução. As frações solúvel, inso-

lúvel e de caldo de cultura sem células foram analisadas por SDS-PAGE. Para a cepa CDS-1 citoplasmática (p484-001), a proteína da tamanho esperado para a hidrolase citoplasmática (44,1 kDa) acumulou-se quase que totalmente na fração solúvel em I24 (24 horas após a indução com ITPG) em todos os três isolados com um rendimento estimado de 0,1 mg/ml. A Figura 8 mostra os resultados representativos da cepa citoplasmática avaliada como EP484-003. Uma banda insignificante foi detectável na fração insolúvel e nenhuma proteína CDS1 foi detectada no caldo de cultura sem célula. Para a cepa periplasmática que expressa a líder Bce nativa de CDS1 (p484-002), a proteína do tamanho esperado para a esterase nativa acumulou-se quase que totalmente na fração solúvel em I24 em todos os três isolados com um rendimento estimado de 0,8 mg/ml. A Figura 8 mostra os resultados representativos para a cepa periplasmática contendo a líder de fusão Bce avaliada como EP484-004. Não ficou claro se a esterase nativa expressa foi inteiramente processada por que o carregamento do gel usado tornou difícil discernir entre o tamanho predito não processado de 47,6 kDa e o tamanho processado de 44,1 kDa. Semelhante aos resultados com a cepa de expressão citoplasmática, uma banda insignificante de tamanho esperado foi detectável na fração insolúvel e nenhuma proteína CDS1 foi detectada no caldo de cultura sem célula. A sequência traduzida da Líder de Bce de interesse está descrita na SEQ ID Nº 8.

Exemplo 4. Identificação e análise de líderes de secreção de *P. fluorescens*

6433 ORFs traduzidas do genoma de MB214 foram analisadas com o programa de predição de peptídeo de sinal, SignalP 2.0 (Nielsen, H. *et al.* Protein Eng., 1977. 10(1): p. 1-6). 1326 foram preditas pelo modelo HMM para conter uma peptidase de sinal. Essas proteínas foram analisadas com PsortB 2.0 (Gardy, J.L. *et al.*, Bioinformatics, 2005. 21(5)p.617-23) e todas aquelas com uma localização final de PsortB identificadas como citoplasmáticas ou da membrana citoplasmática foram removidas restando 891. 82 proteínas, para as quais a probabilidade por SignalP HMM de conter um peptídeo de sinal estava abaixo de 0,79, foram removidas restando 809. O ponto de corte de 0,79 foi escolhido por que era o valor mais alto que não

excluía aprA (RXF04304, conhecida por ser uma proteína extracelular). As sequências amino terminais dessas 809 ORFs traduzidas contendo o peptídeo de sinal como predito pelo algoritmo SignalP Neural Network mais os 7 primeiros aminoácidos da proteína processada foram alinhadas usando
 5 CLUSTALX 1.81 (Thompson, J.D. *et al.* Nucleic Acids Res., 1997. 25(24): p.4876-82).

Huber *et al.* sugeriram que sequências de sinal altamente hidrofóbicas são mais provavelmente serem secretadas cotraducionalmente (Huber, D. *et al.* J. Bacteriol., 2005. 187(9)p:2983-91). Para o propósito de identificação de proteínas secretadas cotraducionalmente, os índices de aminoácidos de Wertz-Scheraga (WS) (Wertz, D.H. e H.A. Scheraga, Macromolecules, 1978. 11(1):p.9-15), foram encontrados serem os melhores. Para esse estudo, esses índices foram obtidos do AAindex no site da web
 10 www.genome.jp/dbjet-bin/www_bget?aax1:WERD780101. Um algoritmo relatado por Boyd (Boyd, D., C. Schierle e J. Beckwiyh, Protein Sci, 1998. 7(1):p.201-5) foi modificado e usado para a classificação das 809 proteínas baseado na hidrofobicidade. O algoritmo rastreia cada sequência calculando a média dos escores WS dentro de uma janela de 12. A região mais hidrofóbica é usada para avaliar o escore WS para a proteína total. Isso rendeu 142
 15 sequências de sinal com escores WS maiores do que 0,69, o ponto de corte definido por Huber *et al.* Essa pequena lista foi comparada com os dados dos experimentos com proteoma completo de 2D-LC realizado pelo Indiana Centers for Applied Protein Sciences (INCAPS). Esses experimentos tentaram identificar e quantificar todas as proteínas expressas em MB214 (descendente de *P. fluorescens* MB101) sob uma variedade de condições de crescimento. Uma proteína que aparece nessa lista com níveis elevados de expressão máxima é provavelmente altamente expressa. Nesses dados, um
 20 escore de prioridade de 1 ou 3 indica alta confiança na identificação da proteína. As proteínas da lista de 142 que foram identificadas nos experimentos do INCAPS com uma prioridade de 1 ou 3 estão listadas na Tabela 8 em
 30 ordem de seus níveis máximos de expressão.

Tabela 8. 7 proteínas únicas da lista de 142 com prioridade 1 ou 3 (indicando alta confiança na identificação) listadas em ordem de níveis máximos de expressão encontrados durante os experimentos de INCAPS.

Prioridade	ID da Proteína	Função Coadjuvante	Max
1	RXF05550.1	proteína da família de tetratricopeptídeo de repetição	377264,2
1	RXF08124.1	proteína quimiotóxica que aceita metilação	134887,4
1	RXF07256.1	proteína TolB	88429.16
3	RXF07256.1_a1	proteína TolB	84020,51
3	RXF074046.2_a1	citocromo c oxidase, subunidade mono-heme, ligada a membrana (ec 1.9.3.1)	79275.3
3	RXF03895.1_a1	asma	50164.08
3	EXF07256.1_pn	proteína TolB	49215,09
1	RXF06792.1	Proteína Hipotética Conservada	47485,35
3	RXF02291.1	proteína de tolerância ao tolueno ttg2C	45703,08

Várias proteínas secretadas cotraducionalmente em *E.coli* foram identificadas. As sequências de várias dessas foram usadas para pesquisar o genoma de MB214 quanto a homólogas. Os genes de *E. coli* eram: Dsba, TorT, SfmC, FocC, CcmH, Yral, TolB, NikA, Flgl. O algoritmo BLAST (Altschul, S.F., et al J. Mol. Biol, 1990.215(3):p.403-10) foi usado para pesquisar um banco de dados de ORFs traduzidas de MB214. As proteínas de MB214 foram colocadas em duas categorias baseadas no grau de homologia que elas mostraram com seus equivalentes de *E. coli*. Proteínas com alta homologia combinaram com escores esperados de $2e^{-84}$ ou melhor. Proteínas com baixa homologia tiveram escores esperados entre $8e^{-17}$ e $5e^{-32}$. Esse método rendeu 11 homólogos em potencial, alguns dos quais se superpuseram aos 7 alvos obtidos acima.

A lista combinada de 18 proteínas únicas foi analisada usando SignalP e 9 alvos finais que foram preditos ter, provavelmente, um sítio de

corte único de peptidase de sinal foram escolhidos para estudos de expressão.

Isolamento e análise de sequência de líderes de secreção

As líderes de secreção identificadas de *P. fluorescens* foram amplificadas a partir do DNA genômico de DC454 (descendente de MB101 de *P. fluorescens*) e clonadas em pCRBLUNTII-TOPO (Invitrogen) para verificação da sequência de DNA. A sequência de DNA e de aminoácidos deduzida de cada líder de secreção de *P. fluorescens* isolada está referida na Tabela 9.

10 Tabela 9. Sequências líder de secreção de *P. fluorescens*

LÍDER	DNA SEQ ID N °	AMINOÁCIDO SEQ ID N°
CupA2	9	10
CupB2	11	12
CupC2	13	14
TolB	49	50
NikA	15	16
FlgI	17	18
ORF5550	19	20
Tgt2C	21	22
ORF8124	23	24

Fusão de líderes de secreção a Gal2 scFv e tioredoxina de *E. coli* e análise de expressão

Cada líder de secreção (Tabela 9) foi fundida em fase com a sequência de Gal2 scFv (Martineau, P. *et al.* 1998 J. Mol. Bio. 280:117) e/ou com a sequência de tioredoxina de *E. coli* (TrxA) (SEQ ID N° 46) usando splicing de extensão superposta PCR (Horton R.M. *et al.* Biotechniques 8:528). Os fragmentos resultantes foram purificados e subsequentemente usados como modelos para uma segunda rodada de PCR para fundir a sequência codificadora da líder de NikA com a sequência de *trxA*. As fusões foram então clonadas no vetor de expressão em *P. fluorescens* pDOW1169 sob o controle do promotor *tac*. Cada construto foi transformado em *P. fluorescens* DC454 e a expressão foi avaliada no formato de alto rendimento. As

culturas foram incubadas em um meio definido com sais minerais suplementado com glicerol 5% em placas com poços de 2 ml de profundidade em um volume de cultura de 0,5 ml. Após um período de crescimento de 24 horas, a proteína recombinante foi induzida com ITPG 0,3 mM e deixada expressar por 24 horas. As culturas foram fracionadas por sonicação e a expressão de proteína e o processamento de líder de secreção foram avaliados por SDS-PAGE e Western blot (Figura 9). Cada uma das líderes testadas, com exceção da líder de Bce, foi encontrada estar parcialmente ou totalmente processada a partir da sequência da proteína Gal2 scFv. Cada uma delas melhorou grandemente a expressão de Gal2 scFv comparado com uma cepa de expressão que codifica Gal2 scFv citoplasmática (nenhuma), indicando que além de direcionar a localização subcelular, essas líderes de secreção também melhoram a expressão global. Não inesperadamente, níveis variáveis de expressão e solubilidade de Gal2 scFv também foram observados. A análise Western confirmou que algumas Gal2 solúveis eram produzidas quando fundidas com CupA2, CupC2, NikA, F1gl e ORF5550 (Figura 10). Embora a expressão da líder de TolB fundida com Gal2 fosse menor do que a observada com outras líderes, a análise Western mostrou que todas as proteínas expressas eram solúveis. A análise N-terminal mostrou que as líderes de TolB, CupA2, CupC2, F1gl, NikA e ORF5550 eram clivadas de Gal2 scFv como esperado (dados não mostrados).

Apesar de não processada a partir de Gal2 scFv, a líder de Bce foi encontrada ser processada a partir de TrxA. A tioredoxina foi descrita como um modelo de proteína para identificação de líderes de secreção cotraducional já que ela se enovela rapidamente no citoplasma (Huber *et al.* 2005 J. Bacteriol, 187:2983). A secreção bem sucedida de TrxA solúvel utilizando a líder de Bce pode indicar que essa líder atua de uma maneira cotraducional para facilitar a secreção periplasmática.

Todas as publicações e pedidos de patente mencionados nessa descrição são indicativas do nível de experiência daqueles versados na técnica a qual essa invenção pertence. Todas as publicações e pedidos de patente estão incorporados aqui por referência na mesma extensão como se

cada publicação ou pedido de patente fosse especificamente e individualmente indicado para ser incorporado por referência.

- Embora a invenção precedente tenha sido descrita em alguns detalhes como meio de ilustração e exemplo com propósitos de clareza de compreensão, será óbvio que certas alterações e modificações podem ser praticadas dentro do escopo das reivindicações anexas.
- 5

REIVINDICAÇÕES

1. Molécula de ácido nucleico isolado que compreende uma sequência codificadora de sinal de secreção para a secreção de um polipeptídeo selecionado do grupo que consiste em uma proteína dissulfeto isomerase C (dsbC), uma proteína mutante que se ligue a fosfato (pbp*), uma proteína dissulfeto isomerase A (dsbA), Bce, CupA2, CupB2, CupC2, Nika, Flgl, uma proteína da família de tetratricopeptídeo repetido (ORF5550), uma proteína de tolerância ao tolueno (Tgt2C) e um polipeptídeo de secreção de proteína quimiotática que aceita metilação (ORF8124).
2. Molécula de ácido nucleico de acordo com a reivindicação 1, em que a dita molécula de ácido nucleico é selecionada do grupo que consiste em:
- a) uma molécula de ácido nucleico compreendendo a sequência de nucleotídeos de SEQ ID N° 5, 1, 3, 7, 9, 11, 13, 15, 17, 19,21 ou 23.
 - b) uma molécula de ácido nucleico compreendendo a sequência de nucleotídeos que tem pelo menos 90% de identidade de sequência com a sequência de nucleotídeos de SEQ ID N° 5, 3, 7, 9, 11, 13, 15, 17,21 ou 23.
 - c) uma molécula de ácido nucleico que codifique um polipolipeptídeo que compreende a sequência de aminoácidos de SEQ ID N° 6, 2, 4, 8, 10, 12, 14, 16, 18, 20,22 ou 24.
 - d) uma molécula de ácido nucleico que compreende uma sequência de nucleotídeos que codifica um polipolipeptídeo que tem pelo menos 90% de identidade de sequência de aminoácidos com a sequência de aminoácidos de SEQ ID N° 6, 4, 8, 10, 12, 14, 16, 18,22 ou 24, em que o dito polipolipeptídeo é um polipeptídeo de secreção.
 - e) uma molécula de ácido nucleico compreendendo uma sequência de nucleotídeos que codifica um polipolipeptídeo que tem pelo menos 96% de identidade de sequência de aminoácidos com a sequência de aminoácidos de SEQ ID N° 20, em que o dito polipolipeptídeo é um polipeptídeo de secreção.
 - f) uma sequência de nucleotídeos que hibridiza sob condições estridentes mais do que substancialmente com a extensão completa da

sequência de nucleotídeos de SEQ ID Nº 5, 3, 7, 9, 11, 13, 15,17 ou 21; e

- g) uma sequência de nucleotídeos que hibridiza sob condições estridentes mais do que substancialmente com a extensão completa de uma sequência de nucleotídeos que codifica uma sequência de aminoácidos selecionada do grupo que consiste em SEQ ID Nº 6, 4, 8, 10, 12, 14, 16,18 e 22.

3. Molécula de ácido nucleico de acordo com a reivindicação 2, em que as ditas condições de hibridização compreendam uma temperatura de cerca de 60°C a cerca de 70°C.

4. Molécula de ácido nucleico de acordo com a reivindicação 2 ou 3, em que as ditas condições de hibridização compreendam uma temperatura de cerca de 68°C.

5. Molécula de ácido nucleico de acordo com a reivindicação 1, em que a dita molécula de ácido nucleico tenha sido ajustada para refletir o códon preferencial de um organismo hospedeiro selecionado para expressar a molécula de ácido nucleico.

6. Vetor compreendendo uma sequência codificadora de sinal de secreção para uma proteína dissulfeto isomerase C (dsbC), uma proteína mutante que se ligue a fosfato (pbp*), uma proteína dissulfeto isomerase A (dsbA), Bce, CupA2, CupB2, CupC2, NikA, Flgl, uma proteína da família de tetratricopeptídeo repetido (ORF5550), uma proteína de tolerância ao tolueno (Tgt2C) e um polipeptídeo de secreção de proteína quimiotática que aceita metilação (ORF8124).

7. Vetor de acordo com a reivindicação 6, em que a dita molécula de ácido nucleico é selecionada do grupo que consiste em:

- a) uma molécula de ácido nucleico compreendendo a sequência de nucleotídeos de SEQ ID Nº 5, 1, 3, 7, 9, 11, 13, 15, 17, 19,21 ou 23;

- b) uma molécula de ácido nucleico compreendendo a sequência de nucleotídeos que tem pelo menos 90% de identidade de sequência com a sequência de nucleotídeos de SEQ ID Nº 5, 3, 7, 9, 11, 13, 15, 17,21 ou 23, em que a dita sequência de nucleotídeos codifica um polipeptídeo de secreção;

- c) uma molécula de ácido nucleico que codifica um polipolipeptí-

deo que compreende a sequência de aminoácidos de SEQ ID Nº 6, 2, 4, 8, 10, 12, 14, 16, 18, 20,22 ou 24;

d) uma molécula de ácido nucleico compreendendo a sequência de nucleotídeos que codifica um polipolipeptídeo que tem pelo menos 90% de identidade de sequência de aminoácidos com a sequência de aminoácidos de SEQ ID Nº 6, 4, 8, 10, 12, 14, 16, 18,22 ou 24, em que o dito polipolipeptídeo é um polipeptídeo de secreção;

e)molécula de ácido nucleico compreendendo uma sequência de nucleotídeos que codifica um polipolipeptídeo que tem pelo menos 96% de identidade de sequência de aminoácidos com a sequência de aminoácidos de SEQ ID Nº 20, em que o dito polipolipeptídeo é um polipeptídeo de secreção;

f) uma sequência de nucleotídeos que hibridiza sob condições estridentes mais do que substancialmente com a extensão completa da sequência de nucleotídeos de SEQ ID Nº 5, 3, 7, 9, 11, 13, 15,17 ou 21; e

g) uma sequência de nucleotídeos que hibridiza sob condições estridentes mais do que substancialmente com a extensão completa de uma sequência de nucleotídeos que codifica uma sequência de aminoácidos selecionada do grupo que consiste em SEQ ID Nº 6, 4, 8, 10, 12, 14, 16,18 e 22.

8. Vetor de acordo com a reivindicação 7, em que as ditas condições de hibridização compreendam uma temperatura de cerca de 60°C a cerca de 70°C.

9. Vetor de acordo com a reivindicação 7 ou 8, em que as ditas condições de hibridização compreendam uma temperatura de cerca de 68°C.

10. Vetor de acordo com a reivindicação 6, em que a dita molécula de ácido nucleico tenha sido ajustada para refletir o códon preferencial de um organismo hospedeiro selecionado para expressar a molécula de ácido nucleico.

11. Vetor de acordo com a reivindicação 6, em que a sequência codificadora de sinal de secreção está operativamente ligada com uma sequência que codifica uma proteína ou polipeptídeo de interesse.

12. Vetor de acordo com a reivindicação 11, em que a proteína ou polipeptídeo de interesse são nativos para um organismo hospedeiro no qual a proteína ou polipeptídeo de interesse são expressos.

5 13. Vetor de acordo com a reivindicação 6, em que a proteína ou polipeptídeo de interesse são nativos para *P. fluorescens*.

14. Vetor de acordo com a reivindicação 11, em que a proteína ou polipeptídeo de interesse são derivados de uma proteína ou polipeptídeo que não são nativos para um organismo hospedeiro no qual a proteína ou polipeptídeo de interesse são expressos.

10 15. Vetor de acordo com a reivindicação 11, em que a proteína ou polipeptídeo de interesse são de um organismo que não é uma Pseudomonad.

16. Vetor de acordo com a reivindicação 6, em que a proteína ou polipeptídeo de interesse são derivados de um organismo eucariótico.

15 17. Vetor de acordo com a reivindicação 16, em que a proteína ou polipeptídeo de interesse são derivados de um organismo de mamífero.

18. Vetor de acordo com a reivindicação 6, compreendendo ainda uma sequência de ligação entre a sequência do polipeptídeo de sinal e a sequência da proteína ou polipeptídeo de interesse.

20 19. Vetor de acordo com a reivindicação 19, em que a sequência de ligação é clivada por uma peptidase de sinal.

20. Vetor de acordo com a reivindicação 6, em que a sequência da proteína ou polipeptídeo de interesse é operativamente ligada a uma segunda sequência de sinal.

25 21. Vetor de acordo com a reivindicação 20, em que a segunda sequência de sinal compreende uma sequência direcionada para um sinal de secreção da membrana externa.

22. Vetor de acordo com a reivindicação 6, em que o vetor compreende ainda um promotor.

30 23. Vetor de acordo com a reivindicação 22, em que o promotor é nativo para uma célula hospedeira bacteriana.

24. Vetor de acordo com a reivindicação 22, em que o promotor

não é nativo para uma célula hospedeira bacteriana.

25. Vetor de acordo com a reivindicação 23, em que o promotor é nativo para *E.coli*.

26. Vetor de acordo com a reivindicação 22, em que o promotor
5 é um promotor induzível.

27. Vetor de acordo com a reivindicação 22, em que o promotor é um promotor *lac* ou um derivado de um promotor *lac*.

28. Célula recombinante compreendendo uma sequência codificadora de sinal de secreção para um polipeptídeo de secreção selecionado
10 do grupo que consiste em uma proteína dissulfeto isomerase C (dsbC), uma proteína mutante que se ligue a fosfato (pbp*), uma proteína dissulfeto isomerase A (dsbA), Bce, CupA2, CupB2, CupC2, NikA, Flgl, uma proteína da família de tetratricopeptídeo repetido (ORF5550), uma proteína de tolerância ao tolueno (Tgt2C) e um polipeptídeo de secreção de proteína quimiotática
15 que aceita metilação (ORF8124).

29. Célula recombinante de acordo com a reivindicação 28, em que a dita sequência codificadora é selecionada do grupo que consiste em:

a) uma molécula de ácido nucleico compreendendo a sequência de nucleotídeos de SEQ ID Nº 5, 1, 3, 7, 9, 11, 13, 15, 17, 19,21 ou 23;

20 b) uma molécula de ácido nucleico compreendendo a sequência de nucleotídeos que tem pelo menos 90% de identidade de sequência com a sequência de nucleotídeos de SEQ ID Nº 5, 3, 7, 9, 11, 13, 15, 17,21 ou 23, em que a dita sequência de nucleotídeos codifica um polipeptídeo de secreção;

25 c) uma molécula de ácido nucleico que codifica um polipolipeptídeo que compreende a sequência de aminoácidos de SEQ ID Nº 6, 2, 4, 8, 10, 12, 14, 16, 18, 20,22 ou 24;

d) uma molécula de ácido nucleico compreendendo a sequência de nucleotídeos que codifica um polipolipeptídeo que tem pelo menos 90%
30 de identidade de sequência de aminoácidos com a sequência de aminoácidos de SEQ ID Nº 6, 4, 8, 10, 12, 14, 16, 18,22 ou 24, em que o dito polipolipeptídeo é um polipeptídeo de secreção;

e) molécula de ácido nucleico compreendendo uma sequência de nucleotídeos que codifica um polipolipeptídeo que tem pelo menos 96% de identidade de sequência de aminoácidos com a sequência de aminoácidos de SEQ ID Nº 20, em que o dito polipolipeptídeo é um polipeptídeo de secreção;

f) uma sequência de nucleotídeos que hibridiza sob condições estridentes mais do que substancialmente com a extensão completa da sequência de nucleotídeos de SEQ ID Nº 5, 3, 7, 9, 11, 13, 15,17 ou 21; e

g) uma sequência de nucleotídeos que hibridiza sob condições estridentes mais do que substancialmente com a extensão completa de uma sequência de nucleotídeos que codifica uma sequência de aminoácidos selecionada do grupo que consiste em SEQ ID Nº 6, 4, 8, 10, 12, 14, 16,18 e 22.

30. Célula de acordo com a reivindicação 28, em que a sequência que codifica o sinal de secreção está em um vetor de expressão.

31. Célula de acordo com a reivindicação 28, em que a sequência que codifica o sinal de secreção está operativamente ligada a uma sequência que codifica uma proteína ou polipeptídeo de interesse.

32. Célula de acordo com a reivindicação 31, em que a célula expressa a proteína ou polipeptídeo de interesse operativamente ligados a um polipeptídeo de sinal de secreção.

33. Célula de acordo com a reivindicação 32, em que a proteína ou polipeptídeo são expressos em um compartimento periplasmático da célula.

34. Célula de acordo com a reivindicação 32, em que uma enzima na célula cliva o polipeptídeo de sinal de secreção da proteína ou polipeptídeo de interesse.

35. Célula de acordo com a reivindicação 28, em que a célula é derivada de um hospedeiro bacteriano.

36. Célula de acordo com a reivindicação 35, em que o hospedeiro é uma *Pseudomonad*.

37. Célula de acordo com a reivindicação 36, em que o hospedeiro é uma *P. fluorescens*.

38. Célula de acordo com a reivindicação 35, em que o hospedeiro é uma *E.coli*.

39. Polipeptídeo isolado que compreende um polipeptídeo de secreção selecionado do grupo que consiste em uma proteína dissulfeto isomerase C (dsbC), uma proteína mutante que se ligue a fosfato (pbp*), uma proteína dissulfeto isomerase A (dsbA), Bce, CupA2, CupB2, CupC2, NikA, Flgl, uma proteína da família de tetratricopeptídeo repetido (ORF5550), uma proteína de tolerância ao tolueno (Tgt2C) e um polipeptídeo de secreção de proteína quimiotática que aceita metilação (ORF8124).

40. Polipeptídeo isolado de acordo com a reivindicação 39, em que o dito polipeptídeo é selecionado do grupo que consiste em:

a) um polipeptídeo que compreende a sequência de aminoácidos de SEQ ID Nº 6, 2, 4, 8, 10, 12, 14, 16, 18, 20,22 ou 24;

b) um polipeptídeo codificado pela sequência de nucleotídeos de SEQ ID Nº 5, 1, 3, 7, 9, 11, 13, 15, 17, 19,21 ou 23;

c) um polipeptídeo compreendendo uma sequência de aminoácidos que tem pelo menos 90% de identidade de sequência com a sequência de aminoácidos de SEQ ID Nº 6, 4, 8, 10, 12, 14, 16, 18,22 ou 24, em que o dito polipeptídeo é um polipeptídeo de sinal de secreção;

d) um polipeptídeo compreendendo uma sequência de aminoácidos pelo menos 96% de identidade de sequência com a sequência de aminoácidos de SEQ ID Nº 20, em que o dito polipeptídeo é um polipeptídeo de sinal de secreção;

e) um polipeptídeo que é codificado por uma sequência de nucleotídeos que é pelo menos 90% idêntica à sequência de nucleotídeos de SEQ ID Nº 5, 3, 7, 9, 11, 13, 15, 17,21 ou 23, em que o dito polipeptídeo é um polipeptídeo de sinal de secreção; e

f) um polipeptídeo codificado por uma sequência de nucleotídeos que hibridiza sob condições estridentes mais do que substancialmente a extensão completa da sequência de nucleotídeos de SEQ ID Nº 5, 3, 7, 9, 11, 13, 15,17 ou 21.

41. Molécula de ácido nucleico de acordo com a reivindicação

40, em que as ditas condições de hibridização compreendem uma temperatura de cerca de 60°C a cerca de 70°C.

42. Molécula de ácido nucleico de acordo com a reivindicação 39, em que as ditas condições de hibridização compreendem uma temperatura de cerca de 68°C.

43. Polipeptídeo de acordo com a reivindicação 39, em que o dito polipeptídeo de sinal de secreção está operativamente ligado a uma proteína ou polipeptídeo de interesse.

44. Polipeptídeo de acordo com a reivindicação 43, em que a proteína ou polipeptídeo de interesse é derivado de um organismo que não é um organismo de *P. fluorescens*.

45. Sistema de expressão para expressão de uma proteína ou polipeptídeo de interesse compreendendo:

a) uma célula hospedeira; e
b) um vetor que compreende uma molécula de ácido nucleico que codifica a proteína ou polipeptídeo de interesse operativamente ligado a um polipeptídeo de sinal de secreção selecionado do grupo que consiste em uma proteína dissulfeto isomerase C (dsbC), uma proteína mutante que se ligue a fosfato (pbp*), uma proteína dissulfeto isomerase A (dsbA), Bce, CupA2, CupB2, CupC2, NikA, Flgl, uma proteína da família de tetratricopeptídeo repetido (ORF5550), uma proteína de tolerância ao tolueno (Tgt2C) e um polipeptídeo de secreção de proteína quimiotática que aceita metilação (ORF8124).

46. Sistema de expressão de acordo com a reivindicação 45, em que o dito polipeptídeo de sinal de secreção é codificado por uma molécula de ácido nucleico selecionada do grupo que consiste em:

a) uma molécula de ácido nucleico compreendendo a sequência de nucleotídeos de SEQ ID Nº 5, 1, 3, 7, 9, 11, 13, 15, 17, 19,21 ou 23;
b) uma molécula de ácido nucleico compreendendo a sequência de nucleotídeos que tem pelo menos 90% de identidade de sequência com a sequência de nucleotídeos de SEQ ID Nº 5, 3, 7, 9, 11, 13, 15, 17,21 ou 23, em que a dita sequência de nucleotídeos codifica um polipeptídeo de secre-

ção;

c) uma molécula de ácido nucleico que codifica um polipolipeptídeo que compreende a sequência de aminoácidos de SEQ ID Nº 6, 2, 4, 8, 10, 12, 14, 16, 18, 20,22 ou 24;

5 d) uma molécula de ácido nucleico compreendendo a sequência de nucleotídeos que codifica um polipolipeptídeo que tem pelo menos 90% de identidade de sequência de aminoácidos com a sequência de aminoácidos de SEQ ID Nº 6, 4, 8, 10, 12, 14, 16, 18,22 ou 24, em que o dito polipolipeptídeo é um polipeptídeo de secreção;

10 e) molécula de ácido nucleico compreendendo uma sequência de nucleotídeos que codifica um polipolipeptídeo que tem pelo menos 96% de identidade de sequência de aminoácidos com a sequência de aminoácidos de SEQ ID Nº 20, em que o dito polipolipeptídeo é um polipeptídeo de secreção;

15 f) uma sequência de nucleotídeos que hibridiza sob condições estridentes mais do que substancialmente com a extensão completa da sequência de nucleotídeos de SEQ ID Nº 5, 3, 7, 9, 11, 13, 15,17 ou 21; e

g) uma sequência de nucleotídeos que hibridiza sob condições estridentes mais do que substancialmente com a extensão completa de
20 uma sequência de nucleotídeos que codifica uma sequência de aminoácidos selecionada do grupo que consiste em SEQ ID Nº 6, 4, 8, 10, 12, 14, 16,18 e 22.

47. Sistema de expressão de acordo com a reivindicação 46, em que as ditas condições de hibridização compreendem uma temperatura de
25 cerca de 60°C a cerca de 70°C.

48. Sistema de expressão de acordo com a reivindicação 46, em que as ditas condições de hibridização compreendem uma temperatura de cerca de 68°C.

49. Sistema de expressão de acordo com a reivindicação 45, em
30 que a célula hospedeira expressa a proteína ou polipeptídeo de interesse operativamente ligados a um polipeptídeo de sinal de secreção.

50. Sistema de expressão de acordo com a reivindicação 49, em

que a proteína ou polipeptídeo de interesse são expressos em um compartimento periplasmático da célula.

51. Sistema de expressão de acordo com a reivindicação 49, em que uma enzima na célula cliva o polipeptídeo de sinal da proteína ou polipeptídeo de interesse.

52. Sistema de expressão de acordo com a reivindicação 45, em que a célula é derivada de um hospedeiro bacteriano.

53. Sistema de expressão de acordo com a reivindicação 45, em que o hospedeiro é uma *Pseudomonad*.

54. Sistema de expressão de acordo com a reivindicação 53, em que o hospedeiro é *P. fluorescens*.

55. Sistema de expressão de acordo com a reivindicação 52, em que o hospedeiro é *E.coli*.

56. Sistema de expressão de acordo com a reivindicação 45, compreendendo ainda um meio de fermentação.

57. Sistema de expressão de acordo com a reivindicação 56, em que o meio de fermentação compreende um indutor químico.

58. Método para expressão de uma proteína recombinante em uma célula hospedeira compreendendo fornecer uma célula hospedeira que compreende um vetor que codifica uma proteína ou polipeptídeo de interesse operativamente ligados a um polipeptídeo de sinal de secreção selecionado do grupo que consiste em uma proteína dissulfeto isomerase C (dsbC), uma proteína mutante que se ligue a fosfato (pbp*), uma proteína dissulfeto isomerase A (dsbA), Bce, CupA2, CupB2, CupC2, NikA, Flgl, uma proteína da família de tetratricopeptídeo repetido (ORF5550), uma proteína de tolerância ao tolueno (Tgt2C) e um polipeptídeo de secreção de proteína quimiotática que aceita metilação (ORF8124).

59. Método de acordo com a reivindicação 58, em que o polipeptídeo de sinal de secreção é codificado por uma molécula de ácido nucleico selecionada do grupo que consiste em:

a) uma molécula de ácido nucleico compreendendo a sequência de nucleotídeos de SEQ ID Nº 5, 1, 3, 7, 9, 11, 13, 15, 17, 19, 21 ou 23;

b) uma molécula de ácido nucleico compreendendo a sequência de nucleotídeos que tem pelo menos 90% de identidade de sequência com a sequência de nucleotídeos de SEQ ID Nº 5, 3, 7, 9, 11, 13, 15, 17,21 ou 23, em que a dita sequência de nucleotídeos codifica um polipeptídeo de secreção;

c) uma molécula de ácido nucleico que codifica um polipolipeptídeo que compreende a sequência de aminoácidos de SEQ ID Nº 6, 2, 4, 8, 10, 12, 14, 16, 18, 20,22 ou 24;

d) uma molécula de ácido nucleico compreendendo a sequência de nucleotídeos que codifica um polipolipeptídeo que tem pelo menos 90% de identidade de sequência de aminoácidos com a sequência de aminoácidos de SEQ ID Nº 6, 4, 8, 10, 12, 14, 16, 18,22 ou 24, em que o dito polipolipeptídeo é um polipeptídeo de secreção;

e) molécula de ácido nucleico compreendendo uma sequência de nucleotídeos que codifica um polipolipeptídeo que tem pelo menos 96% de identidade de sequência de aminoácidos com a sequência de aminoácidos de SEQ ID Nº 20, em que o dito polipolipeptídeo é um polipeptídeo de secreção;

f) uma sequência de nucleotídeos que hibridiza sob condições estridentes mais do que substancialmente com a extensão completa da sequência de nucleotídeos de SEQ ID Nº 5, 3, 7, 9, 11, 13, 15,17 ou 21; e

g) uma sequência de nucleotídeos que hibridiza sob condições estridentes mais do que substancialmente com a extensão completa de uma sequência de nucleotídeos que codifica uma sequência de aminoácidos selecionada do grupo que consiste em SEQ ID Nº 6, 4, 8, 10, 12, 14, 16,18 e 22.

60. Sistema de expressão de acordo com a reivindicação 59, em que as ditas condições de hibridização compreendem uma temperatura de cerca de 60°C a cerca de 70°C.

61. Sistema de expressão de acordo com a reivindicação 59, em que as ditas condições de hibridização compreendem uma temperatura de cerca de 68°C.

62. Método de acordo com a reivindicação 58, em que a célula é cultivada em um meio com sais minerais.

63. Método de acordo com a reivindicação 58, em que a célula é cultivada em alta densidade celular.

64. Método de acordo com a reivindicação 63, em que a célula é cultivada com uma densidade celular de pelo menos 20 g/l.

5 65. Método de acordo com a reivindicação 58, compreendendo ainda purificar a proteína recombinante.

66. Método de acordo com a reivindicação 65, em que a proteína recombinante é purificada por cromatografia de afinidade.

10 67. Método de acordo com a reivindicação 58, em que a ligação operativa da proteína ou polipeptídeo de interesse e o polipeptídeo de sinal de secreção é clivável por uma enzima nativa da célula hospedeira.

68. Método de acordo com a reivindicação 67, em que o polipeptídeo de sinal de secreção é clivado da proteína ou polipeptídeo de interesse durante a expressão.

15 69. Método de acordo com a reivindicação 58, em que a proteína ou polipeptídeo de interesse são nativos do organismo a partir do qual a célula hospedeira é derivada.

70. Método de acordo com a reivindicação 58, em que a proteína ou polipeptídeo de interesse são nativos de um organismo de *P.fluorescens*.

20 71. Método de acordo com a reivindicação 58, em que a proteína ou polipeptídeo de interesse não são nativos do organismo a partir do qual a célula hospedeira é derivada.

25 72. Método de acordo com a reivindicação 58, em que a proteína ou polipeptídeo de interesse são derivados de um organismo que não é uma Pseudomonad.

73. Método de acordo com a reivindicação 58, em que a proteína ou polipeptídeo de interesse são derivados de um organismo eucariótico.

30 74. Método de acordo com a reivindicação 58, em que a proteína recombinante compreende uma sequência que inclui pelo menos dois resíduos de cisteína.

75. Método de acordo com a reivindicação 58, em que pelo menos uma ligação de dissulfeto é formada na proteína recombinante na célula.

76. Método de acordo com a reivindicação 58, compreendendo ainda uma sequência de ligação entre a sequência de polipeptídeo de sinal e a sequência da proteína ou polipeptídeo de interesse.

5 77. Método de acordo com a reivindicação 69, em que pelo menos 50% da proteína ou polipeptídeo de interesse compreendem uma terminação amino nativa.

78. Método de acordo com a reivindicação 77, em que em que pelo menos 80% da proteína ou polipeptídeo de interesse compreendem uma terminação amino nativa.

10 79. Método de acordo com a reivindicação 78, em que em que pelo menos 90% da proteína ou polipeptídeo de interesse compreendem uma terminação amino nativa.

80. Método de acordo com a reivindicação 58, em que pelo menos 50% da proteína recombinante são ativos.

15 81. Método de acordo com a reivindicação 80, em que pelo menos 80% da proteína recombinante são ativos.

82. Método de acordo com a reivindicação 58, em que pelo menos 50% da proteína recombinante são expressos em um compartimento periplasmático.

20 83. Método de acordo com a reivindicação 82, em que pelo menos 75% da proteína recombinante são expressos em um compartimento periplasmático.

25 84. Método de acordo com a reivindicação 83, em que pelo menos 90% da proteína recombinante são expressos em um compartimento periplasmático.

85. Método de acordo com a reivindicação 58, em que a célula hospedeira é uma célula de *Pseudomonad*.

86. Método de acordo com a reivindicação 85, em que a célula é uma célula de *P. fluorescens*.

30 87. Método de acordo com a reivindicação 58, em que a célula é uma célula de *E. coli*.

FIG. 1

PCR de dsbC-skp do modelo pDOW3001

- Iniciadores incorporam sítios de restrição SpeI-HindIII
- Ligado no vetor de expressão pDOW1169

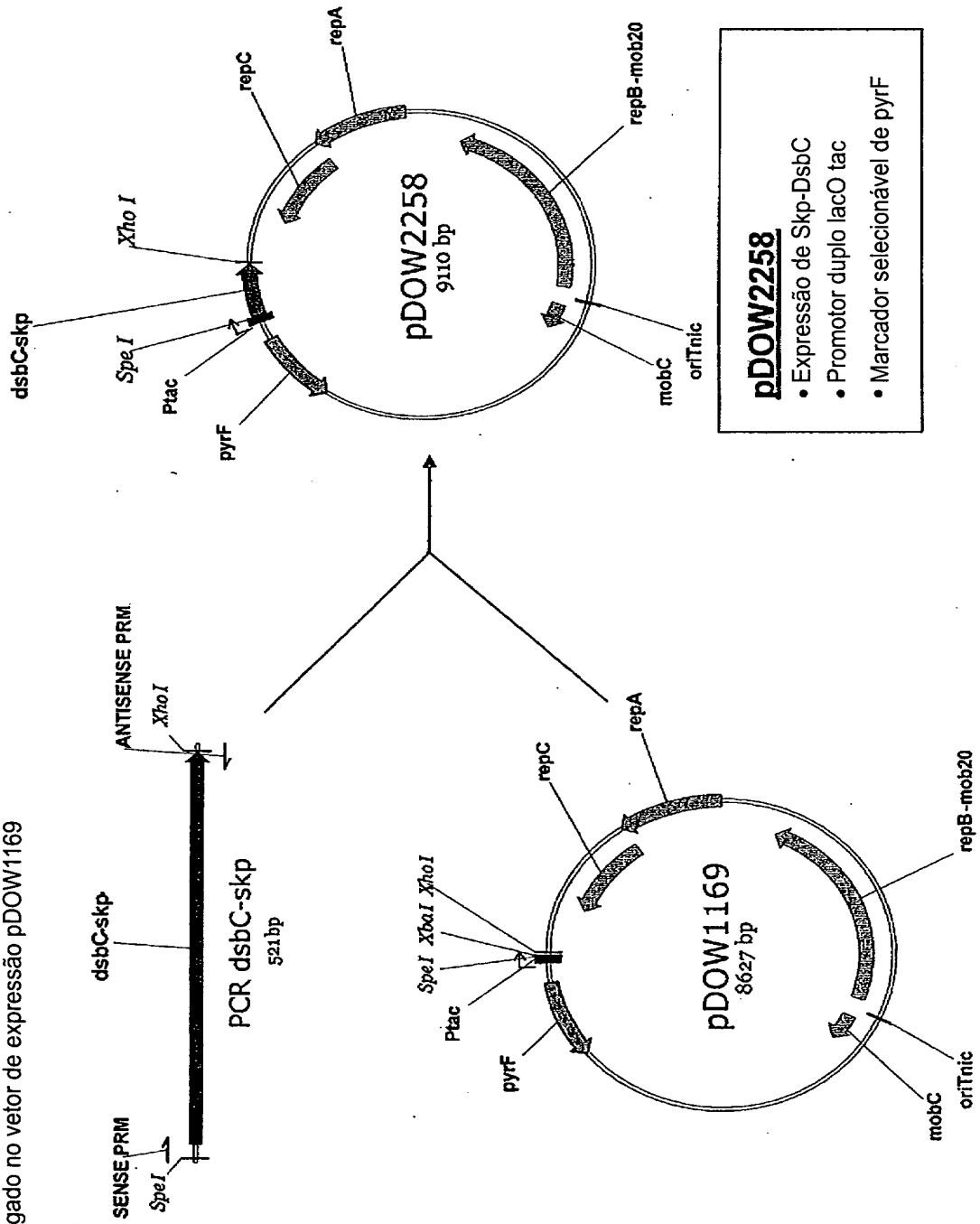
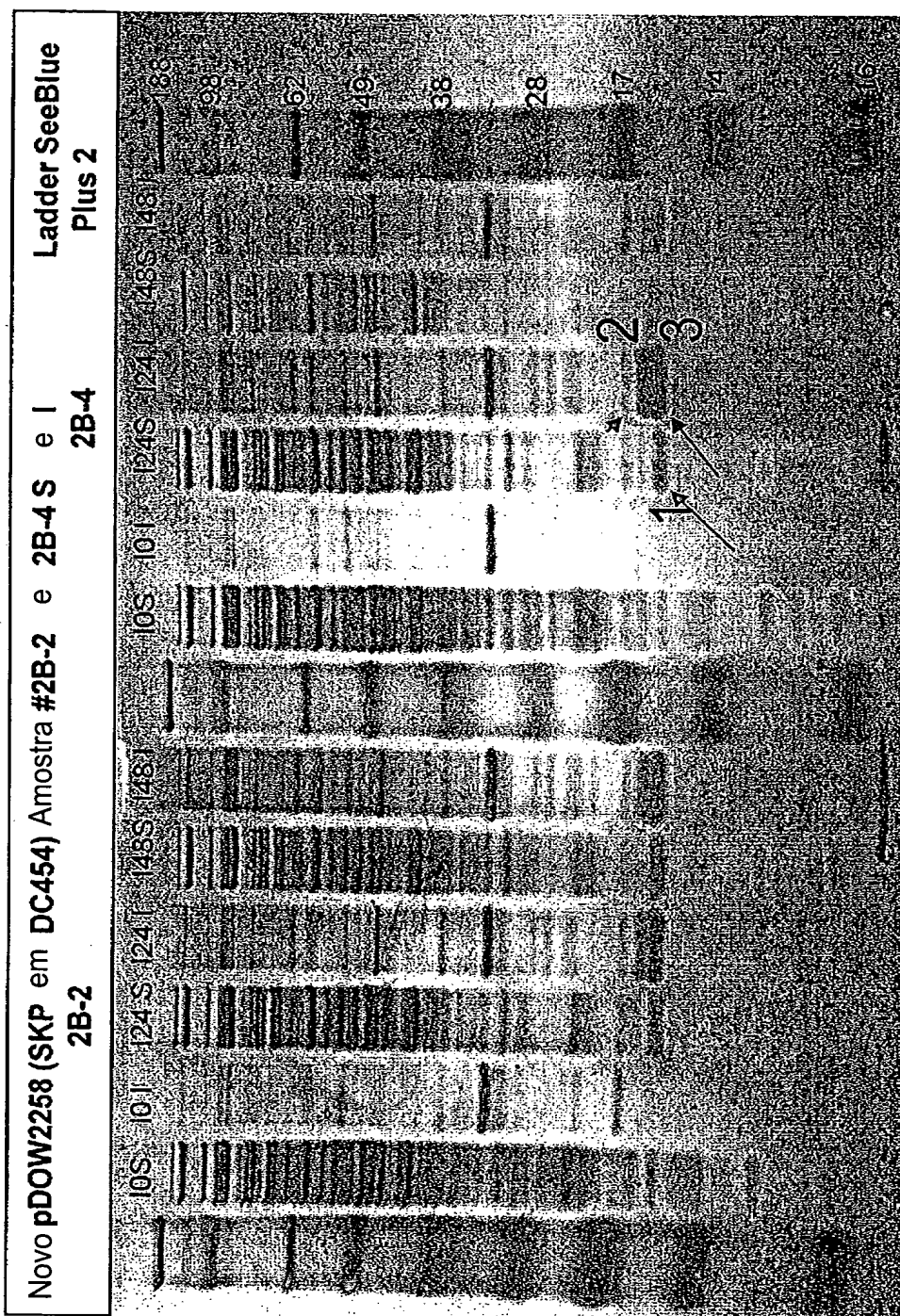


FIG. 2



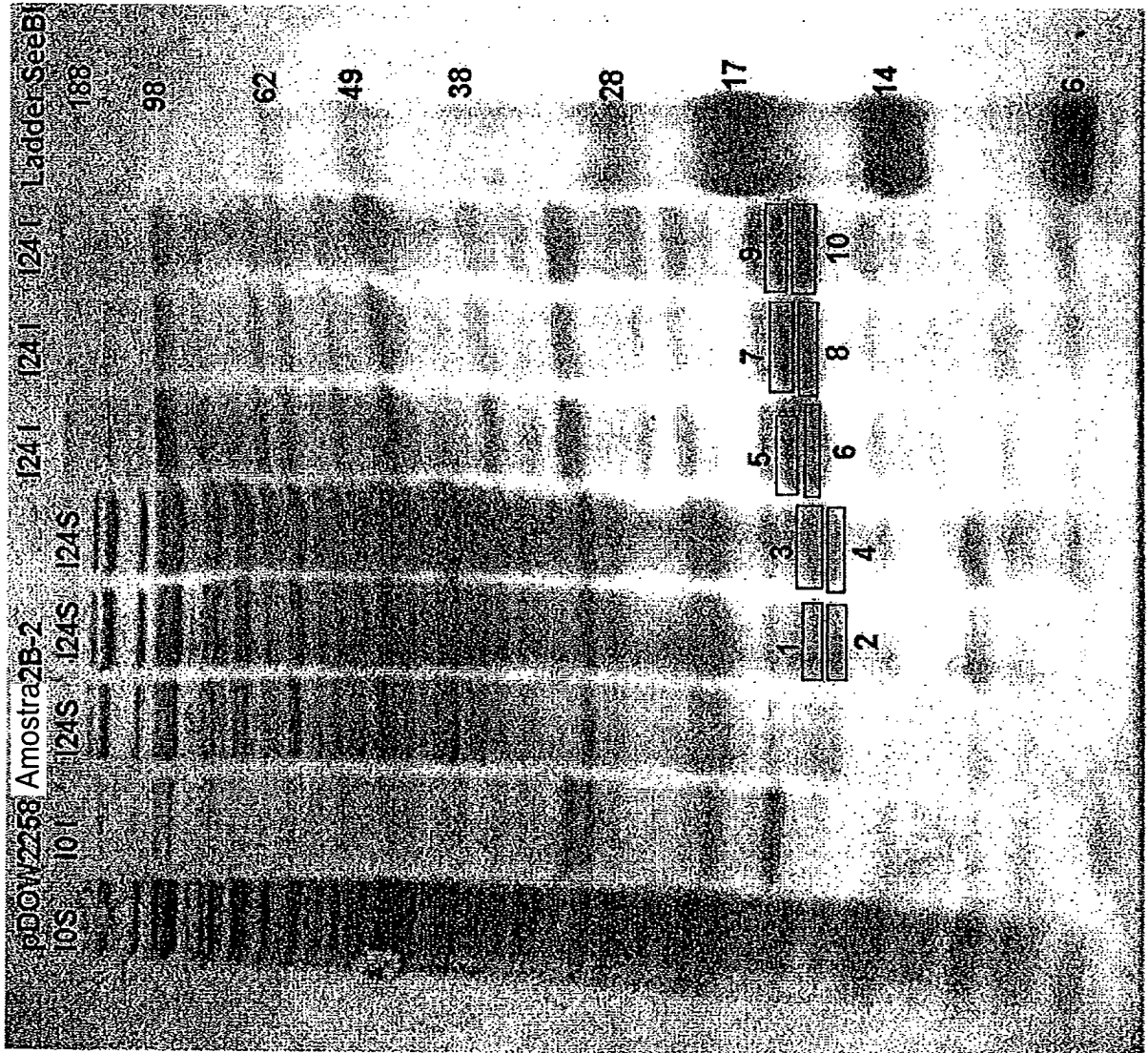
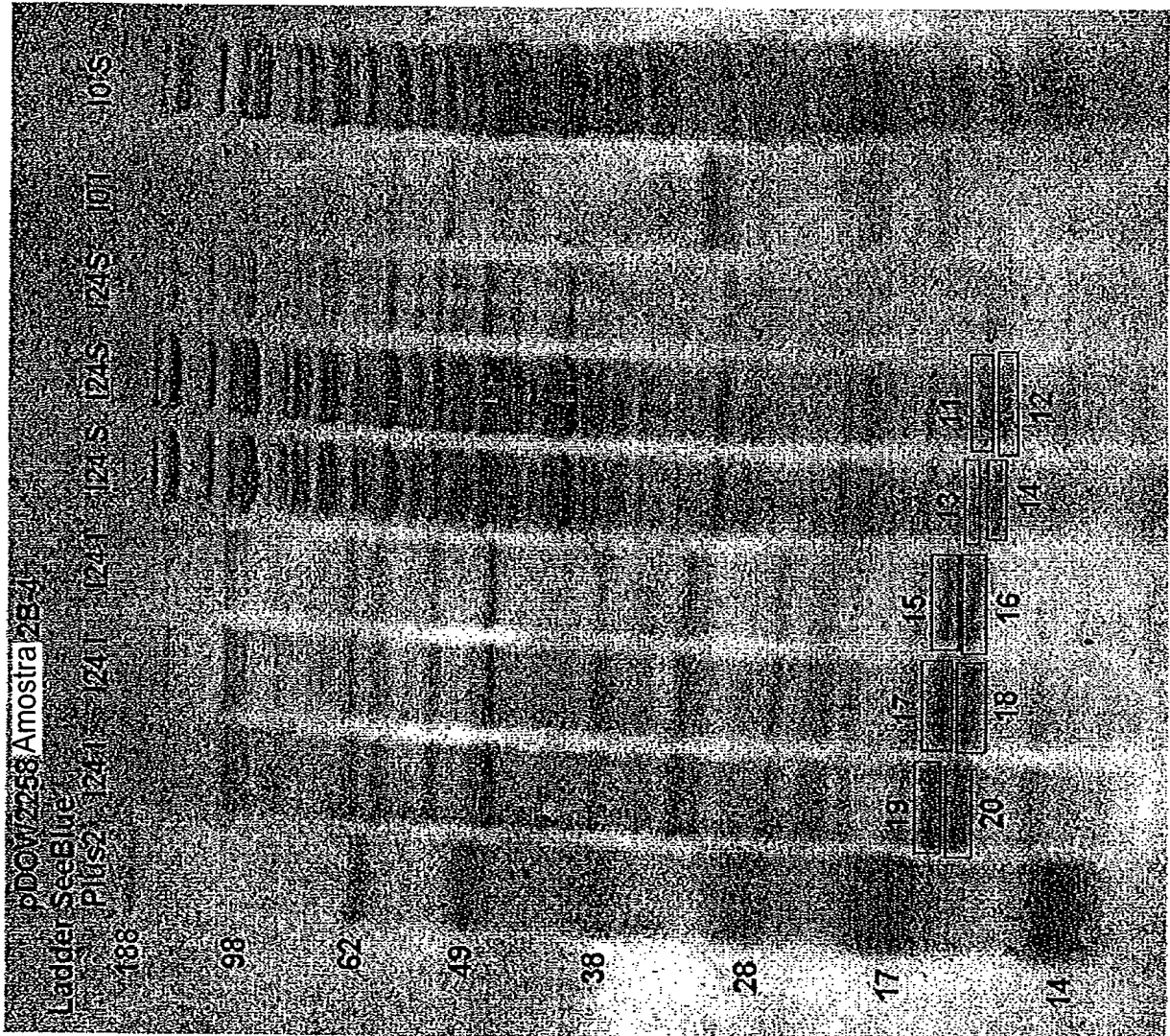
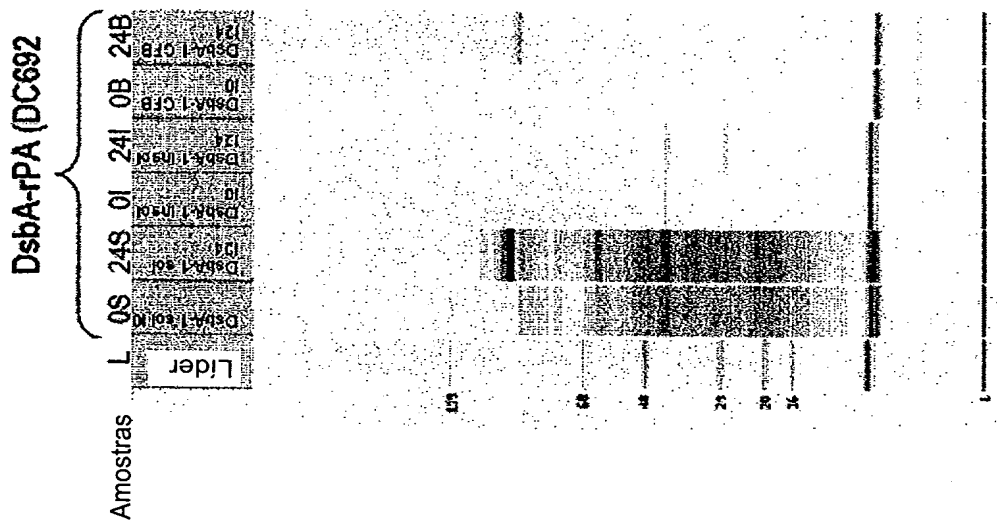


FIG. 3B

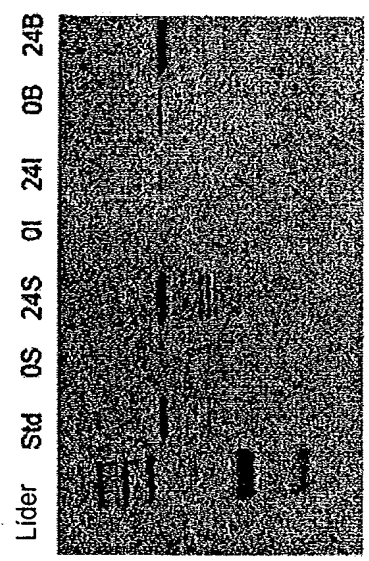




Legenda
S - Solúvel
I - Insolúvel
B - Caldo sem célula
L - Líder

FIG. 4

DsbA-rPA (DC692)



PA

Gene clonado	Alvo de expressão	Cepa hospedeira
DsbA Líder PA83	periplasmic	WT DC454

Clone	Líder	Solúvel (g/L)	Insolúvel (g/L)	Caldo sem célula (g/L)	N-terminus
DC692	DsbA	1.1	0.05	0.23	E V K Q E

FIG. 5

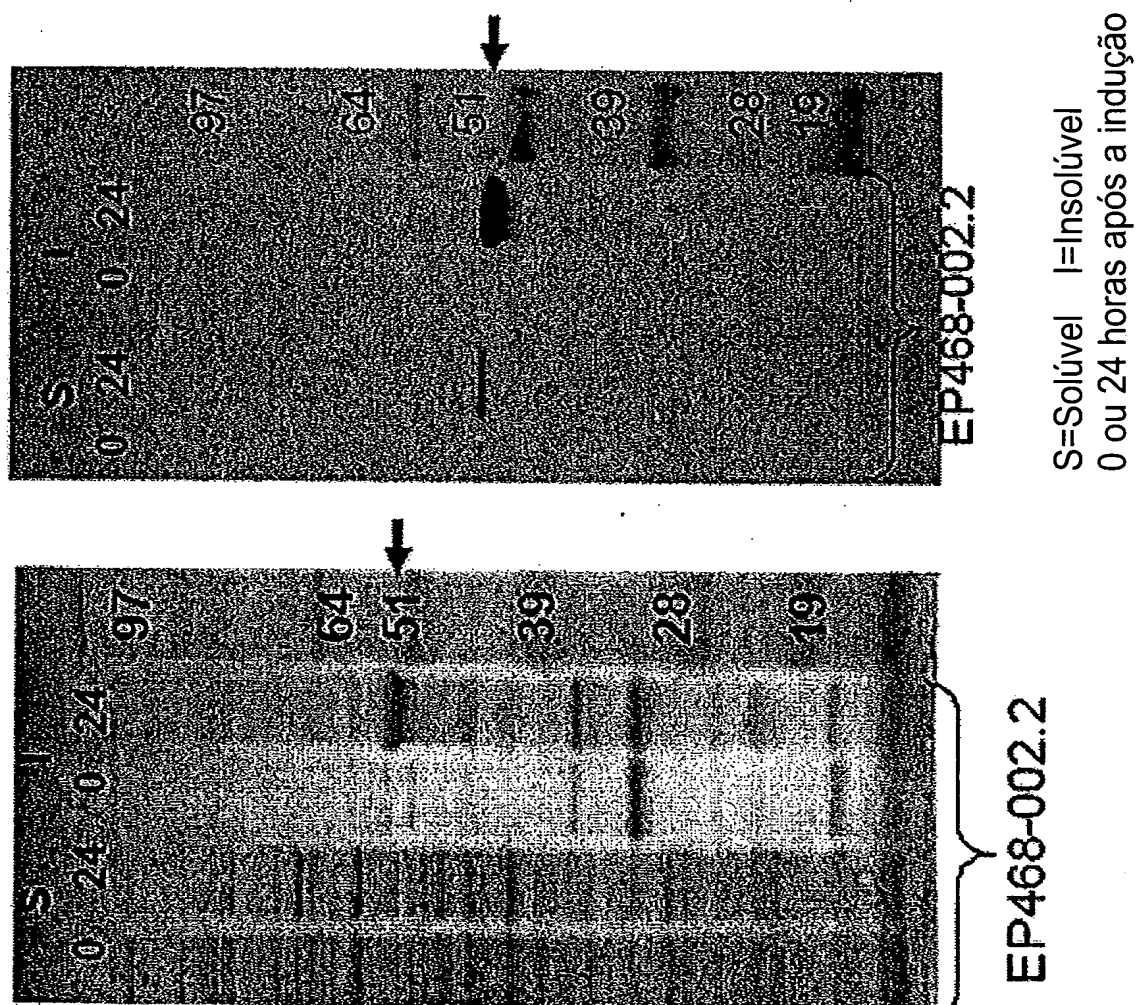


FIG. 6

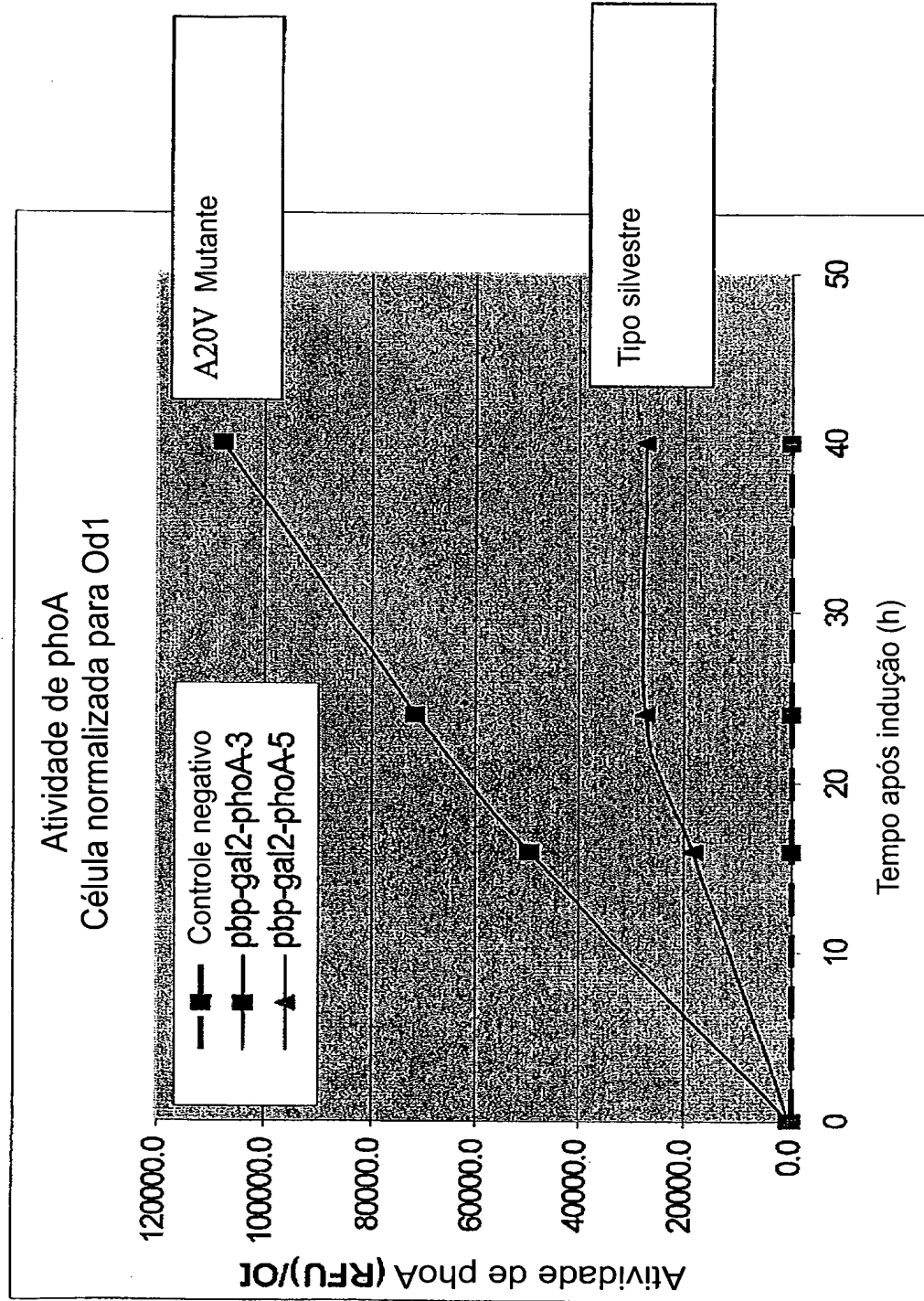


FIG. 7

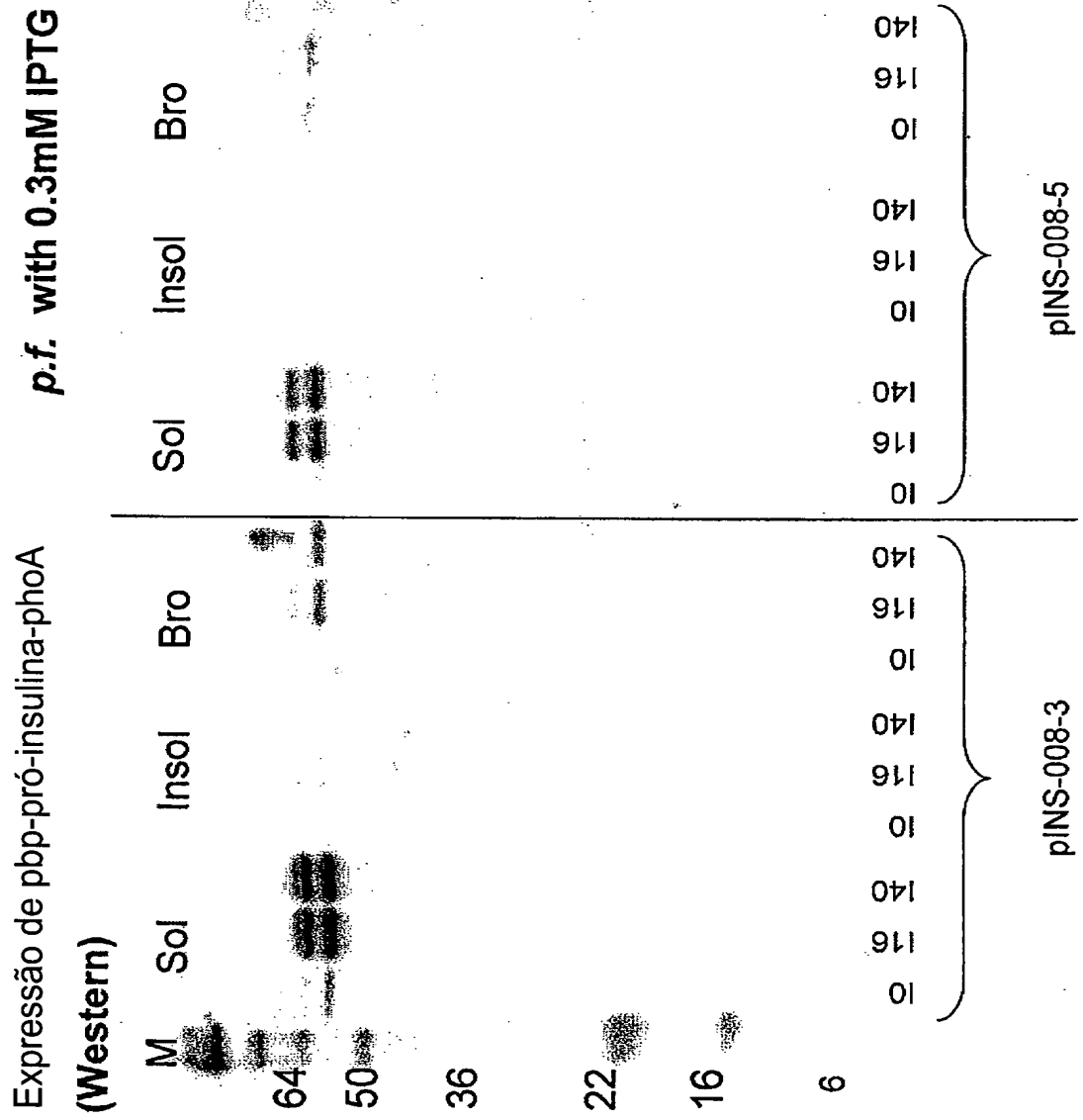


FIG. 8

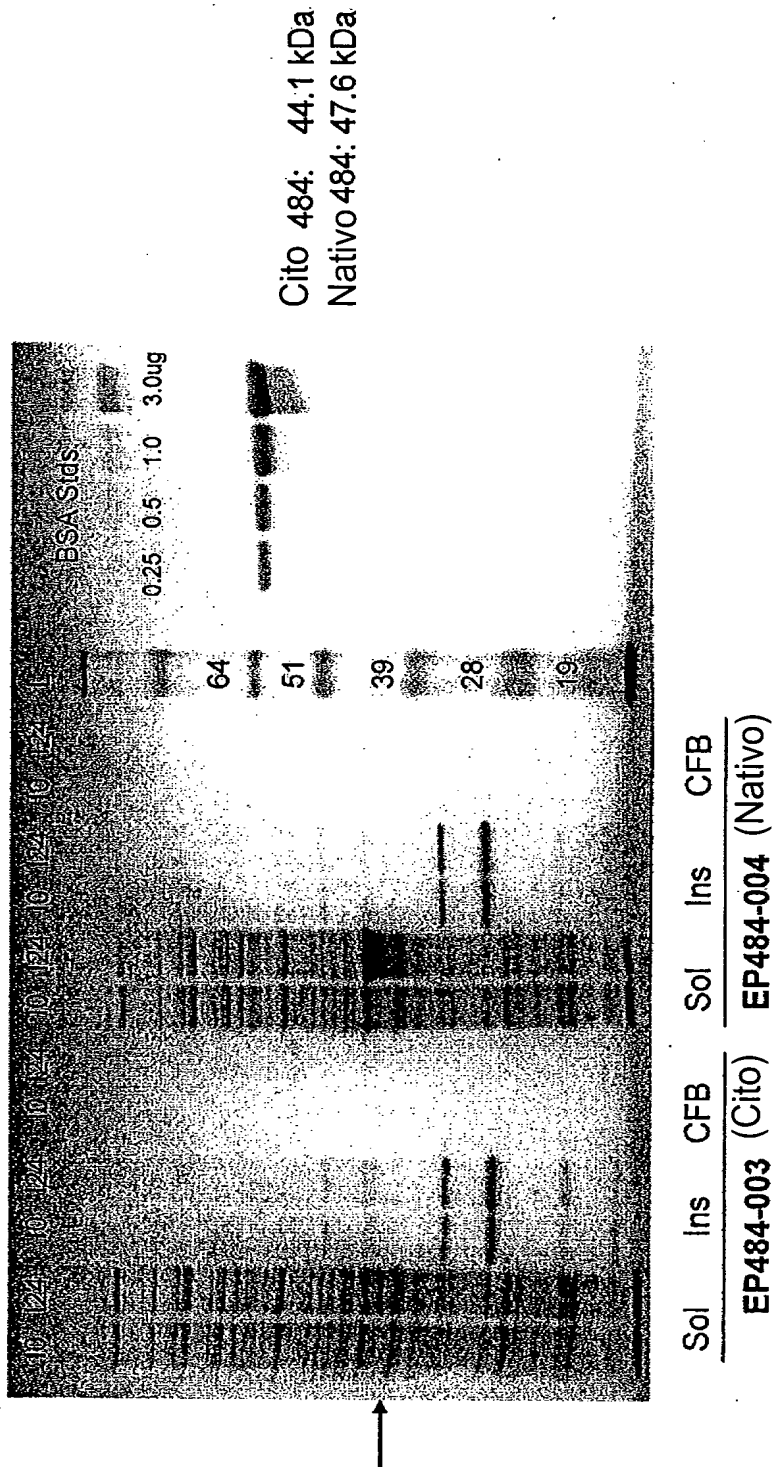


FIG. 9

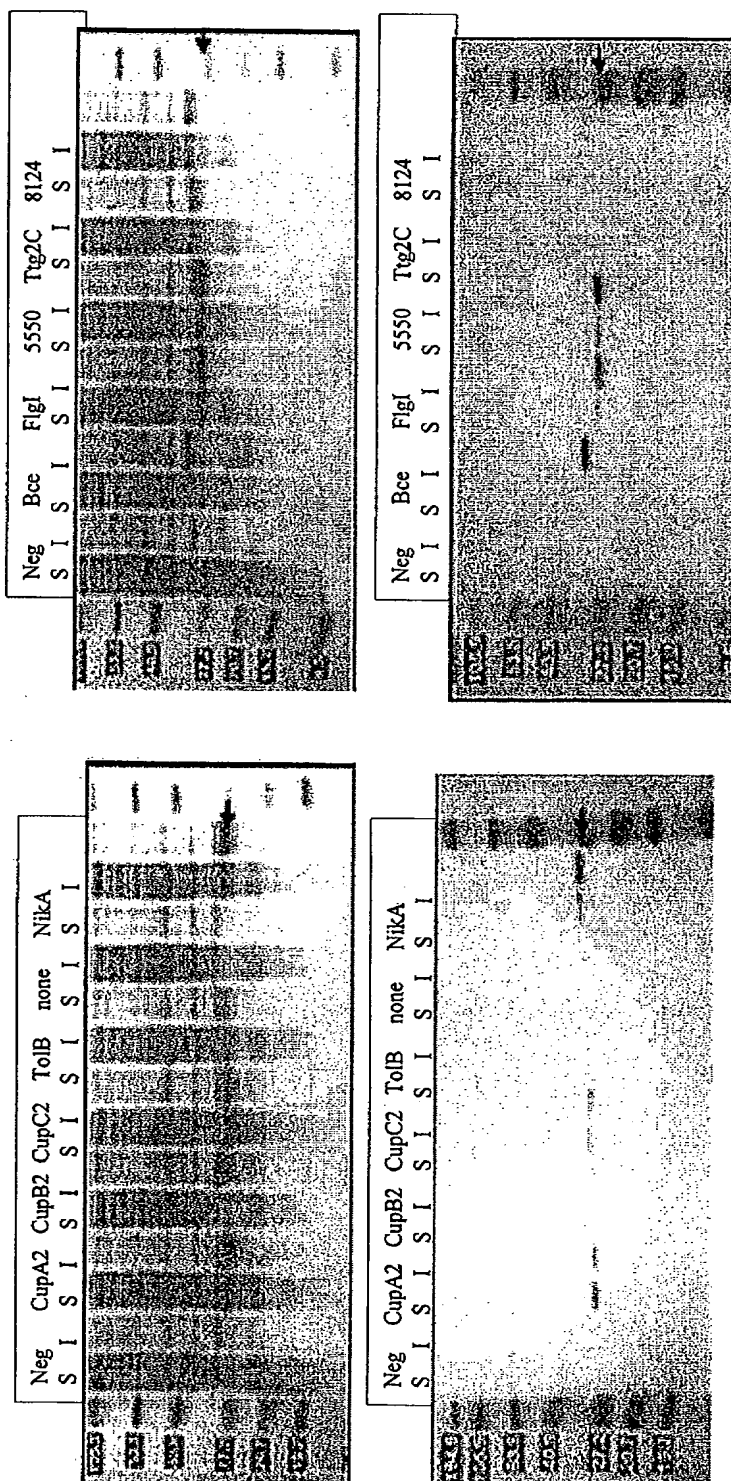
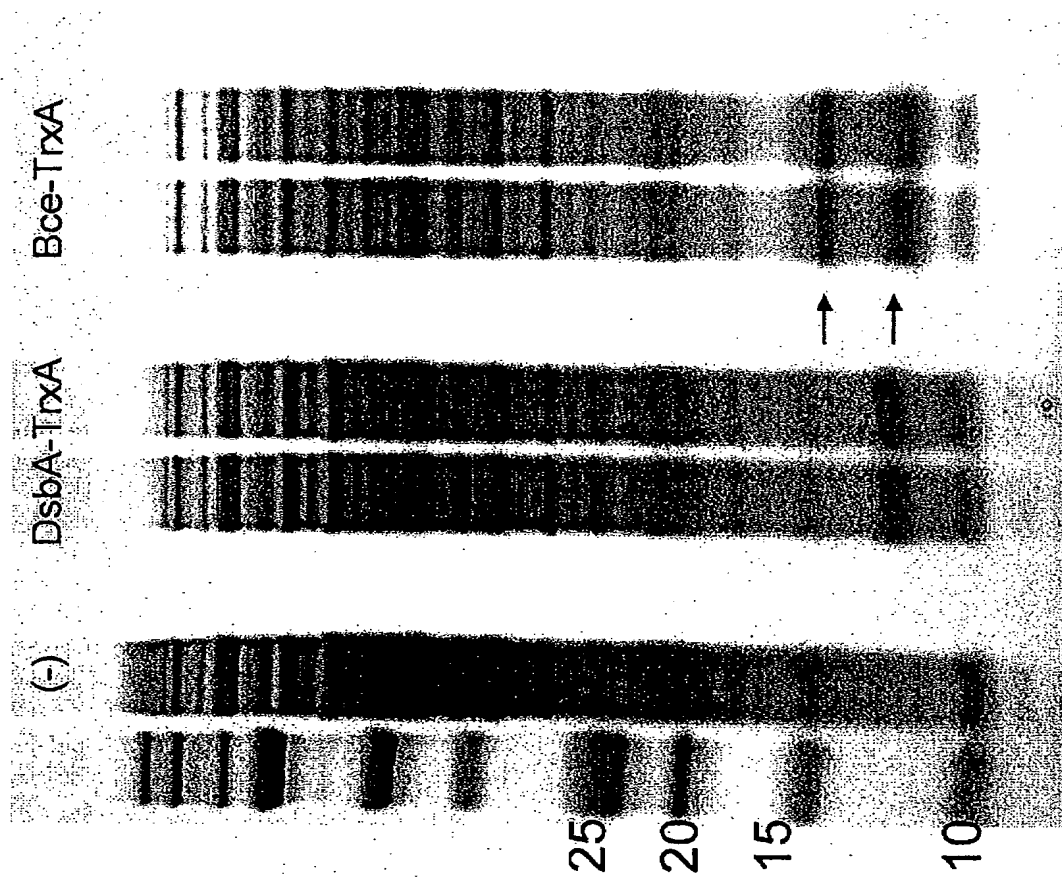


FIG. 10



RESUMO

Patente de Invenção: **"SEQUÊNCIAS-LÍDER BACTERIANAS PARA EXPRESSÃO AUMENTADA"**.

A presente invenção refere-se a composições e métodos para aperfeiçoar a expressão e/ou secreção de proteína ou polipeptídeo de interesse em uma célula hospedeira. São fornecidas composições que compreendem uma sequência codificadora para peptídeo de sinal de secreção bacteriano. As sequências podem ser usadas em construtos de vetor ou sistemas de expressão para transformação ou expressão de uma proteína ou polipeptídeo de interesse em uma célula hospedeira. As composições da invenção são úteis para aumentar o acúmulo de proteínas apropriadamente processadas no espaço periplasmático de uma célula hospedeira ou para aumentar a secreção das proteínas apropriadamente processadas da célula hospedeira. Em particular, moléculas de ácido nucleico isolado que codificam peptídeo de sinal de secreção são fornecidas. Além disso, são abrangidas as sequências de aminoácidos que correspondem às moléculas de ácido nucleico. Em particular, a presente invenção fornece moléculas de ácido nucleico isolado que compreendem sequências de nucleotídeos que codificam as sequências de aminoácidos mostradas em SEQ ID Nº 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22 e 24 e as sequências de nucleotídeos descritas em SEQ ID Nº 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21 e 23, assim como suas variantes e fragmentos.

LISTAGEM DE SEQUÊNCIAS

<110> Coleman, Russell J.
 Retallack, Diane M.
 Schneider, Jane C.
 Ramseier, Thomas M.
 Hershberger, Charles D.
 Lee, Stacey L.
 Resnick, Sol M.

<120> SEQUÊNCIAS-LÍDER BACTERIANAS PARA EXPRESSÃO AUMENTADA

<130> 43292/339376

<160> 50

<170> FastSEQ para Windows Versão 4.0

<210> 1

<211> 72

<212> DNA

<213> Sequência artificial

<220>

<223> Sequência líder da proteína mutante que se liga a fosfato
 (pbp*)

<221> CDS

<222> (1)...(72)

<400> 1

atg	aaa	ctg	aaa	cgt	ttg	atg	gcg	gca	atg	act	ttt	gtc	gct	gct	ggc	48
Met	Lys	Leu	Lys	Arg	Leu	Met	Ala	Ala	Met	Thr	Phe	Val	Ala	Ala	Gly	
1				5					10						15	

gtt	gcg	acc	gtc	aac	gcg	gtg	gcc									72
Val	Ala	Thr	Val	Asn	Ala	Val	Ala									
				20												

<210> 2

<211> 24

<212> PRT

<213> Sequência artificial

<220>

<223> Sequência líder da proteína mutante que se liga a fosfato
 (pbp*)

<400> 2

Met	Lys	Leu	Lys	Arg	Leu	Met	Ala	Ala	Met	Thr	Phe	Val	Ala	Ala	Gly	
1				5					10						15	
Val	Ala	Thr	Val	Asn	Ala	Val	Ala									
				20												

<210> 3

<211> 66

<212> DNA

<213> Pseudomonas fluorescens

<220>

<221> CDS

<222> (1)...(66)

<400> 3

atg cgt aat ctg atc ctc agc gcc gct ctc gtc act gcc agc ctc ttc 48
 Met Arg Asn Leu Ile Leu Ser Ala Ala Leu Val Thr Ala Ser Leu Phe
 1 5 10 15

ggc atg acc gca caa gct 66
 Gly Met Thr Ala Gln Ala
 20

<210> 4
 <211> 22
 <212> PRT
 <213> Pseudomonas fluorescens

<400> 4
 Met Arg Asn Leu Ile Leu Ser Ala Ala Leu Val Thr Ala Ser Leu Phe
 1 5 10 15
 Gly Met Thr Ala Gln Ala
 20

<210> 5
 <211> 63
 <212> DNA
 <213> Pseudomonas fluorescens

<220>
 <221> CDS
 <222> (1)...(63)

<400> 5
 atg cgc ttg acc cag att att gcc gcc gca gcc att gcg ttg gtt tcc 48
 Met Arg Leu Thr Gln Ile Ile Ala Ala Ala Ala Ile Ala Leu Val Ser
 1 5 10 15

acc ttt gcg ctc gcc 63
 Thr Phe Ala Leu Ala
 20

<210> 6
 <211> 21
 <212> PRT
 <213> Pseudomonas fluorescens

<400> 6
 Met Arg Leu Thr Gln Ile Ile Ala Ala Ala Ala Ile Ala Leu Val Ser
 1 5 10 15
 Thr Phe Ala Leu Ala
 20

<210> 7
 <211> 99
 <212> DNA
 <213> Bacillus coagulans

<220>
 <221> CDS
 <222> (1)...(99)

<400> 7
 atg agc aca cga atc ccc cgc cga caa tgg ctg aaa ggc gcc tcg ggc 48
 Met Ser Thr Arg Ile Pro Arg Arg Gln Trp Leu Lys Gly Ala Ser Gly
 1 5 10 15

ctg ctg gcc gcc gcg agc ctg ggc cgg ttg gcc aac cgc gag gcg cgc 96

Leu Leu Ala Ala Ala Ser Leu Gly Arg Leu Ala Asn Arg Glu Ala Arg
 20 25 30

gcc
 Ala

99

<210> 8
 <211> 33
 <212> PRT
 <213> Bacillus coagulans

<400> 8
 Met Ser Thr Arg Ile Pro Arg Arg Gln Trp Leu Lys Gly Ala Ser Gly
 1 5 10 15
 Leu Leu Ala Ala Ala Ser Leu Gly Arg Leu Ala Asn Arg Glu Ala Arg
 20 25 30
 Ala

<210> 9
 <211> 75
 <212> DNA
 <213> Pseudomonas fluorescens

<220>
 <221> CDS
 <222> (1)...(75)

<400> 9
 atg tcg tgc aca cgt gca ttc aaa cca ctg ctg ctg atc ggc ctg gcc 48
 Met Ser Cys Thr Arg Ala Phe Lys Pro Leu Leu Leu Ile Gly Leu Ala
 1 5 10 15
 aca ctg atg tgt tcc cat gca ttc gct 75
 Thr Leu Met Cys Ser His Ala Phe Ala
 20 25

<210> 10
 <211> 25
 <212> PRT
 <213> Pseudomonas fluorescens

<400> 10
 Met Ser Cys Thr Arg Ala Phe Lys Pro Leu Leu Leu Ile Gly Leu Ala
 1 5 10 15
 Thr Leu Met Cys Ser His Ala Phe Ala
 20 25

<210> 11
 <211> 72
 <212> DNA
 <213> Pseudomonas fluorescens

<220>
 <221> CDS
 <222> (1)...(72)

<400> 11
 atg ctt ttt cgc aca tta ctg gcg agc ctt acc ttt gct gtc atc gcc 48
 Met Leu Phe Arg Thr Leu Leu Ala Ser Leu Thr Phe Ala Val Ile Ala
 1 5 10 15

ggc tta ccg tcc acg gcc cac gcg
 Gly Leu Pro Ser Thr Ala His Ala
 20

72

<210> 12
 <211> 24
 <212> PRT
 <213> Pseudomonas fluorescens

<400> 12
 Met Leu Phe Arg Thr Leu Leu Ala Ser Leu Thr Phe Ala Val Ile Ala
 1 5 10 15
 Gly Leu Pro Ser Thr Ala His Ala
 20

<210> 13
 <211> 69
 <212> DNA
 <213> Pseudomonas fluorescens

<220>
 <221> CDS
 <222> (1)...(69)

<400> 13
 atg ccg cct cgt tct atc gcc gca tgt ctg ggg ctg ctg ggc ttg ctc 48
 Met Pro Pro Arg Ser Ile Ala Ala Cys Leu Gly Leu Leu Gly Leu Leu
 1 5 10 15

atg gct acc cag gcc gcc gcc 69
 Met Ala Thr Gln Ala Ala Ala
 20

<210> 14
 <211> 23
 <212> PRT
 <213> Pseudomonas fluorescens

<400> 14
 Met Pro Pro Arg Ser Ile Ala Ala Cys Leu Gly Leu Leu Gly Leu Leu
 1 5 10 15
 Met Ala Thr Gln Ala Ala Ala
 20

<210> 15
 <211> 63
 <212> DNA
 <213> Pseudomonas fluorescens

<220>
 <221> CDS
 <222> (1)...(63)

<400> 15
 atg cgc ctc gct gcc cta ccg cta ttg ctt gcc cct ctc ttt att gcg 48
 Met Arg Leu Ala Ala Leu Pro Leu Leu Ala Pro Leu Phe Ile Ala
 1 5 10 15

ccg atg gcc gtt gcg 63
 Pro Met Ala Val Ala
 20

<210> 16
 <211> 21
 <212> PRT
 <213> Pseudomonas fluorescens

<400> 16
 Met Arg Leu Ala Ala Leu Pro Leu Leu Leu Ala Pro Leu Phe Ile Ala
 1 5 10 15
 Pro Met Ala Val Ala
 20

<210> 17
 <211> 63
 <212> DNA
 <213> Pseudomonas fluorescens

<220>
 <221> CDS
 <222> (1)...(63)

<400> 17
 atg aag ttc aaa cag ctg atg gcg atg gcg ctt ttg ttg gcc ttg agc 48
 Met Lys Phe Lys Gln Leu Met Ala Met Ala Leu Leu Leu Ala Leu Ser
 1 5 10 15
 gct gtg gcc cag gcc 63
 Ala Val Ala Gln Ala
 20

<210> 18
 <211> 21
 <212> PRT
 <213> Pseudomonas fluorescens

<400> 18
 Met Lys Phe Lys Gln Leu Met Ala Met Ala Leu Leu Leu Ala Leu Ser
 1 5 10 15
 Ala Val Ala Gln Ala
 20

<210> 19
 <211> 63
 <212> DNA
 <213> Pseudomonas fluorescens

<220>
 <221> CDS
 <222> (1)...(63)

<400> 19
 atg aat aga tct tcc gcg ttg ctc ctc gct ttt gtc ttc ctc agc ggc 48
 Met Asn Arg Ser Ser Ala Leu Leu Leu Ala Phe Val Phe Leu Ser Gly
 1 5 10 15
 tgc cag gcc atg gcc 63
 Cys Gln Ala Met Ala
 20

<210> 20
 <211> 21
 <212> PRT
 <213> Pseudomonas fluorescens

<400> 20

Met Asn Arg Ser Ser Ala Leu Leu Leu Ala Phe Val Phe Leu Ser Gly
 1 5 10 15
 Cys Gln Ala Met Ala
 20

<210> 21

<211> 99

<212> DNA

<213> *Pseudomonas fluorescens*

<220>

<221> CDS

<222> (1)...(99)

<400> 21

atg caa aac cgc act gtg gaa atc ggt gtc ggc ctt ttc ttg ctg gct 48
 Met Gln Asn Arg Thr Val Glu Ile Gly Val Gly Leu Phe Leu Leu Ala
 1 5 10 15
 ggc atc ctg gct tta ctg ttg ttg gcc ctg cga gtc agc ggc ctt tcg 96
 Gly Ile Leu Ala Leu Leu Leu Leu Ala Leu Arg Val Ser Gly Leu Ser
 20 25 30
 gcc 99
 Ala

<210> 22

<211> 33

<212> PRT

<213> *Pseudomonas fluorescens*

<400> 22

Met Gln Asn Arg Thr Val Glu Ile Gly Val Gly Leu Phe Leu Leu Ala
 1 5 10 15
 Gly Ile Leu Ala Leu Leu Leu Leu Ala Leu Arg Val Ser Gly Leu Ser
 20 25 30
 Ala

<210> 23

<211> 117

<212> DNA

<213> *Pseudomonas fluorescens*

<220>

<221> CDS

<222> (1)...(117)

<400> 23

atg tct ctt cgt aat atg aat atc gcc ccg agg gcc ttc ctc ggc ttc 48
 Met Ser Leu Arg Asn Met Asn Ile Ala Pro Arg Ala Phe Leu Gly Phe
 1 5 10 15
 gcg ttt att ggc gcc ttg atg ttg ttg ctc ggt gtg ttc gcg ctg aac 96
 Ala Phe Ile Gly Ala Leu Met Leu Leu Leu Gly Val Phe Ala Leu Asn
 20 25 30
 cag atg agc aaa att cgt gcg 117
 Gln Met Ser Lys Ile Arg Ala
 35

<210> 24
 <211> 39
 <212> PRT
 <213> Pseudomonas fluorescens

<400> 24
 Met Ser Leu Arg Asn Met Asn Ile Ala Pro Arg Ala Phe Leu Gly Phe
 1 5 10 15
 Ala Phe Ile Gly Ala Leu Met Leu Leu Leu Gly Val Phe Ala Leu Asn
 20 25 30
 Gln Met Ser Lys Ile Arg Ala
 35

<210> 25
 <211> 30
 <212> DNA
 <213> Sequência artificial

<220>
 <223> iniciador de oligonucleotide

<400> 25
 aattactagt aggaggtaca ttatgcgctt 30

<210> 26
 <211> 30
 <212> DNA
 <213> Sequência artificial

<220>
 <223> iniciador de oligonucleotide

<400> 26
 tatactcgag ttatttaacc tgtttcagta 30

<210> 27
 <211> 5
 <212> PRT
 <213> Sequência artificial

<220>
 <223> Primeiros 5 aminoácidos da Sequência predita da proteína para a forma processada de dsbC-Skp

<400> 27
 Ala Asp Lys Ile Ala
 1 5

<210> 28
 <211> 10
 <212> PRT
 <213> Sequência artificial

<220>
 <223> Primeiros 10 aminoácidos da Sequência predita da proteína para a forma não processada de dsbC-Skp

<400> 28
 Met Arg Leu Thr Gln Ile Ile Ala Ala Ala
 1 5 10

<210> 29

<211> 10

<212> PRT

<213> Sequência artificial

<220>

<223> Primeiros 10 aminoácidos da Sequência predita da proteína para a forma processada de dsbC-Skp

<400> 29

Ala Asp Lys Ile Ala Ile Val Asn Met Gly
1 5 10

<210> 30

<211> 63

<212> DNA

<213> Pseudomonas fluorescens

<220>

<221> CDS

<222> (1)...(63)

<400> 30

atg aag aag tcc acc ttg gct gtg gct gta acg ttg ggc gca atc gcc 48
Met Lys Lys Ser Thr Leu Ala Val Ala Val Thr Leu Gly Ala Ile Ala
1 5 10 15

cag caa gca ggc gct 63
Gln Gln Ala Gly Ala
20

<210> 31

<211> 21

<212> PRT

<213> Pseudomonas fluorescens

<400> 31

Met Lys Lys Ser Thr Leu Ala Val Ala Val Thr Leu Gly Ala Ile Ala
1 5 10 15
Gln Gln Ala Gly Ala
20

<210> 32

<211> 72

<212> DNA

<213> Pseudomonas fluorescens

<220>

<221> CDS

<222> (1)...(72)

<400> 32

atg aaa ctg aaa aac acc ttg ggc ttg gcc att ggt tct ctt att gcc 48
Met Lys Leu Lys Asn Thr Leu Gly Leu Ala Ile Gly Ser Leu Ile Ala
1 5 10 15

gct act tct ttc ggc gtt ctg gca 72
Ala Thr Ser Phe Gly Val Leu Ala
20

<210> 33

<211> 24

<212> PRT

<213> *Pseudomonas fluorescens*

<400> 33

```

Met Lys Leu Lys Asn Thr Leu Gly Leu Ala Ile Gly Ser Leu Ile Ala
 1             5             10             15
Ala Thr Ser Phe Gly Val Leu Ala
                20

```

<210> 34

<211> 72

<212> DNA

<213> *Pseudomonas fluorescens*

<220>

<221> CDS

<222> (1)...(72)

<400> 34

```

atg aaa ctg aaa cgt ttg atg gcg gca atg act ttt gtc gct gct ggc 48
Met Lys Leu Lys Arg Leu Met Ala Ala Met Thr Phe Val Ala Ala Gly
 1             5             10             15

gtt gcg acc gcc aac gcg gtg gcc 72
Val Ala Thr Ala Asn Ala Val Ala
                20

```

<210> 35

<211> 24

<212> PRT

<213> *Pseudomonas fluorescens*

<400> 35

```

Met Lys Leu Lys Arg Leu Met Ala Ala Met Thr Phe Val Ala Ala Gly
 1             5             10             15
Val Ala Thr Ala Asn Ala Val Ala
                20

```

<210> 36

<211> 60

<212> DNA

<213> *Pseudomonas fluorescens*

<220>

<221> CDS

<222> (1)...(60)

<400> 36

```

atg ttt gcc aaa ctc gtt gct gtt tcc ctg ctg act ctg gcg agc ggc 48
Met Phe Ala Lys Leu Val Ala Val Ser Leu Leu Thr Leu Ala Ser Gly
 1             5             10             15

cag ttg ctt gct 60
Gln Leu Leu Ala
                20

```

<210> 37

<211> 20

<212> PRT

<213> *Pseudomonas fluorescens*

<400> 37

Met Phe Ala Lys Leu Val Ala Val Ser Leu Leu Thr Leu Ala Ser Gly
 1 5 10 15
 Gln Leu Leu Ala
 20

<210> 38
 <211> 51
 <212> DNA
 <213> Pseudomonas fluorescens

<220>
 <221> CDS
 <222> (1)...(51)

<400> 38
 atg atc aaa cgc aat ctg ctg gtt atg ggc ctt gcc gtg ctg ttg agc 48
 Met Ile Lys Arg Asn Leu Leu Val Met Gly Leu Ala Val Leu Leu Ser
 1 5 10 15
 gct 51
 Ala

<210> 39
 <211> 17
 <212> PRT
 <213> Pseudomonas fluorescens

<400> 39
 Met Ile Lys Arg Asn Leu Leu Val Met Gly Leu Ala Val Leu Leu Ser
 1 5 10 15
 Ala

<210> 40
 <211> 69
 <212> DNA
 <213> Pseudomonas fluorescens

<220>
 <221> CDS
 <222> (1)...(69)

<400> 40
 atg cag aac tat aaa aaa ttc ctt ctg gcc gcg gcc gtc tcg atg gcg 48
 Met Gln Asn Tyr Lys Lys Phe Leu Leu Ala Ala Ala Val Ser Met Ala
 1 5 10 15
 ttc agc gcc acg gcc atg gca 69
 Phe Ser Ala Thr Ala Met Ala
 20

<210> 41
 <211> 23
 <212> PRT
 <213> Pseudomonas fluorescens

<400> 41
 Met Gln Asn Tyr Lys Lys Phe Leu Leu Ala Ala Ala Val Ser Met Ala
 1 5 10 15
 Phe Ser Ala Thr Ala Met Ala
 20

<210> 42
 <211> 93
 <212> DNA
 <213> *Pseudomonas fluorescens*

<220>
 <221> CDS
 <222> (1)...(93)

<400> 42
 atg atc cgt gac aac cga ctc aag aca tcc ctt ctg cgc ggc ctg acc 48
 Met Ile Arg Asp Asn Arg Leu Lys Thr Ser Leu Leu Arg Gly Leu Thr
 1 5 10 15
 ctc acc cta ctc agc ctg acc ctg ctc tcg ccc gcg gcc cat tct 93
 Leu Thr Leu Leu Ser Leu Thr Leu Leu Ser Pro Ala Ala His Ser
 20 25 30

<210> 43
 <211> 31
 <212> PRT
 <213> *Pseudomonas fluorescens*

<400> 43
 Met Ile Arg Asp Asn Arg Leu Lys Thr Ser Leu Leu Arg Gly Leu Thr
 1 5 10 15
 Leu Thr Leu Leu Ser Leu Thr Leu Leu Ser Pro Ala Ala His Ser
 20 25 30

<210> 44
 <211> 5
 <212> PRT
 <213> Sequência artificial

<220>
 <223> Sequência de aminoácidos de N-terminal de azurina
 e ibp

<400> 44
 Ala Gln Val Gln Leu
 1 5

<210> 45
 <211> 1317
 <212> DNA
 <213> *Bacillus coagulans*

<220>
 <221> CDS
 <222> (1)...(1317)

<400> 45
 atg agc aca cga atc ccc cgc cga caa tgg ctg aaa ggc gcc tcg ggc 48
 Met Ser Thr Arg Ile Pro Arg Arg Gln Trp Leu Lys Gly Ala Ser Gly
 1 5 10 15
 ctg ctg gcc gcc gcg agc ctg ggc cgg ttg gcc aac cgc gag gcg cgc 96
 Leu Leu Ala Ala Ala Ser Leu Gly Arg Leu Ala Asn Arg Glu Ala Arg
 20 25 30

gcc gcc gaa gcg agc gcc gcc gcc ccg ctc gac act ggc tcg ctg ggc Ala Ala Glu Ala Ser Ala Ala Ala Pro Leu Asp Thr Gly Ser Leu Gly 35 40 45	144
gcc tcg ccg cgc gcg acg ctc gac gcc tgc ctg caa aaa gcc gtc gac Ala Ser Pro Arg Ala Thr Leu Asp Ala Cys Leu Gln Lys Ala Val Asp 50 55 60	192
gac ggc acg ctc aag agc gtg gtg gcg atg gcc gcc acc gag cgc ggc Asp Gly Thr Leu Lys Ser Val Val Ala Met Ala Ala Thr Glu Arg Gly 65 70 75 80	240
ctc gcc tac cag ggc gcg cgc ggc ccg gcc aac gcg gcc ggc gag ccg Leu Ala Tyr Gln Gly Ala Arg Gly Pro Ala Asn Ala Ala Gly Glu Pro 85 90 95	288
atc ggc ccc gat acg gtg ttc tgg atg ctg tcg atg acc aag gcg atc Ile Gly Pro Asp Thr Val Phe Trp Met Leu Ser Met Thr Lys Ala Ile 100 105 110	336
acc gcc acc gcc tgc atg cag ctg atc gag cag ggc cgg ctc ggc ctc Thr Ala Thr Ala Cys Met Gln Leu Ile Glu Gln Gly Arg Leu Gly Leu 115 120 125	384
gac cag ccc gcc gcc gag atc ctg ccg caa ctg aag gcg ccg cag gtg Asp Gln Pro Ala Ala Glu Ile Leu Pro Gln Leu Lys Ala Pro Gln Val 130 135 140	432
ctg gag ggc ttc gac gcc gcc ggc cag ccc agg ctg cgc ccg gcg cgc Leu Glu Gly Phe Asp Ala Ala Gly Gln Pro Arg Leu Arg Pro Ala Arg 145 150 155 160	480
cgc gcg atc acg gtg cgc cac ctg ctc acg cat acc tcg ggc tat acc Arg Ala Ile Thr Val Arg His Leu Leu Thr His Thr Ser Gly Tyr Thr 165 170 175	528
tac agc atc tgg agc gag gcg ctg ggc cgc tac gaa cag gtc acg ggc Tyr Ser Ile Trp Ser Glu Ala Leu Gly Arg Tyr Glu Gln Val Thr Gly 180 185 190	576
atg ccc gac atc ggc tac tcg ctg aac ggc gcc ttc gcg gcc ccg ctc Met Pro Asp Ile Gly Tyr Ser Leu Asn Gly Ala Phe Ala Ala Pro Leu 195 200 205	624
gaa ttc gag ccc ggc gag cgc tgg caa tac ggc atc ggc atg gat tgg Glu Phe Glu Pro Gly Glu Arg Trp Gln Tyr Gly Ile Gly Met Asp Trp 210 215 220	672
gtg ggc aag ctg gtg gag gcg gtg acc gac cag tcg ctg gaa gtg gcg Val Gly Lys Leu Val Glu Ala Val Thr Asp Gln Ser Leu Glu Val Ala 225 230 235 240	720
ttc cgc gag cgg atc ttc gcg ccg ctc ggc atg cac gat acg ggc ttc Phe Arg Glu Arg Ile Phe Ala Pro Leu Gly Met His Asp Thr Gly Phe 245 250 255	768
ctg atc ggc agc gcg caa aag cgc cgc gtc gcc acc ctg cat cgg cgc Leu Ile Gly Ser Ala Gln Lys Arg Arg Val Ala Thr Leu His Arg Arg 260 265 270	816
cag gcc gat ggc tcg ctg acg ccg gaa ccc ttc gag acc aac cag cgg Gln Ala Asp Gly Ser Leu Thr Pro Glu Pro Phe Glu Thr Asn Gln Arg 275 280 285	864
ccc gag ttc ttc atg ggc ggc ggc ggg ctg ttc agc acc ccg cgc gac Pro Glu Phe Phe Met Gly Gly Gly Gly Leu Phe Ser Thr Pro Arg Asp 290 295 300	912

290	295	300	
tac ctc gcc ttc ctg	cag atg ctg ctg aac ggc	ggc gcc tgg cgc ggc	960
Tyr Leu Ala Phe Leu	Gln Met Leu Leu Asn Gly	Gly Ala Trp Arg Gly	
305	310	315 320	
gag cgg ctg ctg cgg	ccc gac acc gtg gcg agc atg ttc	cgc aac cag	1008
Glu Arg Leu Leu Arg	Pro Asp Thr Val Ala Ser Met Phe	Arg Asn Gln	
	325	330 335	
atc ggc gat ctt cag	gtt cgc gaa atg aag acc gcc	cag ccg gcc tgg	1056
Ile Gly Asp Leu Gln	Val Arg Glu Met Lys Thr Ala	Gln Pro Ala Trp	
	340	345 350	
tcg aac agc ttc gac	caa ttc ccc ggc gcg acg cac	aag tgg ggg ctg	1104
Ser Asn Ser Phe Asp	Gln Phe Pro Gly Ala Thr His	Lys Trp Gly Leu	
	355	360 365	
tcc ttc gat ctc aac	agc gag ccg ggg ccg cac	ggg cgc gcc gcc ggc	1152
Ser Phe Asp Leu Asn	Ser Glu Pro Gly Pro His	Gly Arg Gly Ala Gly	
	370	375 380	
tcg ggt agc tgg gcc	ggc ctg ctg aac acc tac ttc	tgg atc gat ccc	1200
Ser Gly Ser Trp Ala	Gly Leu Leu Asn Thr Tyr Phe	Trp Ile Asp Pro	
	385	390 395 400	
gcc aag cgc gtg acg	ggg gcg ctg ttc acg cag atg	ctg ccg ttc tac	1248
Ala Lys Arg Val Thr	Gly Ala Leu Phe Thr Gln Met	Leu Pro Phe Tyr	
	405	410 415	
gac gcg cgc gtg gtc	gat ctc tac ggg cgc ttc gag	cgc ggg ctc tac	1296
Asp Ala Arg Val Val	Asp Leu Tyr Gly Arg Phe Glu	Arg Gly Leu Tyr	
	420	425 430	
gac ggg ctg ggc cgc	gcc tga		1317
Asp Gly Leu Gly Arg	Ala *		
	435		

<210> 46

<211> 438

<212> PRT

<213> Bacillus coagulans

<400> 46

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

Tyr Ser Ile Trp Ser Glu Ala Leu Gly Arg Tyr Glu Gln Val Thr Gly
 180 185 190
 Met Pro Asp Ile Gly Tyr Ser Leu Asn Gly Ala Phe Ala Ala Pro Leu
 195 200 205
 Glu Phe Glu Pro Gly Glu Arg Trp Gln Tyr Gly Ile Gly Met Asp Trp
 210 215 220
 Val Gly Lys Leu Val Glu Ala Val Thr Asp Gln Ser Leu Glu Val Ala
 225 230 235 240
 Phe Arg Glu Arg Ile Phe Ala Pro Leu Gly Met His Asp Thr Gly Phe
 245 250 255
 Leu Ile Gly Ser Ala Gln Lys Arg Arg Val Ala Thr Leu His Arg Arg
 260 265 270
 Gln Ala Asp Gly Ser Leu Thr Pro Glu Pro Phe Glu Thr Asn Gln Arg
 275 280 285
 Pro Glu Phe Phe Met Gly Gly Gly Gly Leu Phe Ser Thr Pro Arg Asp
 290 295 300
 Tyr Leu Ala Phe Leu Gln Met Leu Leu Asn Gly Gly Ala Trp Arg Gly
 305 310 315 320
 Glu Arg Leu Leu Arg Pro Asp Thr Val Ala Ser Met Phe Arg Asn Gln
 325 330 335
 Ile Gly Asp Leu Gln Val Arg Glu Met Lys Thr Ala Gln Pro Ala Trp
 340 345 350
 Ser Asn Ser Phe Asp Gln Phe Pro Gly Ala Thr His Lys Trp Gly Leu
 355 360 365
 Ser Phe Asp Leu Asn Ser Glu Pro Gly Pro His Gly Arg Gly Ala Gly
 370 375 380
 Ser Gly Ser Trp Ala Gly Leu Leu Asn Thr Tyr Phe Trp Ile Asp Pro
 385 390 395 400
 Ala Lys Arg Val Thr Gly Ala Leu Phe Thr Gln Met Leu Pro Phe Tyr
 405 410 415
 Asp Ala Arg Val Asp Leu Tyr Gly Arg Phe Glu Arg Gly Leu Tyr
 420 425 430
 Asp Gly Leu Gly Arg Ala
 435

<210> 47
 <211> 324
 <212> DNA
 <213> Escherichia coli

<220>
 <221> CDS
 <222> (1)...(324)

<400> 47
 agc gat aaa att att cac ctg act gac gac agt ttt gac acg gat gta 48
 Ser Asp Lys Ile Ile His Leu Thr Asp Asp Ser Phe Asp Thr Asp Val
 1 5 10 15
 ctc aaa gcg gac ggg gcg atc ctc gtc gat ttc tgg gca gag tgg tgc 96
 Leu Lys Ala Asp Gly Ala Ile Leu Val Asp Phe Trp Ala Glu Trp Cys
 20 25 30
 ggt ccg tgc aaa atg atc gcc ccg att ctg gat gaa atc gct gac gaa 144
 Gly Pro Cys Lys Met Ile Ala Pro Ile Leu Asp Glu Ile Ala Asp Glu
 35 40 45
 tat cag ggc aaa ctg acc gtt gca aaa ctg aac atc gat caa aac cct 192
 Tyr Gln Gly Lys Leu Thr Val Ala Lys Leu Asn Ile Asp Gln Asn Pro
 50 55 60
 ggc act gcg ccg aaa tat ggc atc cgt ggt atc ccg act ctg ctg ctg 240
 Gly Thr Ala Pro Lys Tyr Gly Ile Arg Gly Ile Pro Thr Leu Leu Leu
 65 70 75 80
 ttc aaa aac ggt gaa gtg gcg gca acc aaa gtg ggt gca ctg tct aaa 288

Phe Lys Asn Gly Glu Val Ala Ala Thr Lys Val Gly Ala Leu Ser Lys
 85 90 95

ggt cag ttg aaa gag ttc ctc gac gct aac ctg gcg
 Gly Gln Leu Lys Glu Phe Leu Asp Ala Asn Leu Ala
 100 105

324

<210> 48
 <211> 108
 <212> PRT
 <213> Escherichia coli

<400> 48
 Ser Asp Lys Ile Ile His Leu Thr Asp Asp Ser Phe Asp Thr Asp Val
 1 5 10 15
 Leu Lys Ala Asp Gly Ala Ile Leu Val Asp Phe Trp Ala Glu Trp Cys
 20 25 30
 Gly Pro Cys Lys Met Ile Ala Pro Ile Leu Asp Glu Ile Ala Asp Glu
 35 40 45
 Tyr Gln Gly Lys Leu Thr Val Ala Lys Leu Asn Ile Asp Gln Asn Pro
 50 55 60
 Gly Thr Ala Pro Lys Tyr Gly Ile Arg Gly Ile Pro Thr Leu Leu Leu
 65 70 75 80
 Phe Lys Asn Gly Glu Val Ala Ala Thr Lys Val Gly Ala Leu Ser Lys
 85 90 95
 Gly Gln Leu Lys Glu Phe Leu Asp Ala Asn Leu Ala
 100 105

<210> 49
 <211> 63
 <212> DNA
 <213> Pseudomonas fluorescens

<220>
 <221> CDS
 <222> (1)...(63)

<400> 49
 atg aga aac ctt ctt cga gga atg ctt gtc gtt att tgc tgt atg gca
 Met Arg Asn Leu Leu Arg Gly Met Leu Val Val Ile Cys Cys Met Ala
 1 5 10 15

ggg ata gcg gcg gcg
 Gly Ile Ala Ala Ala
 20

63

<210> 50
 <211> 21
 <212> PRT
 <213> Pseudomonas fluorescens

<400> 50
 Met Arg Asn Leu Leu Arg Gly Met Leu Val Val Ile Cys Cys Met Ala
 1 5 10 15
 Gly Ile Ala Ala Ala