República Federativa do Brasil
Ministério da Economia
Instituto Nacional da Propriedade Industrial

(21) BR 112020008369-0 A2

(22) Data do Depósito: 17/10/2018

(43) Data da Publicação Nacional: 03/11/2020

(54) Título: SEQUÊNCIAS LÍDER BACTERIANAS PARA A EXPRESSÃO DE PROTEÍNA PERIPLÁSMICA

(51) Int. Cl.: C07H 21/04; C07K 1/00; C07K 14/00.

(30) Prioridade Unionista: 27/10/2017 US 62/578,304.

(71) Depositante(es): PFENEX INC..

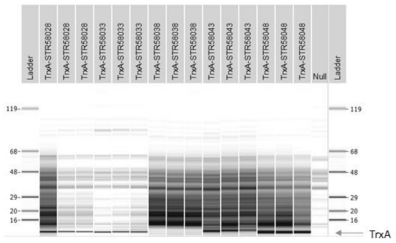
(72) Inventor(es): COLEMAN, RUSSELL J..

(86) Pedido PCT: PCT US2018056376 de 17/10/2018

(87) Publicação PCT: WO 2019/083795 de 02/05/2019

(85) Data da Fase Nacional: 27/04/2020

(57) Resumo: São providas aqui sequências líder bacterianas para a expressão periplásmica de proteínas heterólogas, proteínas de fusão compreendendo as sequências líder bacterianas e métodos para a expressão das mesmas.



SEQUÊNCIAS LÍDER BACTERIANAS PARA A EXPRESSÃO DE PROTEÍNA PERIPLÁSMICA

REFERÊNCIA CRUZADA COM PEDIDOS RELACIONADOS

[0001] Este pedido reivindica o benefício do Pedido Provisório US nº 62/578.304, depositado em 27 de outubro de 2017, o qual é aqui incorporado como referência.

LISTAGEM DE SEQUÊNCIAS

[0002] O presente pedido contém uma Listagem de Sequências a qual foi apresentada eletronicamente no formato ASCII e é aqui incorporada como referência em sua totalidade. A dita cópia ASCII, criada em 16 de outubro de 2018, é nomeada como 38194-752_601_SL.txt é tem 13.698 bytes de tamanho.

ANTECEDENTES DA INVENÇÃO

[0003] Mais de 150 proteínas e polipeptídeos produzidos por recombinação foram aprovados pelo Food e Drug Administration (FDA) dos EUA para uso em fármacos e vacinas biotecnológicos, com outros 370 em ensaios clínicos. Diferentemente de terapias de moléculas pequenas que são produzidas por meio de síntese química, proteínas e polipeptídeos são mais frequentemente produzidos em células vivas. No entanto, os métodos atuais de produção de proteínas recombinantes em bactérias frequentemente produzem proteínas inapropriadamente dobradas, agregadas ou inativas, e muitos tipos de proteínas requerem modificações secundárias que são ineficientemente obtidas pela utilização dos métodos conhecidos.

[0004] Um problema primário com os métodos conhecidos recai na formação de corpos de inclusão feitos de proteínas agregadas no citoplasma, o que pode ocorrer quando uma quantidade excessiva da proteína é acumulada na célula. Um outro problema na produção de proteína recombinante é o estabelecimento de conformação secundária e terciária das proteínas expressas. Uma barreira é que o citoplasma bacteriano resiste ativamente à formação de ligação dissulfeto, que frequentemente está por trás do dobramento apropriado da proteína (Derman *et al*., 1993, Science, 1744-7). Como resultado, muitas proteínas recombinantes,

particularmente aquelas de origem eucariótica, são dobradas de forma inapropriada e inativas quando produzidas em bactérias.

[0005] Numerosas tentativas foram desenvolvidas para aumentar a produção de proteínas dobradas apropriadamente, solúveis e/ou ativas em sistemas recombinantes. Por exemplo, os investigadores alteraram as condições de fermentação, variaram a força do promotor ou utilizaram proteínas chaperonas superexpressas, que frequência auxiliam na prevenção da formação de corpos de inclusão.

[0006] Uma abordagem alternativa para aumentar a coleta de proteínas apropriadamente dobradas, solúveis e/ou ativas é secretar a proteína do meio intracelular. A forma mais comum de secreção de polipeptídeos com uma sequência sinal envolve o sistema Sec. O sistema Sec é responsável pela exportação de proteínas com um líder de secreção no terminal N do sistema Sec através da membrana citoplasmática.

[0007] Estratégias foram desenvolvidas para excretar proteínas da célula para o sobrenadante. Outras estratégias para uma expressão aumentada são direcionadas para atingir a proteína no periplasma. Algumas investigações focalizam na secreção do tipo não Sec. No entanto, a maioria das pesquisas tem focalizado na secreção de proteínas exógenas com um sistema de secreção do tipo Sec. Um número de sinais de secreção foi descrito para o uso na expressão de polipeptídeos ou proteínas recombinantes.

[0008] As estratégias que se baseiam em sequências de sinal para com o objetivo de retirar as proteínas do citoplasma frequentemente produzem uma proteína processada inapropriadamente. Isto é particularmente verdadeiro para sinais de secreção amino-terminal tais como os que levam à secreção por meio do sistema Sec. As proteínas são processadas por meio deste sistema frequentemente ou retêm uma parte do sinal de secreção, requerem um elemento de ligação que é frequentemente clivado de forma inapropriada ou são truncadas no terminal.

[0009] Como fica claro a partir da técnica descrita acima, muitas estratégias foram desenvolvidas para direcionar proteínas para o periplasma de uma célula

hospedeira. No entanto, as estratégias conhecidas não resultarem em uma produção consistentemente alta de proteínas recombinantes apropriadamente processadas e ativas, que são frequentemente purificadas para uso terapêutico. Uma limitação importante nas estratégias anteriores foi a expressão de proteínas com sequências de sinal de secreção deficientes no sistema celular inadequado.

[0010] Permanece uma necessidade por sistemas de expressão aperfeiçoados em larga escala capazes de secretar e processar apropriadamente polipeptídeos recombinantes de maneira a os produzir em forma apropriadamente processada.

SUMÁRIO DA INVENÇÃO

[0011] São providos aqui polipeptídeos apresentando uma primeira parte e uma segunda parte, onde a primeira parte apresenta uma sequência de aminoácidos homóloga a uma ou mais das SEQ ID NOS: 1-3 e a segunda parte apresenta uma sequência de aminoácidos de uma proteína ou polipeptídeo de interesse, e onde a primeira parte e a segunda parte são operacionalmente ligadas. Em algumas realizações, a primeira parte do polipeptídeo não é nativa para a proteína ou polipeptídeo de interesse. Em algumas realizações, a proteína de interesse é selecionada de um anticorpo ou derivado de anticorpo, uma enzima, uma citocina, uma quimiocina, um fator de crescimento e um antígeno de vacina. Em algumas realizações, o anticorpo ou derivado de anticorpo é um scFv, um Fab, um anticorpo humanizado, um anticorpo modificado, um anticorpo de domínio único, um anticorpo heteroespecífico, um anticorpo trivalente, um anticorpo biespecífico, um anticorpo de cadeia única, um fragmento Fab, um anticorpo linear, um diacorpo ou um anticorpo de cadeia completa. Em algumas realizações, o polipeptídeo compreende adicionalmente um ligante. Em algumas realizações, o polipeptídeo compreende adicionalmente um domínio de clivagem. Em algumas realizações, a primeira parte é um peptídeo líder que direciona a expressão da proteína de interesse para o periplasma de uma célula hospedeira procariótica. Em algumas realizações, a célula hospedeira é selecionada e uma célula de *Pseudomonad* ou uma célula de *E. coli*.

[0012] São também providos aqui polipeptídeos compreendendo um peptídeo líder e uma proteína ou polipeptídeo de interesse, onde o peptídeo líder apresenta uma sequência de aminoácidos selecionada de uma ou mais das SEQ ID NOS: 1-3. Em algumas realizações, o peptídeo líder não é nativo para a proteína ou polipeptídeo de interesse. Em algumas realizações, a proteína ou polipeptídeo de interesse é selecionado de um anticorpo ou derivado de anticorpo, uma enzima, uma citocina, uma quimiocina, um fator de crescimento e um antígeno de vacina. Em algumas realizações, o anticorpo ou derivado de anticorpo é um scFv, um Fab, um anticorpo humanizado, um anticorpo modificado, um anticorpo de domínio único, um anticorpo heteroespecífico, um anticorpo trivalente, um anticorpo biespecífico, um anticorpo de cadeia única, um fragmento Fab, um anticorpo linear, um diacorpo ou um anticorpo de cadeia completa. Em algumas realizações, o polipeptídeo compreende adicionalmente um ligante. Em algumas realizações, o polipeptídeo compreende adicionalmente um domínio de clivagem. Em algumas realizações, a primeira parte é um peptídeo líder que direciona a expressão da proteína ou polipeptídeo de interesse para o periplasma de uma célula hospedeira procariótica. Em algumas realizações, a célula hospedeira é selecionada de uma célula de *Pseudomonad* ou uma célula de *E. coli*.

[0013] Adicionalmente são providos aqui polinucleotídeos que codificam um polipeptídeo apresentado uma primeira parte e um segunda parte, onde a primeira parte é codificada por uma sequência de ácidos nucleicos selecionada de uma ou mais das SEQ ID NOS: 4-6 e a segunda parte é codificada por uma sequência de ácidos nucleicos de uma proteína ou polipeptídeo de interesse, e onde a primeira parte e a segunda parte são operacionalmente ligadas. Em algumas realizações, a primeira parte do polipeptídeo não é nativa para a proteína ou polipeptídeo de interesse. Em algumas realizações, a proteína ou polipeptídeo de interesse é selecionado de um anticorpo ou derivado de anticorpo, uma enzima, uma citocina, uma quimiocina, um fator de crescimento e um antígeno de vacina. Em algumas realizações, o anticorpo ou derivado de anticorpo é um scFv, a Fab, um anticorpo humanizado, um anticorpo modificado, um anticorpo de domínio único,

um anticorpo heteroespecífico, um anticorpo trivalente, um anticorpo biespecífico, um anticorpo de cadeia única, um fragmento Fab, um anticorpo linear, um diacorpo ou um anticorpo de cadeia completa. Em algumas realizações, o polinucleotídeo compreende adicionalmente um ácido nucleico que codifica ligante. Em algumas realizações, o polinucleotídeo compreende adicionalmente um ácido nucleico que codifica um domínio de clivagem. Em algumas realizações, a primeira parte é um peptídeo líder que direciona a expressão da proteína ou polipeptídeo de interesse para o periplasma de uma célula hospedeira procariótica. Em algumas realizações, a célula hospedeira é selecionada de uma célula de *Pseudomonad* ou uma célula de *E. coli*.

[0014] Também são providos aqui vetores para a expressão de proteína compreendendo um ácido nucleico que codifica um peptídeo líder, onde o ácido nucleico apresenta uma sequência de ácidos nucleicos selecionada de uma ou mais das SEQ ID NOS: 4-6. Em algumas realizações, o vetor compreende adicionalmente um ligante. Em algumas realizações, o vetor compreende adicionalmente um domínio de clivagem. Em algumas realizações, o peptídeo líder direciona a expressão de uma proteína ou polipeptídeo de interesse para o periplasma de uma célula hospedeira procariótica. Em algumas realizações, a célula hospedeira é selecionada de uma célula de *Pseudomonad* ou uma célula de *E. coli*.

[0015] Também são providos aqui métodos para a produção de uma proteína ou polipeptídeo de interesse em uma cultura de célula procariótica. Em algumas realizações, o método compreende: (a) a cultura de células procarióticas compreendendo um ácido nucleico que codifica a proteína ou polipeptídeo de interesse e um peptídeo líder em um meio de crescimento de cultura celular; e (b) isolamento da proteína ou polipeptídeo de interesse do periplasma das células procarióticas, onde o peptídeo líder compreende uma sequência de aminoácidos selecionada de uma ou mais das SEQ ID NOS: 1-3. Em algumas realizações, a proteína ou polipeptídeo de interesse é selecionado de um anticorpo ou derivado de anticorpo, uma enzima, uma citocina, uma quimiocina, um fator de

crescimento e um antígeno de vacina. Em algumas realizações, o anticorpo ou derivado de anticorpo é um scFv, um Fab, um anticorpo humanizado, um anticorpo modificado, um anticorpo de domínio único, um anticorpo heteroespecífico, um anticorpo trivalente, um anticorpo biespecífico, um anticorpo de cadeia única, um fragmento Fab, um anticorpo linear, um diacorpo ou um anticorpo de cadeia completa. Em algumas realizações, o ácido nucleico codifica um ligante. Em algumas realizações, o ácido nucleico codifica um domínio de clivagem. Em algumas realizações, as células procarióticas são selecionadas de uma célula de *Pseudomonad* ou uma célula de *E. coli*. Em algumas realizações, a expressão da proteína ou polipeptídeo de interesse induzida pela adição de IPTG ao meio de cultura. Em algumas realizações, as células procarióticas são cultivadas a um pH de cerca de 5.0 a cerca de 8,0. Em algumas realizações, as células procarióticas são cultivadas a uma temperatura de cerca de 22ºC a cerca de 33ºC.

[0016] Em algumas realizações, os métodos aqui compreendem um método para a produção de uma proteína ou polipeptídeo de interesse em uma cultura de célula procariótica, o método compreendendo: (a) o cultivo de células procarióticas compreendendo um ácido nucleico que codifica a proteína ou polipeptídeo de interesse e um peptídeo líder em um meio de crescimento de cultura celular; e (b) isolamento da proteína ou polipeptídeo de interesse do periplasma das células procarióticas, onde o peptídeo líder é codificado por uma sequência de ácidos nucleicos selecionada de uma de uma ou mais das SEQ ID NOS: 4-6. Em algumas realizações, a proteína ou polipeptídeo de interesse é selecionado de um anticorpo ou derivado de anticorpo, uma enzima, uma citocina, uma quimiocina, um fator de crescimento e um antígeno de vacina. Em algumas realizações, o anticorpo ou derivado de anticorpo é um scFv, um Fab, um anticorpo humanizado, um anticorpo modificado, um anticorpo de domínio único, um anticorpo heteroespecífico, um anticorpo trivalente, um anticorpo biespecífico, um anticorpo de cadeia única, um fragmento Fab, um anticorpo linear, um diacorpo ou um anticorpo de cadeia completa. Em algumas realizações, o ácido

nucleico codifica a ligante. Em algumas realizações, o ácido nucleico codifica um domínio de clivagem. Em algumas realizações, as células procarióticas são selecionadas de uma célula de *Pseudomonad* ou uma célula de *E. coli*. Em algumas realizações, a expressão da proteína ou polipeptídeo de interesse é induzida pela adição de IPTG ao meio de cultura. Em algumas realizações, as células procarióticas são cultivadas a um pH de cerca de 5,0 a cerca de 8,0. Em algumas realizações, as células procarióticas são cultivadas a uma temperatura de cerca de 22ºC a cerca de 33ºC.

[0017] Em algumas realizações, os métodos aqui compreendem um método para a expressão de uma proteína ou polipeptídeo de interesse no periplasma de uma célula procariótica, o método compreendendo o cultivo de uma célula procariótica compreendendo um ácido nucleico que codifica a proteína ou polipeptídeo de interesse e um peptídeo líder em um meio de crescimento de cultura celular, onde o peptídeo líder compreende uma sequência de aminoácidos selecionada de uma ou mais das SEQ ID NOS: 1-3. Em algumas realizações, a proteína ou polipeptídeo de interesse é selecionado de um anticorpo ou derivado de anticorpo, uma enzima, uma citocina, uma quimiocina, um fator de crescimento e um antígeno de vacina. Em algumas realizações, o anticorpo ou derivado de anticorpo é um scFv, um Fab, um anticorpo humanizado, um anticorpo modificado, um anticorpo de domínio único, um anticorpo heteroespecífico, um anticorpo trivalente, um anticorpo biespecífico, um anticorpo de cadeia única, um fragmento Fab, um anticorpo linear, um diacorpo ou um anticorpo de cadeia completa. Em algumas realizações, o ácido nucleico codifica um ligante. Em algumas realizações, o ácido nucleico codifica um domínio de clivagem. Em algumas realizações, as células procarióticas são selecionadas de uma célula de *Pseudomonad* ou uma célula de *E. coli*. Em algumas realizações, a expressão da proteína ou polipeptídeo de interesse é induzida pela adição de IPTG ao meio de cultura. Em algumas realizações, as células procarióticas são cultivadas a um pH de cerca de 5,0 a cerca de 8,0. Em algumas realizações, as células procarióticas são cultivadas a uma temperatura de cerca de 22ºC a cerca de 33ºC.

[0018] Em algumas realizações, métodos aqui compreendem um método para a expressão de uma proteína ou polipeptídeo de interesse no periplasma de uma célula procariótica, o método compreendendo o cultivo de células procarióticas compreendendo um ácido nucleico que codifica a proteína ou polipeptídeo de interesse e um peptídeo líder em um meio de crescimento de cultura celular, onde o peptídeo líder é codificado por uma sequência de ácidos nucleicos selecionados de uma ou mais das SEQ ID NOS: 4-6. Em algumas realizações, a proteína ou polipeptídeo de interesse é selecionado de um anticorpo ou derivado de anticorpo, uma enzima, uma citocina, uma quimiocina, um fator de crescimento e um antígeno de vacina. Em algumas realizações, o anticorpo ou derivado de anticorpo é um scFv, um Fab, um anticorpo humanizado, um anticorpo modificado, um anticorpo de domínio único, um anticorpo heteroespecífico, um anticorpo trivalente, um anticorpo biespecífico, um anticorpo de cadeia única, um fragmento Fab, um anticorpo linear, um diacorpo ou um anticorpo de cadeia completa. Em algumas realizações, o ácido nucleico codifica a ligante. Em algumas realizações, o ácido nucleico codifica um domínio de clivagem. Em algumas realizações, as células procarióticas são selecionadas de uma célula de *Pseudomonad* ou uma célula de *E. coli*. Em algumas realizações, a expressão da proteína ou polipeptídeo de interesse é induzida pela adição de IPTG ao meio de cultura. Em algumas realizações, as células procarióticas são cultivadas a um pH de cerca de 5,0 a cerca de 8,0. Em algumas realizações, as células procarióticas são cultivadas a uma temperatura de cerca de 22ºC a cerca de 33ºC.

INCORPORAÇÃO POR REFERÊNCIA

[0019] Todas as publicações, patentes e pedidos de patente mencionados neste relatório são aqui incorporados como referência na mesma extensão em que cada publicação, patente ou pedido de patente individual fosse especificamente e individualmente indicado como incorporado como referência.

BREVE DESCRIÇÃO DOS DESENHOS

[0020] As novas características do relatório são apresentadas com particularidade nas reivindicações anexas. Uma melhor compreensão das características e

vantagens do presente relatório será obtida por referência à descrição detalhada a seguir que apresenta realizações ilustrativas, nas quais os princípios do relatório são utilizados e aos desenhos anexos nos quais:

[0021] A **FIG. 1** mostra imagens de Gel SDS-CGE para a expressão de TrxA.

[0022] A **FIG. 2** mostra imagens de Gel SDS-CGE para a expressão de Gal2.

[0023] A **FIG. 3** mostra imagens de Gel SDS-CGE para a expressão de mrPA.

[0024] A**FIG. 4. Sequências Exemplo de Crisantaspase.** Uma sequência de ácidos nucleicos exemplificativa de codificação de crisantaspase (SEQ ID NO: 8) é mostrada com a correspondente sequência de aminoácidos (SEQ ID NO: 7). A FIG. 4 mostra também uma sequência de nucleotídeo de comprimento completo incluindo sítios de clonagem como SEQ ID NO: 20.

[0025] A **FIG. 5. Imagens de Gel SDS-CGE – Tela do Plasmídeo de Expressão de Crisantaspase.** Amostras solúveis de sonicado de caldo total de crescimento em pequena escala de crisantaspase de DC454 (painel superior) e DC441 (painel inferior) foram analisadas por SDS-CGE reduzido. A pista na extrema esquerda mostra o marcador de peso molecular ladder 1 (painel superior PM ladder de 48 KD, 29 KD; painel inferior PM ladder de 48 KD, 29 KD, 21 KD) e a pista na extrema direita mostra os mesmos ladders. A partir da esquerda para a direita (pistas 1 a 46), começando imediatamente à direita do ladder 1 estão as pistas mostrando os padrões de expressão observados quando os seguintes peptídeos líderes de secreção foram fundidos ao terminal N da proteína crisantaspase (RBS alto exceto se indicado diferentemente): sem líder; DsbD; Líder A; DsbA; DsbA-Meio RBS; Azu; Azu-Meio RBS; Lao; Ibp-S31A; TolB; DC432 ineficaz (cepa hospedeira do tipo selvagem contendo como vetor apenas plasmídeo); Tpr; Ttg2C; FlgI; CupC2; CupB2; Pbp; PbpA20V; DsbC; Líder B; Líder C; DC432 ineficaz; Líder D; Líder E; Líder F; Líder G; Líder H; PorE; Líder I; Líder J; Líder K; Líder L; DC432 ineficaz; Líder M; Líder N; Líder O; 5193; Líder P; Líder Q; Líder R; 8484; Líder S; Líder T; DC432 ineficaz. As setas para a direita da imagem do gel indicam a migração da proteína crisantaspase alvo (35 kDa).

[0026] A **FIG. 6. Imagens de Gel SDS-CGE – Tela do Plasmídeo de Expressão de Asparaginase de *E. coli*.** Amostras solúveis (painel superior) de sonicado de caldo total de crescimento em pequena escala (0,5 ml) de asparaginase e insolúveis (painel inferior) foram analisadas por SDS-CGE reduzido. A pista na extrema esquerda mostra o marcador de peso molecular ladder (painel superior PM ladder de 119 kDa, 68 KDa, 48 kDa, 29 kDa, 21 kDa, 16 kDa; painel inferior PM ladder de 119 kDa, 68 KDa, 48 kDa, 29 kDa, 21 kDa, 16 kDa) e a pista na extrema direita mostra os mesmos ladders. A partir da direita iniciando imediatamente à direita do ladder 1 estão as pistas mostrando os padrões de expressão observados em Null, STR55467, STR55689, STR55559, STR55561, STR55569, STR55575, STR55555, STR55571, STR55560, STR55570, STR55572, STR55601, STR55585, STR55592, STR55501 e controles: L-Asparaginase de *E. coli* Sigma 1000 μg/ml, L-Asparaginase de *E. coli* Sigma 500 μg/ml, L-Asparaginase de *E. coli* Sigma 250 μg/ml, L-Asparaginase de *E. coli* Sigma 125 μg/ml e L-Asparaginase de *E. coli* Sigma 62,5 μg/ml. As setas à direita das imagens do gel indicam a migração da proteína asparaginase alvo (35 kDa).

[0027] A **FIG. 7. Imagens de Gel SDS-CGE – Análise da Expressão de Crisantaspase em Frasco de Agitação.** É mostrada a expressão sob diferentes condições de crescimento tal como medida por eletroforese em gel de capilaridade reduzida solúvel (SDS-CGE)**.** A partir da esquerda para a direita são as pistas mostrando os padrões de expressão observados nas seguintes amostras: Ladder 1 (marcadores de peso molecular 68, 48, 29, 21 e 16 KD); STR55987 em I0 (expressão citoplasmática sem líder); STR55987 em I24 (expressão citoplasmática sem líder); STR55987 em I24 (expressão citoplasmática sem líder); STR55987 em I24 (expressão citoplasmática sem líder); STR55979 em I0 (Líder O); STR55979 em I24 (Líder O); STR55979 em I24 (Líder O); STR55979 em I24 (Líder O); STR55980 em I0 (8484 Líder); STR55980 em I24 (Líder 8484); STR55980 em I24 (Líder 8484); STR55980 em I24 (Líder 8484); STR55982em I0 (plasmídeo Null); STR55982 em I24 (plasmídeo Null);

STR55982 em I24 (plasmídeo Null); STR55982 em I24 (plasmídeo Null); Ladder 2 (os mesmos marcadores que no Ladder 1); Sigma *E. coli* AspG 1,000 ug/ml (*E. coli* Asp2 padrão); Sigma *E. coli* AspG 500 ug/ml; Sigma *E. coli* AspG 250 ug/ml; Sigma *E. coli* AspG 125 ug/ml; Sigma *E. coli* AspG 62,5 ug/ml; e Ladder 3 (os mesmos marcadores que no Ladder 1), onde as amostras I0 são tomadas no momento de indução e I24 amostras são tomadas 24 horas após a indução. As setas à direita indicam a migração de L-Asp2 de *E. coli* (35 KD).

[0028] A **FIG. 8** mostra o resultado de LC-MS para a análise das para expressão da proteína.

DESCRIÇÃO DETALHADA DA INVENÇÃO

Visão Geral

[0029] São providas composições e métodos para a produção de altos níveis de proteínas ou polipeptídeos recombinantes apropriadamente processados em uma célula hospedeira. Em particular, são providos novos sinais de secreção os quais promovem o direcionamento da proteína ou polipeptídeo recombinante de interesse para o periplasma de bactérias Gram-negativas ou para o meio extracelular. Os líderes de secreção periplásmica, descritos aqui, possibilitam o transporte de proteínas através da membrana interna para o espaço periplásmico em bactérias gram negativas. Para a expressão recombinante, a expressão periplásmica permite a formação de ligações dissulfeto no periplasma e, em alguns casos, possibilita um alto nível de expressão da proteína recombinante. A expressão para o espaço periplásmico pode também possibilitar uma recuperação/purificação mais eficiente da proteína recombinante. Para os propósitos do presente relatório, um “sinal de secreção”, “líder de secreção”, “polipeptídeo sinal de secreção”, “peptídeo sinal” ou “sequência líder” se referem a uma sequência peptídica (ou o polinucleotídeo que codifica a sequência peptídica) que é útil no direcionamento de uma proteína ou polipeptídeo de interesse para o periplasma de bactérias Gram-negativas ou para o espaço extracelular. As sequências de sinal de secreção da presente invenção incluem os líderes de secreção selecionados dos sinais de secreção AnsB, 8484 e 5193 e

fragmentos e variantes destes. As sequências de aminoácidos para os sinais de secreção são apresentadas nas SEQ ID NOS: 1-3. Exemplos de sequências de nucleotídeos que codificam as SEQ ID NOS: 1-3 e úteis nos presentes métodos são providos nas SEQ ID NOS: 4-6, respectivamente (Tabela 1). Tal como de conhecimento dos técnicos no assunto, uma sequência de aminoácidos pode ser codificada por diferentes sequências de nucleotídeos devido à redundância no código genético. A presente invenção inclui, desta forma, o uso de peptídeos ou proteínas que apresentam as mesmas sequências de aminoácidos, mas são codificados por diferentes sequências de nucleotídeos. São também providos aqui fragmentos e variantes destas sequências de sinal de secreção que podem direcionar a expressão periplásmica de uma proteína recombinante ou polipeptídeo de interesse operacionalmente ligado.

Tabela 1: Sequências de aminoácidos e de ácidos nucleicos

Nome do líder	Sequência de aminoácidos	Sequência de codificação de DNA
AnsB	MKSALKNVIPGALALLLL FPVAAQA (SEQ ID NO: 1)	ATGAAATCTGCATTGAAGAACGTTATTCC GGGCGCCCTGGCCCTTCTGCTGCTATTCCC CGTCGCCGCCCAGGCC (SEQ ID NO: 4)
8484	MRQLFFCLMLMVSLTAH A (SEQ ID NO: 2)	ATGCGACAACTATTTTTCTGTTTGATGCTG ATGGTGTCGCTCACGGCGCACGCC (SEQ ID NO: 5)
5193	MQSLPFSALRLLGVLAVM VCVLLTTPARA (SEQ ID NO: 3)	ATGCAAAGCCTGCCGTTCTCTGCGTTACGC CTGCTCGGTGTGCTGGCAGTCATGGTCTGC GTGCTGTTGACGACGCCAGCCCGTGCC (SEQ ID NO: 6)

[0030] Os métodos descritos aqui provêm aperfeiçoamentos de métodos atuais para a produção de proteínas recombinantes em bactérias que frequentemente produzem proteínas inapropriadamente dobradas, agregadas ou inativas. Adicionalmente, muitos tipos de proteínas requerem modificações secundárias que são alcançadas de forma ineficiente utilizando-se os métodos conhecidos. Os métodos aqui aumentam a coleta de proteínas apropriadamente dobradas, solúveis e/ou ativas pela secreção da proteína do meio intracelular. Em bactérias Gram-

negativas, uma proteína secretada do citoplasma frequentemente acaba no espaço periplásmico, fixada na membrana externa ou no caldo extracelular. Os métodos evitam também a formação de corpos de inclusão, os quais são feitos de proteínas agregadas. A secreção para o espaço periplásmico tem também o efeito de facilitar a formação apropriada de formação de ligação dissulfeto (Bardwell *et al*., 1994, "Pathways of Disulfide Bond Formation in Proteins In Vivo", em Phosphate in Microorganisms: Celular and Molecular Biology, eds. Torriani-Gorini *et al*., pp. 270-5 e Manoil, 2000, Methods in Enzymol. 326:35-47, ambos incorporados aqui como referência). Os benefícios da secreção de uma proteína recombinante incluem o isolamento mais eficiente da proteína; o dobramento apropriado e formação de ligação dissulfeto da proteína transgênica, levando a um aumento na percentagem da proteína na forma solúvel e/ou ativa; formação reduzida de corpos de inclusão e toxicidade reduzida para a célula hospedeira. O potencial para a secreção da proteína de interesse no meio de cultura, em alguns casos, promove a produção da proteína em cultura contínua, em vez de em cultura em batelada.

[0031] Bactérias Gram-negativas evoluíram inúmeros sistemas para a exportação ativa de proteínas através de suas membranas duplas. Estas rotas de secreção incluem, por exemplo: a rota ABC (Tipo I), a rota Path/Fla (Tipo III) e a rota Path/Vir (Tipo IV) para a translocação em uma etapa através tanto do plasma quanto da membrana externa; as rotas Sec (Tipo II), Tat, MscL e Holins para a translocação através da membrana plasmática; e a rotas Sec-plus-porina usher de fímbria (FUP), Sec-plus-auto transportadora (AT), Sec-plus-secreção de dois parceiros (TPS), Sec-plus-de ramo terminal principal (MTB) e Tat-plus-MTB para a translocação em duas etapas através do plasma e membranas externas. Nem todas as bactérias apresentam todas estas rotas de secreção.

[0032] Três sistemas de proteína (tipos I, III e IV) secretam proteínas através de ambas as membranas em uma etapa única acoplada a energia. Quatro sistemas (Sec, Tat, MscL e Holins) secretam apenas através da membrana interna e quatro

outros sistemas (MTB, FUP, AT e TPS) secretam apenas através da membrana externa.

[0033] Em alguns casos, as sequências de sinal aqui utilizam o sistema de secreção Sec. O sistema Sec é responsável pela exportação de proteínas com o líder de secreção do terminal N através das membranas cito plasmáticas (ver, Agarraberes e Dice, 2001, Biochim Biophys Acta. 1513:1-24; Muller *et al*., 2001, Prog Nucleic Acid Res Mol. Biol. 66:107-157). Os complexes de proteína da família Sec são encontrados universalmente em procariotos e eucariotos. O sistema Sec bacteriano consiste em proteínas de transporte, uma proteína chaperona (SecB) ou partícula de reconhecimento de sinal (SRP) e peptidases sinal (SPase I e SPase II). O complexo de transporte Sec em *E. coli* consiste em três proteínas de membrana internas integrais, SecY, SecE e SecG e a ATPase, SecA citoplasmática. A SecA recruta complexos SecY/E/G para formar o canal de translocação ativo. A proteína chaperona SecB se liga à cadeia polipeptídica nascente para prevenir seu dobramento e a direciona para a SecA. A cadeia polipeptídica linear é subsequentemente transportada através do canal SecYEG e, após a clivagem do peptídeo sinal, a proteína é dobrada no periplasma. Três proteínas auxiliares (SecD, SecF e YajC) formam um complexo que não é essencial para a secreção, mas estimulam a secreção em até dez vezes sob muitas condições, particularmente em temperaturas baixas.

[0034] As proteínas que são transportadas para o periplasma, isto é, por meio de um sistema de secreção tipo II, são também frequentemente exportadas para o meio extracelular em uma etapa adicional. Os mecanismos são geralmente por meio de um auto transportador, um sistema de secreção de dois parceiros, um sistema de ramo terminal principal ou uma porina usher de fímbria.

[0035] Dos doze sistemas de secreção conhecidos em bactérias Gram-negativas, oito são conhecidos por utilizarem peptídeos sinais de direcionamento encontrados como parte da proteína expressa. Estes peptídeos sinais interagem com as proteínas dos sistemas de secreção de tal forma que a célula direciona apropriadamente a proteína para seu destino apropriado. Cinco destes oito

sistemas de secreção baseados em peptídeo sinal são aqueles que envolvem o sistema Sec. Estes cinco são referidos como envolvidos na translocação de membrana citoplasmática dependente de Sec e seus peptídeos sinais operativos, em alguns casos, são referidos como sinais de secreção dependentes de Sec. Uma das questões no desenvolvimento de um sinal de secreção apropriado é a de garantir que o sinal seja apropriadamente expresso e clivado a partir da proteína expressa.

[0036] Os peptídeos sinais para a rota Sec geralmente apresentam os seguintes três domínios: (i) uma região n carregada positivamente, (ii) uma região hidrofóbica e (iii) uma região c não carregada, mas polar. O sítio de clivagem para a peptidase sinal é localizado na região c. No entanto, o grau de conversão e comprimento da sequência sinal, bem como a posição do sítio de clivagem, frequentemente variam entre diferentes proteínas.

[0037] Uma assinatura da exportação da proteína dependente de Sec é a presença de uma sequência sinal no terminal amino curta (cerca de 30 aminoácidos) principalmente hidrofóbica na proteína exportada. A sequência sinal auxilia na exportação da proteína e é retirada por clivagem por uma peptidase sinal periplásmica quando a proteína exportada alcança o periplasma. Um peptídeo sinal Sec de terminal N típico contém um domínio N com pelo menos um resíduo de arginina ou lisina, seguido por um domínio que contém um trecho de resíduos hidrofóbicos e um domínio C contendo o sítio de clivagem para as peptidases sinal.

[0038] Sistemas de produção de proteína bacteriana foram desenvolvidos nos quais a os construtos de proteína transgênica são engenheirados como proteínas de fusão contendo tanto uma proteína de interesse quanto um sinal de secreção em uma tentativa de se direcionar a proteína para fora do citoplasma.

[0039] Foi demonstrado que a *P. fluorescens* é uma plataforma aperfeiçoada para a produção de uma variedade de proteínas e vários sinais de secreção eficientes foram identificados deste organismo (ver, Patente US 7.985.564, “Expression Systems with Sec-System Secretion,” incorporada aqui como referência em sua

totalidade). A *P. fluorescens* produz proteínas exógenas em uma forma processada corretamente a um nível mais alto que o tipicamente observado em outros sistemas bacterianos de expressão e transporta estas proteínas em um nível mais alto para o periplasma da célula, levando à recuperação aumentada da proteína recombinante totalmente processada. Por esta razão, em uma realização, é provido um método para a produção de proteína exógena em uma célula de *P. fluorescens* pela expressão da proteína alvo operacionalmente ligada a um sinal de secreção.

[0040] As sequências de sinal de secreção aqui, em alguns casos, são úteis em *Pseudomonas*. O sistema de *Pseudomonas* oferece vantagens para a expressão comercial de polipeptídeos e enzimas, em comparação com outros sistemas bacterianos de expressão. Em particular, a *P. fluorescens* foi identificada como um sistema de expressão vantajoso. A *P. fluorescens* engloba um grupo de saprófitos não patogênicos comuns que colonizam ambientes de solo, água e superfície vegetal. Enzimas comerciais derivadas de *P. fluorescens* têm sido utilizadas para reduzir contaminação ambiental, como aditivos em detergente e para hidrólise estéreo seletiva. A *P. fluorescens* é também utilizada em agricultura para controlar patógenos. A Patente US 4.695.462, "Cellular Encapsulation of Biological Pesticides," descreve a expressão de proteínas bacterianas recombinantes em *P. fluorescens*.

Composições

Líderes de Secreção

[0041] Em uma realização aqui, é provido um peptídeo, onde o peptídeo é um novo líder de secreção ou sinal útil para o direcionamento de uma proteína ou polipeptídeo de interesse para o periplasma de bactérias Gram-negativas ou para o espaço extracelular. Em uma realização, o peptídeo apresenta uma sequência de aminoácidos que é ou é substancialmente homóloga a um sinal de secreção AnsB, 8484 ou 5193 ou um fragmento ou variante destes. A invenção provê também um polipeptídeo compreendendo um peptídeo de sinal de secreção da invenção fundido a uma proteína ou polipeptídeo de interesse alvo e construtos de

expressão que produzem uma proteína de fusão compreendendo um peptídeo de sinal de secreção e um polipeptídeo de interesse. Nas realizações, o peptídeo de sinal de secreção é operacionalmente ligado ao polipeptídeo de interesse.

[0042] Nas realizações, a sequência de sinal de secreção é homóloga ou substancialmente homóloga a um peptídeo de sinal de secreção apresentado em qualquer uma das SEQ ID NOS: 1-3 ou é codificado por uma sequência de polinucleotídeo apresentada em qualquer uma das SEQ ID NOS: 4-6. Em uma outra realização, a sequência de sinal de secreção compreende pelo menos os aminoácidos 2-25 da SEQ ID NO: 1, pelo menos os aminoácidos 2-18 da SEQ ID NO: 2 ou pelo menos os aminoácidos 2-29 da SEQ ID NO: 3. Em ainda uma outra realização, a sequência de sinal de secreção compreende um fragmento de uma das SEQ ID NOS: 1-3, que é truncado por 1, 2, 3, 4, 5, 6, 7, 8, 9 ou 10 aminoácidos do terminal amino, mas retém a atividade biológica, isto é, a atividade de sinal de secreção.

[0043] Em uma realização, a sequência de aminoácidos do peptídeo é uma variante de um dado peptídeo original, onde a sequência da variante é obtida pela substituição de até ou cerca de 30% dos resíduos de aminoácido do peptídeo original com outro(s) resíduo(s) de aminoácido(s), incluindo até cerca de 1%, 2%, 3%, 4%, 5%, 6%, 7%, 8%, 9%, 10%, 11%, 12%, 13%, 14%, 15%, 16%, 17%, 18%, 19%, 20%, 21%, 22%, 23%, 24%, 25%, 26%, 27%, 28%, 29% ou 30%, desde que a variante retenha a função desejada do peptídeo original. Um aminoácido variante com homologia substancial será pelo menos cerca de 70%, pelo menos cerca de 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% homóloga ao peptídeo original. Uma sequência de aminoácidos variante pode ser obtida de várias maneiras incluindo substituições, deleções, truncamentos e inserções de um ou mais aminoácidos de qualquer uma das SEQ ID NOS: 1-3, incluindo 1 ou mais, 1-5, 1-10, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 ou mais substituições, deleções, inserções de aminoácidos ou combinações destas.

[0044] Por “substancialmente homóloga”, “substancialmente idêntica” ou “substancialmente similar” pretende-se dizer uma sequência de aminoácidos ou nucleotídeos que apresenta cerca de ou pelo menos cerca de 60%, cerca de ou pelo menos cerca de 65%, cerca de ou pelo menos cerca de 70%, cerca de ou pelo menos cerca de 75%, cerca de ou pelo menos cerca de 80%, cerca de ou pelo menos cerca de 85%, cerca de ou pelo menos cerca de 81%, cerca de ou pelo menos cerca de 82%, cerca de ou pelo menos cerca de 83%, cerca de ou pelo menos cerca de 84%, cerca de ou pelo menos cerca de 85%, cerca de ou pelo menos cerca de 86%, cerca de ou pelo menos cerca de 87%, cerca de ou pelo menos cerca de 88%, cerca de ou pelo menos cerca de 89%, cerca de ou pelo menos cerca de 90%, cerca de ou pelo menos cerca de 91%, cerca de ou pelo menos cerca de 92%, cerca de ou pelo menos cerca de 93%, cerca de ou pelo menos cerca de 94%, cerca de ou pelo menos cerca de 95%, cerca de ou pelo menos cerca de 96%, cerca de ou pelo menos cerca de 97%, cerca de ou pelo menos cerca de 98% ou cerca de ou pelo menos cerca de 99% ou mais de identidade de sequência em comparação com uma sequência de referência utilizando um programa de alinhamento adequado descrito aqui ou conhecido na técnica utilizando-se parâmetros padrão. Um técnico no assunto irá reconhecer que estes valores podem ser apropriadamente ajustados para determinar a correspondente identidade das proteínas codificadas por duas sequências de nucleotídeos levando-se em conta a degenerescência de códon, similaridade de aminoácido, posicionamento do quadro de leitura e semelhantes.

[0045] Nas realizações, um peptídeo, proteína ou polipeptídeo utilizado na presente invenção pode incluir uma ou mais modificações de um resíduo de aminoácido “não essencial”. Neste contexto, um resíduo de aminoácido “não essencial” é um resíduo que pode ser alterado, por exemplo, deletado, substituído ou derivatizado, na nova sequência de aminoácidos sem abolir ou substancialmente reduzir a atividade (por exemplo, a atividade agonista) do peptídeo original, proteína ou polipeptídeo (também chamado de peptídeo, proteína ou polipeptídeo “análogo” ou “de referência”). Em algumas realizações,

um peptídeo, proteína ou polipeptídeo pode incluir uma ou mais modificações de um resíduo de aminoácido "essencial". Neste contexto, um resíduo de aminoácido "essencial" é um resíduo que quando alterado, por exemplo, deletado, substituído ou derivatizado, na nova sequência de aminoácidos a atividade do peptídeo, proteína ou polipeptídeo de referência é substancialmente reduzida ou abolida. Em tais realizações, em que um resíduo de aminoácido essencial é alterado, o peptídeo, proteína ou polipeptídeo modificado pode apresentar uma atividade do peptídeo, proteína ou polipeptídeo original. As substituições, inserções e deleções podem ser na extremidade terminal N ou terminal C ou podem ser em partes internas do peptídeo, proteína ou polipeptídeo. Como exemplo, o peptídeo, proteína ou polipeptídeo pode incluir 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 ou mais substituições, tanto de maneira consecutiva quanto espaçadas por todo a molécula do peptídeo, proteína ou polipeptídeo. Separadamente ou em combinação com as substituições, o peptídeo, proteína ou polipeptídeo pode incluir 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 ou mais inserções, novamente ou de maneira consecutiva ou espaçadas por toda a molécula do peptídeo, proteína ou polipeptídeo. O peptídeo, proteína ou polipeptídeo, separadamente ou em combinação com as substituições e/ou inserções, pode incluir também 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 ou mais deleções, novamente ou de maneira consecutiva ou espaçadas por toda a molécula do peptídeo, proteína ou polipeptídeo. O peptídeo, proteína ou polipeptídeo, separadamente ou em combinação com as substituições, inserções e/ou deleções, pode incluir também 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 ou mais adições de aminoácidos.

[0046] Substituições incluem substituições conservativas de aminoácido. Uma "substituição conservativa de aminoácido" é uma na qual o resíduo de aminoácido é substituído por um resíduo de aminoácido apresentando uma cadeia lateral similar ou características físico-químicas similares (por exemplo, características eletrostáticas, ligação de hidrogênio, características isostéricas, hidrofóbicas). Os aminoácidos podem ser de ocorrência natural ou não natural. Famílias de resíduos de aminoácidos apresentando cadeias laterais similares são conhecidas na técnica. Estas famílias incluem aminoácidos com cadeias laterais

básicas (por exemplo lisina, arginina, histidina), cadeias laterais ácidas (por exemplo, ácido aspártico, ácido glutâmico), cadeias laterais polares não carregadas (por exemplo, glicina, asparagina, glutamina, serina, treonina, tirosina, metionina, cisteína), cadeias laterais não polares (por exemplo, alanina, valina, leucina, isoleucina, prolina, fenilalanina, triptofano), cadeias laterais β-ramificadas (por exemplo, treonina, valina, isoleucina) e cadeias laterais aromáticas (por exemplo, tirosina, fenilalanina, triptofano, histidina). As substituições podem incluir também alterações não conservativas.

[0047] Os peptídeos, proteínas ou polipeptídeos variantes englobados aqui são biologicamente ativos, isto é, continuam a apresentar a atividade biológica desejada do peptídeo, proteína ou polipeptídeo original; por exemplo, um peptídeo líder de secreção variante retém a atividade de sinal de secreção. Por "retém a atividade" quer-se dizer que a variante apresentará cerca de ou pelo menos cerca de 30%, cerca de ou pelo menos cerca de 35%, cerca de ou pelo menos cerca de 40%, cerca de ou pelo menos cerca de 45%, cerca de ou pelo menos cerca de 50%, cerca de ou pelo menos cerca de 55%, cerca de ou pelo menos cerca de 60%, cerca de ou pelo menos cerca de 65%, cerca de ou pelo menos cerca de 70%, cerca de ou pelo menos cerca de 75%, cerca de ou pelo menos cerca de 80%, cerca de ou pelo menos cerca de 85%, cerca de ou pelo menos cerca de 81%, cerca de ou pelo menos cerca de 82%, cerca de ou pelo menos cerca de 83%, cerca de ou pelo menos cerca de 84%, cerca de ou pelo menos cerca de 85%, cerca de ou pelo menos cerca de 86%, cerca de ou pelo menos cerca de 87%, cerca de ou pelo menos cerca de 88%, cerca de ou pelo menos cerca de 89%, cerca de ou pelo menos cerca de 90%, cerca de ou pelo menos cerca de 91%, cerca de ou pelo menos cerca de 92%, cerca de ou pelo menos cerca de 93%, cerca de ou pelo menos cerca de 94%, cerca de ou pelo menos cerca de 95%, cerca de ou pelo menos cerca de 96%, cerca de ou pelo menos cerca de 97%, cerca de ou pelo menos cerca de 98% ou cerca de ou pelo menos cerca de 99%, cerca de ou pelo menos cerca de 100%, cerca de ou pelo menos cerca de 110%, cerca de ou pelo menos cerca de 125%, cerca de ou pelo

menos cerca de 150%, cerca de ou pelo menos cerca de 200% ou mais de atividade, por exemplo, a atividade de sinal de secreção, do peptídeo, proteína ou polipeptídeo original.

Polinucleotídeos

[0048] O relatório inclui também um ácido nucleico com uma sequência que codifica um novo sinal de secreção útil para o direcionamento de uma proteína ou polipeptídeo de interesse para o periplasma de bactérias Gram-negativas ou para o espaço extracelular. Em uma realização, o polinucleotídeo isolado codifica uma sequência peptídica substancialmente homóloga a um peptídeo de sinal de secreção AnsB, 8484 ou 5193. Em uma outra realização, o presente relatório provê um ácido nucleico que codifica uma sequência peptídica substancialmente homóloga a pelo menos os aminoácidos 2-25 da SEQ ID NO: 1, pelo menos os aminoácidos 2-18 da SEQ ID NO: 2 ou pelo menos os aminoácidos 2-29 da SEQ ID NO: 2 ou provê um ácido nucleico substancialmente homólogo a qualquer uma das sequências de nucleotídeos apresentadas como SEQ ID NOS: 4-6, incluindo variantes e fragmentos biologicamente ativos destas. Em uma outra realização, a sequência de ácidos nucleicos é cerca de ou pelo menos cerca de 60%, cerca de ou pelo menos cerca de 65%, cerca de ou pelo menos cerca de 70%, cerca de ou pelo menos cerca de 75%, cerca de ou pelo menos cerca de 80%, cerca de ou pelo menos cerca de 85%, cerca de ou pelo menos cerca de 81%, cerca de ou pelo menos cerca de 82%, cerca de ou pelo menos cerca de 83%, cerca de ou pelo menos cerca de 84%, cerca de ou pelo menos cerca de 85%, cerca de ou pelo menos cerca de 86%, cerca de ou pelo menos cerca de 87%, cerca de ou pelo menos cerca de 88%, cerca de ou pelo menos cerca de 89%, cerca de ou pelo menos cerca de 90%, cerca de ou pelo menos cerca de 91%, cerca de ou pelo menos cerca de 92%, cerca de ou pelo menos cerca de 93%, cerca de ou pelo menos cerca de 94%, cerca de ou pelo menos cerca de 95%, cerca de ou pelo menos cerca de 96%, cerca de ou pelo menos cerca de 97%, cerca de ou pelo menos cerca de 98% ou cerca de ou pelo menos cerca de

99% ou mais idêntica a qualquer uma das sequência de ácidos nucleicos apresentadas como SEQ ID NOS: 4-6.

[0049] Nas realizações, os peptídeos de sinal de secreção aqui são codificados por uma sequência de nucleotídeos substancialmente homóloga a qualquer uma das sequências de nucleotídeos apresentadas como SEQ ID NOS: 4-6. As correspondentes sequências de sinal de secreção peptídicas apresentando identidade substancial às sequências de sinal de secreção da presente invenção podem ser identificadas utilizando-se qualquer método apropriado conhecido na técnica, por exemplo, PCR, métodos de hibridização ou como descritos na literatura. Ver, 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. Uma sequência de nucleotídeos variante pode incluir uma sequência de nucleotídeos derivada sinteticamente que foi gerada, por exemplo, pela utilização de mutagênese sítio dirigida. Nas realizações, uma sequência mutada ainda codifica os peptídeos de sinal de secreção descritos aqui. Os peptídeos de sinal de secreção variantes são biologicamente ativos, isto é, continuam a possuir a atividade biológica desejada da proteína nativa, isto é, retêm a atividade de sinalização de. Por "retém a atividade" quer-se significar que a variante ainda apresenta cerca de 30%, cerca de ou pelo menos cerca de 35%, cerca de ou pelo menos cerca de 40%, cerca de ou pelo menos cerca de 45%, cerca de ou pelo menos cerca de 50%, cerca de ou pelo menos cerca de 55%, cerca de ou pelo menos cerca de 60%, cerca de ou pelo menos cerca de 65%, cerca de ou pelo menos cerca de 70%, cerca de ou pelo menos cerca de 75%, cerca de ou pelo menos cerca de 80%, cerca de ou pelo menos cerca de 85%, cerca de ou pelo menos cerca de 81%, cerca de ou pelo menos cerca de 82%, cerca de ou pelo menos cerca de 83%, cerca de ou pelo menos cerca de 84%, cerca de ou pelo menos cerca de 85%, cerca de ou pelo menos cerca de 86%, cerca de ou pelo menos cerca de 87%, cerca de ou pelo menos cerca de 88%, cerca de ou pelo menos cerca de 89%, cerca de ou pelo menos cerca de 90%,

cerca de ou pelo menos cerca de 91%, cerca de ou pelo menos cerca de 92%, cerca de ou pelo menos cerca de 93%, cerca de ou pelo menos cerca de 94%, cerca de ou pelo menos cerca de 95%, cerca de ou pelo menos cerca de 96%, cerca de ou pelo menos cerca de 97%, cerca de ou pelo menos cerca de 98% ou cerca de ou pelo menos cerca de 99%, cerca de ou pelo menos cerca de 100%, cerca de ou pelo menos cerca de 110%, cerca de ou pelo menos cerca de 125%, cerca de ou pelo menos cerca de 150%, cerca de ou pelo menos cerca de 200% ou mais da atividade do peptídeo de sinal de secreção original. Qualquer método apropriado para a medição da atividade do peptídeo, proteína ou polipeptídeo, por exemplo, atividade de sinal de secreção. Tais métodos são bem conhecidos na técnica, com exemplos discutidos aqui.

[0050] O técnico no assunto irá observar adicionalmente que alterações, em alguns casos, são introduzidas por mutação nas sequências de nucleotídeos providas aqui levando a alterações na sequência de aminoácidos dos peptídeos de sinal de secreção codificados, sem alterar a atividade biológica dos peptídeos de sinal de secreção. Desta forma, moléculas variantes isoladas de ácido nucleico são frequentemente criadas pela introdução de uma ou mais substituições, adições ou deleções de nucleotídeos na correspondente sequência de nucleotídeos descrita aqui, de tal forma que uma ou mais substituições, adições ou deleções de aminoácidos são introduzidas na proteína codificada. As mutações podem ser introduzidas por qualquer técnica padrão, por exemplo, mutagênese sítio dirigida e mutagênese mediada por PCR.

Homologia de Ácido Nucleico e Aminoácido

[0051] A homologia de ácido nucleico e sequência de aminoácidos é determinada de acordo com qualquer método adequado conhecido na técnica, incluindo, mas não se limitando, àqueles descritos aqui.

[0052] Por exemplo, alinhamentos e pesquisas por sequências similares podem ser realizados utilizando-se o programa U.S. National Center for Biotechnology Information (NCBI, Bethesda, MD), MegaBLAST. O uso deste programa com opções para identidade percentual estabelecida em, por exemplo, 70% para

sequências de aminoácidos ou estabelecida em, por exemplo, 90% para sequências de nucleotídeos, irá identificar aquelas sequências com 70% ou 90% ou acima de identidade de sequência com a sequência consultada. Outro software conhecido na técnica está também disponível para alinhamento e/ou pesquisa de sequências similares, por exemplo, sequências pelo menos 70% ou 90% idêntica a uma corrente de informação contendo uma sequência de sinal de secreção aqui. Por exemplo, um alinhamento de sequência de comparação para identificar sequências pelo menos 70% ou 90% idênticas a uma sequência de consulta é frequentemente realizado pelo uso, por exemplo, dos programas GAP, BESTFIT, BLAST, FASTA e TFASTA disponíveis no GCG Sequence Analysis Software Package (disponibilizado pelo Genetics Computer Group, University of Wisconsin Biotechnology Center, 1710 University Avenue, Madison, Wis. 53705), com os parâmetros padrão tal como ali especificados, mais um parâmetro para a extensão da identidade de sequência estabelecida na percentagem desejada. Também, por exemplo, o programa CLUSTAL (disponível no pacote de software PC/Gene da Intelligenetics, Mountain View, Cal.) pode ser utilizado.

[0053] Estes e outros métodos de alinhamento de sequência são bem conhecidos na técnica e podem ser conduzidos por alinhamento manual, por inspeção visual ou aplicação manual ou automática de um algoritmo de alinhamento de sequência, tal como qualquer um daqueles incorporados nos programas descritos acima. Vários algoritmos úteis incluem, por exemplo: o método de pesquisa por 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 em J. Molec. Biol. 147:195-97 (1981); o método de alinhamento de homologia descrito em 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).

[0054] O GAP versão 10, que utiliza o algoritmo de Needleman e Wunsch (1970) *supra*, pode ser utilizado para determinar identidade ou similaridade de sequência utilizando os seguintes parâmetros: % identidade e % similaridade para uma sequência de nucleotídeos utilizando Peso GAP de 50 e Comprimento Peso de 3 e a matriz de classificação wsgapdna.cmp; % identidade ou % similaridade para uma sequência de aminoácidos utilizando peso GAP de 8 e comprimento peso de 2 e o programa de classificação BLOSUM62. Programas equivalentes ou similares podem ser também utilizados como será entendido por um técnico no assunto. Por exemplo, um programa de comparação de sequência pode ser utilizado o qual, para quaisquer duas sequências em questão, gera um alinhamento apresentando as combinações idênticas de resíduo de nucleotídeo e uma identidade percentual de sequência quando em comparação com o alinhamento correspondente gerado pelo GAP versão 10. Nas realizações, a comparação de sequência é realizada por toda a sequência de consulta ou da sequência em questão ou ambas.

Condições de Hibridização

[0055] Em um outro aspecto aqui, é provido um ácido nucleico que se hibridiza a um ácido nucleico isolado com uma sequência que codifica um peptídeo com uma sequência substancialmente similar a um peptídeo de sinal de secreção AnsB, 8484 ou 5193. Em certas realizações, o ácido nucleico de hibridização irá se ligar sob condições de alta rigidez. Em várias realizações, a hibridização ocorre substancialmente em todo o comprimento da sequência de nucleotídeos que codifica o peptídeo de sinal de secreção, por exemplo, substancialmente por todo o comprimento de uma ou mais das SEQ ID NOS: 4-6. Uma molécula de ácido nucleico se hibridiza a "substancialmente todo o comprimento" de uma sequência de nucleotídeos que codifica um sinal de secreção descrito aqui quando a molécula de ácido nucleico se hibridiza a pelo menos 80% de todo o comprimento de uma ou mais das SEQ ID NOS: 4-6, pelo menos 85%, pelo

menos 90% ou pelo menos 95% de todo o comprimento. A não ser que especificado diferentemente, "substancialmente todo o comprimento" refere-se a pelo menos 80% do comprimento total da sequência de nucleotídeos que codifica um sinal de secreção onde o comprimento é medido em nucleotídeos contíguos (por exemplo, se hibridiza a pelo menos 60 nucleotídeos contíguos da SEQ ID NO: 4, pelo menos 43 nucleotídeos contíguos da SEQ ID NO: 5 ou pelo menos 43 nucleotídeos contíguos da SEQ ID NO: 6, etc.).

[0056] Em um método de hibridização, toda ou parte da sequência de nucleotídeos que codifica o peptídeo de sinal de secreção, em alguns casos, é utilizada para pesquisar bibliotecas de cDNA ou genômicas. Os métodos para a construção de tais bibliotecas de cDNA e genômicas são em geral conhecidos na técnica e são 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 rotuladas 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 enzimático. As sondas para hibridização são frequentemente feitas pela rotulagem de oligonucleotídeos sintéticos com base na sequência de nucleotídeos conhecida que codifica um peptídeo de sinal de secreção descrito aqui. Iniciadores degenerados desenhados com base de nucleotídeos ou resíduo de aminoácidos conservados na sequência de nucleotídeos ou sequência de aminoácidos codificada são algumas vezes adicionalmente utilizados. A sonda tipicamente compreende uma região de sequência de nucleotídeos que se hibridiza sob condições rigorosas a pelo menos cerca de 10, pelo menos cerca de 11, pelo menos cerca de 12, pelo menos cerca de 13, pelo menos cerca de 14, pelo menos cerca de 15, pelo menos cerca de 16, pelo menos cerca de 17, pelo menos cerca de 18, pelo menos cerca de 19, pelo menos cerca de 20 ou mais nucleotídeos consecutivos de uma sequência de nucleotídeos que codifica um peptídeo de sinal de secreção aqui ou um fragmento ou variante desta. Os métodos para a

preparação de sondas para hibridização são em geral conhecidos na técnica e são descritos em Sambrook e Russell, 2001, aqui incorporado como referência.

[0057] Nas técnicas de hibridização, toda ou parte de uma sequência de nucleotídeos conhecida é utilizada como uma sonda que hibridiza seletivamente outras sequências de nucleotídeos correspondentes presentes em uma população de fragmentos de DNA genômico clonado ou fragmentos de cDNA (isto é, bibliotecas genômicas ou de cDNA) provenientes 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 rotuladas com um grupo detectável tal como ^{32}P ou qualquer outro marcador detectável. Desta forma, por exemplo, as sondas para hibridização são frequentemente feitas pela rotulagem de oligonucleotídeos sintéticos com base na sequência de nucleotídeos que codifica um peptídeo de sinal de secreção aqui. Os métodos para a preparação de sondas para hibridização e para a construção de bibliotecas de cDNA e genômicas são em geral conhecidos na técnica e são descritos em Sambrook *et al*., 1989, Molecular Cloning: A Laboratory Manual, 2ª ed., Cold Spring Harbor Laboratory Press, Plainview, New York.

[0058] A sequência completa de nucleotídeos que codifica um peptídeo de sinal de secreção descrito aqui ou uma ou mais partes desta, pode ser utilizada como uma sonda capaz de se hibridizar especificamente a sequências de nucleotídeos correspondentes e RNAs mensageiros que codificam peptídeos de sinal de secreção. Estas sondas podem incluir sequências que são únicas e apresentam preferivelmente pelo menos cerca de 10 nucleotídeos de comprimento ou pelo menos cerca de 15 nucleotídeos de comprimento. Tais sondas podem ser utilizadas para amplificar as sequências de nucleotídeos que codificam um peptídeo de sinal de secreção correspondentes provenientes dei um organismo escolhido por PCR. Esta técnica pode ser utilizada para isolar sequências de codificação adicionais provenientes de um organismo desejado ou como um ensaio de diagnóstico para determinar a presença de sequências de codificação em um organismo. As técnicas de hibridização incluem o rastreamento de

hibridização de bibliotecas de DNA plaqueadas (ou placas ou colônias; ver, por exemplo, Sambrook *et al*., 1989.

[0059] A hibridização de tais sequências pode ser conduzida sob condições rigorosas. Por "condições rigorosas" ou "condições rigorosas de hibridização" quer-se dizer condições sob as quais uma sonda irá se hibridizar à sua sequência alvo em um grau detectável maior que outras sequências (por exemplo, pelo menos 2 vezes em relação ao fundo). As condições rigorosas são dependentes de sequência e serão diferentes em diferentes circunstâncias. Pelo controle do rigor das condições de hibridização e/ou lavagem, as sequências alvo que são 100% complementares à sonda ´são frequentemente identificadas (sondagem homóloga). Alternativamente, as condições rigorosas são algumas vezes ajustadas de maneira a permitir alguma incompatibilidade nas sequências de tal forma que graus mais baixos de similaridade são detectados (sondagem heteróloga). Geralmente, uma sonda é menor que cerca de 1000 nucleotídeos de comprimento, preferivelmente menor que 500 nucleotídeos de comprimento.

[0060] Tipicamente, as condições rigorosas serão aquelas nas quais a concentração salina é menor que cerca de 1,5 M de íon Na, tipicamente cerca de 0,01 a 1,0 M de concentração de íon Na (ou outros sais) em pH 7,0 a 8,3 e a temperatura é de pelo menos cerca de 60°C. Nas realizações, a temperatura é de cerca de 68°C. As condições rigorosas podem ser também obtidas pela adição de agentes de desestabilização tais como formamida. Condições de baixo rigor podem incluir, por exemplo, a hibridização com uma solução tampão de 30 a 35% de formamida, 1 M de NaCl, 1% SDS (dodecil sulfato de sódio) a 37°C e uma lavagem em 1X a 2X SSC (20X SSC = 3,0 M de NaCl/0,3 M de citrato de trissódio) a de 50 a 55°C. Condições de rigor moderado típicas incluem a hibridização em 40 a 45% de formamida, 1,0 M de NaCl, 1% SDS a 37 C e uma lavagem em de 0,5X a 1X de SSC a d 55 a 60°C. Condições de alto rigor típicas incluem a hibridização em 50% de formamida, 1 M de NaCl, 1% SDS a 37°C e uma lavagem em de 0,1X SSC a de 60 a 68°C. Opcionalmente, tampões de lavagem podem compreender cerca de 0,1% a cerca de 1% SDS. A duração da

hibridização é geralmente menor que cerca de 24 horas e tipicamente cerca de 4 a cerca de 12 horas.

[0061] A especificidade é tipicamente função das lavagens após a hibridização, os fatores críticos sendo a força iônica e a temperatura da solução de lavagem final. Para híbridos DNA-DNA, a Tm é frequentemente aproximada a partir da equação de Meinkoth e Wahl (1984) Anal. Biochem. 138:267-284: Tm = 81,5°C + 16,6 (log M) + 0,41 (%GC) - 0,61 (% form) - 500/L; onde M é a molaridade dos cátions monovalentes, %GC é a percentagem dos nucleotídeos guanosina e citosina no DNA, % form é a percentagem de formamida na solução de hibridização e L é o comprimento do híbrido em pares de bases. A Tm é a temperatura (sob força iônica e pH definidos) na qual 50% de uma sequência complementar alvo se hibridiza a uma sonda perfeitamente combinada. A Tm é reduzida em cerca de 1°C para cada 1% de não combinação; desta forma, Tm, hibridização e/ou as condições de lavagem são algumas vezes ajustadas para a hibridização a sequências da entidade desejada. Por exemplo, se sequências com >90% de identidade são procuradas, a Tm é frequentemente reduzida em 10 °C. Geralmente, as condições rigorosas são selecionadas para serem cerca de 5°C abaixo do ponto de fusão térmico (Tm) para a sequência específica e seu complemento a uma força iônica e pH definidos. No entanto, condições rigorosas severas frequentemente utilizam uma hibridização e/ou lavagem 1, 2, 3 ou 4°C abaixo do ponto de fusão térmico (Tm); condições rigorosas moderadas utilizam frequentemente uma hibridização e/ou lavagem 6, 7, 8, 9 ou 10°C abaixo do ponto de fusão térmico (Tm); condições rigorosas baixas frequentemente utilizam uma hibridização e/ou lavagem 11, 12, 13, 14, 15 ou 20°C abaixo do ponto de fusão térmico (Tm). Pela utilização da equação, composições de hibridização e lavagem e Tm desejados, os técnicos no assunto irão entender que variações no rigor da hibridização e/ou soluções de lavagem são inerentemente descritas. Se o grau desejado de não combinação resultar em uma Tm de menos de 45°C (solução aquosa) ou 32°C (solução de formamida), é preferível aumentar a concentração de SSC de tal forma que uma é utilizada uma temperatura mais alta.

Um guia para a hibridização de ácidos nucleicos é encontrado em Tijssen, 1993, Laboratory Techniques in Biochemistry and Molecular Biology—Hybridization with Nucleic Acid 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. Ver Sambrook *et al*., 1989.

Otimização de Códon

[0062] Em uma realização, as composições e métodos aqui compreendem a expressão de uma proteína ou polipeptídeo recombinante de interesse a partir de um construto que foi otimizado para utilização de códon em uma cepa de interesse. Nas realizações, a cepa é uma célula hospedeira de *Pseudomonas*, por exemplo, *Pseudomonas fluorescens*. Métodos para a otimização de códon para aumentar a expressão em hospedeiros bacterianos são conhecidos na técnica e descritos na literatura. Por exemplo, otimização de códons para a expressão em uma cepa hospedeira de *Pseudomonas* é descrita, por exemplo, no Pedido de Patente Publicado US nº 2007/0292918, "Codon Optimization Method" aqui incorporado como referência em sua totalidade.

[0063] Em sistemas heterólogos de expressão, a etapa de otimização pode aumentar a capacidade do hospedeiro em produzir a proteína exógena. A expressão da proteína é governada por uma quantidade de fatores incluindo aqueles que afetam a transcrição, o processamento e estabilidade do mRNA e iniciação da tradução. As etapas de otimização de polinucleotídeo podem incluir etapas para aumentar a capacidade do hospedeiro em produzir a proteína exógena, bem como as etapas para auxiliar o pesquisador em desenhar eficientemente os construtos de expressão. As estratégias de otimização podem incluir, por exemplo, a modificação das regiões de iniciação de tradução, alteração dos elementos estruturais do mRNA e o uso de diferentes vieses de códon. Métodos para a otimização de sequência de ácidos nucleicos de maneira a aumentar a expressão de uma proteína heteróloga em um hospedeiro bacteriano são conhecidos na técnica e descritos na literatura. Por exemplo, a otimização de códons para expressão em uma cepa hospedeira de *Pseudomonas* é descrita, por

exemplo, no Pedido de Patente Publicado US nº 2007/0292918 “Codon Optimization Method” aqui incorporado como referência em sua totalidade.

[0064] A otimização é direcionada para qualquer uma de um número de características da sequência de um gene heterólogo. Como um exemplo específico, uma pausa de tradução rara induzida por códon frequentemente resulta na expressão da proteína heteróloga. Uma pausa de tradução rara induzida por códon inclui a presença de códons no polinucleotídeo de interesse que são raramente utilizados no organismo hospedeiro que apresentam um efeito negativo sobre a tradução da proteína devido à escassez no pool de tRNA disponível. Um método para aumentar a tradução ótima no organismo hospedeiro inclui a realização de otimização de códon a qual resulta algumas vezes na remoção de raros códons do hospedeiro da sequência sintética de polinucleotídeo.

[0065] A iniciação de tradução alternada algumas vezes também resulta na expressão reduzida da proteína heteróloga. A iniciação de tradução alternada inclui uma sequência sintética de polinucleotídeo inadvertidamente contendo motivos capazes de funcionar como um sítio de ligação de ribossomo (RBS). Estes sítios, em alguns casos, resultam na iniciação de tradução de uma proteína truncada de um sítio interno do gene. Um método para a redução da possibilidade de produção de uma proteína truncada, que frequentemente é difícil de ser removida durante a purificação, inclui a eliminação de sequências de RBS putativas internas de uma sequência otimizada de polinucleotídeo.

[0066] A derrapagem de polimerase induzida por repetição frequentemente resulta na expressão reduzida da proteína heteróloga. A derrapagem de polimerase induzida por repetição envolve repetições de sequência de nucleotídeos que foram mostradas provocar a derrapagem ou gaguejamento da DNA polimerase que algumas vezes resulta em mutações “frameshift”. Tais repetições frequentemente também provocam a derrapagem da RNA polimerase. Em um organismo com um desvio de conteúdo G+C alto, ocorre algumas vezes um grau mais alto de repetições compostas de nucleotídeo G ou C. Desta forma, um método para a

redução da possibilidade de induzir a derrapagem da RNA polimerase, inclui a alteração das repetições estendidas de nucleotídeos G ou C.

[0067] Estruturas secundárias de interferência também algumas vezes resultam na redução da expressão de proteína heteróloga. As estruturas secundárias frequentemente sequestram a sequência de RBS ou o códon de iniciação e foram correlacionadas com uma redução na expressão de proteína. As estruturas do tipo "stem loop" estão também frequentemente na pausa e atenuação transcricionais. Uma sequência de polinucleotídeo otimizada usualmente contém estruturas secundárias mínimas na RBS e regiões de codificação genética da sequência de nucleotídeos de maneira a permitir a transcrição e tradução aumentadas.

[0068] Uma outra característica que algumas vezes afeta a expressão de proteína heteróloga é a presença de sítios de restrição. Pela remoção dos sítios de restrição que poderiam interferir com a subsequente sub-clonagem de unidades de transcrição em vetores de expressão hospedeiros, é otimizada uma sequência de polinucleotídeo.

[0069] Por exemplo, o processo de otimização frequentemente é iniciado pela identificação da sequência de aminoácidos desejada a ser expressa de forma heteróloga pelo hospedeiro. A partir da sequência de aminoácidos, é desenhado um polinucleotídeo ou DNA candidato. Durante o desenho da sequência de DNA sintética, a frequência de utilização de códon é frequentemente comparada com a utilização de códon do organismo hospedeiro de expressão e os códons raros do hospedeiro são removidos da sequência sintética. Adicionalmente, a sequência de DNA sintética candidata é algumas vezes modificadas de maneira a remover enzimas de sítios de restrição indesejadas e adicionar ou remover quaisquer sequências sinal desejadas, ligantes ou regiões não traduzidas. A sequência de DNA sintética é frequentemente analisada para a presença de estrutura secundária que possa interferir com o processo de tradução, tal como repetições G/C e estruturas stem-loop. Antes da sequência de DNA candidata ser sintetizada, o desenho da sequência otimizada é frequentemente verificado para se determinar se a sequência codifica corretamente a sequência de aminoácidos desejada.

Finalmente, a sequência de DNA candidata é sintetizada utilizando-se técnicas de síntese de DNA, tais como as conhecidas na técnica.

[0070] Em uma outra realização aqui, a utilização genérica de códon em um organismo hospedeiro, tal como *P. fluorescens*, é frequentemente utilizada para otimizar a expressão da sequência de polinucleotídeo heteróloga. A percentagem e distribuição de códons que raramente seriam considerados como preferidos para um aminoácido particular no sistema de expressão hospedeiro são avaliadas. Valores de 5% e 10% de utilização são frequentemente utilizados como valores de corte para a determinação de códons raros. Por exemplo, os códons listados na **Tabela 2** apresentam uma ocorrência calculada de menos de 5% no genoma de *P. fluorescens* MB214 e seriam geralmente evitados em um gene otimizado expresso em um *P. fluorescens* hospedeiro.

Tabela 2. Códons que ocorrem abaixo de 5% em *P. fluorescens* MB214

Aminoácido(s)	Códon(s) Utilizado(s)	% Ocorrência
G Gly	GGA	3,26
I Ile	ATA	3,05
L Leu	CTA	1,78
	CTT	4,57
	TTA	1,89
R Arg	AGA	1,39
	AGG	2,72
	CGA	4,99
S Ser	TCT	4,28

[0071] O presente relatório contempla o uso de qualquer sequência de codificação de polipeptídeo ou proteína de interesse, incluindo qualquer sequência que tenha sido otimizada para expressão na célula hospedeira de *Pseudomonas* sendo utilizada. As sequências contempladas para uso são frequentemente otimizadas a qualquer grau conforme desejado, incluindo, mas não se limitando a, otimização para eliminar: códons que ocorrem abaixo de 5% na célula hospedeira de *Pseudomonas*, códons que ocorrem abaixo de 10% na célula hospedeira de *Pseudomonas*, uma pausa de tradução induzida por códon raro, uma sequência de RBS interna putativa, uma repetição estendida de nucleotídeos G ou C, uma estrutura secundária de interferência, um sítio de restrição ou combinações destes.

[0072] Além disto, a sequência de aminoácidos de qualquer líder de secreção útil na prática dos métodos providos aqui é codificada por qualquer sequência de ácidos nucleicos apropriada. A otimização de códon para expressão em *E. coli* é descrita, por exemplo, por Welch, *et al*., 2009, PLoS One, "Design Parameters to Control Synthetic Gene Expression in Escherichia coli," 4(9): e7002, Ghane, *et al*., 2008, Krishna R. *et al*., (2008) Mol Biotechnology "Optimization of the AT-content of Codons Immediately Downstream of the Initiation Codon and Evaluation of Culture Conditions for High-level Expression of Recombinant Human G-CSF in *Escherichia coli*" 38:221-232.

Sistemas de Expressão

[0073] Os métodos aqui, em alguns casos, compreendem a expressão de polipeptídeos compreendendo uma proteína ou polipeptídeo de interesse operacionalmente ligados a um peptídeo de sinal de secreção selecionado do grupo consistindo em uma sequência de sinal de secreção AnsB, 8484 ou 5193 ou uma sequência que seja substancialmente homóloga à sequência de sinal de secreção peptídica descrita aqui nas SEQ ID NOS: 1-3. Nas realizações, a sequência de sinal de secreção peptídica é codificada por uma sequência de nucleotídeos apresentada como SEQ ID NOS: 4-6, a partir de um construto de expressão em uma célula hospedeira de *Pseudomonas*. O construto de expressão, em alguns casos, é um plasmídeo. Em algumas realizações, um plasmídeo que codifica a sequência do polipeptídeo ou proteína de interesse compreende um marcador de seleção e as células hospedeiras que mantêm o plasmídeo são crescidas sob condições seletivas. Em algumas realizações, o plasmídeo não compreende um marcador de seleção. Em algumas realizações, o construto de expressão é integrado no genoma de célula hospedeira.

[0074] Os métodos para a expressão de proteínas heterólogas, incluindo sequências reguladoras (por exemplo, promotores, líderes de secreção e sítios de ligação a ribossomo) úteis nos métodos da invenção em cepas hospedeiras, incluindo cepas hospedeiras de *Pseudomonas*, são descritos, por exemplo, na Patente US nº 7.618.799, "Bacterial leader sequences for increased expression",

na Patente US nº 7.985.564, “Expression systems with Sec-system secretion,” nas Patentes US nºs 9.394.571 e 9.580.719, ambas intituladas “Method for Rapidly Screening Microbial Hosts to Identify Certain Strains with Improved Yield and/or Quality in the Expression of Heterologous Proteins”, na Patente US nº 9.453.251, “Expression of Mammalian Proteins in *Pseudomonas fluorescens*,” na Patente US nº 8,603,824, “Process for Improved Protein Expression by Strain Engineering”, e na Patente US nº 8.530.171, “High Level Expression of Recombinant Toxin Proteins,” cada uma incorporada aqui como referência em sua totalidade. Nas realizações, um líder de secreção utilizado no contexto da presente invenção é um líder de secreção tal como descrito em qualquer uma das Patentes US nºs 7.618.799, 7.985.564, 9.394.571, 9.580.719, 9.453.251, 8.603.824 e 8.530.171. Estas patentes descrevem também cepas hospedeiras bacterianas úteis na prática dos métodos aqui, que foram engenheiradas para superexpressar moduladores de dobramento ou onde foram introduzidas mutações na protease de maneira a aumentar a expressão de proteína heteróloga.

[0075] Nas realizações, uma cepa de expressão utilizada nos métodos da invenção é qualquer cepa de expressão descrita no Exemplo 4, tal como listadas na Tabela 13. Nas realizações, uma cepa de expressão utilizada nos métodos da invenção é uma cepa de expressão microbiana apresentando um fenótipo de fundo de uma cepa de expressão descrita no Exemplo 4, tal como listada na Tabela 13. Nas realizações, uma cepa de expressão utilizada nos métodos da invenção é uma cepa de expressão microbiana apresentando um fenótipo de fundo de uma cepa de expressão descrita no Exemplo 4, tal como listada na Tabela 13 e onde a cepa expressa a asparaginase recombinante em uma fusão com o respectivo líder de secreção tal como listado na Tabela 13. Nas realizações, uma cepa de expressão utilizada nos métodos da invenção é uma cepa de expressão microbiana apresentando um fenótipo de fundo das cepas de expressão STR57864, STR57865, STR57866, STR57860, STR57861, STR57862, STR57863 descritas no in Exemplo 4, tal como listadas na Tabela 13, exceto que a cepa de expressão não é um superexpressor modulador de dobramento. Nas realizações, uma cepa

de expressão utilizada nos métodos da invenção é uma cepa de expressão microbiana apresentando um fenótipo de fundo das cepas de expressão STR57864, STR57865, STR57866, STR57860, STR57861, STR57862, STR57863 descritas no Exemplo 4, tal como listadas na Tabela 13, cultivadas sem manitol.

Promotores

[0076] Os promotores utilizados de acordo com os métodos aqui podem ser promotores constitutivos ou promotores regulados. Exemplos comuns de promotores regulados úteis incluem aqueles da família derivada do promotor lac (isto é, o promotor lacZ), especialmente os promotores tac e trc descritos na Patente US nº 4.551.433 de DeBoer, bem como Ptac16, Ptac17, PtacII, PlacUV5 e o promotor lac T7. Em uma realização, o promotor não é derivado da célula hospedeira do organismo. Em certas realizações, o promotor é derivado de um organismo *E. coli*.

[0077] Sequências promotoras induzíveis são utilizadas para regular a expressão de polipeptídeos ou proteínas de interesse de acordo com os métodos aqui. Nas realizações, promotores induzíveis úteis nos métodos aqui incluem aqueles da família derivada do promotor lac (isto é o promotor lacZ), especialmente os promotores tac e trc descritos na Patente US nº 4.551.433 de DeBoer, bem como Ptac16, Ptac17, PtacII, PlacUV5 e o promotor lac T7. Em uma realização, o promotor não é derivado de célula hospedeira do organismo. Em certas realizações, o promotor é derivado de um organismo *E. coli*. Em algumas realizações, um promotor lac é utilizado para regular a expressão do polipeptídeo ou proteína de interesse a partir de um plasmídeo. No caso de derivados do promotor lac ou membros da família, por exemplo, o promotor tac, um indutor é o IPTG (isopropil-β-D-1-tiogalactopiranosídeo, também chamado de "isopropiltiogalactosídeo"). Em certas realizações, o IPTG é adicionado à cultura para induzir a expressão do polipeptídeo ou proteína de interesse a partir de um promotor lac em uma célula hospedeira de *Pseudomonas*.

[0078] Exemplos comuns de promotores do tipo não lac úteis nos sistemas de expressão de acordo com os métodos aqui incluem, por exemplo, os listados na **Tabela 3**.

Tabela 3. Exemplos of promotores não lac	
Promotor	**Indutor**
P_R	Alta temperatura
P_L	Alta temperatura
Pm	Alquil- ou halo-benzoatos
Pu	Alquil- ou halo-toluenos
Psal	Salicilatos
P_{BAD}	Arabinose

[0079] Ver, por exemplo: J. Sanchez-Romero & V. De Lorenzo (1999) Manual of Industrial Microbiology and Biotechnology (A. Demain & J. Davies, eds.) pp. 460-74 (ASM Press, Washington, D.C.); H. Schweizer (2001) Current Opinion in Biotechnology, 12:439-445; R. Slater & R. Williams (2000 Molecular Biology and Biotechnology (J. Walker & R. Rapley, eds.) pp. 125-54 (The Royal Society of Chemistry, Cambridge, UK); e L.-M. Guzman, *et al*., 1995, J. Bacteriol. 177(14): 4121–4130, todos incorporados aqui como referência. Um promotor apresentando a sequência de nucleotídeos de um promotor nativo para a célula hospedeira bacteriana selecionada pode ser utilizada para controlar a expressão do transgene que codifica a proteína recombinante ou polipeptídeo de interesse, por exemplo, um promotor de operon de antranilato ou benzoato de *Pseudomonas* (Pant, Pben). Promotores em tandem podem ser também utilizados onde mais de um promotor são covalentemente ligados entre si, os mesmos ou diferentes quanto a sequência, por exemplo, um promotor em tandem Pant-Pben (interpromotor híbrido) ou um promotor em tandem Plac-Plac ou sendo derivado do mesmo organismo ou de organismos diferentes.

[0080] Os promotores regulados utilizam proteínas reguladoras de promotor de maneira a controlar a transcrição do gene do qual o promotor é uma parte. Onde é

utilizado um promotor regulado, uma proteína reguladora de promotor correspondente também será parte de um sistema de expressão de acordo com os métodos aqui. Exemplos de proteínas reguladoras de promotor incluem: proteínas ativadoras, por exemplo, proteína ativadora de catabólito de *E. coli*, proteína MalT; ativadores transcricionais da família AraC; proteínas repressoras, por exemplo, proteínas de *E. coli LacI*; e proteína reguladoras de função dupla, por exemplo, proteína NagC de *E. coli*. Muitos pares promotor regulado/proteína reguladora de promotor são conhecidos na técnica. Em uma realização, o construto de expressão para a(s) proteína(s) alvo e a proteína heteróloga de interesse está sob o controle do mesmo elemento regulador.

[0081] As proteínas reguladoras de promotor interagem com um composto efetor, isto é, um composto que se associa de forma reversível ou irreversível com a proteína reguladora de tal forma a possibilitar que a proteína ou se libere ou se ligue a pelo menos uma região reguladora de transcrição de DNA do gene que está sob o controle do promotor, permitindo ou bloqueando, desta forma, a ação de uma enzima transcriptase na iniciação da transcrição do gene. Os compostos efetores são classificados ou como indutores ou como co-repressores e estes compostos incluem compostos efetores nativos e compostos indutores injustificados. Muitos trios promotor regulado/proteína reguladora de promotor /composto efetor são conhecidos na técnica. Embora, em alguns casos, um composto efetor seja utilizado em toda a cultura celular ou fermentação, em uma realização na qual é utilizado o promotor regulado, após o crescimento de uma quantidade ou densidade da biomassa da célula hospedeira, a adição direta ou indireta de um composto efetor apropriado à cultura resulta na expressão do(s) gene(s) desejado(s) que codifica(m) a proteína ou polipeptídeo de interesse.

[0082] Nas realizações em que é utilizado um promotor da família, um gene *lacI* está algumas vezes presente no sistema. O gene *lacI*, que é normalmente um gene expresso constitutivamente, codifica a proteína repressora Lac proteína *LacI*, que se liga ao operador lac dos promotores da família lac. Desta forma, onde é

utilizado um promotor da família lac, o gene *lacI* é algumas vezes também incluído e expresso no sistema de expressão.

[0083] Sistemas de promotor úteis em *Pseudomonas* são descritos na literatura, por exemplo, no Pedido de Patente Publicado US nº 2008/0269070, também referenciado acima.

Outros elementos Reguladores

[0084] Nas realizações, o vetor de expressão contém uma sequência de ligação a ribossomo ótima. A força de modulação de tradução pela alteração da região de iniciação de tradução de uma proteína de interesse pode ser utilizada para aumentar a produção de proteínas citoplasmáticas heterólogas que se acumulam principalmente como corpos de inclusão devido a uma taxa de tradução que é muito rápida. A secreção de proteínas heterólogas para o espaço periplásmico de células bacterianas pode ser também aumentada pela otimização em vez de maximização dos níveis de tradução da proteína de tal forma que a taxa de tradução é sincronizada com a taxa de secreção da proteína.

[0085] A região de iniciação de tradução foi definida como a sequência que se estende imediatamente a montante do sítio de ligação ribossômico (RBS) a aproximadamente 20 nucleotídeos a jusante do códon de iniciação (McCarthy *et al*. (1990) Trends in Genetics 6:78-85, incorporado aqui como referência em sua totalidade). Em procariotos, sequências de RBS alternativas podem ser utilizadas para otimizar os níveis de tradução de proteínas heterólogas pelo provimento de taxas de tradução que são reduzidas em relação aos níveis de tradução que utilizando as sequências de RBS canônicas ou consenso (AGGAGG; SEQ ID NO: 11) descritas por Shine e Dalgarno (Proc. Natl. Acad. Sci. USA 71:1342-1346, 1974). Por “taxa de tradução” ou “eficiência de tradução” quer dizer a taxa de tradução de mRNA em proteínas dentro das células. Na maioria dos procariotos, a sequência de Shine-Dalgarno auxilia com a ligação e posicionamento do componente ribossômico 30S em relação ao códon de início no mRNA por meio da interação com uma região rica em pirimidina do RNA ribossômico 16S. A RBS (também chamada aqui de sequência de Shine-

Dalgarno) é localizada no mRNA a jusante do início de transcrição e a montante do início da tradução, tipicamente de 4 a 14 nucleotídeos a montante do códon de início e mais tipicamente de 8 a 10 nucleotídeos a montante do códon de início. Tendo em vista o papel da sequência de RBS na tradução, há uma relação direta entre a eficiência de tradução e a eficiência (ou força) da sequência de RBS.

[0086] Em algumas realizações, a modificação da sequência de RBS resulta em uma redução na taxa de tradução da proteína heteróloga. Esta redução na taxa de tradução pode corresponder a um aumento no nível de proteína ou polipeptídeo apropriadamente processado por grama de proteína produzida ou por grama de proteína do hospedeiro. A taxa de tradução reduzida pode ser correlacionada também com um nível aumentado de proteína ou polipeptídeo recuperável produzido por grama de recombinante ou por grama de proteína da célula hospedeira. A taxa de tradução reduzida pode corresponder também a qualquer combinação de uma expressão aumentada, atividade aumentada, solubilidade aumentada ou translocação aumentada (por exemplo, para um compartimento periplásmico ou secretada para o espaço extracelular). Nesta realização, o termo "aumentado" é relativo ao nível de proteína ou polipeptídeo que é produzido, apropriadamente processado, solúvel e/ou recuperável quando a proteína ou polipeptídeo de interesse é expresso nas mesmas condições e onde a sequência de nucleotídeos que codifica o polipeptídeo compreende a sequência de RBS canônica. Similarmente, o termo "reduzido" é relativo à taxa de tradução da proteína ou polipeptídeo de interesse onde o gene que codifica a proteína ou polipeptídeo compreende a sequência de RBS canônica. A taxa de tradução pode ser reduzida em pelo menos cerca de 5%, pelo menos cerca de 10%, pelo menos cerca de 15%, pelo menos cerca de 20%, cerca de 25%, cerca de 30%, cerca de 35%, cerca de 40%, cerca de 45%, cerca de 50%, cerca de 55%, cerca de 60%, cerca de 65%, cerca de 70, pelo menos cerca de 75% ou mais ou em pelo menos cerca de 2 vezes, cerca de 3 vezes, cerca de 4 vezes, cerca de 5 vezes, cerca de 6 vezes, cerca de 7 vezes ou mais.

[0087] Em algumas realizações, as variantes da sequência de RBS descritas aqui podem ser classificadas como resultando em eficiência de tradução alta, média ou baixa. Em uma realização, as sequências são ranqueadas de acordo com o nível de atividade de tradução em comparação com a atividade de tradução da sequência de RBS canônica. Uma sequência de RBS alta apresenta cerca de 60% a cerca de 100% da atividade da sequência canônica. Uma sequência de RBS média apresenta cerca de 40% a cerca de 60% da atividade da sequência canônica. Uma sequência de RBS baixa apresenta menos de cerca de 40% da atividade da sequência canônica.

[0088] Exemplos de sequências de RBS são mostrados na **Tabela 4**. As sequências foram selecionadas para força de tradução utilizando-se COP-GFP como um gene repórter e ranqueadas de acordo com a percentagem de fluorescência de RBS consenso. Cada RBS variante foi colocada em uma de três classificações gerais de fluorescência: Alta (“Hi” - 100% de fluorescência de fluorescência de RBS Consenso), Média (“Med” - 46-51% de fluorescência de RBS Consenso) e Baixa (“Lo” - 16-29% de fluorescência de RBS Consenso).

Tabela 4. Sequências de RBS

RBS	Sequência	Força de Ligação	SEQ ID NO:
Consenso	AGGAGG	Alta	11
RBS2	GGAGCG	Med	12
RBS34	GGAGCG	Med	13
RBS41	AGGAGT	Med	14
RBS43	GGAGTG	Med	15
RBS48	GAGTAA	Baixa	16
RBS1	AGAGAG	Baixa	17
RBS35	AAGGCA	Baixa	18
RBS49	CCGAAC	Baixa	19

[0089] Um construto de expressão útil na prática dos métodos aqui inclui, em adição à sequência de codificação de proteína, os seguintes elementos reguladores

operacionalmente ligados a esta: um promotor, um sítio de ligação ribossômica (RBS), um terminador de transcrição e sinais de início e de interrupção de tradução. RBSs úteis são obtidos a partir de qualquer espécie útil como células hospedeiras nos sistemas de expressão de acordo com, por exemplo, o Pedido de Patente Publicado US nº 2008/0269070 e Pedido de Patente US Série nº 12/610.207. Muitos RBSs específicos e uma variedade de RBSs consenso são conhecidos, por exemplo, aqueles descritos e referenciados por D. Frishman *et al.*, Gene 234(2):257-65 (8 de julho de. 1999); e B. E. Suzek *et al.*, Bioinformatics 17(12):1123-30 (dezembro de 2001). Em adição, RBSs nativos ou sintéticos podem ser utilizados, por exemplo, aqueles descritos em: EP 0207459 (RBSs sintéticos); O. Ikehata *et al.*, Eur. J. Biochem. 181(3):563-70 (1989). Exemplos adicionais de métodos, vetores e elementos de tradução e transcrição e outros elementos úteis nos métodos aqui são descritos, por exemplo: na Patente US nº 5.055.294 de Gilroy e Patente US nº 5.128.130 de Gilroy *et al.*; Patente US nº 5.281.532 de Rammler *et al.*; Patentes US nºs 4.695.455 e 4.861.595 de Barnes *et al.*; Patente US nº 4.755.465 de Gray *et al.*; e Patente US nº 5.169.760 de Wilcox.

Cepas Hospedeiras

[0090] Hospedeiros bacterianos, incluindo *Pseudomonads* e organismos bacterianos proximamente relacionados são contemplados para uso na prática dos métodos aqui. Em certas realizações, a célula hospedeira de *Pseudomonad* é *Pseudomonas fluorescens*. Em alguns casos, a célula hospedeira é uma célula de *E. coli*.

[0091] As células hospedeiras e construtos úteis na prática dos métodos aqui são identificados ou obtidos utilizando-se reagentes e métodos conhecidos na técnica e descritos na literatura, por exemplo, no Pedido de Patente Publicado US nº 2009/0325230, "Protein Expression Systems" aqui incorporado como referência em sua totalidade. Esta publicação descreve a produção de um polipeptídeo recombinante pela introdução de um construto de ácido nucleico em uma célula hospedeira auxotrófica de *Pseudomonas fluorescens* compreendendo um gene

lacI cromossômico inerte. O construto de ácido nucleico compreende uma sequência de nucleotídeos que codifica o polipeptídeo recombinante operacionalmente ligado a um promotor capaz de direcionar a expressão do ácido nucleico na célula hospedeira e compreende também uma sequência de nucleotídeos que codifica um marcador de seleção auxotrófico. O marcador de seleção auxotrófico é um polipeptídeo que restaura a prototrofia da célula hospedeira auxotrófica. Nas realizações, a célula é auxotrófica para prolina, uracila ou combinações destas. Nas realizações, a célula hospedeira é derivada de MB101 (ATCC depósito PTA-7841). O Pedido de Patente Publicado US nº 2009/0325230, "Protein Expression Systems" e em Schneider, *et al.*, 2005, "Auxotrophic markers *pyrF* and *proC*, in some cases, replace antibiotic markers on protein production plasmids in high-cell-density *Pseudomonas fluorescens* fermentation" Biotechnol. Progress 21(2): 343-8, ambos incorporados aqui como referência em sua totalidade, descrevem a produção de uma cepa hospedeira auxotrófica para uracila que foi construída pela deleção do gene *pyrF* na cepa MB101. O gene *pyrF* foi clonado a partir da cepa MB214 (ATCC depósito PTA-7840) para gerar um plasmídeo que complementa a deleção do *pyrF* de maneira a restaurar a prototrofia. Em realizações particulares, é utilizado um sistema de marcador de seleção auxotrófico *pyrF-proC* duplo em uma célula hospedeira de *P. fluorescens*. Uma cepa hospedeira com *pyrF* deletado produzida como descrito é frequentemente utilizada como a base para a introdução de outras alterações genômicas desejadas, incluindo aquelas descritas aqui como úteis na prática dos métodos aqui.

[0092] Nas realizações, a célula hospedeira é da ordem *Pseudomonadale*s. Onde a célula hospedeira é da ordem *Pseudomonadale*s, pode ser um membro da família *Pseudomonadaceae*, incluindo o gênero *Pseudomonas*. Hospedeiros Proteobacterianos Gama incluem membros da espécie Escherichia coli e membros da espécie *Pseudomonas fluorescens*. As células hospedeiras da ordem *Pseudomonadales*, da família *Pseudomonadaceae* ou do gênero *Psudomonas* são identificáveis por um técnico no assunto e são descritas na literatura (por

exemplo, Manual of Systematics of Archaea and Bacteria de Bergey (publicação online, 2015)). Adicionalmente, em tais cepas as proteases podem ser desativadas e construtos de superexpressão moduladores de dobramento introduzidos utilizando-se métodos bem conhecidos na técnica.

[0093] Deve ser entendido por um técnico no assunto que a produção de uma cepa hospedeira útil nos métodos da presente invenção pode ser gerada utilizando-se uma célula hospedeira disponível publicamente, por exemplo, *P. fluorescens* MB101, por exemplo, pela desativação do gene *pyrF*, utilizando-se qualquer um de muitos métodos apropriados conhecidos na técnica e descritos na literatura. Deve ser entendido também que um plasmídeo restaurador de prototrofia pode ser transformado na cepa, por exemplo, um plasmídeo contendo o gene *pyrF* da cepa MB214 utilizando-se qualquer dos muitos métodos apropriados conhecidos na técnica e descritos na literatura. Adicionalmente, em tais cepas as proteases podem ser desativadas e construtos de superexpressão moduladores de dobramento podem ser introduzidos utilizando-se métodos bem conhecidos na técnica.

[0094] Outros organismos *Pseudomonas* podem ser também úteis. *Pseudomonads* e espécies proximamente relacionadas incluem Proteobactérias Gram-negativas Subgrupo 1, que incluem o grupo de Proteobactérias pertencente às famílias e/ou gêneros descritos no Manual of Systematics of Archaea and Bacteria de Bergey (publicação online, 2015). A **Tabela 5** apresenta estas famílias e gêneros de organismos.

Tabela 5. Famílias e Gêneros Listados na Parte, "Gram-Negative Aerobic Rods and Cocci" (no Manual of Systematics of Archaea and Bacteria de Bergey (publicação online, 2015))

Família I. Pseudomonaceae	Gluconobacter *Pseudomonas* Xanthomonas Zoogloea
Família II. Azotobacteraceae	Azomonas Azotobacter Beijerinckia Derxia

Família III. Rhizobiaceae	Agrobacterium Rhizobium
Família IV. Methylomonadaceae	Methylococcus Methylomonas
Família V. Halobactériasceae	Halobacterium Halococcus
Outros Gêneros	Acetobacter Alcaligenes Bordetella Brucella Francisella Thermus

[0095] *Pseudomonas* e bactérias proximamente relacionadas são geralmente parte do grupo definido como "Proteobactérias Gram(-) Subgrupo 1" ou "Bastões Aeróbicos e Cocos Gram-Negativos" (Manual of Systematics of Archaea and Bacteria de Bergey (publicação online, 2015)). Cepas hospedeiras de *Pseudomonas* são descritas na literatura, por exemplo, no Pedido de Patente Publicado US No. 2006/0040352, ciado acima.

[0096] "Proteobactérias Gram-negativas Subgrupo 1" incluem também Proteobactérias que seriam classificadas sob este título de acordo com os critérios utilizados na classificação. O título inclui também grupos que foram previamente classificados nesta seção, mas que o são maism tais como os gêneros *Acidovorax*, *Brevundimonas*, *Burkholderia*, *Hydrogenophaga*, *Oceanimonas*, *Ralstonia* e *Stenotrophomonas*, o gênero *Sphingomonas* (e o gênero *Blastomonas*, derivado deste), que foi criado pelo reagrupamento de organismos pertencendo (e previamente chamados de espécie) ao gênero *Xanthomonas*, o gênero *Acidomonas*, que foi criado pelo reagrupamento de organismos pertencendo ao gênero *Acetobacter* tal como definido no Manual of Systematics of Archaea and Bacteria de Bergey (publicação online, 2015). Em adição os hospedeiros incluem células do gênero *Pseudomonas*, *Pseudomonas enalia* (ATCC 14393), *Pseudomonas nigrifaciensi* (ATCC 19375) e *Pseudomonas putrefaciens* (ATCC 8071), que foram reclassificados respectivamente como *Alteromonas haloplanktis*, *Alteromonas nigrifaciens* e *Alteromonas putrefaciens*.

Similarmente, por exemplo, *Pseudomonas acidovorans* (ATCC 15668) e *Pseudomonas testosteroni* (ATCC 11996) desde então foram reclassificadas como *Comamonas acidovorans* e *Comamonas testosteroni*, respectivamente; e *Pseudomonas nigrifaciens* (ATCC 19375) e *Pseudomonas piscicida* (ATCC 15057) foram reclassificadas respectivamente como *Pseudoalteromonas nigrifaciens* e *Pseudoalteromonas piscicida*. "Proteobactérias Gram-negativas Subgrupo 1" inclui também Proteobactérias classificadas como pertencendo a qualquer uma das famílias: *Pseudomonadaceae*, *Azotobacteraceae* (agora também chamadas pelo sinônimo, de "grupo Azotobacter" de *Pseudomonadaceae*), *Rhizobiaceae* e *Methylomonadaceae* (agora frequentemente chamadas pelo sinônimo "Methylococcaceae"). Consequentemente, em adição a estes gêneros de outra forma descritos aqui, gêneros Proteobacterianos adicionais que recaem em "Proteobactérias Gram-negativas Subgrupo 1" incluem: 1) bactérias do grupo *Azotobacter* do gênero *Azorhizophilus*; 2) bactérias da família *Pseudomonadaceae* dos gêneros *Cellvibrio*, *Oligella* e *Teredinibacter*; 3) bactérias da família *Rhizobiaceae* dos gêneros *Chelatobacter*, *Ensifer*, *Liberibacter* (também chamados de "Candidatus Liberibacter") e *Sinorhizobium*; e 4) bactérias da família *Methylococcaceae* dos gêneros *Methylobacter*, *Methylocaldum*, *Methylomicrobium*, *Methylosarcina* e *Methylosphaera*.

[0097] A célula hospedeira, em alguns casos, é selecionada de "Proteobactérias Gram-negativas Subgrupo 16". "Proteobactérias Gram-negativas Subgrupo 16" é definido como o grupo de Proteobactérias das seguintes espécies: *Pseudomonas* (com ATCC ou outros números de depósito de cepa(s) típica(s) mostrados entre parêntesis): *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 mendocina* (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 9446, 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); *Pseudomonasflectens* (ATCC 12775); *Pseudomonas azotoformans*; *Pseudomonas brenneri*; *Pseudomonas cedrella*; *Pseudomonas corrugata* (ATCC 29736); *Pseudomonas extremorientalis*; *Pseudomonas fluorescens* (ATCC 35858); *Pseudomonas gessardii*; *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 halodenitrificans*; *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 monteilii* (ATCC 700476); *Pseudomonas mosselii*; *Pseudomonas oryzihabitans* (ATCC 43272); *Pseudomonas* plecoglossicida (ATCC 700383); *Pseudomonas putida* (ATCC

12633); *Pseudomonas* reactans; *Pseudomonas spinosa* (ATCC 14606); *Pseudomonas balearica*; *Pseudomonas luteola* (ATCC 43273);. *Pseudomonas stutzeri* (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*. Em uma realização, a célula hospedeira para a expressão do polipeptídeo ou proteína de interesse é *Pseudomonas fluorescens*.

[0098] A célula hospedeira, em alguns casos, é selecionada de "Proteobactérias Gram-negativas Subgrupo 17". "Proteobactérias Gram-negativas Subgrupo 17" é definido como o grupo de Proteobactérias conhecido na técnica como "*Pseudomonad*s fluorescentes" incluindo aquelas pertencentes, por exemplo, às seguintes espécies de *Pseudomonas*: *Pseudomonas azotoformans*; *Pseudomonas brenneri*; *Pseudomonas cedrella*; *Pseudomonas cedrina*; *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 tolaasii*; e *Pseudomonas veronii*.

Proteases

[0099] Em uma realização, os métodos providos aqui compreendem a utilização de uma célula hospedeira de *Pseudomonas*, compreendendo uma ou mais mutações (por exemplo, uma deleção parcial ou completa) em um ou mais genes de protease, de maneira a produzir os polipeptídeos ou proteínas de interesse. Em algumas realizações, uma mutação em um gene de protease facilita a geração de polipeptídeos ou proteínas de interesse solúveis.

[00100] Genes de protease alvo típicos incluem as proteases classificadas como Aminopeptidases; Dipeptidases; Dipeptidil-peptidases e tripeptidil peptidases; Peptidil-dipeptidases; carboxipeptidases do tipo serina; Metalocarboxipeptidases; carboxipeptidases do tipo cisteína; Omegapeptidases; Serina proteinases; Cisteína proteinases; proteinases aspárticas; Metalo proteinases; ou Proteinases de mecanismo desconhecido.

[00101] As aminopeptidases incluem citosol aminopeptidase (leucil aminopeptidase), alanil aminopeptidase de membrana, cistinil aminopeptidase, tripeptídeo aminopeptidase, prolil aminopeptidase, arginil aminopeptidase, glutamil aminopeptidase, x-pro aminopeptidase, leucil aminopeptidase bacteriana, aminopeptidase termofílica, clostridial aminopeptidase, citosol alanil aminopeptidase, lisil aminopeptidase, x-trp aminopeptidase, triptofanil aminopeptidase, metionil aminopeptidas, aminopeptidase d-estereoespecífica, aminopeptidase ey. As dipeptidases incluem x-his dipeptidase, x-arg dipeptidase, x-metil-his dipeptidase, cys-gly dipeptidase, glu-glu dipeptidase, pro-x dipeptidase, x-pro dipeptidase, met-x dipeptidase, dipeptidase não estereoespecífica, citosol dipeptidase não específica, dipeptidase de membrana, beta-ala-his dipeptidase. As dipeptidil-peptidases e tripeptidil peptidases incluem dipeptidil-peptidase i, dipeptidil-peptidase ii, dipeptidil peptidase iii, dipeptidil-peptidase iv, dipeptidil-dipeptidase, tripeptidil-peptidase I, tripeptidil-peptidase II. As peptidil-dipeptidases incluem peptidil-dipeptidase a e peptidil-dipeptidase b. As carboxipeptidases do tipo serina incluem pro-x lisossômica carboxipeptidase, D-ala-D-ala carboxipeptidase do tipo serina, carboxipeptidase C, carboxipeptidase D. As metalocarboxipeptidases incluem carboxipeptidase a, carboxipeptidase B, lisina(arginina) carboxipeptidase, gly-X carboxipeptidase, alanina carboxipeptidase, muramoilpentapeptídeo carboxipeptidase, carboxipeptidase h, glutamato carboxipeptidase, carboxipeptidase M, muramoiltetrapeptídeo carboxipeptidase, zinco d-ala-d-ala carboxipeptidase, carboxipeptidase A2, pro-x carboxipeptidase de membrana, tubulinil-tyr carboxipeptidase, carboxipeptidase t. As omegapeptidases incluem acilaminoacil-

peptidase, peptidil-glicinamidase, piroglutamil-peptidase I, beta-aspartil-peptidase, piroglutamil-peptidase II, n-formilmetionil-peptidase, pteroilpoli-[gama]-glutamato carboxipeptidase, gama-glu-X carboxipeptidase, acilmuramoil-ala peptidase. As serina proteinases incluem quimotripsina, quimotripsina c, metridina, tripsina, trombina, fator de coagulação Xa, plasmina, enteropeptidase, acrosina, protease alpha-lítica, glutamil, endopeptidase, catepsina G, fator de coagulação viia, fator de coagulação ixa, cucumisi, prolil oligopeptidase, fator de coagulação xia, braquiurina, calicreína de plasma, calicreína de tecido, elastase pancreática, elastase de leucócito, fator de coagulação xiia, quimase, componente complemento c1r55, componente complemento c1s55, convertase da rota do complemento clássica c3/c5, fator de complemento I, fator de complemento D, convertase da rota do complemente alternativa c3/c5, cerevisina, hipodermina C, lisil endopeptidase, endopeptidase 1a, gama-reni, venombina ab, leucil endopeptidase, triptase, escutelarina, kexina, subtilisina, orizina, endopeptidase k, termomicolina, termitase, endopeptidase SO, ativador de plasminogênio T, proteína C, endopeptidase pancreática E, elastase pancreática ii, serina endopeptidase específica de IGA, U-plasminogênio, ativador, venombina A, furina, mieloblastina, semenogelase, granzima A ou proteinase de linfócito T citotóxica 1, granzima B ou proteinase de linfócito T citotóxica 2, estreptogrisina A, treptogrisina B, glutamil endopeptidase II, oligopeptidase B, fator de coagulação de límulo c, fator de coagulação de límulo, enzima da coagulação de límulo, omptina, repressor lexa, peptidase líder bacteriana I, togavirina, flavirina. As cisteína proteinases incluem catepsina B, papaína, ficina, quimopapaina, asclepaina, clostripaina, estreptopaina, actinídeo, catepsina 1, catepsina H, calpaina, catepsina t, glicil endopeptidase, procoagulante de câncer, catepsina S, picornaina 3C, picornaina 2A, caricaina, ananaina, bromelaina de caule, bromelaina de fruto, legumaina, histolisaina, enzima conversora de interleucina 1-beta. As proteinases aspárticas incluem pepsina A, pepsina B, gastricsina, quimosina, catepsina D, neopentesina, renina, retropepsina, enzima conversora de pro-opiomelanocortina, aspergilopepsina I, aspergilopepsina II, penicilopepsina,

rizopuspepsina, endotiapepsina, mucoropepsina, candidapepsina, sacaropepsina, rodotorulapepsin, fisaropepsina, acrocilindropepsina, poliporopepsina, picnoporopepsina, scitalidopepsina a, scitalidopepsina b, xantomonapepsina, catepsina e, barrierpepsina, peptidase líder bacteriana I, pseudomonapepsina, plasmepsina. As metalo proteinases incluem atrolisina a, colagenase microbiana, leucolisina, colagenase intersticial, neprilisina, envelisina, metaloendopeptidase iga-específica, procolagênio N-endopeptidase, thimet oligopeptidase, neurolisina, estromelisina 1, meprina A, procolagênio C-endopeptidase, peptidil-lys metaloendopeptidase, astacina, estromelisina, 2, matrilisina gelatinase, aeromonolisina, pseudolisina, termolisina, bacilolisina, aureolisina, cocolisina, micolisina, metaloendopeptidase beta-lítica, peptidil-asp metaloendopeptidase, neutrofil colagenase, gelatinase B, leishmanolisina, sacarolisina, autolisina, deuterolisina, serralisina, atrolisina B, atrolisina C, atroxase, atrolisina E, atrolisina F, adamalisina, horrilisina, ruberlisina, botropasin, botrolisina, opiolisina, trimerelisina I, trimerelisina II, mucrolisina, pitrilisina, insulisina, O-sialoglicoproteína endopeptidase, russelisina, peptidase mitocondrial intemediária, dactilisina, nardilisina, magnolisina, meprina B, peptidase de processamento mitocondrial, macrófago elastase, coriolisina, toxilisina. As proteinases de mecanismo desconhecido incluem termopsina e o complexo multicatalítico de endopeptidase.

[00102] Certas proteases apresentam tanto atividade de protease e quanto atividade semelhante a chaperon. Quando estas proteases afetam negativamente a produção da proteína e/ou sua qualidade, frequentemente são úteis para especificamente deletar sua atividade de protease e são superexpressas quando sua atividade de chaperona pode afetar positivamente a produção da proteína e/ou sua qualidade. Estas proteases incluem, mas não se limitam a: membros da família Hsp100(Clp/Hsl) RXF04587,1 (clpA), RXF08347.1, RXF04654.2 (clpX), RXF04663.1, RXF01957.2 (hslU), RXF01961.2 (hslV); Membro da família da peptidil-prolil cis-trans isomerase RXF05345,2 (ppiB); membro da família da metalopeptidase M20RXF04892.1 (aminohidrolase); membros da família da

metalopeptidase M24 RXF04693.1 (metionina aminopeptidase) e RXF03364,1 (methionine aminopeptidase); e membro da família da serina peptidase S26 peptidase sinal I RXF01181.1 (peptidase sinal).

[00103] Estas e outras proteases e moduladores de dobramento são conhecidos na técnica e descritos na literatura, por exemplo, na Patente US nº 8.603.824. Por exemplo, a Tabela D da patente descreve Tig (tig, fator de acionamento, FKBP tipo ppiase (ec 5.2.1.8) RXF04655, UniProtKB - P0A850 (TIG_ECOLI)). O documento WO 2008/134461 e a Patente US nº 9.394.571, intitulada "Method for Rapidly Screening Microbial Hosts to Identify Certain Strains with Improved Yield and/or Quality in the Expression of Heterologous Proteins" e incorporados aqui como referência em sua totalidade, descrevem Tig (RXF04655.2, SEQ ID NO: 34 alí), LepB (RXF01181.1, SEQ ID NO: 56 alí), DegP1 (RXF01250, SEQ ID NO: 57 alí), AprA (RXF04304.1, SEQ ID NO: 86 alí), Prc1 (RXF06586.1, SEQ ID NO: 120 alí), DegP2, (RXF07210.1, SEQ ID NO: 124 alí), Lon (RXF04653, SEQ ID NO: 92 alí); DsbA (RXF01002.1, SEQ ID NO: 25 alí) e DsbC (RXF03307.1, SEQ ID NO: 26 alí). Estas sequências e as para outras proteases e moduladores de dobramento também são apresentados na Patente US nº 9.580.719 (Tabela da SEQ ID NOS nas colunas 93-98 desta). Por exemplo, a Patente US nº 9.580.719 provê a sequência que codifica HslU (RXF01957.2) e HslV (RXF01961.2) como SEQ ID NOS 18 e 19, respectivamente.

Rastreamentos de Alto Rendimento

[00104] Nas realizações, um rastreamento de alto rendimento é conduzido para determinar as condições ótimas para a expressão de uma proteína ou polipeptídeo recombinante de interesse. As condições que podem ser variadas no rastreamento incluem, por exemplo, a célula hospedeira, base genética da célula hospedeira (por exemplo, deleções de diferentes proteases), tipo de promotor em um construto de expressão, tipo de líder de secreção fundido ao polipeptídeo ou proteína de interesse codificada, temperatura de crescimento, OD de indução quando é utilizado um promotor induzível, quantidade adicionada de indutor (por

exemplo, a quantidade de IPTG utilizado para a indução quando um promotor lacZ ou derivado deste é utilizado), duração da indução da proteína, temperatura de crescimento após a adição de um agente indutor para uma cultura, taxa de agitação da cultura, método de seleção para a manutenção do plasmídeo, volume de cultura em um vaso e método de lise celular.

[00105] Em algumas realizações, uma biblioteca (ou "conjunto") de cepas hospedeiras é provida, onde cada cepa (ou "população de células hospedeiras") na biblioteca foi geneticamente modificada para modular a expressão de um ou mais genes alvo na célula hospedeira. Uma "cepa hospedeira ótima" ou "sistema de expressão opcional" é frequentemente identificada ou selecionada com base na quantidade, qualidade e/ou localização da proteína de interesse expressa em comparação com outras populações de células hospedeiras fenotipicamente distintas no conjunto. Desta forma, uma cepa hospedeira ótima é a cepa que produz o polipeptídeo de interesse de acordo com uma especificação desejada. Embora a especificação desejada varie dependendo do polipeptídeo sendo produzido, a especificação inclui a qualidade e/ou a quantidade da proteína, seja a proteína sequestrada ou secretada, o dobramento da proteína e semelhantes. Por exemplo, a cepa hospedeira ótima ou o sistema de expressão ótimo produz um rendimento, caracterizado pela quantidade de proteína heteróloga solúvel, a quantidade de proteína heteróloga recuperável, a quantidade de proteína heteróloga apropriadamente processada, a quantidade de proteína heteróloga apropriadamente dobrada, a quantidade de proteína heteróloga ativa e/ou a quantidade total de proteína heteróloga, a um certo nível absoluto ou um certo nível relativo da produzida por uma cepa indicadora, isto é, uma cepa utilizada para comparação.

[00106] Métodos de rastreamento de hospedeiros microbianos para identificar cepas com rendimento aumentado e/ou qualidade aumentada na expressão de proteínas heterólogas são descritos, por exemplo, no Pedido de Patente Publicado US nº 20080269070.

Condições de Crescimento Bacteriano

[00107] Condições de crescimento úteis nos métodos aqui frequentemente compreendem uma temperatura de cerca de 4°C a cerca de 42°C e um pH de cerca de 5,7 a cerca de 8,8. Quando é utilizado um construto de expressão com um promotor lacZ ou derivado deste, a expressão é frequentemente induzida pela adição de IPTG a uma cultura a uma concentração final de cerca de 0,01 mM a cerca de 1,0 mM.

[00108] O pH da cultura é algumas vezes mantido utilizando-se tampões de pH e métodos conhecidos dos técnicos no assunto. O controle do pH durante a cultura é também frequentemente obtido utilizando-se amônia aquosa. Nas realizações, o pH da cultura é de cerca de 5,7 a cerca de 8,8. Em certas realizações, o pH é de cerca de 5,7, 5,8, 5,9, 6,0, 6,1, 6,2, 6,3, 6,4, 6,5, 6,6, 6,7, 6,8, 6,9, 7,0, 7,1, 7,2, 7,3, 7,4, 7,5, 7,6, 7,7, 7,8, 7,9, 8,0, 8,1, 8,2, 8,3, 8,4, 8,5, 8,6, 8,7 ou 8,8. Em outras realizações, o pH é de cerca de 5,7 a 5,9, 5,8 a 6,0, 5,9 a 6,1, 6,0 a 6,2, 6,1 a 6,3, 6,2 a 6,5, 6,4 a 6,7, 6,5 a 6,8, 6,6 a 6,9, 6,7 a 7,0, 6,8 a 7,1, 6,9 a 7,2, 7,0 a 7,3, 7,1 a 7,4, 7,2 a 7,5, 7,3 a 7,6, 7,4 a 7,7, 7,5 a 7,8, 7,6 a 7,9, 7,7 a 8,0, 7,8 a 8,1, 7,9 a 8,2, 8,0 a 8,3, 8,1 a 8,4, 8,2 a 8,5, 8,3 a 8,6, 8,4 a 8,7 ou 8,5 a 8,8. Em ainda outras realizações, o pH é de cerca de 5,7 a 6,0, 5,8 a 6,1, 5,9 a 6,2, 6,0 a 6,3, 6,1 a 6,4 ou 6,2 a 6,5. Em certas realizações, o pH é de cerca de 5,7 a cerca de 6,25.

[00109] Nas realizações, a temperatura de crescimento é mantida a cerca de 4°C a cerca de 42°C. Em certas realizações, a temperatura de crescimento é de cerca de 4°C, cerca de 5°C, cerca de 6°C, cerca de 7°C, cerca de 8°C, cerca de 9°C, cerca de 10°C, cerca de 11°C, cerca de 12°C, cerca de 13°C, cerca de 14°C, cerca de 15°C, cerca de 16°C, cerca de 17°C, cerca de 18°C, cerca de 19°C, cerca de 20°C, cerca de 21°C, cerca de 22°C, cerca de 23°C, cerca de 24°C, cerca de 25°C, cerca de 26°C, cerca de 27°C, cerca de 28°C, cerca de 29°C, cerca de 30°C, cerca de 31°C, cerca de 32°C, cerca de 33°C, cerca de 34°C, cerca de 35°C, cerca de 36°C, cerca de 37°C, cerca de 38°C, cerca de 39°C, cerca de 40°C, cerca de 41°C ou cerca de 42°C. Em outras realizações, a temperatura de crescimento é

mantida a cerca de 25°C a cerca de 27°C, cerca de 25°C a cerca de 28°C, cerca de 25°C a cerca de 29°C, cerca de 25°C a cerca de 30°C, cerca de 25°C a cerca de 31°C, cerca de 25°C a cerca de 32°C, cerca de 25°C a cerca de 33°C, cerca de 26°C a cerca de 28°C, cerca de 26°C a cerca de 29°C, cerca de 26°C a cerca de 30°C, cerca de 26°C a cerca de 31°C, cerca de 26°C a cerca de 32°C, cerca de 27°C a cerca de 29°C, cerca de 27°C a cerca de 30°C, cerca de 27°C a cerca de 31°C, cerca de 27°C a cerca de 32°C, cerca de 26°C a cerca de 33°C, cerca de 28°C a cerca de 30°C, cerca de 28°C a cerca de 31°C, cerca de 28°C a cerca de 32°C, cerca de 29°C a cerca de 31°C, cerca de 29°C a cerca de 32°C, cerca de 29°C a cerca de 33°C, cerca de 30°C a cerca de 32°C, cerca de 30°C a cerca de 33°C, cerca de 31°C a cerca de 33°C, cerca de 31°C a cerca de 32°C, cerca de 30°C a cerca de 33°C ou cerca de 32°C a cerca de 33°C. Em outras realizações, a temperatura é alterada durante a cultura. Em certas realizações, a temperatura é mantida a cerca de 30ºC a cerca de 32ºC antes de um agente para induzir a expressão a partir do construto que codifica o polipeptídeo ou proteína de interesse ser adicionado à cultura e a temperatura é reduzida para cerca de 25ºC a cerca de 27ºC após a adição de um agente para induzir a expressão, por exemplo, IPTG à cultura. Em uma realização, a temperatura é mantida a cerca de 30ºC antes da adição de um agente para induzir a expressão a partir do construto que codifica o polipeptídeo ou proteína de interesse à cultura e a temperatura é reduzida para cerca de 25ºC após a adição de um agente para induzir a expressão à cultura.

Indução

[00110] Como descrito aqui, promotores induzíveis são frequentemente utilizados no construto de expressão para controlar a expressão do polipeptídeo ou proteína de interesse, por exemplo, um promotor lac. No caso de derivados ou membros da família do promotor lac, por exemplo, o promotor tac, o composto efetor é um indutor, tal como um indutor injustificado como o IPTG (isopropil-β-D-1-tiogalactopiranosídeo, também chamado de "isopropiltiogalactosídeo"). Nas realizações, um derivado do promotor lac é utilizado e a expressão do

polipeptídeo ou proteína de interesse é induzida pela adição de IPTG para uma concentração final de cerca de 0,01 mM a cerca de 1,0 mM, quando a densidade celular tiver atingido um nível identificado por uma OD575 de cerca de 25 a cerca de 160. Nas realizações, a OD575 no momento da indução da cultura para o polipeptídeo ou proteína de interesse é de cerca de 25, cerca de 50, cerca de 55, cerca de 60, cerca de 65, cerca de 70, cerca de 80, cerca de 90, cerca de 100, cerca de 110, cerca de 120, cerca de 130, cerca de 140, cerca de 150, cerca de 160, cerca de 170 cerca de 180. Em outras realizações, the OD575 é de cerca de 80 a cerca de 100, cerca de 100 a cerca de 120, cerca de 120 a cerca de 140, cerca de 140 a cerca de 160. Em outras realizações, a OD575 é de cerca de 80 a cerca de 120, cerca de 100 a cerca de 140 ou cerca de 120 a cerca de 160. Em outras realizações, a OD575 é de cerca de 80 a cerca de 140 ou cerca de 100 a 160. A densidade celular é frequentemente medida por outros métodos e expressa em outras unidades, por exemplo, em células por unida de volume. Por exemplo, uma OD575 de cerca de 25 a cerca de 160 de uma cultura de *Pseudomonas fluorescens* é equivalente a aproximadamente 4 x 10^{10} a cerca de 1,6 x 10^{11} unidades formadoras de colônia por mL ou 11 a 70 g/L de peso seco de células. Nas realizações, a expressão do polipeptídeo ou proteína de interesse é induzida pela adição de IPTG para uma concentração final de cerca de 0,01 mM a cerca de 1,0 mM, quando a densidade celular tiver atingido peso úmido de cerca de 0,05 g/g a cerca de 0,4 g/g. Nas realizações, o peso úmido é de cerca de 0,05 g/g, cerca de 0,1 g/g, cerca de 0,15 g/g, cerca de 0,2 g/g, cerca de 0,25 g/g, cerca de 0,30 g/g, cerca de 0,35 g/g, cerca de 0,40 g/g, cerca de 0,05 g/g a cerca de 0,1 g/g, cerca de 0,05 g/g a cerca de 0,15 g/g, cerca de 0,05 g/g a cerca de 0,20 g/g, cerca de 0,05 g/g a cerca de 0,25 g/g, cerca de 0,05 g/g a cerca de 0,30 g/g, cerca de 0,05 g/g a cerca de 0,35 g/g, cerca de 0,1 g/g a cerca de 0,40 g/g, cerca de 0,15 g/g a cerca de 0,40 g/g, cerca de 0,20 g/g a cerca de 0,40 g/g, cerca de 0,25 g/g a cerca de 0,40 g/g, cerca de 0,30 g/g a cerca de 0,40 g/g ou cerca de 0,35 g/g a cerca de 0,40 g/g. Nas realizações, a densidade celular no momento da indução da cultura é equivalente à densidade celular tal como especificado aqui pela

absorbância a OD575, independentemente do método utilizado para a determinação da densidade celular ou das unidades de medida. Um técnico no assunto saberá como fazer a conversão apropriada para qualquer cultura celular.

[00111] Nas realizações, a concentração final de IPTG da cultura é de cerca de 0,01 mM, cerca de 0,02 mM, cerca de 0,03 mM, cerca de 0,04 mM, cerca de 0,05 mM, cerca de 0,06 mM, cerca de 0,07 mM, cerca de 0,08 mM, cerca de 0,09 mM, cerca de 0,1 mM, cerca de 0,2 mM, cerca de 0,3 mM, cerca de 0,4 mM, cerca de 0,5 mM, cerca de 0,6 mM, cerca de 0,7 mM, cerca de 0,8 mM, cerca de 0,9 mM ou cerca de 1 mM. Em outras realizações, a concentração final de IPTG da cultura é de cerca de 0,08 mM a cerca de 0,1 mM, cerca de 0,1 mM a cerca de 0,2 mM, cerca de 0,2 mM a cerca de 0,3 mM, cerca de 0,3 mM a cerca de 0,4 mM, cerca de 0,2 mM a cerca de 0,4 mM, cerca de 0,08 a cerca de 0,2mM ou cerca de 0,1 to 1 mM.

[00112] Nas realizações em que um promotor tipo não lac é utilizado, como descrito aqui e na literatura, outros indutores ou efetores são frequentemente utilizados. Em uma realização, o promotor é um promotor constitutivo.

[00113] Após a adição de um agente de indução, as culturas são frequentemente crescidas por um período de tempo, por exemplo cerca de 24 horas, durante este tempo o polipeptídeo ou proteína de interesse é expressa. Após a adição de um agente de indução, uma cultura é frequentemente crescida por cerca de 1 hr, cerca de 2 hr, cerca de 3 hr, cerca de 4 hr, cerca de 5 hr, cerca de 6 hr, cerca de 7 hr, cerca de 8 hr, cerca de 9 hr, cerca de 10 hr, cerca de 11 hr, cerca de 12 hr, cerca de 13 hr, cerca de 14 hr, cerca de 15 hr, cerca de 16 hr, cerca de 17 hr, cerca de 18 hr, cerca de 19 hr, cerca de 20 hr, cerca de 21 hr, cerca de 22 hr, cerca de 23 hr, cerca de 24 hr, cerca de 36 hr ou cerca de 48 hr. Após um agente de indução ser adicionado a uma cultura, a cultura é crescida por cerca de 1 a 48 horas, cerca de 1 a 24 horas, cerca de 10 a 24 horas, cerca de 15 a 24 horas ou cerca de 20 a 24 horas. As culturas celulares são frequentemente concentradas por centrifugação e o pélete de cultura é novamente suspendido em um tampão ou solução apropriada para o procedimento de lise subsequente.

[00114] Nas realizações, as células são rompidas utilizando-se equipamento para rompimento celular mecânico de alta pressão (os quais estão disponíveis comercialmente, por exemplo, Microfluidics Microfluidizer, Constant Cell Disruptor, homogeneizador Niro-Soavi ou homogeneizador APV-Gaulin). As células expressando os polipeptídeos ou proteínas de interesse são frequentemente rompidas, por exemplo, pela utilização de sonicação. Qualquer método apropriado conhecido na técnica para a lise de células é frequentemente utilizado para liberar a fração solúvel. Por exemplo, nas realizações, são frequentemente utilizados reagentes de lise celular química e/ou enzimática, tais como enzima lítica de parede celular e EDTA, são frequentemente utilizados. O uso de culturas congeladas ou previamente armazenadas está também contemplado nos métodos aqui. As culturas são algumas vezes normalizadas para OD antes da lise. Por exemplo, as células são frequentemente normalizadas para uma OD600 de cerca de 10, cerca de 11, cerca de 12, cerca de 13, cerca de 14, cerca de 15, cerca de 16, cerca de 17, cerca de 18, cerca de 19 ou cerca de 20.

[00115] A centrifugação é realizada utilizando-se qualquer equipamento e método apropriados. A centrifugação de cultura celular ou lisado para os propósitos de separar uma fração solúvel de uma fração insolúvel é bem conhecida na técnica. Por exemplo, as células lisadas são algumas vezes centrifugadas a 20,800 × g por 20 minutos (a 4ºC) e os sobrenadantes são removidos utilizando-se manuseio de líquido manual ou automático. A fração em pélete (insolúvel) é novamente suspendida em uma solução tamponada, por exemplo, solução salina tamponada com fosfato (PBS), pH 7,4. A ressuspensão é frequentemente conduzida utilizando-se, por exemplo, equipamento tal como impulsores conectados a um misturador aéreo, por agitação com barras magnéticas, agitadores de balanço etc.

[00116] Uma “fração solúvel”, isto é, o sobrenadante solúvel obtido após a centrifugação de um lisado e uma “fração insolúvel”, isto é, o pélete obtido após a centrifugação de um lisado, resultam da lise e centrifugação das culturas.

Formato de Fermentação

[00117] Em uma realização, é utilizada fermentação nos métodos para a produção de polipeptídeos e proteínas de interesse aqui. O sistema de expressão de acordo com o presente relatório é cultivado em qualquer formato de fermentação. Por exemplo, modos de fermentação em batelada, descontínua alimentada, semi-contínua e contínua podem ser empregados aqui.

[00118] Nas realizações, o meio de fermentação pode ser selecionado entre meios ricos, meio mínimos e meios com sais minerais. Em outras realizações, é selecionado ou o meio mínimo ou o meio com sais minerais. Em certas realizações, um meio com sais minerais é selecionado.

[00119] O meio com sais minerais consiste em sais minerais e uma fonte de carbono tal como, por exemplo, glicose, sacarose ou glicerol. Exemplos de meio com sais minerais incluem, por exemplo, meio M9, médio para *Pseudomonas* (ATCC 179) e meio de Davis e Mingioli (ver, B D Davis & E S Mingioli (1950) J. Bact. 60:17-28). Os sais minerais utilizados para se obter o meio com sais minerais incluem os selecionados entre, por exemplo, fosfatos de potássio, sulfato ou cloreto de amônio, sulfato ou cloreto de magnésio e minerais traço tais como cloreto de cálcio, borato e sulfatos de ferro, cobre, manganês e zinco. Tipicamente, nenhuma fonte orgânica de nitrogênio, tais como peptona, triptona, aminoácidos ou um extrato de levedura, é incluída em um meio com sais minerais. Em vez disso, é utilizada uma fonte inorgânica de nitrogênio e esta pode ser selecionada entre, por exemplo, sais de amônio, amônia aquosa e amônia gasosa. Um meio com sais minerais tipicamente irá conter glicose ou glicerol como fonte de carbono. Em comparação com meios com sais minerais, os meios mínimos frequentemente contêm sais minerais e uma fonte de carbono, mas são frequentemente suplementados, por exemplo, com baixos níveis de aminoácidos, vitaminas, peptonas ou outros ingredientes, embora estes sejam adicionados em níveis mínimos. O meio é frequentemente preparado utilizando-se os métodos descritos na técnica, por exemplo, no Pedido de Patente Publicado US nº 2006/0040352, citado e incorporado como referência acima. Detalhes dos

procedimentos de cultivo e meios com sais minerais úteis nos métodos aqui são descritos por Riesenberg, D. *et al.*, 1991, "High cell density cultivation of Escherichia coli at controlled specific growth rate", J. Biotechnol. 20 (1):17-27.

[00120] A fermentação pode ser realizada em qualquer escala. Os sistemas de expressão de acordo com o presente relatório são úteis para a expressão de proteína recombinante em qualquer escala. Desta forma, por exemplo, podem ser utilizados volumes de fermentação de escala em microlitro, escala em mililitro, escala em centilitro e escala em decilitro e escala de1 Litro e volumes maiores de fermentação são frequentemente utilizados.

[00121] Nas realizações, o volume de fermentação é de ou é acima de cerca de 1 Litro. Nas realizações, o volume de fermentação é de cerca de 0,5 litro a cerca de 100 litros. Nas realizações, o volume de fermentação é de cerca de 1 litro, cerca de 2 litros, cerca de 3 litros, cerca de 4 litros, cerca de 5 litros, cerca de 6 litros, cerca de 7 litros, cerca de 8 litros, cerca de 9 litros ou cerca de 10 litros. Nas realizações, o volume de fermentação é de cerca de 0,5 litro a cerca de 2 litros, cerca de 0,5 litro a cerca de 5 litros, cerca de 0,5 litro a cerca de 10 litros, cerca de 0,5 litro a cerca de 25 litros, cerca de 0,5 litro a cerca de 50 litros, cerca de 0,5 litro a cerca de 75 litros, cerca de 10 litros a cerca de 25 litros, cerca de 25 litros a cerca de 50 litros ou cerca de 50 litros a cerca de 100 litros Em outras realizações, o volume de fermentação é de ou é 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, 10000 Litros ou 50000 Litros.

Análise da Proteína

[00122] Nas realizações, os polipeptídeos e proteínas de interesse produzidos pelos métodos providos aqui são analisados. Os polipeptídeos e proteínas de interesse podem ser analisados por qualquer método apropriado conhecido dos técnicos no assunto, por exemplo, por interferometria de biocamada, SDS-PAGE, Western blot, Far Western blot, ELISA, espectrometria de absorbância ou de massa (por exemplo, espectrometria de massa em tandem).

[00123] Em algumas realizações, a concentração e/ou quantidades dos polipeptídeos ou proteínas de interesse geradas são determinadas, por exemplo, pelo ensaio de Bradford, absorbância, marcação de Coomassie, espectrometria de massa etc.

[00124] O rendimento da proteína nas frações insolúvel e solúvel tal como descritas aqui é frequentemente determinado por métodos conhecidos pelos técnicos no assunto, por exemplo, por eletroforese em gel capilar (CGE), SDS-PAGE e análise Western blot. As frações solúveis são frequentemente avaliada, por exemplo, utilizando-se interferometria de biocamada.

[00125] Medidas úteis de rendimento de proteína incluem, por exemplo, a quantidade de proteína recombinante por volume de cultura (por exemplo, gramas ou miligramas de proteína/litro de cultura), percentagem ou fração da proteína recombinante medida no pélete insolúvel obtido após a lise (por exemplo, quantidade de proteína recombinante no sobrenadante do extrato/quantidade de proteína na fração insolúvel), percentagem ou fração de proteína ativa (por exemplo, quantidade de proteína ativa/quantidade de proteína utilizada no ensaio), percentagem ou fração proteína de célula total (TCP), quantidade de proteína/célula e percentagem de biomassa seca.

[00126] Nas realizações, os métodos aqui são utilizados para se obter um rendimento de polipeptídeo ou proteína de interesse solúvel de cerca de 20% a cerca de 90% de proteína de célula total. Em certas realizações, o rendimento de polipeptídeo ou proteína de interesse solúvel é de cerca de 20% de proteína de célula total, cerca de 25% de proteína de célula total, cerca de 30% de proteína de célula total, cerca de 31% de proteína de célula total, cerca de 32% de proteína de célula total, cerca de 33% de proteína de célula total, cerca de 34% de proteína de célula total, cerca de 35% de proteína de célula total, cerca de 36% de proteína de célula total, cerca de 37% de proteína de célula total, cerca de 38% de proteína de célula total, cerca de 39% de proteína de célula total, cerca de 40% de proteína de célula total, cerca de 41% de proteína de célula total, cerca de 42% de proteína de célula total, cerca de 43% de proteína de célula total, cerca de 44% de proteína de

célula total, cerca de 45% de proteína de célula total, cerca de 46% de proteína de célula total, cerca de 47% de proteína de célula total, cerca de 48% de proteína de célula total, cerca de 49% de proteína de célula total, cerca de 50% de proteína de célula total, cerca de 51% de proteína de célula total, cerca de 52% de proteína de célula total, cerca de 53% de proteína de célula total, cerca de 54% de proteína de célula total, cerca de 55% de proteína de célula total, cerca de 56% de proteína de célula total, cerca de 57% de proteína de célula total, cerca de 58% de proteína de célula total, cerca de 59% de proteína de célula total, cerca de 60% de proteína de célula total, cerca de 65% de proteína de célula total, cerca de 70% de proteína de célula total, cerca de 75% de proteína de célula total, cerca de 80% de proteína de célula total, cerca de 85% de proteína de célula total ou cerca de 90% de proteína de célula total. Em algumas realizações, o rendimento de polipeptídeo ou proteína de interesse solúvel é de cerca de 20% a cerca de 25% de proteína de célula total, cerca de 20% a cerca de 30% de proteína de célula total, cerca de 20% a cerca de 35% de proteína de célula total, cerca de 20% a cerca de 40% de proteína de célula total, cerca de 20% a cerca de 45% de proteína de célula total, cerca de 20% a cerca de 50% de proteína de célula total, cerca de 20% a cerca de 55% de proteína de célula total, cerca de 20% a cerca de 60% de proteína de célula total, cerca de 20% a cerca de 65% de proteína de célula total, cerca de 20% a cerca de 70% de proteína de célula total, cerca de 20% a cerca de 75% de proteína de célula total, cerca de 20% a cerca de 80% de proteína de célula total, cerca de 20% a cerca de 85% de proteína de célula total, cerca de 20% a cerca de 90% de proteína de célula total, cerca de 25% a cerca de 90% de proteína de célula total, cerca de 30% a cerca de 90% de proteína de célula total, cerca de 35% a cerca de 90% de proteína de célula total, cerca de 40% a cerca de 90% de proteína de célula total, cerca de 45% a cerca de 90% de proteína de célula total, cerca de 50% a cerca de 90% de proteína de célula total, cerca de 55% a cerca de 90% de proteína de célula total, cerca de 60% a cerca de 90% de proteína de célula total, cerca de 65% a cerca de 90% de proteína de célula total, cerca de 70% a cerca de 90% de proteína de célula total, cerca de 75% a cerca de 90% de proteína de

célula total, cerca de 80% a cerca de 90% de proteína de célula total, cerca de 85% a cerca de 90% de proteína de célula total, cerca de 31% a cerca de 60% de proteína de célula total, cerca de 35% a cerca de 60% de proteína de célula total, cerca de 40% a cerca de 60% de proteína de célula total, cerca de 45% a cerca de 60% de proteína de célula total, cerca de 50% a cerca de 60% de proteína de célula total, cerca de 55% a cerca de 60% de proteína de célula total, cerca de 31% a cerca de 55% de proteína de célula total, cerca de 31% a cerca de 50% de proteína de célula total, cerca de 31% a cerca de 45% de proteína de célula total, cerca de 31% a cerca de 40% de proteína de célula total, cerca de 31% a cerca de 35% de proteína de célula total, cerca de 35% a cerca de 55% de proteína de célula total ou cerca de 40% a cerca de 50% de proteína de célula total.

[00127] Nas realizações, os métodos aqui são utilizados para se obter um rendimento de polipeptídeo ou proteína de interesse solúvel de cerca de 1 grama por litro a cerca de 50 gramas por litro. Em certas realizações, o rendimento de polipeptídeo ou proteína de interesse solúvel é de cerca de 1 grama por litro, cerca de 2 gramas por litro, cerca de 3 gramas por litro, cerca de 4 gramas por litro, cerca de 5 gramas por litro, cerca de 6 gramas por litro, cerca de 7 gramas por litro, cerca de 8 gramas por litro, cerca de 9 gramas por litro, cerca de 10 grama por litro, cerca de 11 gramas por litro, cerca de 12 gramas por litro, cerca de 13 gramas por litro, cerca de 14 gramas por litro, cerca de 15 gramas por litro, cerca de 16 gramas por litro, cerca de 17 gramas por litro, cerca de 18 gramas por litro, cerca de 19 gramas por litro, cerca de 20 gramas por litro, cerca de 21 gramas por litro, cerca de 22 gramas por litro, cerca de 23 gramas por litro cerca de 24 gramas por litro, cerca de 25 gramas por litro, cerca de 26 gramas por litro, cerca de 27 gramas por litro, cerca de 28 gramas por litro, cerca de 30 gramas por litro, cerca de 35 gramas por litro, cerca de 40 gramas por litro, cerca de 45 gramas por litro, cerca de 50 gramas por litro, cerca de 1 grama por litro a cerca de 5 gramas por litro, cerca de 1 grama a cerca de 10 gramas por litro, cerca de 10 grama por litro a cerca de 12 gramas por litro, cerca de 10 gramas por litro a cerca de 13 gramas por litro, cerca de 10 gramas por litro a cerca de 14 gramas por litro, cerca

de 10 gramas por litro a cerca de 15 gramas por litro, cerca de 10 gramas por litro a cerca de 16 gramas por litro, cerca de 10 gramas por litro a cerca de 17 gramas por litro, cerca de 10 gramas por litro a cerca de 18 gramas por litro, cerca de 10 gramas por litro a cerca de 19 gramas por litro, cerca de 10 gramas por litro a cerca de 20 gramas por litro, cerca de 10 gramas por litro a cerca de 21 gramas por litro, cerca de 10 gramas por litro a cerca de 22 gramas por litro, cerca de 10 gramas por litro a cerca de 23 gramas por litro, cerca de 10 gramas por litro a cerca de 24 gramas por litro, cerca de 10 gramas por litro a cerca de 25 gramas por litro, cerca de 10 gramas por litro a cerca de 30 gramas por litro, cerca de 10 gramas por litro a cerca de 40 gramas por litro, cerca de 10 gramas por litro a cerca de 50 gramas por litro, cerca de 10 grama por litro a cerca de 12 gramas por litro, cerca de 12 gramas por litro a cerca de 14 gramas por litro, cerca de 14 gramas por litro a cerca de 16 gramas por litro, cerca de 16 gramas por litro a cerca de 18 gramas por litro, cerca de 18 gramas por litro a cerca de 20 gramas por litro, cerca de 20 gramas por litro a cerca de 22 gramas por litro, cerca de 22 gramas por litro a cerca de 24 gramas por litro, cerca de 23 gramas por litro a cerca de 25 gramas por litro, cerca de 10 gramas por litro a cerca de 25 gramas por litro, cerca de 11 gramas por litro a cerca de 25 gramas por litro, cerca de 12 gramas por litro a cerca de 25 gramas por litro, cerca de 13 gramas por litro a cerca de 25 gramas por litro, cerca de 14 gramas por litro a cerca de 25 gramas por litro, cerca de 15 gramas por litro a cerca de 25 gramas por litro, cerca de 16 gramas por litro a cerca de 25 gramas por litro, cerca de 17 gramas por litro a cerca de 25 gramas por litro, cerca de 18 gramas por litro a cerca de 25 gramas por litro, cerca de 19 gramas por litro a cerca de 25 gramas por litro, cerca de 20 gramas por litro a cerca de 25 gramas por litro, cerca de 21 gramas por litro a cerca de 25 gramas por litro, cerca de 22 gramas por litro a cerca de 25 gramas por litro, cerca de 23 gramas por litro a cerca de 25 gramas por litro ou cerca de 24 gramas por litro a cerca de 25 gramas por litro. Nas realizações, o rendimento de proteína recombinante solúvel é de cerca de 10 gramas por litro a cerca de 13 gramas por litro, cerca de 12 gramas por litro a cerca de 14 gramas por litro, cerca

de 13 gramas por litro a cerca de 15 gramas por litro, cerca de 14 gramas por litro a cerca de 16 gramas por litro, cerca de 15 gramas por litro a cerca de 17 gramas por litro, cerca de 16 gramas por litro a cerca de 18 gramas por litro, cerca de 17 gramas por litro a cerca de 19 gramas por litro, cerca de 18 gramas por litro a cerca de 20 gramas por litro, cerca de 20 gramas por litro a cerca de 22 gramas por litro, cerca de 22 gramas por litro a cerca de 24 gramas por litro ou cerca de 23 gramas por litro a cerca de 25 gramas por litro. Nas realizações, rendimento de proteína recombinante solúvel é de cerca de 10 gramas por litro a cerca de 25 gramas por litro, cerca de 12 gramas por litro a cerca de 24 gramas por litro, cerca de 14 gramas por litro a cerca de 22 gramas por litro, cerca de 16 gramas por litro a cerca de 20 gramas por litro ou cerca de 18 gramas por litro a cerca de 20 gramas por litro. Nas realizações, rendimento de proteína extraída é de cerca de 5 gramas por litro a cerca de 15 gramas por litro, cerca de 5 gramas por litro a cerca de 25 gramas por litro, cerca de 10 gramas por litro a cerca de 15 gramas por litro, cerca de 10 gramas por litro a cerca de 25 gramas por litro, cerca de 15 gramas por litro a cerca de 20 gramas por litro, cerca de 15 gramas por litro a cerca de 25 gramas por litro ou cerca de 18 gramas por litro a cerca de 25 gramas por litro.

[00128] Nas realizações, a quantidade de polipeptídeo ou proteína de interesse solúvel detectada na fração solúvel é de cerca de 10% a cerca de 100% da quantidade de polipeptídeo ou proteína de interesse solúvel total produzida. Nas realizações, esta quantidade é de cerca de 10%, cerca de 15%, cerca de 20%, cerca de 25%, cerca de 30%, cerca de 35%, cerca de 40%, cerca de 45%, cerca de 50%, cerca de 55%, cerca de 60%, cerca de 65%, cerca de 70%, cerca de 75%, cerca de 80%, cerca de 85%, cerca de 90%, cerca de 95% ou cerca de 99% ou cerca de 100% da quantidade de polipeptídeo ou proteína de interesse solúvel produzida. Nas realizações, esta quantidade é de cerca de 10% a cerca de 20%, 20% a cerca de 50%, cerca de 25% a cerca de 50%, cerca de 25% a cerca de 50%, cerca de 25% a cerca de 95%, cerca de 30% a cerca de 50%, cerca de 30% a cerca de 40%, cerca de 30% a cerca de 60%, cerca de 30% a cerca de 70%, cerca de 35% a cerca de 50%, cerca de 35% a cerca de 70%, cerca de 35% a cerca de 75%,

cerca de 35% a cerca de 95%, cerca de 40% a cerca de 50%, cerca de 40% a cerca de 95%, cerca de 50% a cerca de 75%, cerca de 50% a cerca de 95%, cerca de 70% a cerca de 95% ou cerca de 80 a cerca de 100% da quantidade de polipeptídeo ou proteína de interesse solúvel produzida.

[00129] Em algumas realizações, a quantidade de polipeptídeo ou proteína de interesse solúvel é expressa como uma percentagem da proteína total solúvel produzida em uma cultura. Dados expressos em termos de polipeptídeo ou proteína de interesse solúvel pelo/volume de cultura celular a uma dada densidade celular podem ser convertidos para dados expressos como percentagem de proteína recombinante da proteína de célula total. Está dentro da capacidade de um técnico no assunto converter o rendimento volumétrico de proteína para % de proteína de célula total, por exemplo, conhecendo a quantidade de proteína de célula total por volume de cultura celular a uma dada densidade celular. Este número pode ser determinado se é conhecido) o peso/volume de célula da cultura a uma dada densidade celular e 2) a percentagem de peso de célula compreendido pela proteína total. Por exemplo, a uma OD550 de 1,0, o peso seco de *E. coli* é reportado como sendo de 0,5 grama/litro ("Production of Heterologous Proteins from Recombinant DNA Escherichia coli in Bench Fermentors" Lin, N.S. e Swartz, J.R., 1992, METHODS: A Companion to Methods in Enzymology 4: 159-168). Uma célula bacteriana compreende polissacarídeos, lipídeos e ácidos nucleicos, bem como proteínas. Uma célula de *E. coli* é reportada como sendo cerca de 52,4 a 55% de proteína por referências incluindo, mas não se limitando a, Da Silva, N.A., *et al.*, 1986, "Theoretical Growth Yield Estimates for Recombinant Cells," Biotechnology and Bioengineering, Vol. XXVIII: 741-746, que estima que a proteína representa até 52,4% em peso da célula de *E. coli*s e "Escherichia coli and Salmonella typhimurium Cellular and Molecular Biology", 1987, Ed. em Chief Frederick C. Neidhardt, Vol. 1, pp. 3-6, que reporta um teor de proteína em *E. coli* como 55% de peso seco de célula. Pela utilização das medições acima (isto é, um peso seco de célula de 0,5 grama/litro e proteína como 55% em peso celular), é calculada a quantidade de proteína de célula total

por volume de cultura celular a uma A550 de 1,0 para *E. coli* como 275 μg de proteína de célula total/ml/A550. Um cálculo da proteína de célula total por volume de cultura celular com base no peso úmido de célula pode utilizar, por exemplo, a determinação por Glazyrina, *et al.* (Microbial Cell Factories 2010, 9:42, incorporado aqui como referência) que uma A600 de 1,0 para *E. coli* resultou em um peso úmido de célula de 1,7 gramas/litro e um peso seco de célula de 0,39 gramas/litro. Por exemplo, utilizando-se esta comparação de peso úmido de célula com peso seco de célula e proteína como 55% do peso seco de célula como descrito acima, a quantidade de proteína de célula total por volume de cultura celular a uma A600 de 1,0 para *E. coli* pode ser calculada como 215 μg de proteína de célula total/ml/A600. Para *Pseudomonas fluorescens*, a quantidade de proteína de célula total por volume de cultura celular a uma dada densidade celular é similar à encontrada para *E. coli*. A *P. fluorescens*, como a *E. coli*, é uma bactéria gram-negativa, em forma de bastão. O peso seco de célula da *P. fluorescens* ATCC 11150 tal como reportado por Edwards, *et al.*, 1972, "Continuous Culture of *Pseudomonas fluorescens* with Sodium Maleate as a Carbon Source" Biotechnology and Bioengineering, Vol. XIV, páginas 123-147, é de 0,5 grama/litro/A500. Este é o mesmo peso reportado por Lin, *et al.*, para *E. coli* a uma A550 de 1,0. Medidas de dispersão de luz feitas a 500 nm e a 550 nm são esperadas serem muito similares. A percentagem de peso celular compreendido por proteína de célula total para *P. fluorescens* HK44 é descrita como 55%, por exemplo, por Yarwood, *et al.*, julho 2002, "Noninvasive Quantitative Measurement of Bacterial Growth in Porous Media under Unsaturated-Flow Conditions," Applied and Environmental Microbiology 68(7):3597-3605. Esta percentagem é similar ou a mesma das dadas para *E. coli* pelas referências descritas acima.

[00130] Nas realizações, a quantidade de polipeptídeo ou proteína de interesse solúvel produzida é de cerca de 0,1% a cerca de 95% da proteína solúvel total produzida em uma cultura. Nas realizações, esta quantidade é maior que cerca de 0,1%, 0,5%, 1%, 5%, 10%, 15%, 20%, 25%, 30%, 40%, 45%, 50%,

55%, 60%, 65%, 70%, 75%, 80%, 85%, 90% ou 95% da proteína solúvel total produzida em uma cultura. Nas realizações, esta quantidade é de cerca de 0,1%, 0,5%, 1%, 5%, 10%, 15%, 20%, 25%, 30%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90% ou 95% da proteína solúvel total produzida em uma cultura. Nas realizações, esta quantidade é de cerca de 5% a cerca de 95%, cerca de 10% a cerca de 85%, cerca de 20% a cerca de 75%, cerca de 30% a cerca de 65%, cerca de 40% a cerca de 55%, cerca de 1% a cerca de 95%, cerca de 5% a cerca de 30%, cerca de 1% a cerca de 10%, cerca de 10% a cerca de 20%, cerca de 20% a cerca de 30%, cerca de 30% a cerca de 40%, cerca de 40% a cerca de 50%, cerca de 50 a cerca de 60%, cerca de 60% a cerca de 70% ou cerca de 80% a cerca de 90% da proteína solúvel total produzida em uma cultura.

[00131] Nas realizações, a quantidade de polipeptídeo ou proteína de interesse solúvel produzida é de cerca de 0,1% a cerca de 50% do peso seco de célula (DCW). Nas realizações, esta quantidade é maior que cerca de 0,1%, 0,5%, 1%, 5%, 10%, 15%, 20%, 25%, 30%, 40%, 45% ou 50% of DCW. Nas realizações, esta quantidade é de cerca de 0,1%, 0,5%, 1%, 5%, 10%, 15%, 20%, 25%, 30%, 40%, 45% ou 50% de PSC. Nas realizações, esta quantidade é de cerca de 5% a cerca de 50%, cerca de 10% a cerca de 40%, cerca de 20% a cerca de 30%, cerca de 1% a cerca de 20%, cerca de 5% a cerca de 25%, cerca de 1% a cerca de 10%, cerca de 10% a cerca de 20%, cerca de 20% a cerca de 30%, cerca de 30% a cerca de 40% ou cerca de 40% a cerca de 50% da proteína solúvel total produzida em uma cultura.

[00132] Nas realizações, a quantidade de um polipeptídeo ou proteína de interesse solúvel heteróloga em relação ao líder de secreção produzido utilizando-se os métodos da invenção é maior que a quantidade da proteína nativa em relação ao líder de secreção, quando produzido sob condições substancialmente similares, por exemplo, condições de crescimento, incluindo a célula hospedeira. Nas realizações, um polipeptídeo ou proteína de interesse solúvel heteróloga operacionalmente ligada ao líder de secreção AnsB, 8484 ou 5193 de acordo com os presentes métodos, é produzido em uma quantidade maior que a quantidade da

proteína nativa para o respectivo líder. Nas realizações, o polipeptídeo ou proteína de interesse heteróloga solúvel operacionalmente ligada ao líder de secreção AnsB, 8484 ou 5193 é produzido em um rendimento que é de cerca de 1,2 vezes a cerca de 5 vezes maior, cerca de 1,2 vezes a cerca de 2,5 vezes, cerca de 1,2 vezes a cerca de 2 vezes, cerca de 1,2 vezes a cerca de 1,5 vezes, cerca de 1,4 vezes a cerca de 3 vezes maior, cerca de 1,4 vezes a cerca de 2,5 vezes, cerca de 1,4 vezes a cerca de 2 vezes, cerca de 1,5 vezes a cerca de 3 vezes, cerca de 1,5 vezes a cerca de 2,5 vezes, cerca de 1,5 vezes a cerca de 2 vezes, cerca de 2 vezes a cerca de 3 vezes, cerca de 1,2 vezes, cerca de 1,3 vezes, cerca de 1,4 vezes, cerca de 1,5 vezes, cerca de 1,6 vezes, cerca de 1,7 vezes, cerca de 1,8 vezes, cerca de 1,9 vezes, cerca de 2 vezes, cerca de 2,5 vezes, cerca de 3 vezes, cerca de 4 vezes ou cerca de 5 vezes maior, que a quantidade da proteína nativa em relação ao líder produzido sob condições substancialmente similares.

[00133] Nas realizações, uma quantidade de um polipeptídeo ou proteína de interesse solúvel operacionalmente ligada ao líder de secreção produzido utilizando-se os métodos da invenção é maior que quantidade do polipeptídeo ou proteína de interesse produzida sem um líder de secreção ou com um líder de secreção diferente, quando produzido sob condições substancialmente similares, por exemplo, condições de crescimento. As condições de crescimento podem incluir, por exemplo, o meio de cultura e a célula hospedeira utilizados. Nas realizações, um polipeptídeo ou proteína de interesse heteróloga solúvel operacionalmente ligado ao líder de secreção AnsB, 8484 ou 5193 de acordo com os presentes métodos, é produzido em uma quantidade maior que a quantidade da proteína de interesse produzida sem um líder de secreção ou com um líder de secreção diferente. Nas realizações, o polipeptídeo ou proteína de interesse heteróloga solúvel operacionalmente ligado ao líder de secreção AnsB, 8484 ou 5193 é produzido em um rendimento que é de cerca de 1,2 vezes a cerca de 20 vezes maior, cerca de 1,2 vezes a cerca de 2,5 vezes, cerca de 1,2 vezes a cerca de 2 vezes, cerca de 1,2 vezes a cerca de 1,5 vezes, cerca de 1,4 vezes a cerca de 3 vezes maior, cerca de 1,4 vezes a cerca de 2,5 vezes, cerca de 1,4 vezes a cerca de

2 vezes, cerca de 1,5 vezes a cerca de 3 vezes, cerca de 1,5 vezes a cerca de 2,5 vezes, cerca de 1,5 vezes a cerca de 2 vezes, cerca de 2 vezes a cerca de 5 vezes, cerca de 2 vezes a cerca de 10 vezes, cerca de 5 vezes a cerca de 15 vezes, cerca de 10 vezes a cerca de 20 vezes, cerca de 1,2 vezes, cerca de 1,3 vezes, cerca de 1,4 vezes, cerca de 1,5 vezes, cerca de 1,6 vezes, cerca de 1,7 vezes, cerca de 1,8 vezes, cerca de 1,9 vezes, cerca de 2 vezes, cerca de 2,5 vezes, cerca de 3 vezes, cerca de 4 vezes, cerca de 5 vezes maior, cerca de 10 vezes maior, cerca de 15 vezes maior ou cerca de 20 vezes maior que a quantidade da proteína produzida sem um líder de secreção ou com um líder de secreção diferente sob condições substancialmente similares.

Solubilidade e Atividade

[00134] A “solubilidade” e a “atividade” de uma proteína, embora sejam qualidades relacionadas, são geralmente determinadas por meios diferentes. A solubilidade de uma proteína, particularmente uma proteína hidrofóbica, indica que os resíduos de aminoácidos hidrofóbicos estão inapropriadamente localizados na parte externa da proteína dobrada. A atividade da proteína, que é frequentemente avaliada utilizando-se diferentes métodos, por exemplo, como descrito abaixo, é um outro indicador da conformação apropriada da proteína. “Solúvel, ativa ou ambas” tal como utilizado aqui, refere-se a uma proteína que é determinada como solúvel, ativa ou tanto solúvel quanto ativa, por métodos conhecidos pelos técnicos no assunto.

Ensaio de Atividade

[00135] Os ensaios para a avaliação da atividade de peptídeos, polipeptídeos ou proteínas de interesse são conhecidos na técnica e incluem, mas não se limitam a, ensaios fluorimétricos, colorimétricos, quimioluminescentes, espectrofotométricos e outros ensaios enzimáticos disponíveis a um técnico no assunto. Estes ensaios são utilizados para comparar a atividade de uma preparação de um peptídeo, polipeptídeo ou proteína de interesse com uma preparação comercial ou outra de um peptídeo, polipeptídeo ou proteína.

[00136] Nas realizações, a atividade é representada pela percentagem de proteína ativa no sobrenadante do extrato em comparação com a quantidade total ensaiada. Isto se baseia na quantidade de proteína determinada como ativa pelo ensaio em relação à quantidade total de proteína utilizada no ensaio. Em outras realizações, a atividade é representada pelo nível de atividade % da proteína em comparação com um padrão, por exemplo, proteína nativa. Isto se baseia na quantidade de proteína ativa na amostra de sobrenadante do extrato em relação à quantidade de proteína ativa em uma amostra padrão (onde a mesma quantidade de proteína de cada amostra é utilizada no ensaio).

[00137] Nas realizações, cerca de 40% a cerca de 100% do peptídeo, polipeptídeo ou proteína de interesse, é determinada como sendo ativa. Nas realizações, cerca de 40%, cerca de 50%, cerca de 60%, cerca de 70%, cerca de 80%, cerca de 90% ou cerca de 100% do peptídeo, polipeptídeo ou proteína de interesse é determinado como sendo ativo. Nas realizações, cerca de 40% a cerca de 50%, cerca de 50% a cerca de 60%, cerca de 60% a cerca de 70%, cerca de 70% a cerca de 80%, cerca de 80% a cerca de 90%, cerca de 90% a cerca de 100%, cerca de 50% a cerca de 100%, cerca de 60% a cerca de 100%, cerca de 70% a cerca de 100%, cerca de 80% a cerca de 100%, cerca de 40% a cerca de 90%, cerca de 40% a cerca de 95%, cerca de 50% a cerca de 90%, cerca de 50% a cerca de 95%, cerca de 50% a cerca de 100%, cerca de 60% a cerca de 90%, cerca de 60% a cerca de 95%, cerca de 60% a cerca de 100%, cerca de 70% a cerca de 90%, cerca de 70% a cerca de 95%, cerca de 70% a cerca de 100% ou cerca de 70% a cerca de 100% do peptídeo, polipeptídeo ou proteína de interesse é determinada como sendo ativa.

[00138] Em outras realizações, cerca de 75% a cerca de 100% do peptídeo, polipeptídeo ou proteína de interesse é determinado como sendo ativo. Nas realizações, cerca de 75% a cerca de 80%, cerca de 75% a cerca de 85%, cerca de 75% a cerca de 90%, cerca de 75% a cerca de 95%, cerca de 80% a cerca de 85%, cerca de 80% a cerca de 90%, cerca de 80% a cerca de 95%, cerca de 80% a cerca de 100%, cerca de 85% a cerca de 90%, cerca de 85% a cerca de 95%, cerca de

85% a cerca de 100%, cerca de 90% a cerca de 95%, cerca de 90% a cerca de 100% ou cerca de 95% a cerca de 100% do peptídeo, polipeptídeo ou proteína de interesse é determinado como sendo ativo.

[00139] Proteínas de Interesse

[00140] Os métodos e composições aqui são úteis para a produção de níveis altos de proteína ou polipeptídeo recombinante de interesse apropriadamente processado em um sistema de expressão celular. Nas realizações, a proteína ou polipeptídeo de interesse (também referido aqui como "proteína alvo" ou "polipeptídeo alvo") é de qualquer espécie e de qualquer tamanho. Em certas realizações, a proteína ou polipeptídeo de interesse é uma proteína ou polipeptídeo terapeuticamente útil. Em algumas realizações, a proteína é uma proteína de mamífero, por exemplo, uma proteína humana, por exemplo, um fator de crescimento, uma citocina, uma quimiocina, um fator de crescimento ou uma proteína do sangue. Nas realizações, a proteína ou polipeptídeo de interesse é processado de maneira similar ao da proteína ou polipeptídeo de referência. Em certas realizações, a proteína ou polipeptídeo não inclui um sinal de secreção na sequência de codificação. Nas realizações, a proteína ou polipeptídeo de interesse é menor que 100 kD, menor que 50 kD ou menor que 30 kD em tamanho. Nas realizações, a proteína ou polipeptídeo de interesse contém pelo menos cerca de 5, 10, 15, 20, 30, 40, 50 ou 100 aminoácidos.

[00141] A informação de sequência extensiva requerida para técnicas de genética molecular e de engenharia genética é amplamente disponível publicamente. O acesso a sequências de nucleotídeos completas de mamífero, bem como humanas, genes, sequências de cDNA, sequências de aminoácidos e genomas são frequentemente obtidos a partir do GenBank no website www.ncbi.nlm.nih.gov/Entrez. Informação adicional é também algumas vezes obtida a partir do GeneCards, uma enciclopédia eletrônica que integra informação sobre genes e seus produtos e aplicações biomédicas do Weizmann Institute of Science Genome and Bioinformatics, a informação de sequência de nucleotídeos é também algumas vezes obtida a partir do EMBL Nucleotide Sequence Database

ou o DNA Databank do Japão (DDBJ); locais adicionais para informação sobre sequências de aminoácidos incluem a página de fonte de informação sobre proteína de Georgetown e Swiss-Prot.

[00142] Exemplos de proteínas que são frequentemente expressas nas composições e métodos aqui incluem, mas não se limitam a, moléculas tais como, por exemplo, renina, um hormônio de crescimento, incluindo hormônio de crescimento humano; hormônio de crescimento bovino; fator de liberação de hormônio de crescimento; hormônio da paratiroide ; hormônio estimulante da tiroide; lipoproteínas; α-1-antitripsina; cadeia A da insulina; cadeia B da insulina; proinsulina; trombopoietina; hormônio estimulante de folículo; calcitonina; hormônio luteinizante; glucagon; fatores de coagulação tais como o fator VIIIC, fator IX, fator tissular e fator de von Willebrands; fatores anticoagulantes tais como Proteína C; fator natriurético atrial; surfactante pulmonar; um ativador de plasminogênio, tal como uroquinase ou ativador de plasminogênio do tipo urina ou tecido humano (t-PA); bombesina; trombina; fator de crescimento hematopoiético; fator alfa e beta da necrose tumoral; encefalinase; uma albumina do soro tal como albumina de soro humano; substância inibidora mulleriana; cadeia A da relaxina; cadeia B da relaxina; prorrelaxina; polipeptídeo associado a gonadotropina de camundongo; uma proteína microbiana tal como beta-lactamase; DNAse; inibina; activina; fator de crescimento endotelial vascular (VEGF); receptores para hormônios ou fatores de crescimento; integrina; proteína A ou D; fatores reumatoides; um fator neurotrófico tal como o fator neurotrófico derivado do 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; cardiotropinas (fator da hipertrofia cardíaca) tal como cardiotropina-1 (CT-1); fator de crescimento derivado de plaqueta (PDGF); fator de crescimento de fibroblasto tal como aFGF e bFGF; fator de crescimento epidérmico (EGF); fator de crescimento transformante (TGF) tal como TGF-alfa e TGF-β, incluindo TGF-β1, TGF- β2, TGF- β3, TGF- β4 ou TGF- β5; fator de crescimento semelhante a insulina I e II (IGF-I e IGF-II); des(1-3)-IGF-I (IGF-I cerebral), proteínas de

ligação a fator de crescimento semelhante a insulina; proteínas CD tais como CD-3, CD-4, CD-8 e CD-19; eritropoietina; fatores osteoindutores; imunotoxinas; uma proteína morfogênica óssea (BMP); um interferon tal como interferon-alfa, -beta e -gama; fatores estimulantes de 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; receptores de célula T; proteínas de superfície de membrana; fator de aceleração de decaimento; antígeno viral tal como, por exemplo, uma parte do envelope da AIDS; proteínas de transporte; receptores de alojamentos; addressinas; proteínas reguladoras; anticorpos; e fragmentos de qualquer um dos polipeptídeos listados acima.

[00143] Em certas realizações, a proteína ou polipeptídeo é selecionado de 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-18BPa, 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 semelhante a insulina (por exemplo, IGF-1, IGF-2); fatores de necrose tumoral (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 de inibição de leucemia (LIF); fator neurotrófico ciliar (CNTF); oncostatina M; fator de célula tronco (SCF); fatores de crescimento transformantes (por exemplo, TGF-α, TGF- β1, TGF- β2, TGF- β3); superfamília TNF (por exemplo, LIGHT/TNFSF14, STALL-1/TNFSF13B (BLy5, BAFF, THANK), TNFalpha/TNFSF2 e TWEAK/TNFSF12); ou quimiocinas (BCA-1/BLC-1, BRAK/Kec, CXCL16, CXCR3, ENA-78/LIX, Eotaxina-1, Eotaxina-2/MPIF-2, Exodus-2/SLC, Fractalcina/Neurotactina, GROalfa/MGSA, HCC-1, I-TAC, Linfotactina/ATAC/SCM, MCP-1/MCAF, MCP-3, MCP-4, MDC/STCP-1/ABCD-1, MIP-1.quadratura., 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, TECK ou toxinas (por exemplo, ω-agatoxina, μ-agatoxina, aitoxina, alopumilotoxina 267A, ω-atracotoxina-HV1, δ-Atracotoxina-Hv1b, Batracotoxina, Botocetina, Arenobufagina, Bufotalina, Bufotenina Cinobufagina, Marinobufagina, Bungarotoxina, Calcicludina, Calcisptina, Cardiotoxina III, Catrocolastatian C, Caribdotoxina, citotoxina de veneno de cobra, Conotoxina, Equinoidina, Eledoisina, Epibatidina, Fibrolase, Hefutoxina, Histrionicotoxina, Huwentoxina I, Huwentoxina II, J-ACTX-Hv1c Toxinas Tipo Kunitz, Dendrotoxina K, Dendrotoxina 1, Latrotoxina, Margatoxina, Maurotoxina, Onchidal, PhTx3, Pumilotoxina 251D, lectina de cascavel, Robustoxina, Saxitoxina, Scillatoxina, Slotoxina, Stromatoxina, Taicatoxina, Taricatoxina, Tetrodotoxina, Ricina, Gelonina, Aflatoxina, Amatoxina, Citrinina, Citocalasina, Ergotamina, Fumagilina, Fumonisina, Gliotoxina, ácido helvólico, ácido ibotênico, Muscimol, Ocratoxina, Patulina, Esterigmatocitstina, Tricoteceno, Vomitoxina, Zeranol, Zearalenona, Toxina do Antrax, antígeno ciclase protetora de adenilato, rPA, Cry toxina, toxinas de *Bordetella pertussis*, toxinas de *Clostridium botulinum*, toxinas de *Clostridium difficile*, toxinas de *Clostridium perfringens*, toxina do tétano, toxinas da difteria, toxina semelhante a Verotoxina/Shiga, enterotoxina termicamente estável, enterotoxina termicamente lábil, Enterotoxinas, Listeriolisina O, fator do cordão de *Mycobacterium tuberculosis*, exotoxina de *Pseudomonas*, endotoxina de *Salmonella*, exotoxina de *Salmonella*, Shiga toxina, alfa toxina de *Staphilococcus aureus*, beta toxina de *Staphilococcus aureus*, delta toxina de *Staphilococcus aureus*, Exfoliatina toxina, toxina da síndrome do choque tóxico, Enterotoxina, Leucocidina, Estreptolisina S ou toxina da cólera).

[00144] Em uma realização aqui, a proteína de interesse é uma proteína ou polipeptídeo multisubunidade. Proteínas multisubunidade que são expressas incluem proteínas homoméricas e heteroméricas. As proteínas multisubunidade podem incluir duas ou mais subunidades, que podem ser as mesmas ou diferentes. Por exemplo, a proteína pode ser uma proteína homomérica compreendendo 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 ou mais subunidades. A proteína pode ser também uma

proteína heteromérica incluindo 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 ou mais subunidades. Proteínas multisubunidade típicas incluem: receptores incluindo receptores de canal de íon; proteínas de matriz extracelular incluindo condroitina; colágeno; imunomoduladores incluindo proteínas MHC, anticorpos de cadeia completa, derivados de anticorpos e fragmentos de anticorpo; enzimas incluindo RNA polimerases e DNA polimerases; e proteínas de membrana.

[00145] Em uma outra realização, a proteína de interesse é algumas vezes uma proteína do sangue. As proteínas do sangue expressas nesta realização incluem, mas não se limitam a, proteínas portadoras, tais como albumina, incluindo albumina humana e bovina, transferrina, meias moléculas de transferrina recombinante, haptoglobina, fibrinogênio e outros fatores de coagulação, componentes complementos, imunoglobulinas, inibidores de enzima, precursores de substâncias tais como angiotensina e bradiquinina, insulina, endotelina e globulina, incluindo alfa, beta e gamaglobulina e outros tipos de proteínas, polipeptídeos e fragmentos destes encontrados principalmente no sangue de mamíferos. As sequências de aminoácidos para numerosas proteínas do sangues foram reportadas (ver, S. S. Baldwin (1993) Comp. Biochem Physiol. 106b:203-218), incluindo a sequência de aminoácidos para a albumina do soro humano (Lawn, L. M., *et al.*, 1981, Nucleic Acids Research, 6103-6114) e transferrina do soro humano (Yang, F. *et al.*, 1984, Proc. Natl. Acad. Sci. USA, 2752-2756).

[00146] Em uma outra realização, a proteína de interesse é algumas vezes uma enzima ou cofator recombinante. As enzimas e cofatores expressos nesta realização incluem, mas não se limitam a, aldolases, amino oxidases, aminoácido oxidases, aspartases, enzimas dependentes de B12, carboxipeptidases, carboxiesterases, carboxilases, quimiotripsina, enzimas que requerem CoA, cianohidrina sintetases, cistationa sintases, descarboxilases, desidrogenases, álcool desidrogenases, desidratases, diaforases, dioxigenases, enoato redutases, epóxido hidrases, fumerases, galactose oxidases, glicose isomerases, glicose oxidases, glicosiltrasferases, metiltransferases, nitrila hidrases, nucleosídeo

fosforilases, oxidoredutases, oxinitilases, peptidases, glicosiltrasferases, peroxidases, enzimas fundidas a um polipeptídeo terapeuticamente ativo, ativador de plasminogênio tissular; uroquinase, reptilase, estreptoquinase; catalase, superóxido dismutase; DNAse, aminoácido hidrolases (por exemplo, asparaginase, amidohidrolases); carboxipeptidases; proteases, tripsina, pepsina, quimiotripsina, papaína, bromelaína, colagenase; neuramimidase; lactase, maltase, sucrase e arabinofuranosidases.

[00147] Em uma outra realização, a proteína de interesse algumas vezes é um anticorpo ou derivado de anticorpo. Nas realizações, o anticorpo ou derivado de anticorpo é um anticorpo humanizado, anticorpo modificado, anticorpo de domínio único (por exemplo, um nanocorpo, anticorpo de domínio único de tubarão IgNAR ou anticorpo VNAR, anticorpo de domínio único de camelo), anticorpo heteroespecífico, anticorpo trivalente, anticorpo biespecífico (por exemplo, uma molécula BiTE), anticorpo de cadeia única, Fab, fragmento Fab, anticorpo linear, diacorpo, anticorpo de cadeia completa ou um fragmento ou parte destes. Um anticorpo de cadeia única frequentemente inclui as regiões de ligação a antígeno de anticorpos em uma cadeia polipeptídica única dobrada de forma estável. Fragmentos Fab são frequentemente um pedaço de um anticorpo particular. O fragmento Fab usualmente contém o sítio de ligação a antígeno. O fragmento Fab frequentemente contém 2 cadeias: um fragmento de cadeia leve e um de cadeia pesada. Estes fragmentos são usualmente ligados por meio de um ligante ou a uma ligação dissulfeto. Anticorpos e derivado de anticorpos úteis nos presentes métodos são conhecidos na técnica e descritos na literatura, por exemplo, na Patente US 9.580.719 na Tabela 9, referenciada aqui e incorporada como referência.

[00148] A sequência de codificação para a proteína ou polipeptídeo de interesse é frequentemente uma sequência nativa de codificação para o polipeptídeo alvo, se disponível, no entanto será mais frequentemente uma sequência de codificação que foi selecionada, melhorada ou otimizada para uso na célula hospedeira de expressão selecionada, por exemplo, pela síntese do gene

para refletir o viés do uso do códon de uma espécie de *Pseudomonas* tal como *P. fluorescens* ou outro organismo adequado. O(s) gene(s) que resulta(m) terá (terão) sido construído(s) dentro ou será (serão) inserido(s) em um ou mais vetores, que serão então transformados na célula hospedeira de expressão. O ácido nucleico ou um polinucleotídeo dito ser provido em uma "forma capaz de ser expressa" significa um ácido nucleico ou um polinucleotídeo que contém pelo menos um gene que é expresso pela célula hospedeira de expressão selecionada.

[00149] Em certas realizações, a proteína de interesse ou é substancialmente homóloga a, ou é uma proteína nativa, tal como uma proteína nativa de mamífero ou humana. Nestas realizações, a proteína não é encontrada em uma forma concatemérica, mas é ligada apenas a um sinal de secreção e opcionalmente uma sequência etiqueta ("tag") para a purificação e/ou reconhecimento.

[00150] Em outras realizações, a proteína de interesse é uma proteína que é ativa a partir de cerca de 20 a cerca de 42°C. Em uma realização, a proteína é ativa a temperaturas fisiológicas e é desativada quando aquecida para temperaturas altas ou extremas, tais como acima 65°C.

[00151] Em uma realização, a proteína de interesse é uma proteína que é ativa a uma temperatura de cerca de 20 a cerca de 42°C e/ou é desativada quando aquecida para temperaturas altas ou extremas, tais como temperaturas acima 65°C; é substancialmente homóloga a, ou é uma proteína nativa, tal como uma proteína nativa de mamífero ou humana e não expressa a partir de ácido nucleicos na forma concatemérica; e o promotor não é um promotor nativo em *P. fluorescens*, mas é derivado de um outro organismo, tal como *E. coli*.

[00152] Em outras realizações, a proteína, quando produzida, também inclui uma sequência de etiquetagem adicional, por exemplo uma sequência que orienta a proteína para o meio extracelular. Em uma realização, a sequência de etiquetagem adicional é operacionalmente ligada ao terminal carboxi da proteína. Em uma outra realização, a proteína inclui um sinal de secreção para um auto transportador, um sistema de secreção de dois parceiros, sistema de ramificação terminal principal ou uma porina usher fímbria.

[00153] Os exemplos a seguir são fornecidos como ilustração e não como limitação.

EXEMPLOS

[00154] Os exemplos a seguir são fornecidos para o propósito de lustrar as várias realizações do relatório e não têm por finalidade limitar o presente relatório de forma alguma. Os presentes exemplos, juntamente com os métodos descritos aqui são realizações representativas, são exemplificativos e não entendidos como limitações no escopo. Alterações nestas e outros usos que estão englobados no espírito do relatório tal como definido pelo escopo das reivindicações irão ocorrer aos técnicos no assunto.

Exemplo 1: Análise da Expressão de Proteína

[00155] A eficiência de três líderes de secreção de *P. fluorescens*, AnsB, 8484 e 5193, em direcionar a expressão solúvel de três diferentes proteínas recombinantes foi avaliada em uma escala de cultura de 0,5 mL.

[00156] Cada sequência de codificação líder de secreção foi fundida em quadro com genes que codificam tioredoxina 1 K-12 de *E. coli* (Trx-1 ou TrxA, 11,9 kDa), o anticorpo Gal2 de cadeia única (gal2 scFv, 29,6 kDa) e um antígeno protetor recombinante mutante de *Bacillus anthracis* (mrPA, 83,7 kDa). A expressão da proteína a partir destes construtos foi comparada com a expressão de construtos contendo a sequência de codificação líder de secreção Pbp fundida a cada uma das mesmas proteínas recombinantes, bem como a cepas construídas para a expressão citoplasmática de cada proteína. A expressão citoplasmática dos construtos da cepa codificaram o iniciador de metionina seguida por uma alanina, fundida à sequência de codificação TrxA, Gal2 ou mrPA. Os construtos de plasmídeo foram transformados na cepa de *P. fluorescens* DC454 e crescidos em meio seletivo (M9 glicose) de maneira a isolar os clones positivos. Os resultados deste teste estão mostrados na **Tabela 6** abaixo. Imagens SDS-CGE típicas são mostradas nas **FIG. 1** (TrxA), **FIG. 2** (Gal2 scFv) e **FIG. 3** (mrPA).

Tabela 6: Expressão de Proteína			
Cepa ID	**Líder de secreção**	**Proteína**	**Título Solúvel Médio (mg/L)**
STR58028	AnsB	TrxA	98
STR58033	8484	TrxA	96
STR58038	5193	TrxA	Nenhum detectado
STR58043	Pbp	TrxA	112
STR58048	Nenhum (Met-Ala)	TrxA	472
STR58053	AnsB	Gal2_ScFv	234
STR58058	8484	Gal2_ScFv	352
STR58063	5193	Gal2_ScFv	296
STR58068	Pbp	Gal2_ScFv	101
STR58073	Nenhum (Met-Ala)	Gal2_ScFv	59
STR58078	AnsB	mrPA	1131
STR58083	8484	mrPA	1564
STR58088	5193	mrPA	609
STR58093	Pbp	mrPA	487
STR58098	Nenhum (Met-Ala)	mrPA	Nenhum detectado

[00157] Cada cepa foi crescida em triplicata por 24 horas em meio com sais minerais, após o quê foi induzida com IPTG e mais 24 horas de incubação. O caldo foi coletado, diluído três vezes com PBS, sonicado e centrifugado. O sobrenadante foi coletado como a fração solúvel da proteína e analisado por SDS-CGE (utilizando-se o sistema Caliper LabChip® GXII). As bandas induzidas foram quantificadas em comparação com o ladder interno do sistema. As proteínas Gal2 scFv e mrPA foram expressas em níveis mais altos que a TrxA quando fundidas aos líderes AnsB, 8484 e 5193, no entanto, não foi detectada expressão de Gal2 ou mrPA quando as proteínas foram direcionadas para o citoplasma.

Exemplo 2: Rastreamento do Plasmídeo de Expressão Crisantaspase Tier 1 no Formato de 96 Poços

[00158] Em adição, foi avaliada a expressão de proteína a partir de construtos de fusão de L-asparaginase (crisantaspase) de *Erwinia* tipo II com líder.

[00159] Para o rastreamento do plasmídeo de expressão, a sequência de codificação otimizada da proteína crisantaspase foi desenhada e sintetizada para a expressão em *P. fluorescens*. o DNA de codificação para a sequência peptídica de crisantaspase (**FIG. 4**, sequência de aminoácidos apresentada como SEQ ID NO: 7 e sequência de ácidos nucleicos apresentada como SEQ ID NO: 8) foi desenhado para refletir a utilização do códon apropriado para *P. fluorescens*. Uma região do DNA contendo um sítio de enzima de restrição ímpar (SapI ou LguI) foi adicionada a montante da sequência de codificação de crisantaspase desenhada para fusão direta em quadro com a sequência de codificação líder de secreção presente no vetor de expressão. Uma região de DNA contendo 3 códons de interrupção e um sítio de enzima de restrição ímpar (SapI) foi adicionada a jusante da sequência de codificação. Foram construídos plasmídeos contendo o gene otimizado da crisantaspase fundido a 37 diferentes líderes de secreção de *P. fluorescens* e dois sítios de afinidade de ligação a ribossomo (RBS) (**Tabela 7**). Um plasmídeo adicional foi construído para expressar a proteína crisantaspase sem um líder periplásmico de maneira a se localizar a proteína crisantaspase dentro do citoplasma da célula. A expressão do alvo foi acionada a partir do promotor Ptac e a tradução foi iniciada a partir de um sítio de ligação ribossômica (RBS) de alta atividade. Os 40 plasmídeos resultantes foram transformados em duas cepas hospedeiras de *P. fluorescens*, DC454 (WT) e DC441 (PD), para produzir 80 cepas de expressão para o rastreamento de Tier 1 (estratégia de expressão). O ranqueamento das estratégias de expressão foi baseado nos títulos estimados de SDS-CGE do monómero de crisantaspase. O objetivo principal do rastreamento da estratégia de expressão foi avaliar um grande subconjunto de plasmídeos de expressão e seus elementos genéticos incorporados e para eliminar aquelas estratégias de expressão que produziram a expressão alvo de baixa qualidade.

[00160] As culturas resultantes das 80 transformações (40 estratégias de expressão x 2 cepas hospedeiras) foram crescidas em placas de 96 poços tal como descrito no Exemplo 2. Amostras da fração sonicada do caldo de cultura total colhidas 24 horas após a indução foram analisadas por SDS-CGE. A expressão da proteína induzida consistente com o peso molecular esperado para o monômero de crisantaspase (35 kDa), que também co-migrou com L-Asp de *E. coli* (Produto Sigma #A3809), foi quantificada. A quantificação em SDS-CGE das amostras reduzidas foi completada por comparação com as bandas induzidas em uma curva padrão de L-Asp de *E. coli*. As 20 amostras com maior rendimento do caldo das cepas hospedeiras DC454 e DC441foram ranqueadas (**Tabela 9**) com base nos títulos de monômero de crisantaspase solúvel. A **FIG. 5** mostra as figuras de gel de SDS-CGE geradas a partir da análise das amostras sonicadas da cultura em 96 poços. A **Tabela 9** mostra os títulos 24 horas após a indução (I24) como estimados pela análise SDS-CGE dos sonicados solúveis reduzidos e quantificados pelo ladder interno Labchip® para os plasmídeos expressos nas cepas hospedeiras DC454 (tabela à esquerda) e DC441 (tabela à direita). É também mostrado a fusão do líder de secreção produzida a partir de cada plasmídeo de expressão p743.

[00161] Os seis plasmídeos e líderes de secreção incorporados (p743-013 (FlgI), p743-038 (8484), p743-020 (LolA), p743-018 (DsbC), p743-009 (Ibp-S31A) e p743-034 (5193)) observados no topo das 10 cepas de expressão com maior produção derivadas tanto da cepa hospedeira DC454 quanto da DC441 estão marcados com ** na **Tabela 9**. Adicionalmente, o plasmídeo de expressão p743-001, desenhado para expressão citoplasmática da proteína crisantaspase, ranqueou nos dois rendimentos em solúvel mais altos para ambos os hospedeiros. Para ambas as cepas hospedeiras combinadas, os dez títulos mais alto em solúveis variaram de 525 a 1,523 µg/mL. O rendimento em insolúvel foi baixo para todas as expressões observadas com o maior rendimento em insolúvel atingindo 230 µg/mL utilizando-se o plasmídeo p743-013. A observação dos padrões de formação de banda SDS-CGE (**FIG. 5**) mostrou que o processamento mais

completo de líder de secreção (remoção por exportação para o periplasma) ocorreu na utilização dos plasmídeos de expressão p743-013, p743-033 (líder O), p743-038, p743-009 e p743-018 enquanto foi observado que os plasmídeos p743-016 e p743-017 produzem um produto de truncagem de peso molecular proeminente mais baixo. Os 10 plasmídeos de expressão mostrados na **Tabela 8**, foram escolhidos para o rastreamento de cepa hospedeira subsequente na escala HTP de 96 poços com base no título de solúveis alto estimado por SDS-CGE. As dez estratégias de expressão selecionadas foram então combinadas com 24 cepas hospedeiras únicas o que poderia influenciar adicionalmente o título e a qualidade da proteína crisantaspase.

Tabela 7: Plasmídeos de Expressão

Plasmídeo ID	Força de RBS	Líder de Secreção	Vetor RpC
p743-001	Alta	Nenhum: (Met-Ala)	pDOW5277
p743-002	Alta	DsbD	pDOW3949
p743-003	Alta	Líder A	pFNX3952
p743-004	Alta	DsbA	pDOW5206
p743-005	Med	DsbA	pDOW5207
p743-006	Alta	Azu	pDOW5209
p743-007	Med	Azu	pDOW5210
p743-008	Alta	Lao	pDOW5217
p743-009	Alta	Ibp-S31A	pDOW5220
p743-010	Alta	TolB	pDOW5223
p743-011	Alta	Tpr	pDOW5226
p743-012	Alta	Ttg2C	pDOW5232
p743-013	Alta	FlgI	pDOW5235
p743-014	Alta	CupC2	pDOW5238
p743-015	Alta	CupB2	pDOW5241
p743-016	Alta	Pbp	pDOW5201
p743-017	Alta	PbpA20V	pDOW5259
p743-018	Alta	DsbC	pDOW5262
p743-019	Alta	Líder B	pFNX3941
p743-020	Alta	Líder C	pFNX3942
p743-021	Alta	Líder D	pFNX3943
p743-022	Alta	Líder E	pFNX3944

p743-023	Alta	Líder F	pFNX3947
p743-024	Alta	Líder G	pFNX3948
p743-025	Alta	Líder H	pFNX3950
p743-026	Alta	PorE	pDOW5256
p743-027	Alta	Líder I	pFNX3959
p743-028	Alta	Líder J	pFNX3957
p743-029	Alta	Líder K	pFNX3958
p743-030	Alta	Líder L	pFNX4202
p743-031	Alta	Líder M	pFNX4203
p743-032	Alta	Líder N	pFNX4204
p743-033	Alta	Líder O	pFNX4205
p743-034	Alta	5193	pFNX4206
p743-035	Alta	Líder P	pFNX4207
p743-036	Alta	Líder Q	pFNX4208
p743-037	Alta	Líder R	pFNX4209
p743-038	Alta	8484	pFNX4210
p743-039	Alta	Líder S	pFNX4211
p743-040	Alta	Líder T	pFNX4212
p743-041	Alta	AnsB	pFNX3968
p743-042	Alta	Nenhum	pDOW5271

Tabela 8: Plasmídeos de Expressão Selecionados paro rastreamento de Cepa

Plasmídeo ID	Força de RBS	Líder de Secreção
p743-042	Alta	Nenhum
p743-009	Alta	Ibp-S31A
p743-017	Alta	Pbp-A20V
p743-013	Alta	FlgI
p743-018	Alta	DsbC
p743-020	Alta	Líder C
p743-033	Alta	Líder O
p743-034	Alta	5193
p743-038	Alta	8484
p743-041	Alta	AnsB

[00162] A **Tabela 9** mostra os títulos após 24 horas da indução (I24) tal como estimados pela análise SDS-CGE de sonicados solúveis reduzidos e quantificados pelo ladder interno Labchip® para plasmídeos expressados nas cepas hospedeiras DC454 (tabela à esquerda) e DC441 (tabela à direita). É também mostrada a fusão do líder de secreção produzida a partir de cada plasmídeo de expressão p743.

Tabela 9: **20 Cepas de expressão mais bem colocadas a partir do rastreamento de Tier 1 (**Seis líderes de secreção observados nos 10 plasmídeos de maior rendimento a partir de ambas as cepas hospedeiras.)**

Cepa ID (Hospedeiro DC454)	Plasmídeo	Líder de secreção	I24 Solúvel (ug/mL)	Cepa ID (Hospedeiro DC441)	Plasmídeo	Líder de secreção	I24 Solúvel (ug/mL)
STR55337	p743-001	Nenhum	1482	STR55429	p743-013	FlgI**	1523
STR55349	p743-013	FlgI**	1314	STR55417	p743-001	Nenhum	1369
STR55369	p743-033	Líder O	969	STR55424	p743-008	Lao	1329
STR55374	p743-038	8484**	966	STR55442	p743-026	PorE	1136
STR55356	p743-020	Líder C**	852	STR55434	p743-018	DsbC**	946
STR55354	p743-018	DsbC**	832	STR55436	p743-020	Líder C**	727
STR55345	p743-009	Ibp-S31A**	810	STR55454	p743-038	8484**	639
STR55370	p743-034	5193**	674	STR55425	p743-009	Ibp-S31A**	600
STR55347	p743-011	Tpr	657	STR55450	p743-034	5193**	532
STR55348	p743-	Ttg2C	567	STR55432	p743-	Pbp	525

	012				016		
STR55371	p743-035	Líder P	439	STR55427	p743-011	Tpr	492
STR55351	p743-015	CupB2	433	STR55449	p743-033	Líder O	477
STR55346	p743-010	TolB	431	STR55428	p743-012	Ttg2C	384
STR55360	p743-024	Líder G	279	STR55426	p743-010	TolB	363
STR55359	p743-023	Líder F	273	STR55435	p743-019	Líder B	342
STR55340	p743-004	DsbA	270	STR55431	p743-015	CupB2	248
STR55355	p743-019	Líder B	249	STR55433	p743-017	PbpA20V	225
STR55353	p743-017	PbpA20V	239	STR55451	p743-035	Líder P	223
STR55373	p743-037	Líder R	223	STR55420	p743-004	DsbA	222
STR55350	p743-014	CupC2	220	STR55452	p743-036	Líder Q	190

Exemplo 3: Preparação de Construtos de Expressão de Asparaginase de *E. coli*

[00163] Foi avaliada a expressão de proteína a partir de construtos de fusão asparaginase de *E. coli*-líder.

[00164] O gene da L-asparaginase II de *E. coli* A-1-3 foi otimizado para a expressão em *P. fluorescens* e clonado em um conjunto de vetores de expressão para expressão citoplasmática e periplásmica. A sequência de aminoácidos utilizada é descrita como SEQ ID NO: 9. A sequência de ácidos nucleicos utilizada é descrita como SEQ ID NO: 10.

[00165] A expressão foi avaliada utilizando-se uma série de sequências líder de secreção, algumas com uma sequência de RBS alta e algumas com uma sequência de RBS média. Em adição, a expressão citoplasmática foi avaliada sem a utilização de líder.

[00166] Cada construto foi transformado em cepas hospedeiras de *P. fluorescens* DC454 (deficiente em *pyrF*, sem PD ou FMO) e DC441 (deficiente em *pyrF*, PD) e as cepas de expressão resultantes foram avaliadas para a produção de L-asparaginase II de *E. coli* A-1-3 em culturas de 0,5 mL. O caldo total foi sonicado, centrifugado e as frações solúveis foram analisadas por SDS-CGE.

Crescimento e Expressão no Formato de 96 Poços

[00167] Para o rastreamento de plasmídeo de expressão, misturas de ligação para cada um dos plasmídeos de expressão de L-asparaginase II de *E. coli* A-1-3 foram transformadas nas cepas hospedeiras de *P. fluorescens* DC454 e DC441 como se segue. Vinte e cinco microlitros de células competentes foram congeladas e transformadas em uma placa de 96 multipoços Nucleovette® (Lonza) e a mistura de ligação foi adicionada a cada poço. As células foram eletroporadas utilizando-se o sistema Nucleofector™ de 96 poços Shuttle™ (Lonza AG). As células foram então transferidas para placas de 96 poços profundos com 400 µl de maio com sais M9, 1% de glicose e elementos traço. As placas de 96 poços (placas semeadas) foram incubadas a 30°C com agitação por 48 horas. Dez microlitros de cultura de semente foram transferidos em duplicata para placas de 96 poços profundos, cada poço contendo 500 µl de meio HTP, suplementado com elementos traço e 5% de glicerol e incubado como antes por 24 horas. Foi adicionado Isopropil-β-D-1-tiogalactopiranosídeo (IPTG) no ponto no tempo de 24 horas a cada poço para uma concentração final de 0,3 mM, para induzir a expressão das proteínas alvo. Foi adicionado manitol (Sigma) a cada poço para uma concentração final de 1% para induzir a expressão dos moduladores de dobramento nas cepas que superexpressam modulador de dobramento. A densidade celular foi medida por densidade ótica a 600 nm

(OD600) 24 horas após a indução de maneira a se monitorar o crescimento. Vinte e quatro horas após a indução, as células foram coletadas, diluídas 1:3 em 1X PBS para um volume final de 400µl, então foram congeladas. As amostras foram preparadas e analisadas como descrito abaixo.

[00168] Os resultados da expressão para as melhores amostras identificadas no rastreamento de plasmídeo de expressão são mostrados nas Tabelas 10 e 11.

Tabela 10: Rastreamento de Plasmídeo de Expressão em DC454

Nome Amostra	Líder	Resultado (mg/L)
STR55304-2	Lao	2052
STR55312-2	Pbp	2016
STR55305-2	Ibp-S31A	1413
STR55334-2	8484	1322
STR55333-2	Líder R	1073
STR55313-2	PbpA20V	1068
STR55302-2	Azu	884
STR55317-2	Líder D	819
STR55315-2	Líder B	764
STR55310-2	CupC2	751

Tabela 11: Rastreamento de Plasmídeo de Expressão no Hospedeiro DC441

Nome Amostra	Líder	Resultado-Bkg (I0) ug/mL
STR55382-2	Azu	993
STR55384-2	Lao	1422
STR55385-2	Ibp-S31A	1048
STR55392-2	Pbp	1305
STR55393-2	PbpA20V	997
STR55397-2	Líder D	698
STR55413-2	Líder R	1005

STR55414-2	8484	1199

[00169] Para o rastreamento de cepa hospedeira, foram incluídos plasmídeos de expressão selecionados com base nos resultados do rastreamento de plasmídeo de expressão cada um transformado em cada uma das 24 cepas hospedeiras *P. fluorescens* em um conjunto, incluindo cepas do tipo selvagem (WT) ou parental DC454, cepas com deleção de protease (PD), cepas que superexpressam modulador de dobramento (FMO) e cepas que superexpressam deleção de protease mais modulador de dobramento (PD/FMO). Asparaginase d *E. coli* fundida ao líder de secreção de aparaginase de *P. fluorescens* (AnsB) foi incluída no conjunto (sequência de aminoácidos apresentada como SEQ ID NO: 1; sequência de codificação apresentada como SEQ ID NO: 4). Os moduladores de dobramento, quando presentes, foram codificados em um segundo plasmídeo e a expressão foi acionada por um promotor induzível de manitol de nativo de *P. fluorescens*. As transformações no rastreamento de cepa hospedeira foram realizadas como se segue: vinte e cinco microlitros de células competentes da cepa hospedeira de *P. fluorescens* foram congelados e transferidos para uma placa 96-multipoço Nucleovette® e 10 µl de DNA plasmidial (10 ng) foram adicionados a cada poço. As células foram eletroporadas, cultivadas, induzidas no formato HTP e colhidas como descrito para o rastreamento de expressão de plasmídeo acima. As amostras foram preparadas e analisadas como descrito abaixo.

Preparação de Amostras para Análise

[00170] As frações solúveis foram preparadas por sonicação seguida de centrifugação. As amostras de caldo de cultura (400 µL) foram sonicadas com o Cell Lysis Automated Sonication Sistema (CLASS, Scinomix) com 24 sondas "tip horn" sob as seguintes configurações: 20 pulsos por poço a 10 segundos por pulso e potência de 60% com 10 segundos entre cada pulso (Sonics Ultra-Cell). Os lisados foram centrifugados a 5500 x g por 15 minutos (4ºC) e os sobrenadantes foram coletados (fração solúvel).

Análise SDS-CGE

[00171] Amostras de proteína foram analisadas por eletroforese em gel capilar SDS com microchip utilizando um instrumento LabChip GXII (PerkinElmer) com um chip HT Protein Express e correspondentes reagentes (PerkinElmer). As amostras foram preparadas segundo o protocolo do fabricante (Protein User Guide Document No. 450589, Rev. 3). Em resumo, em uma placa de PCR de poço cônico de polipropileno de 96 poços, 4 µL de amostra foram misturados com 14 µL de tampão de amostra, com 70 mM do agente redutor DTT, aquecido.

[00172] A amostra de caldo total retirada 24 horas após a indução foi processada como descrito acima e as frações solúveis foram analisadas por SDS-CGE.

[00173] Um kit de ensaio para atividade de L-asparaginase comercialmente disponível atividade (Sigma) detectou uma atividade significativa de L-asparaginase em amostras do lisado de cultura HTP na cepa de maior rendimento STR55382 (líder Lao) quando em comparação com a amostra Null.

[00174] Os plasmídeos e correspondentes líderes de secreção selecionados no experimento de rastreamento de plasmídeo de expressão incluíram:

[00175] p742-006 (Azu)

[00176] p742-008 (LAO)

[00177] p742-009 (Ibp-S31A)

[00178] p742-016 (Pbp)

[00179] p742-017 (PbpA20V)

[00180] p742-021 (Líder D)

[00181] p742-037 (Líder R)

[00182] p742-038 (8484)

[00183] p742-041 (*P. fluorescens* AnsB).

[00184] As cepas de expressão foram cultivadas e induzidas como descrito acima. A análise SDS-CGE das frações solúveis e frações insolúveis mostrou um alto nível de expressão da asparaginase (**FIG. 6**). Foram observados altos títulos nas cepas de expressão incluindo aqueles apresentados na **Tabela 12**.

Tabela 12: Rastreamento de Cepa Hospedeira					
Cepa de expressão ID	**Plasmídeo (Líder)**	**Cepa hospedeira**	**Solúvel I24 (ug/ml)**	**Insolúvel I24 (ug/ml)**	**Célula Total I24 (ug/ml)**
STR55467	p742-041 (AnsB)	DC454 (pyrF, sem PD, sem FMO)	3603	268	3871
STR55689	p742-009 (Ibp-S31A)	DC542 (pyrF proC, FMO lepB (RXF01181,1)	2503	152	2655
STR55559	p742-009 (Ibp-S31A)	PF1201 (pyrF proC, deficiente em proteases Lon (RXF04653,1), DegP1(RXF01250,2) DegP2 S219A (RXF07210,1 com substituição S219A), Prc1(RXF06586,1) e AprA (RXF04304,1))	2491	174	2665
STR55561	p742-016 (Pbp)	PF1201	2070	484	2554
STR55555	p742-038 (8484)	DC549 (pyrF proC, FMO tig-RXF04655)	1772	94	1865

Exemplo 4: Construção de Cepas

[00185] Foram geradas as seguintes cepas hospedeiras de asparaginase KO de *P. fluorescens*.

[00186] O PF1433 (PyrF, AspG1 e deficiente em AspG2) foi construído pela deleção sequencial dos genes aspG2 e aspG1 na cepa hospedeira DC454 (deficiente em PyrF).

[00187] O PF1434 (PyrF, ProC, AspG1 e deficiente em AspG2) foi construído por deleção sequencial dos genes *aspG1* e *aspG2* na cepa hospedeira

DC455 (*pyrF proC*). A cepa DC455 é a cepa parental de ambas as cepas DC542 e DC549.

[00188] O PF1442 (PyrF, ProC, AspG1, AspG2, Lon, DegP1, DegP2 S219A, Prc1 e deficiente em AprA) foi construído pela deleção sequencial dos genes aspG2 e aspG1 na cepa hospedeira PF1201 (PyrF, ProC, proteases Lon, DegP1, DegP2 S219A, Prc1 e deficiente em AprA).

[00189] O PF1443 (PyrF, ProC, AspG1 e deficiente em AspG2; FMO LepB expresso a partir do pDOW3700) foi construído pela transformação do plasmídeo FMO que codifica LepB pDOW3700 no PF1434.

[00190] O PF1444 (PyrF, ProC, AspG1 e deficiente em AspG2; FMO Tig expresso a partir do pDOW3707) foi construído pela transformação do plasmídeo FMO que codifica Tig pDOW3703 no PF1434.

[00191] O PF1445 (PyrF, ProC, AspG1, AspG2, Lon, DegP1, DegP2, S219A, Prc1 e deficiente em AprA; FMO DsbAC-Skp expresso a partir do pFNX4142) foi construído pela transformação do PF1442 com plasmídeo que codifica DsbAC-Skp pFNX4142.

[00192] As cepas utilizadas estão descritas na **Tabela 13**.

Tabela 13: Antecedentes dos Hospedeiros das Cepas de Expressão Deficientes em Asparaginase

Cepa de expressão ID	**Cepa Hospedeira**	**Plasmídeo de Expressão**	**Fenótipo de Fundo***	**Líder de secreção**
STR57867	PF1433	p742-041	Wild-tipo	AnsB
STR57864	PF1445	p742-009	PD/FMO	Ibp-S31A
STR57865	PF1445	p742-016	PD/FMO	Pbp
STR57866	PF1445	p742-041	PD/FMO	AnsB
STR57860	PF1443	p742-041	FMO	AnsB
STR57861	PF1444	p742-041	FMO	AnsB
STR57862	PF1443	p742-009	FMO	Ibp-S31A
STR57863	PF1444	p742-038	FMO	8484

* no que diz respeito à deficiência/deleção da protease e superexpressão do modulador de dobramento.

Exemplo 5: Fermentação em 2 L e Cálculo da % de Solúvel em TCP das Cepas de Expressão Selecionadas

[00193] As cepas STR57863 e STR57860 descritas no Exemplo 3 foram escalonadas para fermentação em 2 L e cada uma foi examinada em até oito diferentes condições de fermentação. As fermentações em escala de 2 L (aproximadamente 1 L de volume de fermentação final) foram geradas pela inoculação em um frasco de agitação contendo 600 mL de um meio quimicamente definido suplementado com extrato de levedura e glicerol com uma cultura estoque congelada da cepa selecionada. Após 16 a 24 h de incubação sob agitação a 30°C, partes iguais de cada cultura de frasco de agitação foram transferidas de forma asséptica para cada sistema de fermentação multiplex de 8 unidades contendo um meio quimicamente definido desenhado para suportar uma biomassa alta. Nos fermentadores de 2 L, as culturas foram conduzidas sob condições controladas para pH, temperatura e oxigênio dissolvido em um modo de batelada alimentada de glicerol. O processo de fermentação em alta densidade de batelada alimentada consistiu em uma fase de crescimento seguida por uma fase de indução, iniciada pela adição de IPTG e 5 g/L de manitol quando a cultura alcançou a biomassa alvo (peso úmido de célula). As condições durante a fase de indução variaram de acordo com o desenho do experimento. A fase de indução da fermentação foi deixada prosseguir por aproximadamente 24 horas. Amostras analíticas foram retiradas do fermentador para a determinação da densidade celular (densidade ótica a 575 nm) e foram então congeladas para análise subsequente para a determinação do nível de expressão do gene alvo. No ponto no tempo final time point de 24 horas após a indução, todo o caldo da fermentação de cada vaso foi coletado por centrifugação a 15900 x g por de 60 a 90 minutos. A pasta de células e o sobrenadante foram separados e a pasta retida e congelada a -80°C.

[00194] A **Tabela 14** mostra os resultados da expressão com as cepas STR57863 e STR57860 sob várias condições de fermentação. Como mostrado, várias das combinações cepa/condição de fermentação resultaram em >30% de expressão de TCP asparaginase. A proteína de célula total foi calculada como se seque:

[00195] 0,55 de DCW de proteína de célula total x 500 μg/mL de DCW a A550 = 275 μg de proteína de célula total/ml (ou mg/L) a A550 =1

[00196] TCP no ponto no tempo final (I24) = OD575 * 275mg/L de TCP

[00197] % Solúvel de TCP = 100* (título de solúveis/TCP)

Tabela 14: 2 L Resultados da Expressão em Fermentação

Nome Cepa	Pontos de Ajuste			OD575 Final	TCP a I24 (mg/L)	Título Solúveis (mg/L)	Título Insol. (mg/L)	Título Total (mg/L)	Título % solúvel	TCP % solúvel
	wcw g/g	pH	IPTG (mM)							
STR57863	0,4	6,5	0,2	197	54175	4652	745	5397	86,2	8,59
	0,4	7,2	0,08	218	59950	3879	756	4635	83,7	6,47
	0,2	7,2	0,2	217	59675	8518	1565	10083	84,5	14,27
	0,2	6,5	0,2	189	51975	24658	2329	26987	91,4	47,44
	0,2	6,5	0,08	221	60775	2863	837	3700	77,4	4,71
	0,4	7,2	0,2	221	60775	14245	1206	15451	92,2	23,44
	0,4	6,5	0,08	201	55275	8226	1137	9363	87,9	14,88
STR57860	0,4	6,5	0,2	233	64075	18508	2166	20674	89,5	28,88
	0,4	7,2	0,08	218	59950	15241	1404	16645	91,6	25,42
	0,2	7,2	0,2	245	67375	21445	4700	26145	82,0	31,83
	0,2	6,5	0,08	168	46200	15478	3357	18835	82,2	33,50
	0,4	7,2	0,2	206	56650	34283	2226	36509	93,9	60,52
	0,2	7,2	0,08	228	62700	35301	2387	37688	93,7	56,30
	0,4	6,5	0,08	198	54450	30284	1903	32187	94,1	55,62

Exemplo 6: Expressão de Asparaginase em Frasco de Agitação

[00198] A expressão em frasco de agitação (200 mL) foi realizada para se avaliar a produção de proteína a partir de cepas de expressão de crisantaspase, L-asparaginase de *E. coli* tipo II e L-asparaginase de *P. fluorescens* tipo II. Os plasmídeos de expressão de asparaginase foram transformados em cepa hospedeira deficiente em asparaginase PF1433 para produzir as cepas de expressão, como mostradas na **Tabela 15**. Os lisados gerados a partir das

amostras de frasco de agitação foram utilizados para a análise da atividade inicial e confirmação da massa intacta por análise LC-MS.

Tabela 15. Resultados da Expressão em Frasco de Agitação

Cepa	Plasmídeo	Reduzido Solúvel Médio (µg/ml)	% CV	Reduzido Insolúvel Médio (µg/ml)	% CV
STR55976 (asparaginase de *E. coli* A-1-3 tipo II)	p742-041 (AnsB)	1011	5	129	41
STR55977 (asparaginase de *E. coli* A-1-3 tipo II)	p742-009 (Ibp-S31A)	1514	9	253	56
STR55978 (Crisantaspase)	p743-042 (citoplasmática; sem líder)	1464	7	181	30
STR55979 (Crisantaspase)	p743-033 (Líder O)	743	13	195	42
STR55980 (Crisantaspase)	p743-083 (líder8484)	908	2	180	48
STR55981 (L-Asp2 de *P. fluorescens* MB101 nativa)	p744-001 (AnsB)	677	12	200	49

[00199] A **Tabela 15** mostra os títulos médios estimados tal como determinados pela análise SDS-CGE de frações solúveis reduzida e sonicadas insolúveis (dez diferentes repetições) de três cepas de expressão de crisantaspase construídas utilizando-se a cepa hospedeira PF1433 (cepa nativa deficiente em asparaginase, do tipo selvagem) analisadas na escala de volume de trabalho do frasco de agitação de 200 mL. Os títulos de SDS-CGE foram estimados com base na comparação de uma curva padrão de L-Asp de *E. coli* (Sigma). As imagens de gel de SDS-CGE tomadas tanto da análise do sonicado solúvel quanto do insolúvel de cada cepa são mostradas na **FIG. 7**. Pode-se observar que o título da proteína L-Asp2 nativa de *P. fluorescens*, produzida a partir de um construto compreendendo o líder AnsB (STR55981, a proteína de fusão líder de secreção

nativo-asparaginase nativa) foi substancialmente mais baixo que o título da proteína asparaginase de *E. coli* produzida a partir de um construto compreendendo o líder AnsB (STR55976), isto é, onde o líder AnsB foi fundido à proteína asparaginase heteróloga (677 μg/ml em comparação com 1011 μg/ml, uma diferença de cerca de 1,5 vezes).

[00200] Incluído na análise foram os crescimentos em frasco de agitação a partir de duas cepas ineficazes: STR55982 e DC432. A cepa DC432 contém o plasmídeo pDOW1169, que não contém região de codificação de crisantaspase, em uma cepa hospedeira de *P. fluorescens do tipo selvagem*. A STR55982 contém o plasmídeo pDOW1169 na cepa hospedeira PF1433 que contém deleções cromossômicas de ambas as sequências de codificação nativas de asparaginase. Todas as três cepas de expressão de crisantaspase produziram predominantemente a expressão da proteína crisantaspase solúvel, com a cepa STR55978 alcançando os títulos de solúveis mais altos de até 14 g/L. Além disto, nenhuma penalidade de crescimento foi observada na medida em que todas as três cepas de expressão crisantaspase alcançaram uma densidade celular similar (OD600 = 23,0, 27,0 e 27,8) em 24 horas após a indução quando em comparação com as cepas ineficazes STR55982 e DC432, o que produziu uma OD600 de 21,7 e 23,7, respectivamente, em 24 horas após a indução.

[00201] As amostras de sonicado solúvel gerado de cada uma das cinco cepas de expressão em frasco de agitação foram analisadas para a atividade de asparaginase utilizando-se um kit comercial adquirido da Sigma (Asparaginase Activity Assay Kit) de acordo com as instruções do fabricante. Este kit mede a atividade utilizando uma reação enzimática acoplada que produz um produto final colorimétrico proporcional ao aspartato gerado. A asparaginase tipo II de *E. coli* da Sigma (A3809) foi colocada no lisado de STR55982 como um controle negativo (última fila).

[00202] Os resultados da atividade são mostrados na **Tabela 16**.

Tabela 16. Ensaio de Atividade de Asparaginase de Amostras de Cultura de Frasco de Agitação Sonicadas

Descrição da Amostra	Plasmídeo ID	Amostra ID	Título de Asparaginase (mg/ml)*	Fator de Diluição da Amostra	Aspartato Gerado (nmol)	Δ A570 (TF-T0) 20 min
AnsB/*E. coli* Asp2	p742-041	STR55976	0,20	25,000	0,36	0,03
Ibp-S31A/*E. coli* Asp2	p742-009	STR55977	0,30	25,000	0,74	0,04
Crisantaspase expressa no Citoplasma	p743-042	STR55978	0,29	25,000	0,89	0,04
Líder O/Crisantaspase	p743-033	STR55979	0,15	25,000	0,86	0,04
Líder 8484/Crisantaspase	p743-038	STR55980	0,18	25,000	0,88	0,04
AnsB/*P. fluorescens* nativa L-Asp2	p744-001	STR55981	0,14	25,000	0,10	0,02
L-Asp-inef.	plasmídeo vazio	STR55982	0,00	25,000	0,00	0,00
L-asp+inef.	plasmídeo vazio	DC432	0,00	25,000	0,00	0,00
Inef. com 250 µg/ml**	N/A	L-Asp2 Sigma	0,25	25,000	0,53	0,03

*Determinado por SDS-CGE
**Sigma A3809 *E. coli* AspG2 colocada no lisado de STR55982 (Ineficaz deficiente em AspG)

[00203] Embora ambas as amostras de ineficazes não tenham mostrado atividade mensurável no fator de diluição de 1:25000, as amostras de sonicado solúvel das cepas STR55976, STR55977, STR55978, STR55979 e STR55980 diluídas a 1:25000 mostraram atividade comparável com a amostra da cepa ineficaz STR55982 similarmente diluída que recebeu 250 µg/mL de L-asparaginase de *E. coli* da Sigma (A3809). Estes resultados de atividade inicial utilizando-se um kit comercialmente disponível pareceriam indicar que a proteína crisantaspase e a proteína asparaginase de *E. coli* expressa em *P. fluorescens* podem formar a enzima asparaginase ativa tetramérica centro dos sonicados gerados.

[00204] A **Tabela 17** mostra os resultados de massa intacta por LC-MS da análise da proteína crisantaspase de sonicados solúveis produzidos pelas cepas STR55978, STR55979 e STR55980 em frascos de agitação e de asparaginase de *E. coli* de sonicados solúveis produzidos pelas cepas STR55976 e STR55977. O peso molecular observado (35,053 Da) da proteína crisantaspase de cada cepa é consistente com o peso molecular teórico (35054,2 Da) indicando que todas as três cepas geram a sequência de aminoácidos esperada e processamento completo ou remoção do líder de secreção, se presente. O peso molecular observado (34591) da asparaginase *E. coli* de cada cepa é consistente com o peso molecular teórico (34591,96), indicando que ambas as cepas geram a sequência de aminoácidos esperada e que o processamento completo ou remoção do líder de secreção ocorreu. A L-Asp de *E. coli* da Sigma foi analisada como controle. A **FIG. 8** mostra uma leitura de espectrometria de massa para STR55978, indicando a clivagem apropriada da crisantaspase do líder AnsB. Foi mostrado que a Asp2 de *E. coli* expressa em STR55976 e STR55977 por LC-MS é apropriadamente clivada dos líderes AnsB e Ibp-S31A, respectivamente.

Tabela 17. Análise LC-MS de Amostras de Cultura Sonicada em Frasco de Agitação

Nome da Amostra	PM Teór. (Da) – sinal	PM Observado (Da)	PM Obs. - Teór. (Da)
STR55976 AnsB/*E. coli* Asp2	34591,96	34591	-0,96
STR55977 Ibp-S31A/*E. coli* Asp2		34591	-0,96
STR55978 Crisantaspase expressa no Citoplasma	35054,2	35053	-1,2
STR55979 Líder O/Crisantaspase		35053	-1,2
STR55980 líder 8484/Crisantaspase		35053	-1,2
STR55981 AnsB/*P. fluorescens* nativa L-Asp2	36178,4	36177	-1,4
E. coli L-Asp A3809 Sigma	34591,96	34591	-0,96

Tabela 18: Tabela de Sequências Listadas		
Proteína ou Ácido nucleico	**Sequência**	**SEQ ID NO:**
AnsB aminoácido	MKSALKNVIPGALALLLLFPVAAQA	1
8484 aminoácido	MRQLFFCLMLMVSLTAHA	2
5193 aminoácido	MQSLPFSALRLLGVLAVMVCVLLTTPARA	3
AnsB ácido nucleico	ATGAAATCTGCATTGAAGAACGTTATTCCGGGCGCCCTGGCCCT TCTGCTGCTATTCCCCGTCGCCGCCCAGGCC	4
8484 ácido nucleico	ATGCGACAACTATTTTTCTGTTTGATGCTGATGGTGTCGCTCAC GGCGCACGCC	5
5193 ácido nucleico	ATGCAAAGCCTGCCGTTCTCTGCGTTACGCCTGCTCGGTGTGCT GGCAGTCATGGTCTGCGTGCTGTTGACGACGCCAGCCCGTGCC	6
aminoácido de *Erwinia* Crisantaspase	ADKLPNIVILATGGTIAGSAATGTQTTGYKAGALGVDTLINAVP EVKKLANVKGEQFSNMASENMTGDVVLKLSQRVNELLARDDVDG VVITHGTDTVEESAYFLHLTVKSDKPVVFVAAMRPATAISADGP MNLLEAVRVAGDKQSRGRGVMVVLNDRIGSARYITKTNASTLDT FKANEEGILGVIIGNRIYYQNRIDKLHTTRSVFDVRGLTSLPKV DILYGYQDDPEILYDAAIQHGVKGIVYAGMGAGSVSVRGIAGMR KAMEKGVVVIRSTRTGNGIVPPDEELPGLVSDSLNPAHARILLM LALTRTSDPKVIQEYFHTY	7
Uma sequência de ácidos nucleicos optimizada para *P. fluorescens*, que codifica a Crisantaspase de *Erwinia* da SEQ ID NO: 7	GCAGACAAACTCCCTAACATCGTAATCCTCGCAACTGGTGGTAC CATCGCAGGCAGCGCCGCCACCGGCACGCAGACCACTGGCTACA AGGCCGGCGCGCTGGGCGTAGACACGCTGATCAACGCCGTCCCG GAAGTGAAGAAACTGGCCAACGTCAAGGGTGAGCAATTCTCCAA CATGGCCAGCGAGAACATGACTGGCGATGTGGTACTGAAGCTCT CGCAGCGCGTGAACGAACTGCTCGCCCGCGACGACGTGGACGGC GTGGTGATCACCCACGGCACTGATACCGTCGAAGAGTCGGCGTA CTTTCTCCACCTGACCGTGAAGTCCGATAAGCCCGTGGTGTTTG TCGCCGCGATGCGCCCGGCGACCGCCATCAGCGCCGACGGGCCG ATGAATCTGTTGGAAGCCGTGCGCGTGGCGGGTGACAAGCAAAG CCGCGGTCGGGGCGTAATGGTCGTCCTGAACGATCGGATCGGTA GCGCGCGGTACATCACCAAGACGAACGCCTCCACGCTGGACACC TTCAAGGCGAACGAAGAGGGGTACCTGGGGGTGATCATTGGCAA TCGTATCTATTACCAGAACCGCATCGACAAGCTGCACACCACCC GCTCGGTGTTCGACGTGCGCGGTCTGACTAGCCTGCCCAAGGTC GACATCCTGTACGGCTACCAAGACGACCCGGAGTACCTCTACGA CGCGGCGATCCAGCATGGCGTGAAGGGCATCGTCTACGCCGGTA TGGGTGCCGGCTCGGTGTCGGTCCGCGGCATCGCGGGTATGCGC AAGGCCATGGAGAAAGGCGTGGTCGTGATTCGCTCGACCCGGAC TGGCAATGGCATCGTACCGCCCGATGAAGAACTCCCGGGGCTCG TGAGCGATAGCCTCAACCCCGCGCACGCCCGGATCCTGCTGATG CTGGCGCTCACGCGGACCAGCGACCCCAAGGTCATTCAAGAGTA CTTCCACACCTAC	8
Sequência de aminoácidos madura de L-	LPNITILATGGTIAGGGDSATKSNYTAGKVGVENLVNAVPQLKD IANVKGEQVVNIGSQDMNDDVWLTLAKKINTDCDKTDGFVITHG TDTMEETAYFLDLTVKCDKPVVMVGAMRPSTSMSADGPFNLYNA	9

Asparaginase Tipo II de *E. coli* A-1-3 (sem sequência líder de secreção)	VVTAADKASANRGVLVVMNDTVLDGRDVTKTNTTDVATFKSVNY GPLGYIHNGKIDYQRTPARKHTSDTPFDVSKLNELPKVGIVYNY ANASDLPAKALVDAGYDGIVSAGVGNGNLYKTVFDTLATAAKNG TAVVRSSRVPTGATTQDAEVDDAKYGFVASGTLNPQKARVLLQL ALTQTKDPQQIQQIFNQY	
Sequência de ácidos nucleicos otimizada que codifica a sequência de aminoácidos madura de L-Asparaginase Tipo II de *E. coli* A-1-3 da SEQ ID NO: 9	CTCCCTAACATTACTATTCTGGCCACTGGCGGTACGATTGCAGG CGGCGGTGACTCAGCCACCAAGTCGAATTACACCGCCGGTAAGG TCGGTGTCGAAAACCTCGTCAACGCCGTGCCGCAGCTGAAAGAT ATCGCCAACGTCAAGGGCGAGCAAGTGGTGAACATCGGCTCCCA AGATATGAACGATGACGTGTGGCTGACGCTGGCCAAGAAAATCA ACACCGATTGCGACAAGACGGACGGGTTTGTCATCACCCACGGC ACCGACACTATGGAAGAGACTGCCTACTTCCTCGACCTCACGGT GAAGTGCGATAAACCGGTAGTGATGGTGGGCGCCATGCGCCCGA GCACCTCGATGAGCGCGGACGGCCCGTTCAATCTGTACAACGCC GTGGTAACCGCAGCGGACAAGGCGTCCGCGAACCGCGGTGTATT GGTAGTGATGAACGATACGGTGCTCGATGGGCGCGATGTGACCA AGACCAATACCACTGATGTGGCCACCTTCAAGAGCGTGAACTAT GGCCCGCTGGGCTACATCCATAACGGCAAGATCGATTACCAGCG TACTCCCGCCCGGAAGCACACCTCGGACACCCCCTTCGACGTGT CGAAACTGAACGAACTGCCCAAGGTCGGCATCGTCTACAACTAC GCCAATGCGAGCGATCTGCCCGCGAAGGCCCTGGTGGACGCCGG CTACGACGGGATCGTATCGGCGGGTGTGGGCAATGGTAACCTGT ACAAGACCGTGTTTGACACCCTGGCGACGGCGGCGAAGAACGGC ACCGCCGTGGTCCGCAGCAGCCGCGTGCCCACTGGGGCGACCAC CCAAGACGCCGAGGTCGACGACGCGAAGTACGGCTTCGTAGCCA GCGGCACCCTGAACCCGCAAAAGGCCCGGGTCCTGCTGCAGCTG GCGCTCACGCAGACGAAGGACCCGCAGCAAATCCAACAGATCTT CAACCAGTAC	10
Shine-Dalgarno RBS (consenso)	AGGAGG	11
RBS2	GGAGCG	12
RBS34	GGAGCG	13
RBS41	AGGAGT	14
RBS43	GGAGTG	15
RBS48	GAGTAA	16
RBS1	AGAGAG	17
RBS35	AAGGCA	18
RBS49	CCGAAC	19
Uma sequência de ácidos nucleicos otimizada para *P. fluorescens*, que codifica a Crisantaspase de *Erwinia* da SEQ ID NO: 7, incluindo sítios	ATATGCTCTTCAGCCGCAGACAAACTCCCTAACATCGTAATCCT CGCAACTGGTGGTACCATCGCAGGCAGCGCCGCCACCGGCACGC AGACCACTGGCTACAAGGCCGGCGCGCTGGGCGTAGACACGCTG ATCAACGCCGTCCCGGAAGTGAAGAAACTGGCCAACGTCAAGGG TGAGCAATTCTCCAACATGGCCAGCGAGAACATGACTGGCGATG TGGTACTGAAGCTCTCGCAGCGCGTGAACGAACTGCTCGCCCGC GACGACGTGGACGGCGTGGTGATCACCCACGGCACTGATACCGT CGAAGAGTCGGCGTACTTTCTCCACCTGACCGTGAAGTCCGATA AGCCCGTGGTGTTTGTCGCCGCGATGCGCCCGGCGACCGCCATC AGCGCCGACGGGCCGATGAATCTGTTGGAAGCCGTGCGCGTGGC	20

de restrição tal como mostrado na Fig. 4	GGGTGACAAGCAAAGCCGCGGTCGGGGCGTAATGGTCGTCCTGA ACGATCGGATCGGTAGCGCGCGGTACATCACCAAGACGAACGCC TCCACGCTGGACACCTTCAAGGCGAACGAAGAGGGGTACCTGGG GGTGATCATTGGCAATCGTATCTATTACCAGAACCGCATCGACA AGCTGCACACCACCCGCTCGGTGTTCGACGTGCGCGGTCTGACT AGCCTGCCCAAGGTCGACATCCTGTACGGCTACCAAGACGACCC GGAGTACCTCTACGACGCGGCGATCCAGCATGGCGTGAAGGGCA TCGTCTACGCCGGTATGGGTGCCGGCTCGGTGTCGGTCCGCGGC ATCGCGGGTATGCGCAAGGCCATGGAGAAAGGCGTGGTCGTGAT TCGCTCGACCCGGACTGGCAATGGCATCGTACCGCCCGATGAAG AACTCCCGGGGCTCGTGAGCGATAGCCTCAACCCCGCGCACGCC CGGATCCTGCTGATGCTGGCGCTCACGCGGACCAGCGACCCCAA GGTCATTCAAGAGTACTTCCACACCTACTGATAATAGTTCAGAA GAGCATAT	

[00205] Embora as realizações preferidas do presente relatório tenham sido mostradas e descritas aqui, tais realizações são providas apenas como exemplo. Numerosas variações, alterações e substituições irão ocorrer aos técnicos no assunto sem se afastarem do relatório. Deve ser entendido que várias alternativas às realizações do relatório descrito aqui podem ser empregadas na prática dos presentes métodos. É pretendido que as reivindicações a anexas definam o escopo da invenção e que métodos e estruturas dentro do escopo destas reivindicações e seus equivalentes estejam cobertos por estas.

REIVINDICAÇÕES

1. Polipeptídeo compreendendo um peptídeo líder operacionalmente ligado a uma proteína ou polipeptídeo de interesse, **caracterizado** pelo fato do peptídeo líder apresentar a sequência de aminoácidos tal como apresentada nas SEQ ID NO: 1, 2 ou 3.

2. Polipeptídeo de acordo com a reivindicação 1, **caracterizado** pelo fato do peptídeo líder não ser nativo à proteína ou polipeptídeo de interesse.

3. Polipeptídeo de acordo com a reivindicação 1, **caracterizado** pelo fato da proteína ou polipeptídeo de interesse ser selecionado de um anticorpo ou derivado de anticorpo, uma enzima, uma citocina, uma quimiocina, um fator de crescimento e um antígeno de vacina.

4. Polipeptídeo de acordo com a reivindicação 2, **caracterizado** pelo fato do anticorpo ou derivado de anticorpo ser um scFv, a Fab, um anticorpo humanizado, um anticorpo modificado, um anticorpo de domínio único, um anticorpo heteroespecífico, um anticorpo trivalente, um anticorpo biespecífico, um anticorpo de cadeia única, a fragmento Fab, um anticorpo linear, um diacorpo ou um anticorpo de cadeia completa.

5. Polipeptídeo de acordo com a reivindicação 1, **caracterizado** pelo fato de compreender adicionalmente um ligante.

6. Polipeptídeo de acordo com a reivindicação 1, **caracterizado** pelo fato de compreender adicionalmente um domínio de clivagem.

7. Polipeptídeo de acordo com a reivindicação 1, **caracterizado** pelo fato da primeira parte ser um peptídeo líder que direciona a expressão da proteína ou polipeptídeo de interesse para o periplasma de uma célula hospedeira procariótica.

8. Polipeptídeo de acordo com a reivindicação 7, **caracterizado** pelo fato da célula hospedeira procariótica ser selecionada de uma célula de *Pseudomonad* ou uma célula de *E. coli*.

9. Método para a produção de uma proteína ou polipeptídeo de interesse em uma cultura de célula procariótica, **caracterizado** pelo fato de compreender: (a) o cultivo de células procarióticas em um meio de crescimento de cultura celular, onde as células procarióticas compreendem um ácido nucleico que codifica a proteína ou polipeptídeo de interesse operacionalmente ligados a um peptídeo líder; e (b) o isolamento da proteína ou polipeptídeo de interesse do periplasma das células procarióticas, onde o peptídeo líder compreende uma sequência de aminoácidos selecionada das SEQ ID NOS: 1, 2 e 3.

10. Método de acordo com a reivindicação 9, **caracterizado** pelo fato da proteína ou polipeptídeo de interesse ser selecionado de um anticorpo ou derivado de anticorpo, uma enzima, uma citocina, uma quimiocina, um fator de crescimento e um antígeno de vacina.

11. Método de acordo com a reivindicação 10, **caracterizado** pelo fato do anticorpo ou derivado de anticorpo ser um scFv, a Fab, um anticorpo humanizado, um anticorpo modificado, um anticorpo de domínio único, um anticorpo heteroespecífico, um anticorpo trivalente, um anticorpo biespecífico, um anticorpo de cadeia única, um fragmento Fab, um anticorpo linear, um diacorpo ou um anticorpo de cadeia completa.

12. Método de acordo com a reivindicação 9, **caracterizado** pelo fato do ácido nucleico codificar um ligante.

13. Método de acordo com a reivindicação 9, **caracterizado** pelo fato do ácido nucleico codificar um domínio de clivagem.

14. Método de acordo com a reivindicação 9, **caracterizado** pelo fato das células procarióticas serem selecionada de uma célula de *Pseudomonad* ou uma célula de *E. coli*.

15. Método para a expressão de uma proteína ou polipeptídeo de interesse no periplasma de uma célula procariótica, **caracterizado** pelo fato de compreender o cultivo da célula procariótica compreendendo um ácido nucleico que codifica a proteína ou polipeptídeo de interesse operacionalmente ligados a um peptídeo líder em um meio de crescimento de cultura celular, onde o peptídeo

líder compreende uma sequência de aminoácidos selecionada das SEQ ID NOS: 1, 2 e 3.

16. Método de acordo com a reivindicação 15, **caracterizado** pelo fato da proteína ou polipeptídeo de interesse ser selecionado de um anticorpo ou derivado de anticorpo, uma enzima, uma citocina, uma quimiocina, um fator de crescimento e um antígeno de vacina.

17. Método de acordo com a reivindicação 16, **caracterizado** pelo fato do anticorpo ou derivado de anticorpo ser um scFv, a Fab, um anticorpo humanizado, um anticorpo modificado, um anticorpo de domínio único, um anticorpo heteroespecífico, um anticorpo trivalente, um anticorpo biespecífico, um anticorpo de cadeia única, um fragmento Fab, um anticorpo linear, um diacorpo ou um anticorpo de cadeia completa.

18. Método de acordo com a reivindicação 15, **caracterizado** pelo fato do ácido nucleico codificar um ligante.

19. Método de acordo com a reivindicação 15, **caracterizado** pelo fato do ácido nucleico codificar um domínio de clivagem.

20. Método de acordo com a reivindicação 15, **caracterizado** pelo fato das células procarióticas serem selecionadas de uma célula de *Pseudomonad* ou um célula de *E. coli*.

DESENHOS

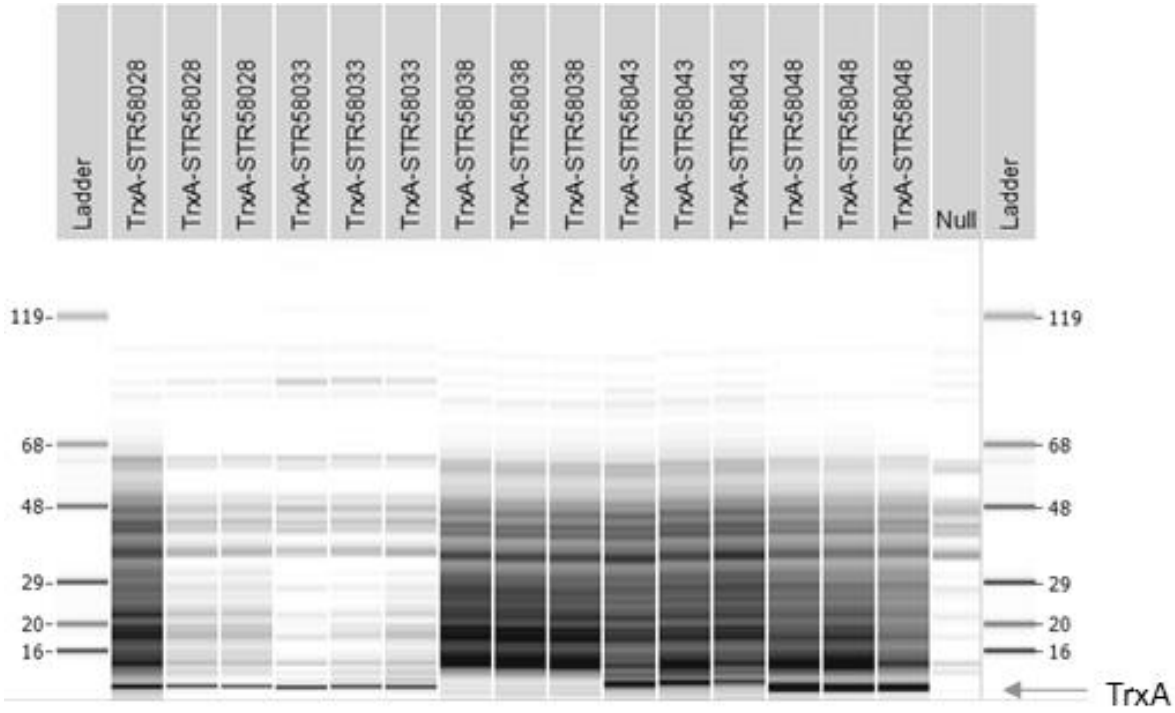


FIG. 1

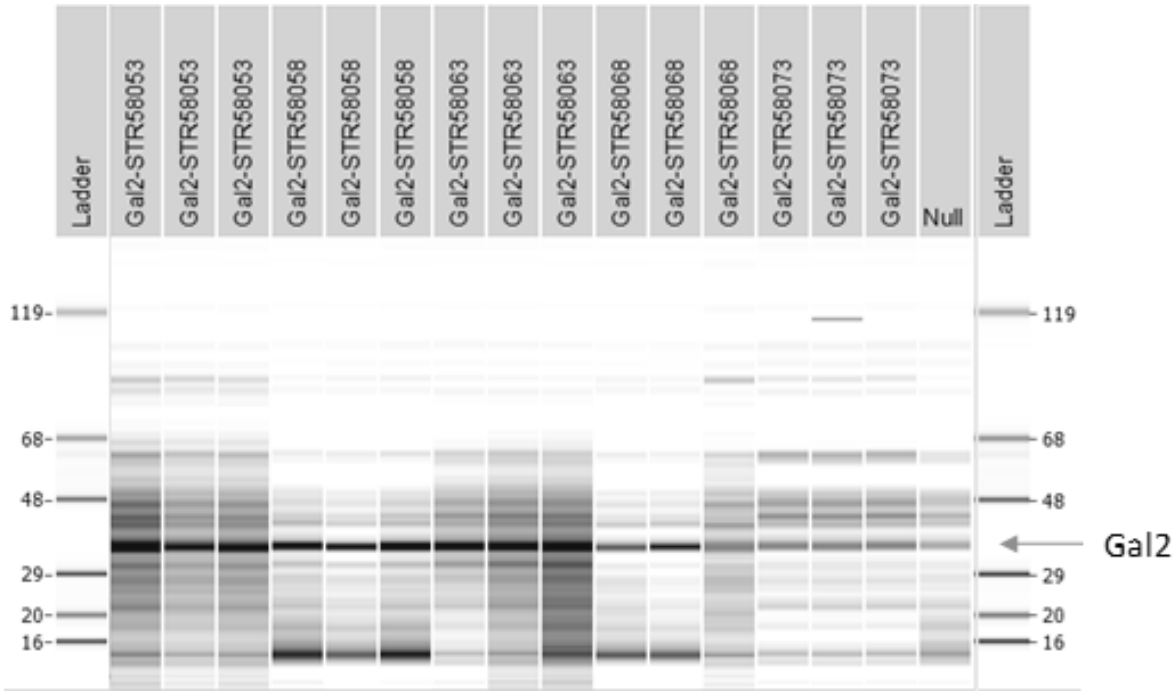

Ladder
Gal2-STR58053
Gal2-STR58053
Gal2-STR58053
Gal2-STR58058
Gal2-STR58058
Gal2-STR58058
Gal2-STR58063
Gal2-STR58063
Gal2-STR58063
Gal2-STR58068
Gal2-STR58068
Gal2-STR58068
Gal2-STR58073
Gal2-STR58073
Gal2-STR58073
Null
Ladder
119
68
48
29
20
16
Gal2

FIG. 2

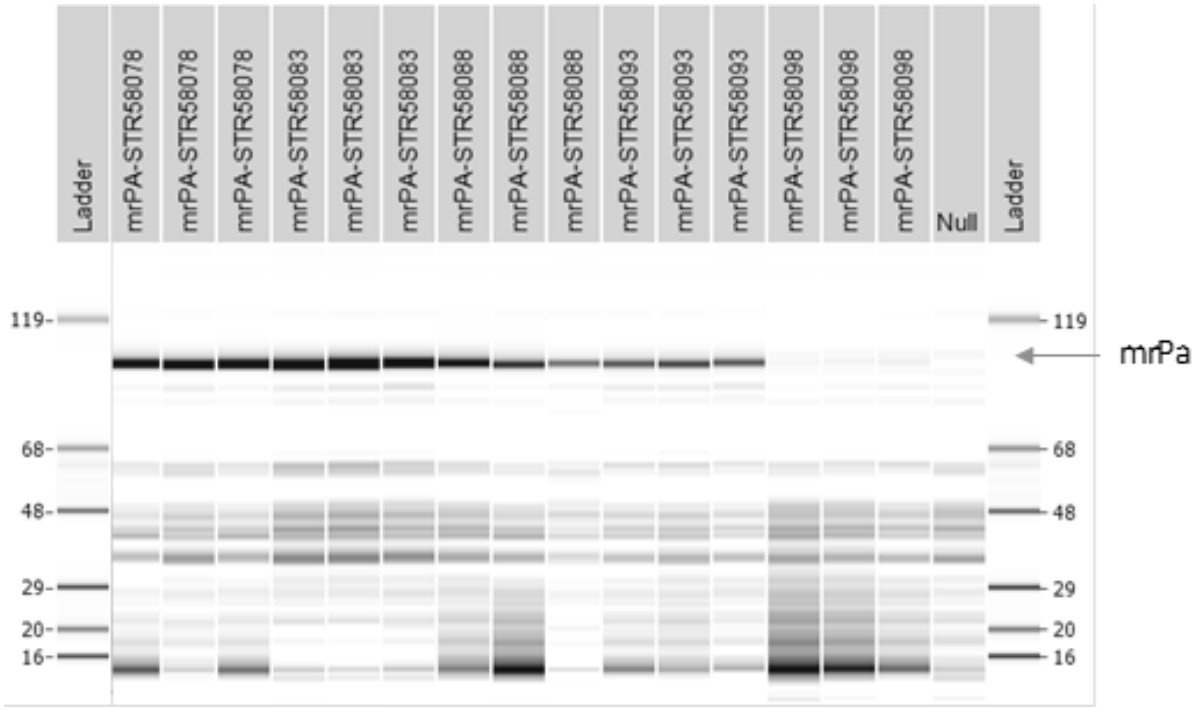

Ladder
mrPA-STR58078
mrPA-STR58078
mrPA-STR58078
mrPA-STR58083
mrPA-STR58083
mrPA-STR58083
mrPA-STR58088
mrPA-STR58088
mrPA-STR58088
mrPA-STR58093
mrPA-STR58093
mrPA-STR58093
mrPA-STR58098
mrPA-STR58098
mrPA-STR58098
Null
Ladder
119
68
48
29
20
16
mrPa

FIG. 3

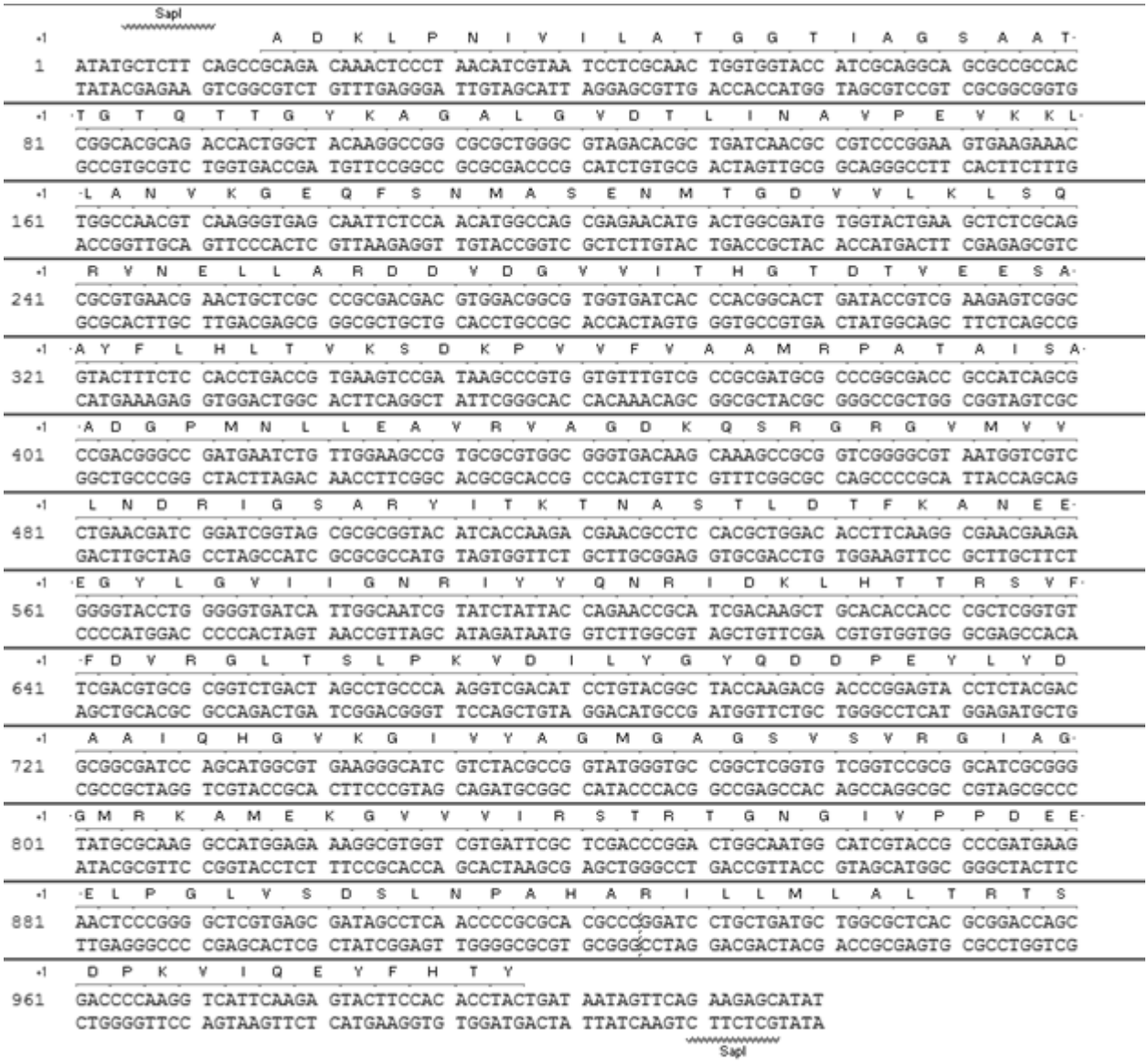

SapI
-1 A D K L P N I V I L A T G G T I A G S A A T
1 ATATGCTCTT CAGCCGCAGA CAAACTCCCT AACATCGTAA TCCTCGCAAC TGGTGGTACC ATCGCAGGCA GCGCCGCCAC
TATACGAGAA GTCGGCGTCT GTTTGAGGGA TTGTAGCATT AGGAGCGTTG ACCACCATGG TAGCGTCCGT CGCGGCGGTG
-1 T G T Q T T G Y K A G A L G V D T L I N A V P E V K K L
81 CGGCACGCAG ACCACTGGCT ACAAGGCCGG CGCGCTGGGC GTAGACACGC TGATCAACGC CGTCCCGGAA GTGAAGAAAC
GCCGTGCGTC TGGTGACCGA TGTTCCGGCC GCGCGACCCG CATCTGTGCG ACTAGTTGCG GCAGGGCCTT CACTTCTTTG
-1 L A N V K G E Q F S N M A S E N M T G D V V L K L S Q
161 TGGCCAACGT CAAGGGTGAG CAATTCTCCA ACATGGCCAG CGAGAACATG ACTGGCGATG TGGTACTGAA GCTCTCGCAG
ACCGGTTGCA GTTCCCACTC GTTAAGAGGT TGTACCGGTC GCTCTTGTAC TGACCGCTAC ACCATGACTT CGAGAGCGTC
-1 R V N E L L A R D D V D G V V I T H G T D T V E E S A
241 CGCGTGAACG AACTGCTCGC CCGCGACGAC GTGGACGGCG TGGTGATCAC CCACGGCACT GATACCGTCG AAGAGTCGGC
GCGCACTTGC TTGACGAGCG GGCGCTGCTG CACCTGCCGC ACCACTAGTG GGTGCCGTGA CTATGGCAGC TTCTCAGCCG
-1 A Y F L H L T V K S D K P V V F V A A M R P A T A I S A
321 GTACTTTCTC CACCTGACCG TGAAGTCCGA TAAGCCCGTG GTGTTTGTCG CCGCGATGCG CCCGGCGACC GCCATCAGCG
CATGAAAGAG GTGGACTGGC ACTTCAGGCT ATTCGGGCAC CACAAACAGC GGCGCTACGC GGGCCGCTGG CGGTAGTCGC
-1 A D G P M N L L E A V R V A G D K Q S R G R G V M V V
401 CCGACGGGCC GATGAATCTG TTGGAAGCCG TGCGCGTGGC GGGTGACAAG CAAAGCCGCG GTCGGGGCGT AATGGTCGTC
GGCTGCCCGG CTACTTAGAC AACCTTCGGC ACGCGCACCG CCCACTGTTC GTTTCGGCGC CAGCCCCGCA TTACCAGCAG
-1 L N D R I G S A R Y I T K T N A S T L D T F K A N E E
481 CTGAACGATC GGATCGGTAG CGCGCGGTAC ATCACCAAGA CGAACGCCTC CACGCTGGAC ACCTTCAAGG CGAACGAAGA
GACTTGCTAG CCTAGCCATC GCGCGCCATG TAGTGGTTCT GCTTGCGGAG GTGCGACCTG TGGAAGTTCC GCTTGCTTCT
-1 E G Y L G V I I G N R I Y Y Q N R I D K L H T T R S V F
561 GGGGTACCTG GGGGTGATCA TTGGCAATCG TATCTATTAC CAGAACCGCA TCGACAAGCT GCACACCACC CGCTCGGTGT
CCCCATGGAC CCCCACTAGT AACCGTTAGC ATAGATAATG GTCTTGGCGT AGCTGTTCGA CGTGTGGTGG GCGAGCCACA
-1 F D V R G L T S L P K V D I L Y G Y Q D D P E Y L Y D
641 TCGACGTGCG CGGTCTGACT AGCCTGCCCA AGGTCGACAT CCTGTACGGC TACCAAGACG ACCCGGAGTA CCTCTACGAC
AGCTGCACGC GCCAGACTGA TCGGACGGGT TCCAGCTGTA GGACATGCCG ATGGTTCTGC TGGGCCTCAT GGAGATGCTG
-1 A A I Q H G V K G I V Y A G M G A G S V S V R G I A G
721 GCGGCGATCC AGCATGGCGT GAAGGGCATC GTCTACGCCG GTATGGGTGC CGGCTCGGTG TCGGTCCGCG GCATCGCGGG
CGCCGCTAGG TCGTACCGCA CTTCCCGTAG CAGATGCGGC CATACCCACG GCCGAGCCAC AGCCAGGCGC CGTAGCGCCC
-1 G M R K A M E K G V V V I R S T R T G N G I V P P D E E
801 TATGCGCAAG GCCATGGAGA AAGGCGTGGT CGTGATTCGC TCGACCCGGA CTGGCAATGG CATCGTACCG CCCGATGAAG
ATACGCGTTC CGGTACCTCT TTCCGCACCA GCACTAAGCG AGCTGGGCCT GACCGTTACC GTAGCATGGC GGGCTACTTC
-1 E L P G L V S D S L N P A H A R I L L M L A L T R T S
881 AACTCCCGGG GCTCGTGAGC GATAGCCTCA ACCCCGCGCA CGCCCGGATC CTGCTGATGC TGGCGCTCAC GCGGACCAGC
TTGAGGGCCC CGAGCACTCG CTATCGGAGT TGGGGCGCGT GCGGGCCTAG GACGACTACG ACCGCGAGTG CGCCTGGTCG
-1 D P K V I Q E Y F H T Y
961 GACCCCAAGG TCATTCAAGA GTACTTCCAC ACCTACTGAT AATAGTTCAG AAGAGCATAT
CTGGGGTTCC AGTAAGTTCT CATGAAGGTG TGGATGACTA TTATCAAGTC TTCTCGTATA
SapI

FIG. 4

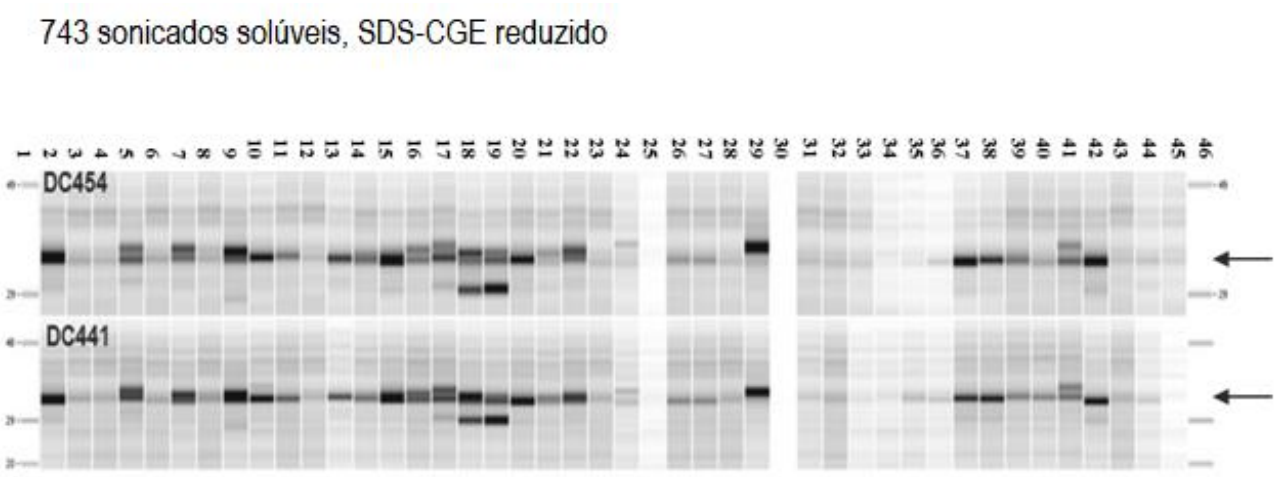


FIG. 5

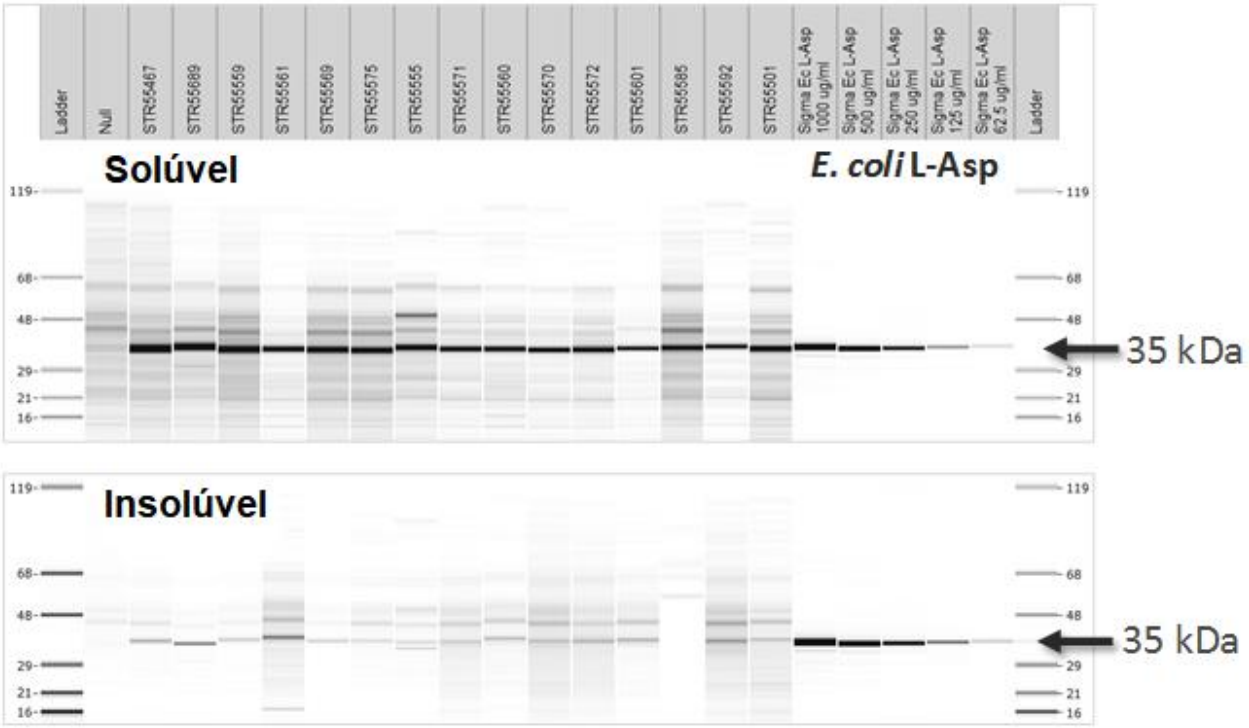

Ladder
Null
STR55467
STR55689
STR55559
STR55561
STR55569
STR55575
STR55555
STR55571
STR55560
STR55570
STR55572
STR55601
STR55585
STR55592
STR55501
Sigma Ec L-Asp 1000 ug/ml
Sigma Ec L-Asp 500 ug/ml
Sigma Ec L-Asp 250 ug/ml
Sigma Ec L-Asp 125 ug/ml
Sigma Ec L-Asp 62.5 ug/ml
Ladder
Solúvel
E. coli L-Asp
119
68
48
29
21
16
35 kDa
Insolúvel
35 kDa

FIG. 6

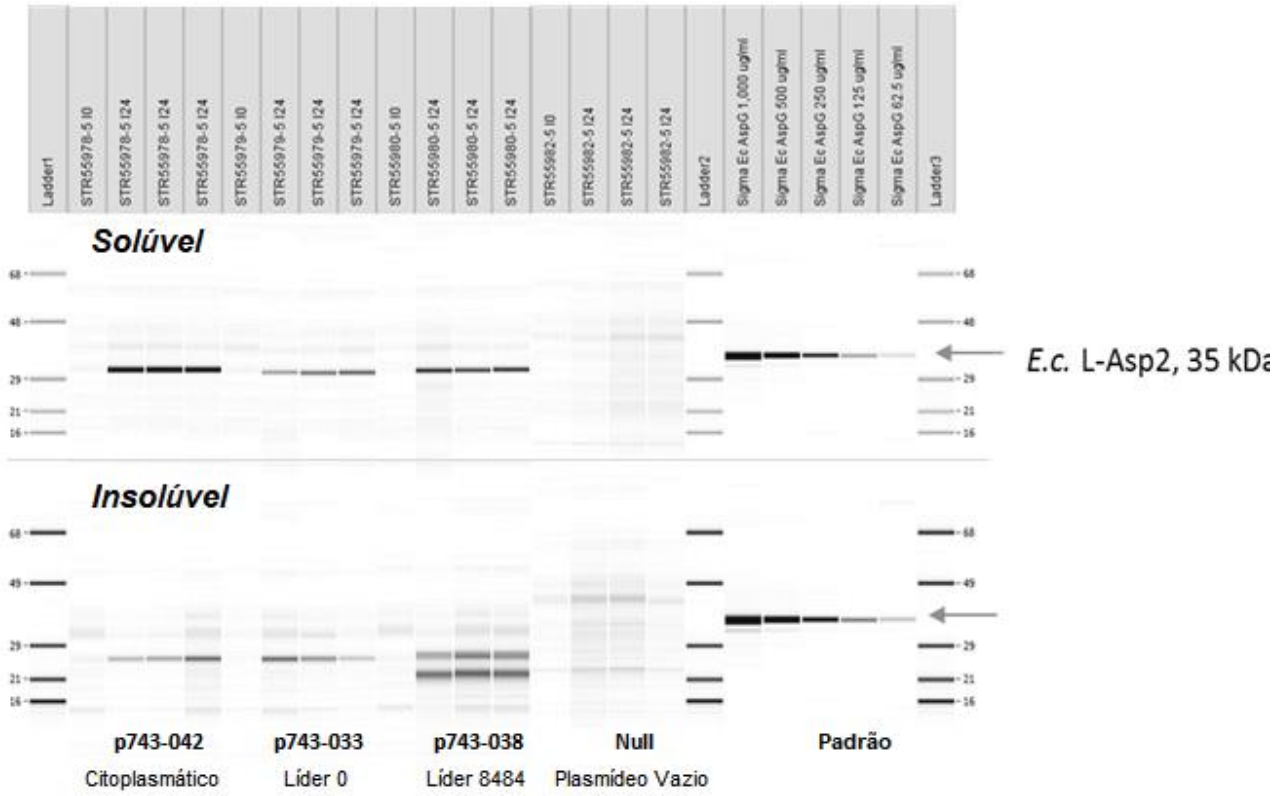

Ladder1
STR555978-5 I0
STR555978-5 I24
STR555978-5 I24
STR555978-5 I24
STR555979-5 I0
STR555979-5 I24
STR555979-5 I24
STR555979-5 I24
STR555980-5 I0
STR555980-5 I24
STR555980-5 I24
STR555980-5 I24
STR555982-5 I0
STR555982-5 I24
STR555982-5 I24
STR555982-5 I24
Ladder2
Sigma Ec AspG 1,000 ug/ml
Sigma Ec AspG 500 ug/ml
Sigma Ec AspG 250 ug/ml
Sigma Ec AspG 125 ug/ml
Sigma Ec AspG 62.5 ug/ml
Ladder3
Solúvel
Insolúvel
E.c. L-Asp2, 35 kDa
p743-042
Citoplasmático
p743-033
Líder 0
p743-038
Líder 8484
Null
Plasmídeo Vazio
Padrão

FIG. 7

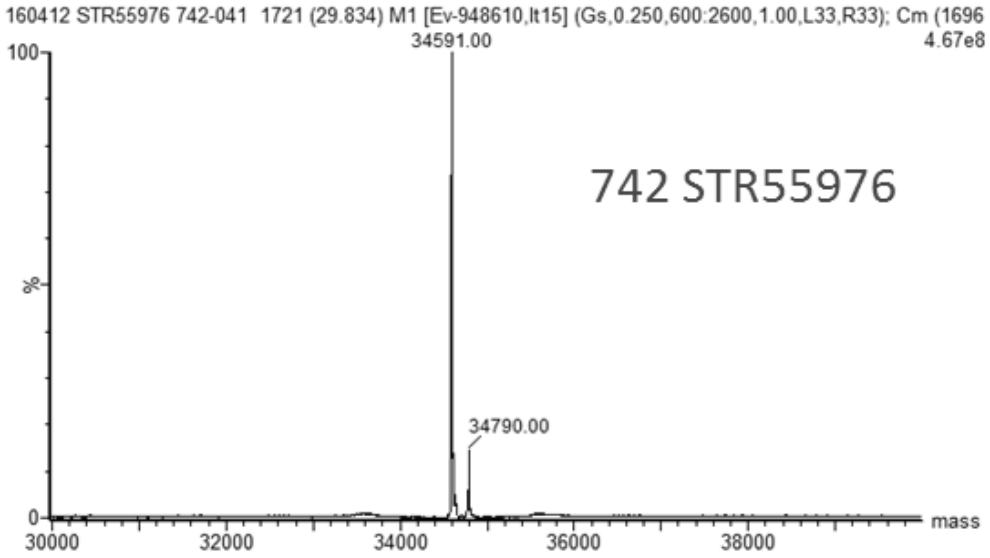

160412 STR55976 742-041 1721 (29.834) M1 [Ev-948610,It15] (Gs,0.250,600:2600,1.00,L33,R33); Cm (1696
34591.00
4.67e8
100
%
0
742 STR55976
34790.00
mass
30000
32000
34000
36000
38000

FIG. 8

RESUMO

SEQUÊNCIAS LÍDER BACTERIANAS PARA A EXPRESSÃO DE PROTEÍNA PERIPLÁSMICA

São providas aqui sequências líder bacterianas para a expressão periplásmica de proteínas heterólogas, proteínas de fusão compreendendo as sequências líder bacterianas e métodos para a expressão das mesmas.

Este anexo apresenta o código de controle da listagem de sequências biológicas.

Código de Controle

Campo 1



Campo 2



Outras Informações:

- Nome do Arquivo: 14_PCT.US2018.056376_LISTAGEM DE SEQUÊNCIAS.txt
- Data de Geração do Código: 27/04/2020
- Hora de Geração do Código: 19:35:10
- Código de Controle:
    - Campo 1: 51E31EE607593006
    - Campo 2: 44A1AE2A6E655622