

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

(21) PI0513826-4 A2



(22) Data de Depósito: 26/07/2005  
(43) Data da Publicação: 22/06/2010  
(RPI 2059)

(51) Int.Cl.:  
C12P 21/06

**(54) Título: PROCESSO PARA EXPRESSÃO DE  
PROTEÍNA MELHORADA ATRAVÉS DE  
ENGENHARIA DE CEPA**

(30) Prioridade Unionista: 26/07/2004 US 60/591,489

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

(72) Inventor(es): Charles H. Squires, Hongfan Jin, Thomas Martin Ramseier

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

(86) Pedido Internacional: PCT US2005026390 de 26/07/2005

(87) Publicação Internacional: WO 2006/014899de 09/02/2006

**(57) Resumo:** A presente invenção refere-se a um processo para melhorar os níveis de produção de proteínas ou peptídeos recombinantes ou melhorar o nível de proteínas ou peptídeos recombinantes ativos expressos em células hospedeiras. A invenção é um processo de comparar dois perfis genéticos de uma célula que expressa uma proteína recombinante e modificar a célula para alterar a expressão de um produto de gene que é sobre-regulado em resposta à expressão da proteína recombinante. O processo pode melhorar a produção de proteína ou pode melhorar a qualidade da proteína, por exemplo, aumentando a solubilidade de uma proteína recombinante.

Descritivo da Patente de Invenção para "PROCESSO PARA EXPRESSÃO DE PROTEÍNA MELHORADA ATRAVÉS DE ENGENHARIA DE CEPA".

REFERÊNCIA CRUZADA AO PEDIDO DE PATENTE RELACIONADO

5 Este pedido de patente reivindica prioridade ao Pedido Provisório US Nº 60/591.489, depositado em 26 de julho de 2004.

CAMPO DA INVENÇÃO

A presente invenção refere-se ao campo de produção de proteína, e em particular é um processo para melhorar os níveis de produção de proteínas ou peptídeos recombinantes ou melhorar o nível de proteínas ou peptídeos recombinantes ativos expressos em células hospedeiras.

ANTECEDENTES

Mais que 155 proteínas e peptídeos recombinantemente produzidos foram aprovados pela U.S. Food and Drug Administration (FDA) para o uso como fármacos e vacinas de biotecnologia, com outros 370 em experimentações clínicas. Diferente das terapêuticas de molécula pequena que são produzidas através de síntese química, as proteínas e peptídeos são mais eficazmente produzidos em células vivas. Em muitos casos, a célula ou organismo foi geneticamente modificado para produzir ou aumentar a produção da proteína.

Quando uma célula é modificada para produzir quantidades grandes de uma proteína alvo, a célula é colocada sob tensão e freqüentemente reage induzindo ou suprimindo outras proteínas. A tensão que uma célula hospedeira sofre durante a produção de proteínas recombinantes pode aumentar a expressão, por exemplo, de proteínas específicas ou co-fatores para causar degradação da proteína recombinante sobre-expressada. A expressão aumentada de proteínas compensatórias pode ser contra-produtiva à meta de expressar níveis altos de proteína recombinante ativa, de comprometimento total. Expressão diminuída ou carência de expressão adequada de outras proteínas pode causar desdobramento e agregação da proteína recombinante. Embora seja conhecido que uma célula sob tensão alterará seu perfil de expressão de proteína, não é conhecido em qualquer exemplo dado

que proteínas específicas serão sobre-reguladas ou sub-reguladas.

#### MICROARRANJOS

Tecnologia de microarranjo pode ser usada para identificar a presença e nível de expressão de um número grande de polinucleotídeos em um ensaio simples. Ver por exemplo patente U.S. Nº 6.040.138, depositada em 15 de setembro de 1995, patente U.S. Nº 6.344.316, depositada em 25 de junho de 1997, patente U.S. Nº 6.261.776, depositada em 15 de abril de 1999, patente U.S. Nº 6.403.957, depositada em 16 de outubro de 2000, patente U.S. Nº 6.451.536, depositada em 27 de setembro de 2000, patente U.S. Nº 6.532.462, depositada em 27 de agosto de 2001, patente U.S. Nº 6.551.784, depositada em 9 de maio de 2001, patente U.S. Nº 6.420.108, depositada em 9 de fevereiro de 1998, patente U.S. Nº 6.410.229, depositada em 14 de dezembro de 1998, patente U.S. Nº 6.576.424, depositada em 25 de janeiro de 2001, patente U.S. Nº 6.687.692, depositada em 2 de novembro de 2000, patente U.S. Nº 6.600.031, depositada em 21 de abril de 1998 e patente U.S. Nº 6.567.540, depositada em 16 de abril de 2001, todas atribuídas a Affymetrix, Inc.

Patente U.S. Nº 6.607.885 de E. I. duPont de Nemours e Co. descreve métodos para perfilar e identificar as alterações de expressão de gene após submeter uma célula bacteriana a expressão alterando as condições mediante comparação de uma primeira e segunda medição de microarranjo.

Wei et al. usou uma análise de microarranjo para investigar perfis de expressão de gene de *E. coli* com indução de gene de lac (Wei Y., et al. (2001) High-density microarray-mediated gene expression profiling of *Escherichia coli*. *J Bacteriol.* 183(2):545-56). Outros grupos também investigaram os perfis transcripcionais regulados após mutação de genes endógenos ou deleção de genes reguladores (Sabina, J., et al (2003) Interfering with Different Steps of Protein Synthesis Explored by Transcriptional Profiling of *Escherichia coli* K-12 *J Bacteriol.* 185:6158-6170; Lee JH (2003) Global analyses of transcriptomes and proteomas of a parent strain and an L-threonine-overproducing mutant strain. *J Bacteriol.* 185(18):5442-51; Kabir

MM, et al. (2003) Gene expression patterns for metabolic pathway in pgi knockout *Escherichia coli* with and without phb genes based on RT-PCR *J Biotechnol.* 105(1-2):11-31; Eymann C., et al. (2002) *Bacillus subtilis* functional genomics: global characterization of the stringent response by proteoma and transcriptome analysis. *JBacteriol.* 184(9):2500-20.

Gill et al. descreve o uso de tecnologia de microarranjo para identificar alterações na expressão de genes relacionados à tensão em *E. coli* após expressão de proteínas de fusão recombinantes de cloranfenicol acetil-transferase (Gill et al. (2001) Genomic Analysis of High-Cell-Density Recombinant *Escherichia coli* Fermentation and "Cell Conditioning" for Improved Recombinant Protein Yield *Biotech. Bioengin.* 72:85-95). O perfil de transcrição de gene de tensão, compreendendo apenas 16% do genoma total, em densidade de célula alta foi usado para avaliar estratégias de "condicionamento de célula" para alterar os níveis das chaperonas, proteases e outras proteínas intracelulares antes da sobre-expressão da proteína recombinante. As estratégias para "condicionamento" envolveram manipulação farmacológica das células, incluindo através de tratamentos de ditiotreitol e etanol.

Asai et al. descreve o uso de análise de microarranjo para identificar genes alvos ativados por sobre-expressão de certos fatores sigma que são tipicamente induzidos após tensões de célula (Asai K., et al. (2003) DNA microarray analysis of *Bacillus subtilis* sigma factors of extracytoplasmic function family. *FEMS Microbiol. Lett.* 220(1):155-60). Células que sobre-expressam fatores sigma como também genes repórteres ligados aos promotores de fator sigma foram usadas para mostrar indução de gene regulada por tensão.

Choi et al. descreveu a análise e sobre-regulação de genes metabólicos que são sub-regulados em culturas de batelada de alta densidade de *E. coli* que expressa proteína de fusão de fator de crescimento similar à insulina humana (IGF-I<sub>f</sub>) (Choi et al. (2003) Enhanced Production of Insulin-Like Growth Factor I Fusion Protein in *Escherichia coli* by Coexpression of the Down-Regulated Genes Identified by Transcriptome Profiling App. *Envir. Microbio.* 69:4737-4742). O foco deste trabalho estava nas alterações me-

tabólicas que ocorrem durante as condições de alta densidade após indução de proteína. Genes que foram sub-regulados após indução de produção de proteína recombinante durante condições de crescimento de densidade alta foram identificados e genes metabólicos específicos que tinham sido 5 sub-regulados foram expressados em células produzindo IGF-I<sub>f</sub> recombinante. O trabalho mostrou que aumentando a produção metabólica de certas bases de nucleotídeo e aminoácidos poderia aumentar a produção de proteína e aquelas taxas de crescimento poderiam ser modificadas aumentando a expressão de uma molécula transportadora metabólica sub-regulada. Estas 10 estratégias foram projetadas para alterar o ambiente celular para reduzir tensões metabólicas associadas à produção de proteína em geral ou com cultura de densidade alta.

### DEGRADAÇÃO DE PROTEÍNA

Degradação indesejada de proteína recombinante apresenta um 15 obstáculo ao uso eficiente de certos sistemas de expressão. A expressão de proteínas exógenas freqüentemente induz respostas de tensão em células hospedeiras que podem ser, por exemplo, defesas naturais a uma fonte de carbono limitada. Todas as células contêm um número grande de genes capazes de produzir proteínas degradativas. Não é possível prognosticar quais 20 proteases serão reguladas por um hospedeiro dado em resposta à expressão de uma proteína recombinante particular. Por exemplo, as bactérias *P. fluorescens* contém até 200 proteases e proteínas relacionadas à protease.

No citoplasma de *E. coli*, a proteólise é em geral realizada por um grupo de proteases e moléculas de co-fator. Etapas de degradação mais 25 precoces são realizadas por cinco Hsps ATP-dependentes: Lon/La FtsH/HflB, ClpAP, ClpXP e ClpYQ/HslUV (Gottesman S (1996) Proteases and their targets in *Escherichia coli*. *Annu. Rev. Genet.* 30:465-506). Junto com FtsH (uma protease associada à membrana interna, o sítio ativo desta opõem-se ao citoplasma), ClpAP e ClpXP são responsáveis pela degradação de proteínas 30 modificadas em seus términos carboxila pela adição da cauda de desestabilização não-polar AANDENYALAA (Gottesman S, et al. (1998) The ClpXP and ClpAP proteases degrade proteins with carboxyl-terminal peptide tails added

by the SsrA-tagging system. *Genes Dev.* 12:1338-1347; Herman C, et al. (1998) Degradation of carbóxi-terminal-tagged cytoplasmic proteins by the *Escherichia coli* protease HflB (FtsH). *Genes Dev.* 12:1348-1355).

Vários métodos foram tomados para evitar degradação durante a produção de proteína recombinante. Um método é produzir cepas hospedeiras que carregam mutações em um gene de protease. Baneyx e Georgiou, por exemplo, utilizaram uma cepa deficiente de protease para melhorar o rendimento de uma proteína de fusão de proteína A-β-lactamase (Baneyx F, Georgiou G. (1991) Construction and characterization of *Escherichia coli* strains deficient in multiple secreted proteases: protease III degrades high-molecular-weight substrates in vivo. *J Bacteriol* 173: 2696-2703). Park et al. usou um método mutacional similar para melhorar a atividade da proteína recombinante 30% comparada com a cepa de origem de *E. coli* (Park S. et al. (1999) Secretory production of recombinant protein by a high cell density culture of a protease negative mutant *Escherichia coli* strain. *Biotechnol. Progr.* 15:164-167). Patentes U.S. Nºs 5.264.365 e 5.264.365 descrevem a construção de *E. coli* deficiente de protease, particularmente múltiplas cepas deficientes de protease, para produzir polipeptídeos proteoliticamente sensíveis. Publicação de PCT Nº WO 90/03438 descreve a produção de cepas de *E. coli* que incluem cepas deficientes de protease ou cepas incluindo um inibidor de protease. Similarmente, Publicação de PCT Nº WO 02/48376 descreve cepas de *E. coli* deficientes em proteases DegP e Prc.

#### DOBRAMENTO DE PROTEÍNA

Outro obstáculo principal na produção de proteínas recombinantes em células hospedeiras é que a célula freqüentemente não é adequadamente equipada para produzir proteína solúvel ou ativa. Embora a estrutura primária de uma proteína seja definida por sua seqüência de aminoácido, a estrutura secundária é definida pela presença de hélices alfas ou folhas beta, e a estrutura ternária por ligações covalentes entre extensões de proteína adjacentes, como ligações de dissulfeto. Ao expressar proteínas recombinantes, particularmente em grande escala, a estrutura secundária e terciária da própria proteína é de importância crítica. Qualquer alteração significativa

em estrutura de proteína pode render uma molécula funcionalmente inativa, ou uma proteína com atividade biológica significativamente reduzida. Em muitos casos, uma célula hospedeira expressa moduladores de dobramento (FMs) que são necessários para produção apropriada de proteína recombinante ativa. Porém, nos níveis altos de expressão em geral requeridos para produzir produtos de biotecnologia utilizáveis, economicamente satisfatórios, uma célula freqüentemente não pode produzir modulador ou moduladores de dobramento nativo suficiente para processar a proteína recombinante.

Em certos sistemas de expressão, superprodução de proteínas exógenas pode ser acompanhada por seu desdobramento e segregação em agregados insolúveis. Em células bacterianas, estes agregados são conhecidos como corpos de inclusão. Em *E. coli*, a rede de moduladores/ chaperonas de dobramento inclui a família de Hsp70. A chaperona de Hsp70 principal, DnaK, impede agregação da proteína e suporta eficazmente o redobramento das proteínas estragadas. A incorporação de proteínas de choque térmico em agregados de proteína pode facilitar a desagregação. Porém, proteínas processadas para corpos de inclusão podem, em certos casos, ser restabelecidas através de processamento adicional da fração insolúvel. Proteínas encontradas em corpos de inclusão tipicamente têm que ser purificadas através de etapas múltiplas, incluindo desnaturação e renaturação. Processos de renaturação típicos para proteínas alvejadas por corpo de inclusão envolvem tentativas de dissolver o agregado em desnaturante concentrado e remoção subsequente do desnaturante por diluição. Agregados são freqüentemente formados novamente neste estágio. O processamento adicional soma custo, não há nenhuma garantia que o redobramento *in vitro* renderá produto biologicamente ativo, e as proteínas restabelecidas podem incluir quantidades grandes de impurezas de fragmento.

Um método para reduzir agregação de proteína é através de engenharia de fermentação, mais comumente reduzindo a temperatura de cultivo (ver Baneyx F (1999) *In vivo folding of recombinant proteins in Escherichia coli*. In *Manual of Industrial Microbiology and Biotechnology*, Ed. Davies et al. Washington, DC: American Society for Microbiology ed. 2:551-565 e

referências nele). A mais recente realização que dobramento de proteína *in vivo* é assistida por chaperonas moleculares, que promove a isomerização apropriada e alvejamento celular de outros polipeptídeos transientemente interagindo com intermediários de dobramentos, e por foldases, que acelera as etapas limitativas de taxa ao longo da via de dobramento, forneceu métodos adicionais para combater o problema de formação de corpo de inclusão (ver por exemplo Thomas JG et al. (1997). Molecular chaperones, folding catalysts and the recovery of active recombinant proteins from *E. coli*: to fold or to refold. *Appl Biochem Biotechnol*, 66:197-238).

Em certos casos, a sobre-expressão de chaperonas foi observada aumentar os rendimentos solúveis de proteínas propensas à agregação (ver Baneyx, F. (1999) Recombinant Protein Expression in *E. coli* *Curr. Opin. Biotech.* 10:411-421 e referências nele). O processo não parece envolver dissolução de corpos de inclusão recombinantes pré-formados mas está relacionado ao dobramento melhorado das cadeias de proteína recentemente sintetizadas. Por exemplo, Nishihara et al. co-expressou *groESL* e *dnaJK/grpE* no citoplasma para melhorar a estabilidade e acumulação de Cryj2 recombinante (um alergênio de pólen de cedro japonês) (Nishihara K, Kanemori M, Kitagawa M, Yanagi H, Yura T. 1998. Chaperone coexpression plasmids: differential and synergistic roles of DnaK-DnaJ-GrpE and GroEL-GroES in assisting folding of an allergen of Japanese cedar pollen, Cryj2, in *Escherichia coli*. *Appl. Environ. Microbiol.* 64:1694). Lee e Olins também co-expressaram *GroESL* e *DnaK* e aumentaram a acumulação de procolagenase humana em dez vezes (Lee S, Olins P. 1992. Effect of overproduction of heat shock chaperones *GroESL* and *DnaK* on human procollagenase production in *Escherichia coli*. *JBC* 267:2849-2852). O efeito benéfico associado a um aumento na concentração intracelular destas chaperonas parece altamente dependente da natureza da proteína superproduzida, e o sucesso não é de forma alguma garantido.

Uma necessidade existe por processos para desenvolvimento de cepas hospedeiras que mostram produção, atividade ou solubilidade melhoradas de proteína ou peptídeo recombinante para reduzir os custos de fa-

bricação e aumentar o rendimento dos produtos ativos.

É, portanto, um objetivo da invenção fornecer processos para melhorar expressão de proteína recombinante em um hospedeiro.

5 É um outro objetivo da invenção fornecer processos que aumentam os níveis de expressão em células hospedeiras expressando proteínas ou peptídeos recombinantes.

É outro objetivo da invenção fornecer processos para aumentar os níveis de proteína solúvel feita em sistemas de expressão recombinante.

10 É ainda outro objetivo da invenção fornecer processos para aumentar os níveis de proteína ativa feita em sistemas de expressão recombinante.

#### SUMÁRIO

Um processo é fornecido para melhorar a expressão de uma proteína ou peptídeo recombinante compreendendo:

15 i) expressar a proteína ou peptídeo recombinante em uma célula hospedeira;

ii) analisar um perfil genético da célula e identificar um ou mais produtos de gene endógeno que são sobre-regulados na expressão ou sobre-expressão da proteína ou peptídeo recombinante; e

20 iii) alterar a expressão de um ou mais produtos de gene endógeno identificados geneticamente modificando a célula.

O processo pode fornecer expressão melhorada como medida por rendimentos melhorados de proteína, ou pode melhorar a recuperação da proteína ativa, por exemplo aumentando a solubilidade da proteína recombinante expressa, ou uma proteína ou peptídeo relacionado.

Usando este processo, pode ser determinado quais das muitas proteínas celulares são "escolhidas" pela célula para compensar a expressão da proteína recombinante estranha, e esta informação pode levar ao desenvolvimento de sistemas de expressão de proteína mais eficazes. Por exemplo, 30 é conhecido que, tipicamente, uma célula seletivamente sobre-regulará uma ou mais proteases para degradar uma proteína recombinante sobre-expres-sada. Porém, pode não ser prognosticado com antecedência qual(is) prote-

ase(s) a célula sobre-regulará para compensar a tensão causada por qualquer proteína recombinante dada. Análise do perfil genético da célula por microarranjo ou tecnologia equivalente pode identificar quais proteases são sobre-reguladas em uma célula dada em resposta à produção de proteína exógena. Esta informação é depois usada para geneticamente modificar a célula para diminuir a expressão destas proteases particulares, poupando outras proteínas que são úteis ou até mesmo necessárias para homeostase de célula.

Como outro exemplo, uma célula pode seletivamente sobre-regular um ou mais moduladores ou co-fatores de dobramentos para aumentar a capacidade de dobramento ou solubilidade da proteína recombinante. Novamente, não pode ser prognosticado com antecedência quais moduladores ou co-fatores de dobramentos serão selecionados em um sistema dado para ajudar no processamento de uma proteína recombinante específica. Analisando o perfil genético através de microarranjo ou tecnologia equivalente permite a identificação dos moduladores ou co-fatores de dobramentos que foram sobre-regulados. Com base nesta informação, a célula é geneticamente modificada para aumentar a expressão dos moduladores ou co-fatores de dobramentos selecionados preferidos pela célula para a proteína recombinante dada. Esta modificação pode aumentar o percentual de proteína ativa se restabelecida, enquanto minimizando o impacto prejudicial em homeostase de célula.

Portanto, o rendimento e/ou atividade e/ou solubilidade da proteína recombinante pode(m) ser aumentado(s) ou modificando o organismo hospedeiro ou aumentando ou diminuindo a expressão de uma proteína compensatória (isto é uma proteína que é sobre-regulada em resposta à tensão de célula dada) de uma maneira que é seletiva e que deixa outros mecanismos benéficos inteiros da célula.

O processo pode ser usado iterativamente até a expressão de proteína recombinante ativa ser otimizado. Por exemplo, usando o processo descrito acima, a célula ou organismo hospedeiro é geneticamente modificado para sobre-regular, sub-regular, knock-in ou knock-out uma ou mais

proteínas compensatórias identificadas. A célula ou organismo hospedeiro assim modificado pode depois ser cultivado para expressar a proteína recombinante, ou uma proteína ou peptídeo relacionado, e proteínas compensatórias adicionais identificadas por meio de microarranjo ou análise equivalente. A célula ou organismo hospedeiro modificado é depois de novo geneticamente modificado para sobre-regular, sub-regular, knock-in ou knock-out as proteínas compensatórias adicionais selecionadas. Este processo pode ser iterado até uma célula ou organismo hospedeiro ser obtido que exibe expressão máxima de ativo e/ou proteína solúvel sem enfraquecimento 5 devido do organismo ou célula hospedeira. Estas etapas por exemplo podem ser repetidas por exemplo, uma, duas, três, quatro, cinco, seis, sete, oito, 10 nove, ou dez ou mais vezes.

Em outra modalidade, o processo também compreende: iv) expressar a proteína ou peptídeo recombinante em uma célula geneticamente 15 modificada. Em ainda outra modalidade, o processo também compreende: v) analisar um segundo perfil genético da célula geneticamente modificada expressando proteína ou peptídeo recombinante e identificando um ou mais produtos de gene adicionais que são diferencialmente expressos na célula modificada que expressa a proteína ou peptídeo recombinante. Em uma outra 20 modalidade, o processo adicionalmente compreende: vi) alterar a expressão de um ou mais produtos de gene adicionais identificados para fornecer uma célula modificada dupla. Opcionalmente, a proteína ou peptídeo recombinante, ou uma proteína ou peptídeo relacionado, pode ser expressado na célula modificada dupla. Os produtos de gene diferencialmente regulados identificados na célula modificada podem ser sobre ou sub-regulados quando 25 comparados à célula hospedeira ou quando comparados à célula modificada que não expressa proteína ou peptídeo recombinante.

Em ainda outra modalidade, o processo também compreende: iv) analisar um segundo perfil genético de uma célula geneticamente modificada expressando a proteína ou peptídeo recombinante e identificar um ou mais 30 produtos de gene adicionais que são diferencialmente expressos na célula modificada que não está expressando a proteína ou peptídeo recombinante.

Em uma outra modalidade, o processo adicionalmente compreende: v) alterar a expressão de um ou mais produtos de gene adicionais identificados na célula modificada para fornecer uma célula modificada dupla. Os produtos de gene diferencialmente regulados identificados na célula modificada podem 5 ser sobre ou sub-regulados quando comparados à célula ou organismo hospedeiro ou quando comparados à célula modificada que não expressa a proteína ou peptídeo recombinante.

Em uma modalidade específica, um processo é fornecido para melhorar a expressão de uma proteína ou peptídeo recombinante compreendendo: i) expressar a proteína ou peptídeo recombinante em uma célula hospedeira; ii) analisar um perfil genético da célula e identificar pelo menos uma protease que é sobre-regulada quando a proteína ou peptídeo recombinante é expressado; e iii) alterar a expressão de uma protease identificada geneticamente modificando a célula ou organismo hospedeiro para reduzir a expressão da protease sobre-regulada. Em uma outra modalidade, o processo compreende alterar a expressão de pelo menos uma segunda protease identificada na célula modificada para fornecer uma célula modificada de protease dupla. Em outra modalidade, o processo também compreende: iv) expressar a proteína ou peptídeo recombinante, ou uma proteína ou peptídeo 15 relacionado, em uma célula modificada de protease. Em outra modalidade, o processo também compreende analisar um segundo perfil genético da célula modificada de protease para identificar um ou mais produtos de gene adicionais que são diferencialmente expressos na célula modificada.

Em outra modalidade, um processo é fornecido para melhorar a expressão de uma proteína ou peptídeo recombinante compreendendo: i) expressar a proteína ou peptídeo recombinante em uma célula hospedeira; ii) analisar um perfil genético da célula e identificar pelo menos um modulador de dobramento sobre-regulado (FM) que é sobre-regulado após sobre-expressão da proteína ou peptídeo recombinante; e iii) alterar a expressão de pelo 20 menos um modulador de dobramento identificado geneticamente modificando a célula para fornecer uma célula modificada de FM. Em uma outra modalidade, o processo compreende alterar a expressão de pelo menos um se-

gundo modulador de dobramento identificado na célula modificada para fornecer uma célula modificada de FM duplo. Em outra modalidade, o processo também compreende: iv) expressar a proteína ou peptídeo recombinante, ou uma proteína ou peptídeo relacionado, em uma célula modificada de FM. Em 5 outra modalidade, o processo também compreende analisar um segundo perfil genético da célula modificada de FM para identificar um ou mais produtos de gene adicionais que são diferencialmente expressos na célula modificada.

O termo "perfil genético" como aqui usado é significado para incluir uma análise de genes em um genoma, mRNA transscrito de genes no genoma (ou o cDNA equivalente), produtos de transcrição que foram modificados por uma célula como variantes de encaixe de genes em sistemas eucarióticos, ou proteínas ou peptídeos translados de genes em um genoma, incluindo proteínas que são modificadas pela célula ou transladada de variantes de encaixe de mRNA transladado do genoma. Um perfil genético é significado incluir mais de um gene ou produto de gene, e tipicamente inclui um grupo de pelo menos 5, 10, 50, 100 ou mais genes ou produtos de gene que são analisados.

Em uma modalidade, o perfil genético analisado pode ser um 20 perfil de transcriptoma, isto é um perfil dos produtos de transcrição de genes do genoma. O processo pode incluir analisar o perfil de transcriptoma usando um microarranjo ou tecnologia equivalente. Nesta modalidade, o microarranjo pode incluir os pares de ligação para pelo menos uma porção do transcriptoma da célula hospedeira, e tipicamente inclui amostras de pares de ligação 25 para os produtos de gene de pelo menos 50% do genoma do organismo. Mais tipicamente, o microarranjo inclui amostras de pelo menos 80%, 90%, 95%, 98%, 99% ou 100% dos pares de ligação para produtos de gene no genoma da célula hospedeira.

Em uma modalidade separada, o microarranjo pode incluir um 30 subconjunto selecionado de pares de ligação para genes ou produtos de gene que representam classes de produtos que são afetadas pela expressão de proteína recombinante. Exemplos não-limitativos incluem proteases putativas

ou conhecidas, co-fatores de proteases ou proteínas similares à protease; moduladores de dobramentos, moduladores co-fatores de dobramentos ou proteínas que podem melhorar o dobramento ou solubilidade da proteína; fatores de transcrição; proteínas envolvidas em estabilidade de ácido nucléico ou iniciação translacional; cinases; receptores extracelulares ou intracelulares; enzimas metabólicas; co-fatores metabólicos; proteínas de envelope; fatores de sigma; proteínas ligadas à membrana; proteínas de transmembrana; proteínas associadas à membrana e genes de manutenção. O perfil genético pode ser analisado medindo a ligação dos genes expressos da célula hospedeira que expressa a proteína ou peptídeo recombinante para o microarranjo. O perfil de transcriptoma pode também ser analisado usando ensaios de não-microarranjo como ensaios de mancha, incluindo ensaios de northern blot, ou colunas revestidas com pares de ligação.

Em outra modalidade, o perfil genético analisado pode ser um perfil de proteoma, isto é um perfil das proteínas produzidas de genes em um organismo dado. O processo pode incluir analisar o perfil de proteoma usando, por exemplo, eletroforese bidimensional. Técnicas como espectrometria de massa em combinação com ferramentas de separação como eletroforese em gel bidimensional ou cromatografia líquida multidimensional, podem também ser usadas no processo. Em eletroforese bidimensional, as proteínas separadas podem incluir proteínas de pelo menos 10% do proteoma do organismo. Mais tipicamente, proteínas de pelo menos 20%, 30%, 40%, 60%, 80% ou 90% das proteínas no proteoma da célula hospedeira são separados e analisados através de técnicas como tingimento de proteínas e/ou espectrometria de massa.

Em modalidade adicional, o perfil de proteoma é analisado usando espectrometria de massa. Há várias técnicas relacionadas que usam cromatografia líquida (LC) acoplada à espectrometria de massa (MS) e espectrometria de massa de tandem (MS/MS) para identificar proteínas e medir sua abundância relativa. Freqüentemente, uma amostra é marcada com um marcador de isótopo pesado que permite a comparação com a outra amostra sem alterar as propriedades químicas. Por exemplo, em uma amostra a cis-

teína de aminoácido pode ser marcada com um marcador que contém oito átomos de hidrogênio. A outra amostra é marcada com um marcador que contém oito átomos de deutério ("pesado") no lugar de (+8 Dáltons). Os dados de MS podem ser usados para encontrar pares de peptídeos de 8 Dáltons separadamente e quantificar a diferença. Dados de MS/MS dos mesmos peptídeos fornecem uma aproximação de seqüência primária, e a proteína ID. Outros experimentos identificam as proteínas *in vivo* mediante cultivo das células com aminoácidos "pesados". Estes tipos de técnicas podem ser usados para identificar milhares de proteínas em um experimento simples e estimar a abundância relativa se presente em ambas as amostras (ver Goodlett DR e Aebersold RH (2001). Mass Spectrometry in Proteomics. Chem Rev 101:269-295). ICAT é um tipo de MS/MS, representa Marcadores de Afinação Codificados de Isótopo (ver Gygi SP, Rist B, Gerber SA, Turecek F, Gelb MH, e Aebersold RH (1999). Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. Nat Biotech 17:994-999).

Em outra modalidade, o processo pode incluir analisar o perfil de proteoma usando, por exemplo, um microarranjo. Nesta modalidade, o arranjo pode incluir os pares de ligação para pelo menos uma porção das proteínas expressas pela célula hospedeira sob condições de crescimento apropriadas, e tipicamente inclui os pares de ligação para proteínas de pelo menos 10% do proteoma do organismo. Mais tipicamente, o microarranjo inclui os pares de ligação para proteínas de pelo menos 20%, 30%, 40%, 60%, 80% ou 90% das proteínas no proteoma da célula hospedeira. Os pares de ligação podem ser anticorpos que podem ser fragmentos de anticorpo como fragmentos de anticorpo de cadeia simples. Em uma modalidade separada, o microarranjo pode incluir os pares de ligação para um subconjunto selecionado de proteínas do proteoma, incluindo, por exemplo, proteínas de protease putativa ou moduladores de dobramentos putativos. O microarranjo pode tipicamente também incluir um conjunto de pares de ligação para proteínas que são usadas como controles. O perfil genético pode ser analisado medindo a ligação das proteínas da célula hospedeira que expressa a proteína ou peptídeo recombinante aos pares de ligação no microarranjo. O perfil de proteoma pode

também ser analisado em um formato de ensaio padrão, como um ensaio de Elisa ou um ensaio de western blot padrão.

As amostras no perfil genético podem ser analisadas individualmente ou agrupadas em agrupamentos. Os agrupamentos podem tipicamente ser agrupados por similaridade em expressão de gene. Em modalidades particulares, os agrupamentos podem ser agrupados como genes que são sobre-regulados a uma extensão similar ou genes que são sub-regulados a uma extensão similar.

O gene sobre-regulado identificado é tipicamente identificado 10 comparando um perfil genético da célula hospedeira que expressa a proteína ou peptídeo recombinante a um perfil genético da célula hospedeira que não expressa a proteína ou peptídeo recombinante. Em uma outra modalidade, uma célula hospedeira que expressa uma proteína homóloga é analisada para a primeira proteína recombinante.

15 O genoma da célula hospedeira que expressa a proteína ou peptídeo recombinante pode ser modificado através de recombinação, por exemplo recombinação homóloga ou recombinação heteróloga. O genoma pode também ser modificado por mutação de um ou mais nucleotídeos em uma estrutura de leitura aberta que codifica um gene, particularmente uma 20 protease identificada. Em outra modalidade, a célula hospedeira é modificada incluindo um ou mais vetores que codificam um inibidor de um gene ou produto de gene identificado, como um inibidor de protease. Em outra modalidade, a célula hospedeira é modificada por inibição de um promotor que pode ser um promotor nativo. Em uma modalidade separada, a célula hospedeira é 25 modificada incluindo um ou mais vetores que codificam um gene, tipicamente um modulador de dobramento ou um co-fator de um modulador de dobramento. Em outra modalidade, a célula hospedeira é modificada intensificando um promotor para um modulador de dobramento identificado ou um co-fator para um modulador de dobramento, incluindo adicionando um promotor e- 30 xógeno ao genoma da célula hospedeira.

A célula hospedeira pode ser qualquer célula capaz de produzir proteína ou peptídeo recombinante. Em uma modalidade, a célula hospedeira

é um procarioto, como uma célula bacteriana incluindo, mas não limitada a uma espécie de *Escherichia* ou uma de *Pseudonaonas*. A célula hospedeira pode ser uma célula de *Pseudomonas* como uma célula de *P. fluorescens*. Em outras modalidades, a célula hospedeira é uma célula de *E. coli*. Em outra 5 modalidade a célula hospedeira é uma célula eucariótica, por exemplo uma célula de inseto, incluindo mas não limitado a uma célula de uma espécie de *Spodoptera*, *Trichoplusia*, *Drosophila* ou *Estignaene*, ou uma célula mamífera, incluindo mas não limitada a uma célula murina, uma célula de hamster, uma de macaco, uma de primata ou uma célula humana. Em outra modalidade, a 10 célula hospedeira é uma célula de planta, incluindo, mas não limitada a, uma célula de tabaco, milho, uma célula de uma espécie de *Arabidopsis*, célula de batata ou arroz. Em outra modalidade, um organismo inteiro é analisado no processo, incluindo mas não limitado a um organismo transgênico.

Em uma modalidade, os genes ou produtos de gene compensatórios sobre-regulados identificados são uma ou mais proteases e/ou um ou mais moduladores de dobramentos. Em certas modalidades, um gene ou produto de gene identificado pode também ser uma subunidade de uma protease ou um modulador de dobramento ou um co-fator de uma protease ou um co-fator de um modulador de dobramento. Em uma modalidade, o gene 15 identificado pode ser selecionado de uma peptidase serina, treonina, cisteína, aspártico ou metalo. Em certas outras modalidades, o gene ou produto de gene identificado pode ser selecionado de *hsIV*, *hsIU*, *clpA*, *clpB* e *clpX*. O gene ou produto de gene identificado pode também ser um co-fator de uma protease. Em outra modalidade, o gene ou produto de gene identificado é um 20 modulador de dobramento. Em certas modalidades, o gene ou produto de gene identificado pode ser selecionado de uma proteína de chaperona, uma foldase, uma peptidil prolil isomerase e uma isomerase de ligação de dissulfeto. Em uma modalidade, o gene ou produto de gene identificado pode ser selecionado de *htpG*, *cbpA*, *dnaJ*, *dnaK* e *fkbP*. Em uma modalidade, um gene 25 ou produto de gene homólogo ao gene sobre-regulado identificado é modificado no genoma do hospedeiro.

O processo pode levar à produção aumentada de proteína ou

peptídeo recombinante em uma célula hospedeira, por exemplo, aumentando a quantidade de proteína por grama de proteína de hospedeiro (proteína de célula total) em uma quantidade dada de tempo, ou aumentando a quantidade de comprimento de tempo durante o qual a célula ou organismo está produzindo a proteína recombinante. A produção aumentada pode otimizar a eficiência da célula ou organismo, por exemplo, diminuindo a despesa de energia, aumentando o uso de recursos disponíveis ou diminuindo os requerimentos para suplementos de crescimento em meios de crescimento. A produção aumentada pode também resultar em um nível aumentado de proteína recuperável ou peptídeo, como proteína solúvel, produzido por grama de recombinante ou por grama de proteína de célula hospedeira.

A invenção também inclui uma célula hospedeira recombinante melhorada que é produzida pelo processo reivindicado.

#### BREVE DESCRIÇÃO DOS DESENHOS

Figura 1 é um gráfico de uma comparação de crescimento (densidade óptica com o passar do tempo) de cepas diferentes de *P. fluorescens*. As células foram induzidas com 0,3 M de IPTG em 24 h após inoculação. As cepas são: DC280 que abriga o vetor vazio pDOW1339, DC240 que produz a enzima de nitrilase citoplasmática solúvel e DC271 que produz a hGH periplasmática parcialmente insolúvel. DC206, a cepa parental de DC280, DC240 e DC271, foi incluída como um controle. As amostras foram tiradas em 0 e 4 h de indução pós-IPTG para isolamento de RNA e profilação da expressão de gene, como indicado pelas setas.

Figura 2 é um gráfico de agrupamento hierárquico de todos os genes de cepas de *P. fluorescens* DC280, DC240 e DC271 em 12 agrupamentos em 4 h após IPTG quando comparados em 0 h IPTG (indicado no fundo da figura). Com base no valor e tendência, os genes foram aglomerados e agrupados usando o algoritmo de agrupamento hierárquico de Spotfire DecisionSite. Linhas pontilhadas indicam pontos de dados que foram filtrados devido à qualidade de manchas fraca ou baixo nível de expressão. O eixo geométrico x representa a comparação de cada cepa; o eixo geométrico y representa o valor de expressão com relação 4 h atrás para antes da indução

de IPTG. Todos os FMs identificados são realçados. Agrupamento 7 mostra 2 genes de FM e 2 de subunidade de protease que são altamente expressos em cepa DC271, que superproduzem a proteína de hGH periplásica. Os genes de FM restantes são agrupados no agrupamento 6.

5 Figura 3 é uma análise de agrupamento hierárquico do agrupamento 6 da Figura 2. No agrupamento 8 novo, dois moduladores de dobramentos, DnaK e DnaJ, foram identificados ambos destes mostraram níveis de expressão mais altos para produção de proteína recombinante periplásica similar ao HslVU, CbpA e HtpG previamente identificado. Agrupamento 6  
10 mostra onde o resto dos FMs estão agrupados.

Figura 4 é um diagrama de Venn mostrando a protease sobre-regulada e FMs dos três conjuntos de experimentos na Tabela 5, 6 e 7. Como resumido na Tabela 5, 6 e 7, a lista de genes foi organizada em esquema de Venn para realçar a sobreposição da lista de gene entre os três  
15 conjuntos de experimentos indicados no canto. Para cada gene, a razão de cada experimento foi exibida com 2 como um corte.

Figura 5 é um gráfico da análise de seqüência dos genes *hsIV* (RXF01961) e *hsIU* (RXF01957) de *P. fluorescens* gerados por Artemis. O diagrama de uso de códon (painel do topo) indica que o limite de gene está correto. Este é confirmado pelos melhores homólogos das seqüências de proteína de HslV e HslU de *P. aeruginosa* como indicado em baixo dos genes  
20 de RXF01961 e RXF01957. O diagrama de contagem de qualidade de Phrap mostra que a qualidade de seqüência é boa, isto é a linha de contagem está acima da linha horizontal indicando uma qualidade melhor que 1 erro em 10  
25 kb (painel mediano). As caixas apontadas brancas abertas abaixo dos genes mostram a localização das sondas geradas para uso nos experimentos de microarranjo de DNA.

Figura 6 é uma ilustração esquemática de uma construção de mutante de *hsIU* onde um produto de PCR de aproximadamente 550 bp de  
30 *hsIU* (caixa azul claro) foi ligado no vetor de clonagem TOPO TA2.1 (círculo). O plasmídeo resultante foi transformado em células de *P. fluorescens* competentes e colônias resistentes à canamicina (kan) foram analisadas em PCR

diagnóstica para confirmar a construção de uma mutação de inserção no gene de *hslU*.

Figura 7 é um gráfico de uns ensaios de curva de crescimento comparando cepa do tipo selvagem com de mutante de *hslU* que superproduz hGH ou pbp::hGH em meio de produção de frasco de agitação. As setas indicam pontos de tempo onde foram tiradas as amostras.

Figura 8 é uma imagem de análise de SDS-PAGE das cepas DC271 e DC373 que expressam pbp::hGH. As amostras de DC271 (tipo selvagem, W) e DC373 (mutante de *hslU*, M) foram tiradas logo antes da indução da proteína (0 h) e depois 4 h, 8 h, 24 h, e 30 h após adição de IPTG. Frações solúveis (S) e insolúveis (I) foram preparadas para cada amostra analisada. A produção de hGH não processado e processado é indicada pelas setas. O marcador (Ma) de peso molecular (MW) é mostrado à direita dos géis.

Figura 9 é uma imagem da análise de SDS-PAGE das cepas DC369 e DC372 que expressam hGH no citoplasma. As amostras foram tiradas de DC369 (tipo selvagem, W) e DC372 (mutante de *hslU*, M) logo antes da indução da proteína (0 h) e depois 4 h, 8 h, 24 h, 30 h e 50 h após adição de IPTG. Frações solúveis (S) e insolúveis (I) foram preparadas para cada amostra analisada. A produção de hGH é indicada por uma seta. O marcador (Ma) de peso molecular (MW) é mostrado à direita dos géis.

Figura 10 é um gráfico de curvas de crescimento de cepas que expressam a proteína de fusão de hGH::COP. As cepas incluem: DC369 que expressa hGH apenas (não fundido com COP) como um controle negativo; HJ104, o tipo selvagem que expressa hGH::COP; HJ105, o mutante de *hslU* que expressa hGH::COP.

Figura 11 é um gráfico das medições da atividade de fluorescência verde para cepas que expressam a proteína de fusão de hGH::COP usando um fluorímetro. Cinco OD600 de cultura de célula foram ensaiadas para cada cepa abrigando hGH ou hGH::COP em pontos de tempo diferentes após indução de IPTG. As cepas testadas incluem: DC369 que expressa hGH apenas (não fundido com COP) como um controle negativo; HJ104, o tipo

selvagem que expressa hGH::COP; HJ105, o mutante de *hslU* que expressa hGH::COP. A tabela inserida mostra o percentual de aumento da fluorescência relativa no mutante de *hslU* comparado ao tipo selvagem em pontos de tempo diferentes após indução de IPTG.

5 Figura 12 é uma representação pictórica do processo de medir abundância relativa de mRNA entre duas amostras.

Figura 13 é uma representação da construção de deleção cromossômica de gene de *hslUV* na cepa *pyrF*-negativa. A. Plasmídeo PDOW2050 contém fragmentos de DNA de 505 bp e 634 bp flanqueando o 10 gene de *hslUV*. Uma vez que plasmídeo de suicídio pDOW2050 não pode replicar em *P. fluorescens*, as células resistentes à tetraciclina apenas serão geradas após um evento de recombinação simples em uma das regiões homólogas que integram o plasmídeo inteiro no genoma. B. Células resistentes à tetraciclina contêm o plasmídeo inteiro integrado no genoma. Estas 15 células também contêm o gene de *pyrF* codificado do plasmídeo. Seleção para células que têm o segundo evento recombinante ocorreu através de células de colocação em placa em placas de ágar suplementadas com FOA, que em cepas *pyrF*-positivas, são convertidas em um composto tóxico. C. A 20 cepa de deleção cromossômica foi confirmada por análise de seqüenciação.

Figura 14 é um gráfico de fluorescência relativa com o passar do tempo para medições de atividade de fluorescência verde para as cepas que expressam a proteína de fusão de hGH::COP usando um fluorímetro. Duplicatas foram usadas tanto para cepa de deleção do tipo selvagem (HJ104) como de *hslUV* (HJ117).

25 Figura 15 são imagens de géis de SDS-PAGE de cepas que expressam hGH com ou sem moduladores de dobramentos GrpE-DnakJ. As amostras foram removidas em vários tempos após indução por IPTG (0, 4, 8, 24 e 48 h), normalizadas em OD600 de 20 e lisadas usando EasyLyse. As frações solúveis (S) insolúveis (I) foram separadas em um Critério de BioRad 30 15% de Tris HCl gel de SDS-PAGE e tingido com Coomassie.

#### DESCRÍÇÃO DETALHADA

Um processo é fornecido para melhorar a expressão de uma

proteína ou peptídeo recombinante compreendendo i) expressar a proteína ou peptídeo recombinante em uma célula hospedeira; ii) analisar um perfil genético da célula e identificar um ou mais produtos de gene sobre-regulados endógenos, incluindo uma ou mais proteases ou moduladores de dobramentos que são sobre-regulados na expressão da proteína ou peptídeo recombinante; e iii) alterar a expressão de um ou mais produtos de gene identificados geneticamente modificando a célula. Em outra modalidade, o processo também compreende expressar a proteína ou peptídeo recombinante em uma célula geneticamente modificada. Em outra modalidade, o processo 10 também compreende analisar um segundo perfil genético da célula geneticamente modificada para identificar um ou mais produtos de gene adicionais que são diferencialmente expressos na célula modificada. Em uma outra modalidade, o processo compreende alterar a expressão de pelo menos um segundo produto de gene identificado na célula modificada para fornecer uma 15 célula modificada dupla. O processo pode fornecer expressão melhorada como medido por rendimentos melhorados de proteína, ou pode melhorar a recuperação de proteína ativa, por exemplo aumentando solubilidade da proteína recombinante expressa.

Mais em geral, a invenção inclui um processo para melhorar a 20 expressão de uma proteína ou peptídeo recombinante em uma célula ou organismo hospedeiro compreendendo:

- i) expressar a proteína ou peptídeo recombinante na célula ou organismo hospedeiro recombinante;
- ii) analisar um perfil genético da célula recombinante para identificar 25 um gene compensatório ou produto de gene que é expressado em um nível mais alto na célula recombinante que em um de uma célula hospedeira que não foi modificada para expressar a proteína recombinante ou uma célula recombinante que não está expressando a proteína recombinante; e
- iii) alterar a expressão do gene compensatório ou produto de gene 30 identificado na célula recombinante através de modificação genética para fornecer uma célula recombinante modificada que alcança um aumento na expressão, atividade ou solubilidade de proteína recombinante.

Ao longo do relatório descritivo, quando uma faixa for fornecida, deveria ser entendido que os componentes são significados ser independentes. Por exemplo, uma faixa de 1-6 significa independentemente 1, 2, 3, 4, 5 ou 6.

5 As etapas do processo são descritas em mais detalhe abaixo.

ETAPA I: MODIFICAÇÃO GENÉTICA DE CÉLULA OU ORGANISMO HOSPEDEIRO PARA EXPRESSAR UMA PROTEÍNA OU PEPTÍDEO RE-COMBINANTE EM UMA CÉLULA HOSPEDEIRA

Na primeira etapa do processo, uma célula hospedeira é modificada para ter a capacidade para expressar uma proteína ou peptídeo recombinante. A célula hospedeira pode ser modificada usando qualquer técnica conhecida na técnica. Por exemplo, a proteína recombinante pode ser expressada de um vetor de expressão que é exógeno ao genoma da célula e que é transfeccionado ou transformado na célula. A construção de vetores de expressão como também técnicas para transfecção ou transformação é descrita abaixo. A célula hospedeira pode também ser modificada para expressar uma proteína ou peptídeo recombinante de uma inserção genômica como descrito abaixo. Um gene que codifica a proteína ou peptídeo recombinante pode ser inserido no genoma da célula ou organismo hospedeiro através de técnicas como recombinação homóloga ou heteróloga. Estas técnicas são descritas abaixo.

A proteína ou peptídeo recombinante pode ser expressado sob o controle de um elemento que requer manipulação adicional da célula. Por exemplo, tratamento químico da célula pode ser requerido para iniciar ou intensificar a expressão de proteína ou de peptídeo. Os elementos de promotor e de repressor que governam a expressão de proteínas ou peptídeos recombinantes nas células hospedeiras são descritos abaixo e são bem conhecidos na técnica. Estes podem incluir elementos de promotor com base no promotor "lac" responsivo a IPTG.

30 SELEÇÃO DE UMA CÉLULA OU ORGANISMO HOSPEDEIRO

O processo da invenção pode ser usado em qualquer sistema hospedeiro dado, incluindo de origem eucariótica ou procariótica. O processo

é em geral limitado apenas pela disponibilidade de informação bastante genética para análise de um perfil genético para identificar um gene identificado. Embora seja em geral típico que as seqüências representativas de uma porcentagem grande do genoma estão disponíveis, por exemplo pelo menos 5 50%, 60%, 70%, 80%, 90%, 95%, 98%, 99% ou 100% das seqüências expressas ou encontradas no genoma, transcriptoma ou proteoma, a invenção pode ser praticada usando apenas uma porção das seqüências no genoma, transcriptoma ou proteoma. Em particular, em circunstâncias quando a informação disponível incluir informação sobre um grupo de seqüências relacionadas, como um grupo metabolicamente ligado, apenas uma porção pequena das seqüências representativas do genoma pode ser usada para o processo da invenção. O processo é também não limitado às proteínas recombinantes particulares sendo expressadas, como um aspecto fundamental 10 do processo a capacidade é racional e iterativamente para sistemas de expressão de projeto com base em técnicas para identificar alterações celulares que ocorrem em uma célula hospedeira na expressão de proteínas ou peptídeos recombinantes e para modular a célula hospedeira usando procedimentos conhecidos na técnica.

A célula hospedeira pode ser qualquer célula capaz de produzir 20 proteína ou peptídeo recombinante. Em uma modalidade, a célula hospedeira é uma célula microbiana, isto é, uma célula de uma bactéria, fungo, levedura ou outros eucariotas unicelulares, procariotos e vírus. Os sistemas comumente usados para produzir proteínas ou peptídeos recombinantes incluem certas células bacterianas, particularmente *E. coli*, por causa de seus 25 requerimentos de crescimento relativamente baratos e capacidade de potencial para produzir proteína em culturas de batelada grandes. Levedura é também usada para expressar proteínas e peptídeos biologicamente relevantes, particularmente para propósitos de pesquisa. Sistemas incluem *Saccharomyces cerevisiae* ou *Pichia pastoris*. Estes sistemas são bem caracterizados, fornecem em geral níveis aceitáveis de expressão de proteína total e são 30 comparativamente rápidos e baratos. Sistemas de expressão de célula de inseto também emergiram como uma alternativa para expressar proteínas

recombinantes em forma biologicamente ativa. Em alguns casos, proteínas corretamente dobradas que são pós-translacionalmente modificadas podem ser produzidas. Sistemas de expressão de célula mamífera, como células de ovário de hamster chinês, também foram usados para a expressão de proteínas recombinantes. Em uma escala pequena, estes sistemas de expressão são freqüentemente eficazes. Certos biológicos podem ser derivados de proteínas mamíferas, particularmente em aplicações de saúde animal ou humana. Em outra modalidade, a célula hospedeira é uma célula de planta, incluindo, mas não limitada a, uma célula de tabaco, milho, uma célula de uma espécie de *Arabidopsis*, célula de batata ou de arroz. Em outra modalidade, um organismo multicelular é analisado ou modificado no processo, incluindo mas não limitado a um organismo transgênico. Técnicas para analisar e/ou modificar um organismo multicelular são em geral com base nas técnicas descritas para modificar as células descritas abaixo.

Em uma modalidade, a célula hospedeira pode ser um procarionto como uma célula bacteriana incluindo, mas não limitada a uma espécie de *Escherichia* ou *Pseudomonas*. Células bacterianas típicas são descritas, por exemplo, em "Biological Diversity: Bacteria and Archaeans", a chapter of the On-Line Biology Book, fornecido por Dr MJ Farabee do Estrella Mountain Community College, Arizona, USA em URL: [http://www.emc.maricopa.edu/faculty/farabee/BIOBK/BioBookDiversity\\_2.html](http://www.emc.maricopa.edu/faculty/farabee/BIOBK/BioBookDiversity_2.html).

Em certas modalidades, a célula hospedeira pode ser uma célula de *Pseudomonas*, e pode tipicamente ser uma célula de *P. fluorescens*. Em outras modalidades, a célula hospedeira pode também ser uma célula de *E. coli*. Em outra modalidade a célula hospedeira pode ser uma célula eucariótica, por exemplo uma célula de inseto, incluindo mas não limitada a uma célula de uma espécies de *Spodoptera*, *Trichoplusia*, *Drosophila* ou *Estigmene*, ou uma célula mamífera, incluindo mas não limitada a uma célula murina, uma célula de hamster, um macaco, um primata ou uma célula humana.

Em certas modalidades, a célula hospedeira é uma célula de *Pseudomonad*, e pode ser por exemplo um organismo de *P. fluorescens*.

Em uma modalidade, a célula hospedeira pode ser um membro

de qualquer da taxa bacteriana. A célula pode, por exemplo, ser um membro de qualquer espécie de eubactéria. O hospedeiro pode ser um membro qualquer um da taxa: *Acidobacteria*, *Actinobacteria*, *Aquificae*, *Bacteroidetes*, *Chlorobi*, *Chlamydiae*, *Choroflexi*, *Chrysogenetes*, *Cyanobacteria*, *Deferribacteres*, *Deinococcus*, *Dictyoglomi*, *Fibrobacteres*, *Firmicutes*, *Fusobacteria*, *Gemmatimonadetes*, *Lentisphaerae*, *Nitrospirae*, *Planctomycetes*, *Proteobacteria*, *Spirochaetes*, *Thermodesulfobacteria*, *Thermomicrobia*, *Thermotogae*, *Thermus (Thermales)* ou *Verrucomicrobia*. Em uma modalidade de uma célula hospedeira eubacteriana, a célula pode ser um membro de qualquer espécie de eubactéria, excluindo *Cyanobacteria*.

O hospedeiro bacteriano pode também ser um membro de qualquer espécie de *Proteobacteria*. Uma célula hospedeira proteobacteriana pode ser um membro de qualquer um da taxa *Alphaproteobacteria*, *Betaproteobacteria*, *Gammaproteobacteria*, *Deltaproteobacteria* ou *Epsilonproteobacteria*. Além disso, o hospedeiro pode ser um membro de qualquer um da taxa *Alphaproteobacteria*, *Betaproteobacteria*, or *Gammaproteobacteria*, e um membro de qualquer espécie de *Gammaproteobacteria*.

Em uma modalidade de um hospedeiro *Gammaproteobacteriano*, o hospedeiro será o membro de qualquer um da taxa *Aeromonadales*, *Alteromonadales*, *Enterobacteriales*, *Pseudomonadales* ou *Xanthomonadales*; ou um membro de qualquer espécie dos *Enterobacteriales* ou *Pseudomonadales*. Em uma modalidade, a célula hospedeira pode ser da ordem *Enterobacteriales*, a célula hospedeira será um membro da família *Enterobacteriaceae*, ou um membro de qualquer um dos gêneros *Erwinia*, *Escherichia* ou *Serratia*; ou um membro do gênero *Escherichia*. Em uma modalidade de uma célula hospedeira da ordem *Pseudomonadales*, a célula hospedeira será um membro da família *Pseudomonadaceae*, até mesmo do gênero *Pseudomonas*. Hospedeiros de proteobacterianos gama incluem os membros das espécies *Escherichia coli* e membros das espécies *Pseudomonas fluorescens*.

Outros organismos de *Pseudomonas* podem também ser úteis. *Pseudomonads* e espécies estritamente relacionadas incluem *Proteobacteria Gram (-) Subgrupo 1*, que inclui o grupo de *Proteobacteria* que pertence às

famílias e/ou gêneros descritos como "Gram-Negative Aerobic Rods and Cocci" por R.E. Buchanan e N.E. Gibbons (eds.), Bergey's Manual of Determinative Bacteriology, págs. 217-289 (8<sup>a</sup> ed., 1974) (The Williams & Wilkins Co., Baltimore, MD, USA) (doravante "Bergey (1974)"). A tabela a seguir

5 apresenta estes famílias e gêneros de organismos.

As famílias e gêneros listados na Parte, "Gram-Negative Aerobic Rods and Cocci" (em Bergey (1974))	
Família I. Pseudomonadaceae	<i>Gluconobacter</i> <i>Pseudomonas</i> <i>Xanthomonas</i> <i>Zoogloea</i>
Família II. Azotobacteraceae	<i>Azomonas</i> <i>Azotobacter</i> <i>Beijerinckia</i> <i>Dexria</i>
Família III. Rhizobiaceae	<i>Agrobacterium</i> <i>Rhizobium</i>
Família IV. Methylomonadaceae	<i>Methylococcus</i> <i>Methylomonas</i>
Família V. Halobacteriaceae	<i>Halobacterium</i> <i>Halococcus</i>
Outros gêneros	<i>Acetobacter</i> <i>Alcaligenes</i> <i>Bordetella</i> <i>Brucella</i> <i>Francisella</i> <i>Thermus</i>

"Proteobacteria Gram(-) Subgrupo 1" também inclui Proteobacte-  
ria que seria classificada neste título de acordo com os critérios usados na

classificação. O título também inclui grupos que foram previamente classificados nesta seção mas não são mais, como os gêneros *Acidovorax*, *Brevundimonas*, *Burkholderia*, *Hydrogenophaga*, *Oceanimonas*, *Ralstonia* e *Stenotrophomonas*, o gênero *Sphingomonas* (e o gênero *Blastomonas*, dele derivado), que foi criado reagrupando organismos pertencentes (e previamente chamados espécie de) ao gênero *Xanthomonas*, ao gênero *Acidomonas*, que foi criado reagrupando os organismos pertencentes ao gênero *Acetobacter* conforme definido em Bergey (1974). Além disso, hospedeiros podem incluir células do gênero *Pseudomonas*, *Pseudomonas enalia* (ATCC 14393), *Pseudomonas nigrifaciens* (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) foram reclassificados 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*. "Proteobacteria Gram(-) Subgrupo 1" também inclui Proteobacteria classificadas como pertencentes a qualquer uma das famílias: Pseudomonadaceae, Azotobacteraceae (agora freqüentemente chamada pelo sinônimo, o "grupo de Azotobacter" de Pseudomonadaceae), Rhizobiaceae e Methylomonadaceae (agora freqüentemente chamada pelo sinônimo, "Methylococcaceae"). Conseqüentemente, além daqueles gêneros do contrário descritos aqui, outros gêneros de Proteobacterial caindo dentro de "Proteobacteria Gram(-) Subgrupo 1" incluem: 1) bactérias do grupo de Azotobacter do gênero *Azorhizophilus*; bactérias da família Pseudomonadaceae dos gêneros *Cellvibrio*, *Oligella* e *Teredinibacter*; 3) bactérias da família Rhizobiaceae dos gêneros *Chelatobacter*, *Ensifer*, *Liberibacter* (também chamado "*Candidatus Liberibacter*") e *Sinorhizobium*; e 4) bactérias da família Methylococcaceae dos gêneros *Methylobacter*, *Methylocaldum*, *Methylomicrobium*, *Methylosarcina* e *Methylosphaera*.

Em outra modalidade, a célula hospedeira é selecionada de

"Proteobacteria Gram (-) Subgrupo 2". "Proteobacteria Gram (-) Subgrupo 2" é definido como o grupo de Proteobacteria dos gêneros a seguir (com os números totais das cepas depositadas publicamente disponíveis listadas no catálogo destes indicadas entre parêntese, todas depositadas na ATCC, exceto como do contrário indicado): Acidomonas (2); *Acetobacter* (93); *Gluconobacter* (37); *Brevundimonas* (23); *Beijerinckia* (13); *Dexia* (2); *Brucella* (4); *Agrobacterium* (79); *Chelatobacter* (2); *Ensifer* (3); *Rhizobium* (144); *Sinorhizobium* (24); *Blastomonas* (1); *Sphingomonas* (27); *Alcaligenes* (88); *Bordetella* (43); *Burkholderia* (73); *Ralstonia* (33); *Acidovorax* (20); *Hydrogenophaga* (9); *Zoogloea* (9); *Methylobacter* (2); *Methylocaldum* (1 em NCIMB); *Methylococcus* (2); *Methylomicrobium* (2); *Methylomonas* (9); *Methylosarcina* (1); *Methylosphaera*; *Azomonas* (9); *Azorhizophilus* (5); *Azotobacter* (64); *Cellvibrio* (3); *Oligella* (5); *Pseudomonas* (1139); *Francisella* (4); *Xanthomonas* (229); *Stenotrophomonas* (50); e *Oceanimonas* (4).

15 Espécies de célula hospedeira exemplares de "Proteobacteria Gram (-) Subgrupo 2" incluem, mas não são limitadas às bactérias a seguir (com os números de depósito de ATCC ou outros de cepa(s) exemplar(es) destas mostrados entre parêntese): *Acidomonas methanolica* (ATCC 43581); *Acetobacter aceti* (ATCC 15973); *Gluconobacter oxidans* (ATCC 19357);  
20 *Brevundimonas diminuta* (ATCC 11568); *Beijerinckia indica* (ATCC 9039 e ATCC 19361); *Dexia gumrnosa* (ATCC 15994); *Brucella melitensis* (ATCC 23456), *Brucella abortus* (ATCC 23448); *Agrobacterium tumefaciens* (ATCC 23308), *Agrobacterium radiobacter* (ATCC 19358), *Agrobacterium rhizogenes* (ATCC 11325); *Chelatobacter heintzii* (ATCC 29600); *Ensifer adhaerens* (ATCC 33212); *Rhizobium leguminosarum* (ATCC 10004); *Sinorhizobium fredii* (ATCC 35423); *Blastomonas natatoria* (ATCC 35951); *Sphingomonas paucimobilis* (ATCC 29837); *Alcaligenes faecalis* (ATCC 8750); *Bordetella pertussis* (ATCC 9797); *Burkholderia cepacia* (ATCC 25416); *Ralstonia pickettii* (ATCC 27511); *Acidovorax facilis* (ATCC 11228); *Hydrogenophaga flava* (ATCC 33667); *Zoogloea ramigera* (ATCC 19544); *Methylobacter luteus* (ATCC 49878); *Methylocaldum gracile* (NCIMB 11912); *Methylococcus capsulatus* (ATCC 19069); *Methylomicrobium agile* (ATCC 35068); *Methylo-*

*nas methanica* (ATCC 35067); *Methylosarcina fibrata* (ATCC 700909); *Methylosphaera hansonii* (ACAM 549); *Azomonas agilis* (ATCC 7494); *Azorhizophilus paspali* (ATCC 23833); *Azotobacter chroococcum* (ATCC 9043); *Cellvibrio mixtus* (UQM 2601); *Oligella urethralis* (ATCC 17960); *Pseudomonas aeruginosa* (ATCC 10145), *Pseudomonas fluorescens* (ATCC 35858); *Francisella tularensis* (ATCC 6223); *Stenotrophomonas maltophilia* (ATCC 13637); *Xanthomonas campestris* (ATCC 33913); e *Oceanimonas doudoroffii* (ATCC 27123).

Em outra modalidade, a célula hospedeira é selecionada de "Proteobacteria Gram(-) Subgrupo 3". "Proteobacteria Gram(-) Subgrupo 3" é definida como o grupo de Proteobacteria dos seguintes gêneros: *Brevundimonas*; *Agrobacterium*; *Rhizobium*; *Sinorhizobium*; *Blastomonas*; *Sphingomonas*; *Alcaligenes*; *Burkholderia*; *Ralstonia*; *Acidovorax*; *Hydrogenophaga*; *Methylobacter*; *Methylocaldum*; *Methylococcus*; *Methylomicrobium*; *Methylomonas*; *Methylosarcina*; *Methylosphaera*; *Azomonas*; *Azorhizophilus*; *Azotobacter*; *Cellvibrio*; *Oligella*; *Pseudomonas*; *Teredinibacter*; *Francisella*; *Stenotrophomonas*; *Xanthomonas*; e *Oceanimonas*.

Em outra modalidade, a célula hospedeira é selecionada de "Proteobacteria Gram(-) Subgrupo 4". "Proteobacteria Gram(-) Subgrupo 4" é definida como o grupo de Proteobacteria dos seguintes gêneros: *Brevundimonas*; *Blastomonas*; *Sphingomonas*; *Burkholderia*; *Ralstonia*; *Acidovorax*; *Hydrogenophaga*; *Methylobacter*; *Methylocaldum*; *Methylococcus*; *Methylomicrobium*; *Methylomonas*; *Methylosarcina*; *Methylosphaera*; *Azomonas*; *Azorhizophilus*; *Azotobacter*; *Cellvibrio*; *Oligella*; *Pseudomonas*; *Teredinibacter*; *Francisella*; *Stenotrophomonas*; *Xanthomonas*; e *Oceanimonas*.

Em outra modalidade, a célula hospedeira é selecionada de "Proteobacteria Gram(-) Subgrupo 5". "Proteobacteria Gram(-) Subgrupo 5" é definida como o grupo de Proteobacteria dos seguintes gêneros: *Methylobacter*; *Methylocaldum*; *Methylococcus*; *Methylomicrobium*; *Methylomonas*; *Methylosarcina*; *Methylosphaera*; *Azomonas*; *Azorhizophilus*; *Azotobacter*; *Cellvibrio*; *Oligella*; *Pseudomonas*; *Teredinibacter*; *Francisella*; *Stenotrophomonas*; *Xanthomonas*; e *Oceanimonas*.

A célula hospedeira pode ser selecionada de "Proteobacteria Gram(-) Subgrupo 6". "Proteobacteria Gram(-) Subgrupo 6" é definida como o grupo de Proteobacteria dos seguintes gêneros: *Brevundimonas*; *Blastomonas*; *Sphingomonas*; *Burkholderia*; *Ralstonia*; *Acidovorax*; *Hydrogenophaga*; 5 *Azomonas*; *Azorhizophilus*; *Azotobacter*; *Cellvibrio*; *Oligella*; *Pseudomonas*; *Teredinibacter*; *Stenotrophomonas*; *Xanthomonas*; e *Oceanimonas*.

A célula hospedeira pode ser selecionada de "Proteobacteria Gram(-) Subgrupo 7". "Proteobacteria Gram(-) Subgrupo 7" é definida como o grupo de Proteobacteria dos seguintes gêneros: *Azomonas*; *Azorhizophilus*; 10 *Azotobacter*; *Cellvibrio*; *Oligella*; *Pseudomonas*; *Teredinibacter*; *Stenotrophomonas*; *Xanthomonas*; e *Oceanimonas*.

A célula hospedeira pode ser selecionada de "Proteobacteria Gram(-) Subgrupo 8". "Proteobacteria Gram(-) Subgrupo 8" é definida como o grupo de Proteobacteria dos seguintes gêneros: *Brevundimonas*; *Blastomonas*; *Sphingomonas*; *Burkholderia*; *Ralstonia*; *Acidovorax*; *Hydrogenophaga*; 15 *Pseudomonas*; *Stenotrophomonas*; *Xanthomonas*; e *Oceanimonas*.

A célula hospedeira pode ser selecionada de "Proteobacteria Gram(-) Subgrupo 9". "Proteobacteria Gram(-) Subgrupo 9" é definida como o grupo de Proteobacteria dos seguintes gêneros: *Brevundimonas*; *Burkholderia*; *Ralstonia*; *Acidovorax*; *Hydrogenophaga*; *Pseudomonas*; *Stenotrophomonas*; e *Oceanimonas*. 20

A célula hospedeira pode ser selecionada de "Proteobacteria Gram(-) Subgrupo 10". "Proteobacteria Gram(-) Subgrupo 10" é definida como o grupo de Proteobacteria dos seguintes gêneros: *Burkholderia*; *Ralstonia*; 25 *Pseudomonas*; *Stenotrophomonas*; e *Xanthomonas*.

Em outra modalidade, a célula hospedeira é selecionada de "Proteobacteria Gram(-) Subgrupo 11". "Proteobacteria Gram(-) Subgrupo 11" é definida como o grupo de Proteobacteria dos seguintes gêneros: *Pseudomonas*; *Stenotrophomonas*; e *Xanthomonas*. A célula hospedeira pode ser 30 selecionada de "Proteobacteria Gram(-) Subgrupo 12". "Proteobacteria Gram(-) Subgrupo 12" é definida como o grupo de Proteobacteria dos seguintes gêneros: *Burkholderia*; *Ralstonia*; *Pseudomonas*. A célula hospedeira

pode ser selecionada de "Proteobacteria Gram(-) Subgrupo 13". "Proteobacteria Gram(-) Subgrupo 13" é definida como o grupo de Proteobacteria dos seguintes gêneros: *Burkholderia*; *Ralstonia*; *Pseudomonas*; e *Xanthomonas*. A célula hospedeira pode ser selecionada de "Proteobacteria Gram(-) Subgrupo 14". "Proteobacteria Gram(-) Subgrupo 14" é definida como o grupo de Proteobacteria dos seguintes gêneros: *Pseudomonas* e *Xanthomonas*. A célula hospedeira pode ser selecionada de "Proteobacteria Gram(-) Subgrupo 15". "Proteobacteria Gram(-) Subgrupo 15" é definida como o grupo de Proteobacteria do gênero *Pseudomonas*.

A célula hospedeira pode ser selecionada de "Proteobacteria Gram(-) Subgrupo 16". "Proteobacteria Gram(-) Subgrupo 16" é definida como o grupo de Proteobacteria das seguintes espécies de *Pseudomonas* (com os números de depósito da ATCC ou outros de cepa(s) exempla(es) mostrada(s) entre parênteses): *Pseudomonas abietaniphila* (ATCC 700689); *Pseudomonas aeruginosa* (ATCC 10145); *Pseudomonas alcaligenes* (ATCC 14909); *Pseudomonas anguilliseptica* (ATCC 33660); *Pseudomonas citronellolis* (ATCC 13674); *Pseudomonas fluorescens* (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 alcaliplaila*; *Pseudomonas alginovora*; *Pseudomonas andersonii*; *Pseudomonas asplenii* (ATCC 23835); *Pseudomonas azelaica* (ATCC 27162); *Pseudomonas beijerinckii* (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); *Pseudomonas flectens* (ATCC 12775); *Pseudomonas azotoformans*;

*Pseudomonas brenneri*; *Pseudomonas cedrella*; *Pseudomonas corrugata* (ATCC 29736); *Pseudomonas exagitaçãoorientalis*; *Pseudomonas fluorescens* (ATCC 35858); *Pseudomonas gessardii*; *Pseudomonas libanensis*; *Pseudomonas mandelii* (ATCC 700871); *Pseudomonas marginalis* (ATCC 10844);  
5     *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 grimonii*; *Pseudomonas halodenitrificans*; *Pseudomonas halophila*; *Pseudomonas hibiscicola* (ATCC 19867); *Pseudomonas huttiensis* (ATCC 14670);  
10    *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);  
15    *Pseudomonas psychrophila*; *Pseudomonas fulva* (ATCC 31418); *Pseudomonas monteilii* (ATCC 700476); *Pseudomonas mosselii*; *Pseudomonas oryzihabitans* (ATCC 43272); *Pseudomonas plecoglossicida* (ATCC 700383);  
20    *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);  
25    *Pseudomonas viridiflava* (ATCC 13223); *Pseudomonas therrnocarboxidovorans* (ATCC 35961); *Pseudomonas thermotolerans*; *Pseudomonas thivervalensis*; *Pseudomonas vancouverensis* (ATCC 700688); *Pseudomonas wisconsinensis*; e *Pseudomonas xiamenensis*.

A célula hospedeira pode ser selecionada da "Proteobacteria Gram(-) Subgrupo 17". "Proteobacteria Gram(-) Subgrupo 17" é definida como

o grupo de Proteobacteria conhecida na técnica como as "Pseudomonads fluorescentes" incluindo aquelas pertencentes, por exemplo, às seguintes espécies de *Pseudomonas*: *Pseudomonas azotoformans*; *Pseudomonas brenneri*; *Pseudomonas cedrella*; *Pseudomonas corrugata*; *Pseudomonas exagitationis*; *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*.

A célula hospedeira pode ser selecionada de "Proteobacteria Gram(-) Subgrupo 18". "Proteobacteria Gram(-) Subgrupo 18" é definida como o grupo de todas as subespécies, variedades, cepas e outras unidades sub-especiais das espécies *Pseudomonas fluorescens*, incluindo aquelas pertencentes, por exemplo, às seguintes (com os números de depósito de ATCC ou outros de cepa(s) exemplar(es) mostrados entre parênteses): *Pseudomonas fluorescens* biotipo A, também chamada biovar 1 ou biovar I (ATCC 13525); *Pseudomonas fluorescens* biotipo B, também chamada biovar 2 ou biovar II (ATCC 17816); *Pseudomonas fluorescens* biotipo C, também chamada biovar 3 ou biovar III (ATCC 17400); *Pseudomonas fluorescens* biotipo F, também chamada biovar 4 ou biovar IV (ATCC 12983); *Pseudomonas fluorescens* biotipo G, também chamada biovar 5 ou biovar V (ATCC 17518); *Pseudomonas fluorescens* biovar VI; *Pseudomonas fluorescens* Pf0-1; *Pseudomonas fluorescens* Pf-5 (ATCC BAA-477); *Pseudomonas fluorescens* SBW25; and *Pseudomonas fluorescens* subsp. *cellulosa* (NCIMB 10462).

A célula hospedeira pode ser selecionada de "Proteobacteria Gram (-) Subgrupo 19". "Proteobacteria Gram (-) Subgrupo 19" é definida como o grupo de todas as cepas de *Pseudomonas fluorescens* biotipo A. Uma cepa típica deste biotipo é a cepa de *P. fluorescens* MB101 (ver patente US Nº 5.169.760 de Wilcox), e derivados desta. Um exemplo de um derivado desta é a cepa de *P. fluorescens* MB214, construída inserindo no lócus de asd cromossômico de MB 101 (gene de aspartato desidrogenase), uma construção

nativo de *E. coli* de Placl-lacl-lacZYA (isto é em que PlacZ foi deletado).

Cepas de *P. fluorescens* adicionais que podem ser usadas na presente invenção incluem *Pseudomonas fluorescens* Migula e *Pseudomonas fluorescens* Loitokitok, enquanto tendo as designações de ATCC a seguir:

5 A célula hospedeira pode ser selecionada de "Proteobacteria Gram(-) Subgrupo 19". "Proteobacteria Gram(-) Subgrupo 19" é definida como o grupo de todas as cepas de *Pseudomonas fluorescens* biotipo A. Uma cepa típica A típica deste biotipo é cepa de *P. fluorescens* MB101 (ver Patente US Nº 5.169.760 de Wilcox), e derivados destes. Um exemplo de um derivado  
10 10 deste é a cepa de *P. fluorescens* MB214, construída inserindo dentro do lócus de *asd* cromossômico de MB 101 (gene de aspartato desidrogenase), a construção nativa de P1acl-lacl-lacZYA de *E. coli* (isto é em que PlacZ foi deletado).

15 Cepas adicionais de *P. Fluorescens* que podem ser usadas na presente invenção incluem *Pseudomonas fluorescens* Migula e *Pseudomonas fluorescens* Loitokitok, tendo as seguintes designações da ATCC: [NCIB 8286]; NRRL B-1244; NCIB 8865 cepa CO1; NCIB 8866 cepa CO2; 1291 [ATCC 17458]; IFO 15837; NCIB 8917; LA; NRRL B-1864; pirrolidina; PW2 [ICMP 3966; NCPPB 967; NRRL B-899]; 13475; NCTC 10038; NRRL B1603  
20 [6; IFO 15840]; 52-1C; CCEB 488-A [BU 140]; CCEB 553 [IEM 15/47j; IAM 1008 [A1-H-27]; IAM 1055 [AHH-23]; 1 [IFO 15842]; 12 [ATCC 25323; NTH 11; den Dooren de Jong 216]; 18 [IFO 15833; WRRL P-7]; 93 [TR-10]; 108 [52-22; IFO 15832]; 143 [IFO 15836; PL]; 149 [2-40-40; IFO 15838]; 182 [IFO 3081; PJ 73]; 184 [IFO 15830]; 185 [W2 LA]; 186 [IFO 15829; PJ 79]; 187 [NCPPB  
25 263]; 188 [NCPPB 316]; 189 [PJ227; 1208]; 191 [IFO 15834; PJ 236; 22/1]; 194 [Klinge R-60; PJ 253]; 196 [PJ 288]; 197 [PJ 290]; 198 [PJ 302]; 201 [PJ 368]; 202 [PJ 372]; 203 [PJ 376]; 204 [IFO 15835; PJ 682]; 205 [PJ 686]; 206 [PJ 692]; 207 [PJ 693]; 208 [PJ 722]; 212 [PJ 832]; 215 [PJ 849]; 216 [PJ 885]; 267 [B-9]; 271 [B-1612]; 401 [C71A; IFO 15831; PJ 187]; NRRL B-3178 [4; IFO  
30 15841]; KY 8521; 3081; 30-21; [IFO 3081]; N; PYR; PW; D946-B83 [BU 2183; FERM-P 3328]; P-2563 [FERM-P 2894; IFO 13658]; IAM-1126 [43F]; M-1; A506 [A5-06]; A505 [A5-05-1]; A526 [A5-26]; B69; 72; NRRL B-4290; PMW6

[NCIB 11615]; SC 12936; AI [IFO 15839]; F 1847 [CDC-EB]; F 1848 [CDC 93]; NCIB 10586; P17; F-12; AmMS 257; PRA25; 6133D02; 6519E01; N1; SC15208; BNL-WVC; NCTC 2583 [NCIB 8194]; H13; 1013 [ATCC 11251; CCEB 295]; IFO 3903; 1062; ou Pf-5.

5 Outros hospedeiros adequados incluem aqueles classificados em outras partes da referência, como Proteobactérias Gram (+). Em uma modalidade, a célula hospedeira é um *E. coli*. A seqüência de genoma para *E. coli* foi estabelecida para MG1655 de *E. coli* (Blattner, et al. (1997) The complete genome sequence of *Escherichia coli* K-12 *Science* 277(5331): 1453-74) e  
10 microarranjo de DNA estão comercialmente disponíveis para K12 de *E. coli* (MWG Inc, High Point, NC). *E. coli* pode ser cultivado em um meio rico como Luria-Bertani (LB) (10 g/L de triptona, 5 g/L de NaCl, 5 g/L de extrato de levedura) ou um meio mínimo definido como M9 (6 g/L de Na<sub>2</sub>HPO<sub>4</sub>, 3 g/L de KH<sub>2</sub>PO<sub>4</sub>, 1 g/L de NH<sub>4</sub>Cl, 0,5 g/L NaCl, pH 7,4) com uma fonte de carbono  
15 apropriada como 1% de glicose. Habitualmente, uma cultura durante a noite de células de *E. coli* é diluída e inoculadas em meio fresco rico ou mínimo em um frasco agitador ou um fermentador e crescida a 37°C.

Um hospedeiro pode também ser de origem mamífera, como uma célula derivada de um mamífero incluindo qualquer mamífero humano ou  
20 não-humano. Mamíferos podem incluir, mas não são limitados a primatas, macacos, porcino, ovino, bovino, roedores, ungulado, porcos, suínos, ovelha, carneiro, cabras, gado, cervos, mulas, cavalos, macacos, bugios, cachorros, gatos, ratos e camundongos.

Uma célula hospedeira pode também ser de origem de planta.  
25 Qualquer planta pode ser selecionada para a identificação de genes e seqüências reguladoras. Exemplos de alvos de planta adequados para o isolamento de genes e seqüências reguladoras incluiriam mas não limitariam a alfafa, maçã, abricó, Arabidopsis, alcachofra, rúcula, espargos, abacate, banana, cevada, feijões, beterraba, amora-preta, mirtilo, brócolos, couve-de-bruxelas, repolho, óleo de canola, cantalupo, cenoura, mandioca, mamona, couve-flor, aipo, cereja, chicória, coentro, cítrico, clementinas, trevo, coco, café, milho, algodão, oxicoco, pepino, abeto de Douglas, berinjela, en-

dívia, escarola, eucalipto, funcho, figos, alho, cabaço, uva, toronja, melão honeydew, jicama, kiwi, alface, alho-porró, limão, lima, pinheiro-amarelo, linhaça, manga, melão, cogumelo, nectarina, noz, aveia, óleo de palma, óleo de colza, quiabo, azeitona, cebola, laranja, uma planta ornamental, palma,

5 mamão, salsa, pastinaca, ervilha, pêssego, amendoim, pêra, pimenta, caqui, pinheiro, abacaxi, musa, ameixa, romã, álamo, batata, jerimum, marmelo, pinho, radicchio, rabanete, colza, framboesa, arroz, centeio, sorgo, pinheiro do Sul, feijão-soja, espinafre, abóbora, morango, beterraba-açucareira, cana-de-açúcar, girassol, batata-doce, liquidambar, mexerica, chá, tabaco, to-

10 mate, triticale, grama, nabo, uma videira, melancia, trigo, inhames e abobrinha. Em algumas modalidades, plantas úteis no processo são *Arabidopsis*, milho, trigo, feijão-soja e algodão.

Para expressão de uma proteína ou peptídeo recombinante, ou para modulação de um gene compensatório identificado, qualquer promotor de planta pode ser usado. Um promotor pode ser um promotor de RNA polimerase II de planta. Elementos incluídos em promotores de planta podem ser uma caixa de TATA ou caixa de Goldberg-Hogness, tipicamente posicionada aproximadamente 25 a 35 pares de base a montante (5') do sítio de iniciação de transcrição, e a caixa de CCAAT, localizada entre 70 e 100 pares de base a montante. Em plantas, a caixa de CCAAT pode ter uma seqüência de consenso diferente que a seqüência funcionalmente análoga dos promotores mamíferos (Messing et al., Em: *Genetic Engineering of Plants*, Kosuge et al., eds., págs. 211-227, 1983). Virtualmente além disso, todos promotores incluem seqüências de ativação ou intensificadores adicionais a montante (Benoist e Chambon, *Nature* 290:304-310, 1981; Gruss et al., *Proc. Nat. Acad. Sci. USA* 78:943-947, 1981; e Khoury e Gruss, *Cell* 27:313-314, 1983) estendendo-se por volta de -100 bp a -1.000 bp ou mais a montante do sítio de iniciação de transcrição.

#### EXPRESSÃO DE PROTEÍNA OU PEPTÍDEO RECOMBINANTE

30 Como descrito abaixo, uma célula ou organismo hospedeiro pode ser criado para expressar proteína ou peptídeo recombinante usando técnicas padrão. Por exemplo, proteína recombinante pode ser expressada de um

vetor ou de um gene exógeno inserido no genoma do hospedeiro. Vetores que podem ser usados para expressar proteínas exógenas são bem conhecidos na técnica e são descritos abaixo. Genes para expressar proteína ou peptídeo recombinante podem também ser inseridos no genoma usando técnicas como 5 recombinação homóloga ou heteróloga, como descritas abaixo.

A proteína ou peptídeo recombinante pode ser expressado após indução com um composto químico sob expressão de um gene endógeno ou de produto de gene. A proteína recombinante pode também ser expressa quando a célula hospedeira for colocada em um ambiente particular. Ele-10 mentos promotores específicos são descritos abaixo. Estes incluem, mas não são limitados a, promotores que podem ser induzidos sob tratamento da cé-  
lula com químicas como IPTG, benzoato ou antranilato.

#### PROTEÍNAS/PEPTÍDEOS RECOMBINANTES

A célula hospedeira foi projetada para expressar uma proteína ou 15 peptídeo recombinante. Este pode ser de qualquer espécie e de qualquer tamanho. Porém, em certas modalidades, a proteína ou peptídeo recombi-  
nante é uma proteína ou peptídeo terapeuticamente útil. Em algumas moda-  
lidades, a proteína pode ser uma proteína mamífera, por exemplo uma pro-  
teína humana, e pode ser, por exemplo, um fator de crescimento, uma citocina,  
20 um quimiocina ou uma proteína de sangue. A proteína ou peptídeo recom-  
binante pode ser expressado primariamente em uma forma inativa na célula  
hospedeira. Em certas modalidades, a proteína ou peptídeo recombinante é  
menos que 100 kD, menos que 50 kD ou menos que 30 kD em tamanho. Em  
modalidades de ceratina, a proteína ou peptídeo recombinante é um peptídeo  
25 de pelo menos 5, 10, 15, 20, 30, 40, 50 ou 100 aminoácidos.

Vetores de expressão existentes permitem a produção de proteína recombinante em *E. coli*. Para todos estes sistemas de expressão de 30 proteína procedimentos de clonagem rotineiros como descritos mais cedo podem ser seguidos (Sambrook, et al. (2000) *Molecular cloning: A laboratory manual*, terceira edição, Cold Spring Harbor, Nova Iorque, Cold Spring Harbor Laboratory Press).

O sistema expressão Champion® pET fornece um nível alto de

produção de proteína. Expressão é induzida do promotor de T7lac forte. Este sistema tira proveito da atividade alta e especificidade da bacteriófago T7 RNA polimerase para transcrição de nível alto do gene de interesse. O operador de lac localizado na região de promotor fornece regulação mais firme

5 que os vetores baseados em T7 tradicionais, melhorando a estabilidade do plasmídeo e viabilidade da célula (Studier, F. W. e B. A. Moffatt (1986) Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes *Journal of Molecular Biology* 189(1): 113-30; Rosenberg, et al. (1987) Vectors for selective expression of cloned DNAs by T7 RNA poly-  
10 merase *Gene* 56(1): 125-35). O sistema de expressão de T7 usa o promotor de T7 e T7 RNA polimerase (T7 RNAP) para transcrição de nível alto do gene de interesse. Expressão de nível alto é alcançada em sistemas de expressão de T7 porque o T7 RNAP é mais processiva que *E. coli* RNAP nativa e é dedicada à transcrição do gene de interesse. Expressão do gene identificado é  
15 induzida fornecendo uma fonte de T7 RNAP na célula hospedeira. Isto é realizado usando um hospede de *E. coli* de BL21 contendo uma cópia cromossômica do gene de T7 RNAP. O gene de T7 RNAP está sob o controle do promotor lacUV 5 que pode ser induzido por IPTG. T7 RNAP é expressado em indução e transcreve o gene de interesse.

20 O sistema de expressão de pBAD permite expressão firmemente controlada, titulável da proteína recombinante através da presença de fontes de carbono específicas como glicose, glicerol e arabinose (Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter" *Journal of Bacteriology* 177(14): 412130). Os vetores de  
25 pBAD são projetados exclusivamente para dar controle preciso nos níveis de expressão. Expressão de gene heterólogo dos vetores de pBAD é iniciada no promotor de araBAD. O promotor é positiva e negativamente regulado pelo produto do gene de araC. AraC é um regulador transcripcional que forma um complexo com L-arabinose. Na ausência de L-arabinose, o dímero de AraC  
30 bloqueia a transcrição. Para ativação transcripcional máxima dois eventos são requeridos: (i.) L-arabinose ligue a AraC permitindo a transcrição começar. (ii.) O complexo de proteína de ativador de cAMP (CAP)-cAMP liga ao DNA e

estimula a ligação de AraC na localização correta da região de promotor.

O sistema de expressão de *trc* permite nível alto, expressão regulada em *E. coli* do promotor de *trc*. Os vetores de expressão de *trc* foram otimizados para expressão de genes eucarióticos em *E. coli*. O promotor de 5 *trc* é um promotor híbrido forte derivado dos promotores triptofano (*trp*) e lactose (*lac*). Ele é regulado pelo operador de *lacO* e o produto do gene de *lac<sup>Q</sup>* (Brosius, J. (1984) Toxicity of an overproduced foreign gene product in *Escherichia coli* and its use in plasmid vectors for the selection of transcription terminators *Gene* 27(2): 161-72).

10 A invenção também inclui a célula hospedeira recombinante melhorada que é produzida pelo processo reivindicado. Em uma modalidade, a invenção inclui uma célula produzida pelo processo descrito. Em outra modalidade, a invenção inclui uma célula ou organismo hospedeiro que expressa uma proteína recombinante que foi geneticamente modificada para reduzir a expressão de pelo menos duas proteases. Em outras modalidades, 15 a invenção inclui uma célula ou organismo hospedeiro que expressa uma proteína recombinante que foi geneticamente modificada para reduzir a expressão de pelo menos uma protease selecionada do grupo que consiste em produtos de genes de *hsIV*, *hsIU*, *clpX*, *clpA* e *clpB*, e em certas submoda- 20 lidades, a célula ou organismo foi modificado para reduzir a expressão de *HsIV* ou *HsIU*. Em certas modalidades, a célula ou organismo hospedeiro modificado expressa uma proteína derivada mamífera recombinante, e pode expressar proteína derivada que pode ser hormônio de crescimento humano para um humano recombinante. A célula pode ser modificada por qualquer 25 técnica conhecida na técnica, por exemplo por uma técnica em que pelo menos um gene de protease é knock-out do genoma, ou mutando pelo menos um gene de protease para reduzir a expressão de uma protease, ou alterando pelo menos um promotor de pelo menos um gene de protease para reduzir expressão de uma protease.

30 Em outra modalidade, hospedeiro ou organismo que expressam uma proteína recombinante que é apresentada que foi geneticamente modificada para aumentar a expressão de pelo menos um, pelo menos dois mo-

duladores de dobramentos, ou pelo menos três moduladores de dobramentos. Em certas submodalidades, os moduladores de dobramentos não são subunidades de modulador de dobramento. O modulador de dobramento pode ser selecionado do grupo que consiste em produtos de genes de *cbpA*, *htpG*,  
5 *dnaK*, *dnaJ*, *fkbP2*, *groES* e *groEL*, e, em certas submodalidades, pode ser *htpG* ou *cbpA*. A célula ou organismo hospedeiro podem, em um exemplo não-limitativo, expressar uma proteína mamífera, como uma proteína humana. A proteína pode ser hormônio de crescimento humano. O modulador ou moduladores de dobramento podem ser aumentados, por exemplo, incluindo  
10 um vetor de expressão como descrito aqui na célula. A expressão de modulador de dobramento pode também ser aumentada, por exemplo, mutando um promotor de um modulador de dobramento ou subunidade de modulador de dobramento. Uma célula ou organismo hospedeiro que expressa uma proteína recombinante pode também ser geneticamente modificado para aumentar  
15 a expressão de pelo menos um modulador de dobramento e diminuir a expressão de pelo menos uma protease ou proteína de protease. Organismos compreendendo uma ou mais células produzidas pelo processo descrito estão também inclusos na invenção.

ETAPA II: ANÁLISE DE UM PERFIL GENÉTICO PARA IDENTIFICAR UM  
20 GENE OU PRODUTO DE GENE COMPENSATÓRIO QUE É EXPRESSADO  
EM UM NÍVEL MAIS ALTO NA CÉLULA RECOMBINANTE

O processo da invenção inclui analisar um perfil genético da célula recombinante para identificar um gene compensatório ou produto de gene que são expressados a um nível mais alto na célula recombinante que em ou  
25 uma célula hospedeira que não foi modificada para expressar a proteína recombinante ou uma célula recombinante que não estão expressando a proteína recombinante.

Um "perfil genético" como aqui usado pode incluir genes em um genoma, mRNA transcreveram de genes no genoma ou cDNA derivaram de  
30 mRNA transrito de genes no genoma. Um perfil genético pode também incluir produtos de transcrição que foram modificados por uma célula como variantes de encaixe de genes em sistemas eucarióticos, ou proteínas

transladadas dos genes em um genoma, incluindo proteínas que são modificadas pela célula ou transladas de variantes de encaixe de mRNA transladado do genoma. Um perfil genético é significado referir-se somente à análise simultânea de entidades múltiplas, como em um arranjo ou outro sistema de multiplexação, incluindo análise de borrão simultânea múltipla ou cromatografia de coluna com pares de ligação de múltiplos ligados ao pacote. De acordo com a invenção, pelo menos 5, 10, 25, 50, 70, 80, 90 ou 100 ou mais genes ou produtos de gene são analisados simultaneamente.

### TRANSCRIPTOMA

Em uma modalidade, o perfil genético analisado é um perfil de transcriptoma. Um transcriptoma completo refere-se ao conjunto completo de transcrições de mRNA produzidas pelo genoma de uma só vez. Ao contrário do genoma, o transcriptoma é dinâmico e varia consideravelmente em circunstâncias divergentes devido aos padrão diferentes de expressão de gene.

Transcriptômicos, o estudo do transcriptoma, é um meio inclusivo de identificar padrão de expressão de gene. O transcriptoma analisado pode incluir o conjunto conhecido completo de genes transcritos, isto é o conteúdo de mRNA ou cDNA correspondente de uma célula ou organismo hospedeiro de hospedeiro. O cDNA pode ser uma cadeia de nucleotídeos, um polinucleotídeo, nucleotídeo, molécula de ácido nucléico isolados ou qualquer fragmento ou complemento destes que originou recombinante ou sinteticamente e é bifilamentar ou unifilamentar, de codificação e/ou de não-codificação, um exon ou um ítron de uma molécula de DNA genômica, ou combinado com carboidrato, lipídios, proteína ou elementos ou substâncias inorgânicas. A cadeia de nucleotídeo pode ser pelo menos 5, 10, 15, 30, 40, 50, 60, 70, 80, 90 ou 100 nucleotídeos em comprimento. O transcriptoma pode também incluir apenas uma porção do conjunto conhecido de transcrições genéticas. Por exemplo, o transcriptoma pode incluir menos que 98%, 95, 90, 85, 80, 70, 60 ou 50% das transcrições conhecidas em um hospedeiro. O transcriptoma pode também ser alvejado a um conjunto específico de genes.

Em uma modalidade, o processo de triagem pode incluir triagem usando um arranjo ou um microarranjo para identificar um perfil genético. Em

outra modalidade, o perfil de transcriptoma pode ser analisado usando processos conhecidos como hibridização em ensaios de borrão como northern blots. Em outra modalidade, o processo pode incluir processos com base em PCR como RT-PCR que podem quantificar a expressão de um conjunto particular de genes. Em uma modalidade da invenção, um gene identificado, por exemplo uma proteína de modulador de dobramento (FM) ou proteína de protease, isto é uma protease, peptidase, ou polipeptídeo ou co-fator associado, é identificada por um processo de triagem de processamento alto.

O processo pode incluir analisar o perfil de transcriptoma usando um microarranjo ou técnica equivalente. Nesta modalidade, o microarranjo pode incluir pelo menos uma porção do genoma transcrita da célula hospedeira, e tipicamente inclui os pares de ligação para amostras de genes de pelo menos 50% dos genes transcritos do organismo. Mais tipicamente, o microarranjo ou técnica equivalente inclui os pares de ligação para amostras de pelo menos 80%, 90%, 95%, 98%, 99% ou 100% dos genes transcritos no genoma da célula hospedeira. Porém, em uma modalidade separada, o microarranjo pode incluir os pares de ligação para um subconjunto selecionado de genes do genoma, incluindo mas não limitados a, genes de protease putativos ou genes de modulador de dobramento putativos. Um microarranjo ou técnica equivalente pode tipicamente também incluir os pares de ligação para um conjunto de genes que são usados como controles, como genes de manutenção. Um microarranjo ou técnica equivalente pode também incluir genes crescidos em cachos em grupos como genes que codificam para proteínas degradativas, moduladores e co-fatores de dobramentos, proteínas metabólicas como proteínas envolvidas em metabolismo de glicose ou aminoácido ou síntese de nucleobase, fatores de transcrição, fatores de estabilização de ácido nucléico, genes regulados de sinal extracelular como cinases e receptores ou proteínas de andaime.

Um microarranjo é em geral formado ligando um número grande de pares de ligação distintos que podem incluir polinucleotídeos, aptâmeros, químicas, anticorpos ou outras proteínas ou peptídeos a um suporte sólido como um microfatia, lâmina de vidro ou outros, em um padrão definido.

Contatando o microarranjo com uma amostra obtida de uma célula de interesse e detectando a ligação dos pares de ligação expressos na célula que hibridam com as seqüências na fatia, o padrão formado pelos polinucleotídeos de hibridização permite a identificação de genes ou agrupamentos de genes que são expressados na célula. Além disso, onde cada membro ligado ao suporte sólido for conhecido, a identidade dos pares de hibridização com os da amostra de ácido nucléico pode ser identificada. Uma tecnologia de resistência de microarranjo é que permite a identificação da expressão de gene de diferencial simplesmente comparando os padrão de hibridização.

Exemplos de processos de triagem de processamento alto incluem hibridização de mRNA de célula hospedeira ou cDNA substancialmente correspondente, para um/uns arranjo(s) hidrolizável(is) ou microarranjo(s). O arranjo ou microarranjo pode ser um ou mais arranjo(s) de oligômeros ou polímeros de ácido nucléico ou de análogos de ácido nucléico. Em uma modalidade, o(s) arranjo(s) ou microarranjo(s) será(ão) independente ou coletivamente um/uns arranjo(s) ou microarranjo(s) amplo(s) em genoma de célula hospedeira, contendo uma população de oligômeros ou polímeros de ácido nucléico ou de análogos de ácido nucléico cujas seqüências de nucleotídeo são hidrolizáveis em porções representativas de todos os genes conhecidos codificar ou prognosticados como codificando FMs na cepa de célula hospedeira ou todos os genes conhecidos codificar ou prognosticados codificar proteases ou proteínas de protease na cepa de célula hospedeira. Um microarranjo amplo em genoma inclui seqüências que ligam a uma porção representativa de todas das seqüências de estrutura de leitura aberta (ORF) conhecidas ou prognosticadas, como de mRNA ou cDNA correspondente do hospedeiro.

As seqüências de oligonucleotídeo ou análogos no arranjo tipicamente hibridam com as seqüências de mRNA ou de cDNA correspondentes da célula hospedeira e tipicamente compreendem uma seqüência de nucleotídeo complementar a pelo menos uma porção de uma seqüência de mRNA ou cDNA hospedeira, ou uma seqüência homóloga à seqüência de mRNA ou de cDNA hospedeira. Filamentos de DNA simples com seqüências comple-

mentares podem emparelhar entre si e formarem moléculas bifilamentares. Microarranjo em geral aplicam-se ao princípio de hibridização em um formato altamente paralelo. Em vez de um identificado, milhares de identificados potenciais diferentes podem ser arranjados em um suporte sólido de miniatura.

- 5 Em vez de uma única sonda de DNA marcado, uma mistura complexa de moléculas de DNA marcado é usada, preparada do RNA de um tipo de célula ou tecido particular. As abundâncias de moléculas de DNA marcado individuais nesta sonda complexa tipicamente refletem os níveis de expressão dos genes correspondentes. Em um processo simplificado, quando hibridar com o  
10 arranjo, as seqüências abundantes gerarão sinais fortes e seqüências raras gerarão sinais fracos. A resistência do sinal pode representar o nível de expressão de gene na amostra original.

Em uma modalidade, um arranjo ou microarranjo amplo em genoma será usado. Em uma modalidade, o arranjo representa mais que 50%  
15 das estruturas de leitura aberta no genoma do hospedeiro, ou mais que 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% ou 100% das estruturas de leitura aberta conhecidas no genoma. O arranjo pode também representar pelo menos uma porção de pelo menos 50% das seqüências conhecidas codificar proteína na célula hospedeira. Em modalidades separadas, o arranjo representa mais que 50% dos genes ou genes putativos da célula hospedeira, ou mais que 55%, 60%, 65%,  
20 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% ou 100% dos genes conhecidos ou genes putativos. Em uma modalidade, mais de um oligonucleotídeo ou análogo podem ser usados para cada gene  
25 ou seqüência de gene putativo ou estrutura de leitura aberta. Em uma modalidade, estes múltiplos oligonucleotídeos ou análogos representam porções diferentes de um gene conhecido ou seqüência de gene putativo. Para cada gene ou seqüência de gene putativo, de cerca de 1 a cerca de 10000 ou de 1 a cerca de 100 ou de 1 a cerca de 50, 45, 40, 35, 30, 25, 20, 15, 10 ou menos  
30 oligonucleotídeos ou análogos podem estar presentes no arranjo.

Um microarranjo ou um arranjo ou microarranjo amplo em genoma completo pode ser preparado de acordo com qualquer processo co-

nhecido na técnica, com base no conhecimento das(s) seqüênci(a)s do genoma de célula hospedeira, ou das seqüências de codificação propostas no genoma, ou com base no conhecimento das seqüências de mRNA expressas na célula hospedeira ou organismo hospedeiro.

- 5           Para tipos diferentes de células hospedeiras, o mesmo tipo de microarranjo pode ser aplicado. Os tipos de microarranjo incluem microarranjo de DNA complementar (cDNA) (Schena, M., et al. (1995) Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* 270:467-70) e microarranjo de oligonucleotídeo (Lockhart, et al.
- 10          (1996) Expression monitoring by hybridization to high-density oligonucleotide arrays. *Nat Biotechnol* 14:1675-80). Para microarranjo de cDNA, o fragmento de DNA de uma estrutura de leitura aberta parcial ou inteira é impresso nas lâminas. As características de hibridização podem ser diferentes ao longo da lâmina porque porções diferentes das moléculas podem ser impressas em
- 15          localizações diferentes. Para os arranjos de oligonucleotídeo, oligos 2080 mer podem ser sintetizados ou *in situ* (na fatia) ou por síntese convencional seguida por imobilização na fatia, porém em geral todas as sondas são projetadas para ser similares com respeito à temperatura de hibridização e afinidade de ligação (Butte, A. (2002) The use and analysis of microarray data. *Nat Rev Drug Discov* 1:951-60).

Na análise do perfil de transcriptoma, os oligômeros ou polímeros de ácido nucléico ou de análogos de ácido nucléico podem ser RNA, DNA, ou um análogo de RNA ou DNA. Tais análogos de ácido nucléico são conhecidos na técnica e incluem, por exemplo: ácidos nucléicos de peptídeo (PNA); ácidos nucléicos de arabinose; ácidos nucléicos de altritol; ácidos nucléicos ligados (BNA), por exemplo, ácidos nucléicos ligados a 2'-0,4'-C-etileno e ácidos nucléicos ligados a 2'-0,4'-C-metileno; ácidos nucléicos de cicloexenila; ácidos nucléicos com base em nucleotídeo 2',5'-ligado; ácidos nucléicos de morfolino (unidades de morfolino nucleobase-substituídas conectadas, por exemplo, através de ligações de fosforodiamidato); análogos de ácido nucléico substituídos na cadeia principal, por exemplo, ácidos nucléicos 2'-substituídos, em que pelo menos um dos átomos de carbono 2' de um ácido

nucléico do tipo oligo ou polissacarídeo ou análogo é independentemente substituído com, por exemplo, qualquer um de um grupo halo, tio, amino, alifático, oxialifático, tioalifático ou aminoalifático (em que alifático é tipicamente C<sub>1</sub>-C<sub>10</sub> alifático).

5           Oligonucleotídeos ou análogos de oligonucleotídeo no arranjo podem ser de tamanho uniforme e, em uma modalidade, podem ser cerca de 10 a cerca de 1000 nucleotídeos, cerca de 20 a cerca de 1000, 20 a cerca de 500, 20 a cerca de 100, cerca de 20, cerca de 25, cerca de 30, cerca de 40, cerca de 50, cerca de 60, cerca de 70, cerca de 80, cerca de 90 ou cerca de  
10          100 nucleotídeos de comprimento.

O arranjo de sondas de oligonucleotídeo pode ser um arranjo de densidade alta compreendendo mais que cerca de 100, ou mais que cerca de 1.000 ou mais sondas de oligonucleotídeo diferentes. Tais arranjos de densidade alta podem compreender uma densidade de sonda de mais que cerca  
15          de 60, mais em geral maior que cerca de 100, mais em geral maior que cerca de 600, freqüentemente maior que cerca de 1000, mais freqüentemente maior que cerca de 5.000, o mais freqüentemente maior que cerca de 10.000, tipicamente maior que cerca de 40.000 mais tipicamente maior que cerca de 100.000, e em certas circunstâncias é maior que cerca de 400.000 sondas de  
20          oligonucleotídeo diferentes por cm<sup>2</sup> (onde oligonucleotídeos diferentes referem-se aos oligonucleotídeos que têm seqüências diferentes). As sondas de oligonucleotídeo variam de cerca de 5 a cerca de 500, ou cerca de 5 a 50, ou de cerca de 5 a cerca de 45 nucleotídeos, ou de cerca de 10 a cerca de 40 nucleotídeos e o mais tipicamente de cerca de 15 a cerca de 40 nucleotídeos  
25          em comprimento. Arranjos particulares contêm sondas que variam de cerca de 20 a cerca de 25 oligonucleotídeos em comprimento. O arranjo pode compreender mais que 10, ou mais que 50, ou mais que 100, e tipicamente mais de 1000 sondas de oligonucleotídeo específicas para cada gene identificado. Em uma modalidade, o arranjo compreende pelo menos 10 sondas de  
30          oligonucleotídeo diferentes para cada gene. Em outra modalidade, o arranjo tem 20 ou menos oligonucleotídeos complementares a cada gene. Embora uma superfície de arranjo planar seja típica, o arranjo pode ser fabricado em

uma superfície de virtualmente qualquer forma ou até mesmo em superfícies múltiplas.

O arranjo pode também compreender sondas de controle de disparidade. Onde tais controles de disparidade estão presentes, a etapa de quantificação pode compreender calcular a diferença em intensidade do sinal de hibridização entre cada uma das sondas de oligonucleotídeo e sua sonda de controle de disparidade correspondente. A quantificação pode também compreender calcular a diferença média em intensidade do sinal de hibridização entre cada uma das sondas de oligonucleotídeo e sua sonda de controle de disparidade correspondente para cada gene.

Em alguns formatos de ensaio, a sonda de oligonucleotídeo pode ser presa, isto é, através de ligação covalente, a um suporte sólido. Arranjos de oligonucleotídeo podem ser quimicamente sintetizados através de processos de síntese de polímero imobilizado paralela ou por processos de síntese de polímero de direcionada por luz, por exemplo em substratos de poli-L-lisina como lâminas. Arranjos quimicamente sintetizados são vantajosos em que a preparação da sonda não requer clonagem, uma etapa de amplificação de ácido nucléico ou síntese enzimática. O arranjo inclui sondas de teste que são sondas de oligonucleotídeo cada uma destas tem uma seqüência que é complementar a uma subseqüência de um dos genes (ou o mRNA ou o cRNA antissense correspondente) cuja expressão é para ser detectada. Além disso, o arranjo pode conter controles de normalização, controles de disparidade e controles do nível de expressão como descritos aqui.

Um arranjo pode ser projetado para incluir um oligonucleotídeo de hibridização por gene conhecido em um genoma. Os oligonucleotídeos ou pares de ligação equivalentes podem ser 5'-amino modificados para suportar ligação covalente às lâminas revestidas com epóxi. Os oligonucleotídeos podem ser projetados para reduzir hibridização cruzada, por exemplo reduzindo a identidade de seqüência para menos que 25% entre os oligonucleotídeos. Em geral, a temperatura de fundição dos oligonucleotídeos é analisada antes de o projeto do arranjo assegurar teor de GC consistente e Tm, e a

estrutura secundária dos pares de ligação de oligonucleotídeo é otimizada. Para perfilação de transcriptoma, a estrutura secundária é tipicamente minimizada. Em uma modalidade, cada oligonucleotídeo é impresso em pelo menos duas localizações diferentes na lâmina para aumentar a precisão.

- 5 Oligonucleotídeos de controle podem também ser projetados com base nas seqüências de diferentes espécies que a célula ou organismo hospedeiro para mostrar ligação de base.

As amostras no perfil genético podem ser analisadas individualmente ou agrupadas em agrupamentos. Os agrupamentos podem tipicamente ser agrupados por similaridade em expressão de gene. Em uma modalidade, os agrupamentos podem ser individualmente agrupados como genes que são regulados a uma extensão similar em uma célula hospedeira. Os agrupamentos podem também incluir grupos de genes que são regulados a uma extensão similar em uma célula hospedeira recombinante, por exemplo genes que são sobre-regulados ou sub-regulados a uma extensão similar comparada a uma célula hospedeira ou uma célula modificada ou uma inalterada. Os agrupamentos podem também incluir grupos relacionados por estrutura de gene ou de proteína, função ou, no caso de um arranjo de transcriptoma, por colocação ou agrupamento dos pares de ligação aos genes no genoma do hospedeiro. Grupos de pares de ligação ou grupos de genes ou proteínas analisados podem incluir genes selecionados de, mas não limitados a: genes que codificam para proteases putativas ou conhecidas, co-fatores de proteases ou proteínas similares à protease; moduladores de dobramentos, co-fatores de moduladores de dobramentos ou proteínas que poderiam melhorar o dobramento ou solubilidade da proteína; fatores de transcrição; proteínas envolvidas em estabilidade de ácido nucléico ou iniciação translacional; cinases; receptores extracelulares ou intracelulares; enzimas metabólicas; co-fatores metabólicos; proteínas de envelope; fatores de sigma; proteínas ligadas à membrana; proteínas de transmembrana; proteínas associadas à membrana e genes de manutenção.

### PROTEOMA

Em outra modalidade, o perfil genético analisado é um perfil de

proteoma. O proteoma de um hospedeiro é o conjunto completo de proteínas produzidas pelo genoma de uma só vez. O proteoma é em geral muito mais complexo que o genoma ou o transcriptoma porque cada proteína pode ser quimicamente modificada após a síntese. Muitas proteínas são clivadas durante a produção, fosforiladas, acetiladas, metiladas, ou tem grupos carboidrato acrescentados a elas, dependendo da célula hospedeira. O proteoma é também muito dinâmico. Proteômicos, o estudo do proteoma, pode abranger vários aspectos diferentes da estrutura de proteína, expressão de proteína e função. As técnicas para análise de proteoma não são tão diretas quanto às usadas em transcriptônicos. Porém, uma vantagem dos proteômicos é que as moléculas funcionais da célula estão sendo estudadas.

O processo pode incluir técnicas que medem os níveis de expressão de proteína, interações de proteína-proteína, interações de proteína-molécula pequena ou atividades enzimáticas. Em uma modalidade, o proteoma é analisado usando um processo de triagem que inclui medição de tamanho de certas proteínas, tipicamente usando espectrometria de massa. Em uma modalidade, a técnica para analisar o perfil de proteoma inclui hibridização de um anticorpo com uma proteína de interesse. Por exemplo, o processo pode incluir processos de western blot como conhecido na técnica ou pode incluir cromatografia de coluna. O processo pode também incluir processos padrão como triagem de Elisa conhecida na técnica. O processo pode também incluir ligação de pares de ligação modificados de ácido nucléico que podem ser aptâmeros ou podem ser proteína ou os pares de ligação química para proteínas ou fragmentos de peptídeo no proteoma e um processo de triagem pode incluir amplificação dos ácidos nucléicos. O processo pode também incluir compostos químicos que ligam às proteínas ou fragmentos de proteínas em um proteoma e o processo pode incluir medição da ligação através de meios químicos. A medição pode também incluir medição dos produtos de reação em uma reação química, ou por ativação de um fluoróforo. Técnicas como espectrometria de massa em combinação com ferramentas de separação como eletroforese em cromatografia líquida em gel bidimensional ou multidimensional, pode também ser usado no processo.

Tipicamente, o processo inclui uma técnica de triagem de processamento alto.

O processo da invenção pode incluir analisar o perfil de proteoma usando, por exemplo, eletroforese bidimensional. Este é um método para a separação e identificação das proteínas em uma amostra através de deslo-

5 camento em duas dimensões orientadas em ângulos retos um ao outro. Isto permite a amostra separar em uma área maior, aumentando a resolução de cada componente. A primeira dimensão é tipicamente com base na carga de uma molécula particular enquanto a segunda dimensão pode ser com base no tamanho de uma molécula. Na primeira dimensão, as proteínas são resolvi-  
10 das de acordo com seus pontos isoelétricos usando eletroforese de gradiente de pH imobilizado (IPGE), focalização isoelétrica (IEF) ou eletroforese de gradiente de pH de não-equilíbrio. Sob condições padrão de temperatura e concentração de uréia, os pontos de focalização observados da grande maioria das proteínas estreitamente aproxima-se dos pontos isoelétricos prognosticados calculados das composições de aminoácido das proteínas. Em  
15 geral, a primeira etapa após a preparação de uma amostra do hospedeiro inclui operar a amostra junto de um gradiente de pH, um processo conhecido como focalização isoelétrico. Os gradientes de pH podem ser gerados acrescentando anfólitos a um gel de acrilamida. Esta é uma mistura de espé-  
20 cies anfotéricas com uma faixa de valores de pl. Os gradientes de pH podem também ser gerados adicionando Immobiline que é similar aos anfólitos mas foram imobilizados dentro do gel de poliacrilamida produzindo um gradiente de pH imobilizado que não necessita ser pré-focalizado.

A segunda dimensão em eletroforese bidimensional pode ser separação por tamanho de proteínas. As proteínas podem ser separadas de acordo com seu peso molecular aproximado usando poli-acrilamida-eletroforese de dodecil sulfato de sódio (SDS-PAGE). A técnica é extensamente usada e conhecida na técnica. A idéia básica é revestir as proteínas com um detergente (SDS) que reveste todas as proteínas em uma amostra e negati-  
25 vamente as carrega. As proteínas são depois submetidas à eletroforese em gel. Os géis podem tipicamente ser géis de acrilamida e pode estar em um gradiente de densidade. A carga colocada no gel empurra as proteínas a-

través do gel com base no tamanho. Em eletroforese bidimensional, as proteínas separadas podem incluir proteínas de pelo menos 10% do proteoma do organismo. Mais tipicamente, as proteínas de pelo menos 20%, 30%, 40%, 60%, 80% ou 90% das proteínas no proteoma da célula hospedeira são separadas e analisadas através de técnicas como tingimento de proteínas e/ou espectrometria de massa.

O processo da invenção pode também incluir analisar o perfil de proteoma usando um microarranjo. Nesta modalidade, o microarranjo pode incluir os pares de ligação para pelo menos uma porção das proteínas expressas pela célula hospedeira sob condições de crescimento apropriadas, e tipicamente inclui os pares de ligação para proteínas de pelo menos 5% do proteoma do organismo. Mais tipicamente, o microarranjo inclui os pares de ligação para proteínas de pelo menos 10%, 20%, 30%, 40%, 60%, 80% ou 90% das proteínas no proteoma da célula hospedeira. Os pares de ligação podem ser anticorpos que podem ser fragmentos de anticorpo como fragmentos de anticorpo de cadeia simples. Os pares de ligação podem também incluir aptâmeros que são moléculas incluindo ácidos nucléicos que ligam às proteínas específicas ou porções de proteínas. Em uma modalidade separada, o microarranjo pode incluir pares de ligação para um subconjunto selecionado de proteínas do proteoma, incluindo, por exemplo, proteínas de protease putativas ou moduladores de dobramento putativos. O microarranjo pode tipicamente também incluir um conjunto de pares de ligação para proteínas que são usadas como controles. O perfil genético pode ser analisado medindo a ligação das proteínas da célula hospedeira que expressa a proteína ou peptídeo recombinante aos pares de ligação no microarranjo.

O formato de arranjo de proteína mais simples em geral consiste em um número grande de reagentes de captura de proteína ligados às manchas definidas em um material de suporte planar. Este arranjo é depois exposto a uma amostra de proteína complexa. A ligação das proteínas de analito específicas às manchas individuais pode depois ser monitorada usando métodos diferentes. Em casos onde os analitos foram pré-marcados com uma tintura fluorescente, a ligação pode ser diretamente monitorada

usando um escâner de fluorescência. Freqüentemente o anticorpo no formato do tipo sanduíche clássico é usado em que dois reagentes de ligação de proteína simultaneamente ligam ao mesmo antígeno: um anticorpo é imobilizado sobre a superfície, e o outro é fluorescentemente marcado ou conjugado com uma enzima que pode produzir um produto fluorescente, lúminescente ou colorido quando fornecido com o substrato apropriado.

Anticorpos monoclonais ou seus fragmentos de ligação de antígeno são correntemente uma escolha para agentes de captura devido a sua especificidade alta, afinidade e estabilidade. Eles foram usados em uma variedade de ensaios de perfilação de proteína de analito simples clássicos como ensaios imunoabsorventes ligados à enzima (ELISA) desde os anos setenta. Adicionalmente, bibliotecas de exibição de fago dos fragmentos de anticorpo oferecem o potencial para produção de anticorpo em escalas proteômicas. Estas bibliotecas podem ser usadas para isolar os agentes de ligação de afinidade alta contra proteína identificada em um prazo significativamente mais curto que é possível com processos com base em imunização. Exibição de ribossoma e exibição de mRNA são processos adicionais, completamente *in vitro*, que contam ligação física das proteínas de biblioteca a suas seqüências de mRNA de codificação. Tais processos têm sido usados com sucesso para selecionar reagentes de ligação de afinidade alta para proteínas identificadas (Wilson, DS, et al. (2001) *The use of mRNA display to select high-affinity protein-binding peptides* *Proc Natl Acad Sci USA* 98:3750-3755). Vários grupos tomaram um método diferente para desenvolver reagentes de captura de proteína de afinidade alta para biofatias de proteína. Por exemplo, os aptâmeros foram usados, que são moléculas de RNA ou DNA unifilamentares que originam de experimentos de seleção *in vitro* (denominados SELEX: evolução sistemática de ligandos por enriquecimento exponencial) com afinidades altas para proteínas. Um outro desenvolvimento em tecnologias de aptâmero é assim-chamado fotoaptâmeros. Estas moléculas têm um atributo adicional que intensifica sua utilidade como reagentes de captura de proteína. Elas carregam o grupo de reticulação fototativável 5'-bromodeoxiuridina que, quando ativado por luz UV, pode causar

reticulação covalente com as proteínas identificadas ligadas (Petach, H & Gold, L (2002) Dimensionality is the issue: use of photoaptamers in protein microarrays *Curr Opin Biotechnol* 13:309-314). O evento de foto-reticulação fornece uma segunda dimensão de especificidade similar à ligação de um 5 anticorpo de detecção secundário em um imunoensaio em sanduíche.

Uma ampla variedade de substratos de superfície e químicas de ligação foi avaliada pela imobilização de agentes de captura em microarranjo de proteína. Um modo para imobilizar as proteínas em um suporte sólido conta com interações não-covalentes com base em hidrofóbico ou interações 10 de van der Waals, ligação de hidrogênio ou forças eletrostáticas. Exemplos de imobilização eletrostática incluem o uso de materiais como lâminas de vidro revestidas com nitrocelulose e polilisina ou aminopropil silano. Microarranjos de proteína foram também fabricados por meio de adsorção física sobre as superfícies de plástico de placas de 96 cavidades. Um exemplo de ligação 15 covalente de proteínas à superfície foi descrito por MacBeath e Schreiber (MacBeath, G & Schreiber, SL(2000) Printing proteins as microarrays for high-throughput function determination *Science* 289:1760-1763). Devido à afinidade muito alta de estreptavidina para biotina, a imobilização de proteínas biotiniladas sobre as superfícies de estreptavidina pode ser considerada 20 quase covalente (Peluso, P et al. (2003) Optimizing antibody immobilization strategies for the construction of protein microarrays *Anal Biochem* 312:113-124). Outras estratégias foram descritas (X-ray photoelectron spectroscopy and radiometry studies of biotin-derivatized poly(L-lysine)-grafted-poly(ethylene glycol) monolayers on metal oxides (*Langmuir*) 7313-7322; 25 Ruiz-Taylor, LA et al. (2001) Monolayers of derivatized poly(L-lysine)-grafted poly(ethylene glycol) on metal oxides as a class of biomolecular interfaces *Proc Natl Acad Sci U S A* 2001, 98:852-857; Espejo A, Bedford MT. (2004) Protein-domain microarrays Processes Mol Biol. 264:173-81; Zhu, H. et al. (2001) Global analysis of protein activities using proteome chips. *Science* 30 Express).

As amostras no perfil genético podem ser analisadas individualmente ou agrupados em agrupamentos. Os agrupamentos podem tipica-

mente ser agrupados por similaridade em expressão de gene. Em uma modalidade, os agrupamentos podem ser individualmente agrupados como proteínas que são reguladas a uma extensão similar em uma célula hospedeira. Os agrupamentos podem também incluir grupos de proteínas que são 5 reguladas a uma extensão similar em uma célula hospedeira recombinante, por exemplo, que é sobre-regulado ou sub-regulado a uma extensão similar comparada a uma célula hospedeira ou uma célula modificada ou uma inalterada. Os agrupamentos podem também incluir grupos relacionados por estrutura de proteína, função ou processamento. Grupos de pares de ligação 10 de proteína em um arranjo, ou grupos de proteínas analisados em um ensaio diferente como eletroforese bidimensional, podem ser selecionados de, mas não são limitados a: proteases putativas ou conhecidas, co-fatores de proteases ou proteínas similares à protease; moduladores de dobramentos, co-fatores de moduladores de dobramentos ou proteínas que poderiam melhorar 15 o dobramento ou solubilidade da proteína; fatores de transcrição; proteínas envolvidas na estabilidade de ácido nucléico ou iniciação translacional; cinases; receptores extracelulares ou intracelulares; enzimas metabólicas; co-fatores metabólicos; proteínas de envelope; e genes de manutenção.

### METABOLOMA

20 Processos de análise proteômica permitem determinar a abundância e distribuição de muitas proteínas simultaneamente. Porém, as consequências funcionais de alterações para o proteoma são relatadas apenas indiretamente. Outro método é medir os níveis destas moléculas pequenas, ou metabólitos. Um perfil genético analisado no processo da invenção pode 25 desse modo incluir um perfil metabolômico. Processos para analisar o metaboloma de um hospedeiro específico incluem cromatografia de gás, cromatografia líquida de pressão alta e eletroforese capilar para separar metabólitos de acordo com várias propriedades química e físicas. As moléculas podem depois ser identificadas usando processos como espectrometria de 30 massa.

### DETECÇÃO/ANÁLISE

O processo inclui analisar um perfil genético para identificar um

gene ou produto de gene compensatório que é expressado a um nível mais alto na célula recombinante. Em geral, esta etapa inclui o monitoramento da expressão (por exemplo detecção e ou quantificação da expressão) de uma multidão de genes ou produtos de gene. A expressão é em geral monitorada detectando a ligação dos produtos de gene de célula hospedeira para um perfil de transcriptoma, proteoma ou de metaboloma como descrito acima. A análise da ligação pode envolver uma comparação da ligação entre uma célula hospedeira recombinante que expressa proteína ou peptídeo recombinante e uma célula hospedeira pura ou uma célula hospedeira recombinante que não expressa a proteína ou peptídeo.

#### DETECÇÃO

Esta etapa inclui o monitoramento da expressão (por exemplo detecção e ou quantificação da expressão) de uma multidão de genes ou produtos de gene. A expressão é em geral monitorada detectando a ligação dos produtos de gene de célula hospedeira a um perfil de transcriptoma, proteoma ou metaboloma como descrito acima. Tipicamente, pelo menos cerca de 10 genes, ou pelo menos cerca de 100, ou pelo menos cerca de 1000 e ou pelo menos cerca de 10.000 genes diferentes podem ser ensaiados de uma vez. O processo pode envolver fornecer um fundo geral de ácidos nucléicos identificados comprendendo transcrições de RNA de um ou mais dos ditos genes, ou ácidos nucléicos derivados das transcrições de RNA; hibridar o fundo geral de ácidos nucléicos com um arranjo de sondas de oligonucleotídeo imobilizadas em uma superfície onde o arranjo comprehende mais de 100 oligonucleotídeos diferentes e cada oligonucleotídeo diferente está localizado em uma região predeterminada da dita superfície, cada oligonucleotídeo diferente é ligado à superfície através de pelo menos uma ligação covalente, e as sondas de oligonucleotídeo são complementares às transcrições de RNA ou ácidos nucléicos derivados das transcrições de RNA; e quantificar os ácidos nucléicos hibridados no arranjo. Uma representação pictórica de uma técnica para monitorar a expressão de um produto de gene entre duas amostras é descrita na Figura 12.

O processo pode também envolver fornecer um fundo geral de

proteínas celulares. Estes podem ser derivados de lisados celulares que são feitos através de células de lise usando detergentes ou tensoativos; usar lise osmótica; usar alterações térmicas, como ciclos de congelação-descongelamento; usar dispositivos mecânicos ou usar alterações de pressão. Tipicamente as químicas são incluídas no processo de lise de uma célula ou sistema de células que inibem certas proteínas, como proteases, particularmente proteases não-específicas, para limitar a degradação das proteínas.

5 Além disso, lisados de célula são tipicamente mantidos em ou abaixo de 4°C, e podem ser mantidos em ou abaixo de 0°C ou em ou abaixo de 20°C durante o processamento. Os lisados de célula podem ser separados antes de outro processamento, por exemplo por cromatografia de exclusão de tamanho, permuta de íons ou cromatografia de matriz de afinidade como usando HPLC.

10

Tipicamente, o produto genético identificado, mRNA, cDNA, proteína ou metabólito é marcado com um marcador ou sonda detectável. O marcador ou sonda pode ser uma ou mais moléculas fluorescentes ou fluoróforos. Estes podem incluir moléculas comercialmente disponíveis como Cy3 e Cy5 ligados, por exemplo, a nucleotídeos particulares que podem ser incorporados em um cDNA transcritos reverso para fornecer moléculas detectáveis para triagem. Em uma modalidade, uns primeiros fluoróforos são incorporados em uma amostra do hospedeiro e um segundo fluoróforo é incorporado em uma amostra de um hospedeiro expressando proteína ou peptídeo recombinante. Em uma modalidade, o primeiro fluoróforo e segundo fluoróforo emitem diferentes comprimentos de onda de luz. Nesta modalidade, a ligação das amostras do hospedeiro e do hospedeiro expressando proteína recombinante pode ser monitorada no mesmo ensaio. Em outra modalidade, os fluoróforos são excitados com diferentes comprimentos de onda de luz. Em outra modalidade, o primeiro e segundo fluoróforo é excitado ou emite luz no mesmo comprimento de onda. Nesta modalidade, as amostras do hospedeiro e da proteína recombinante expressando hospedeiro são tipicamente monitoradas em ensaios diferentes.

15

20

25

30

O processo pode adicionalmente incluir uma etapa de quantificar a hibridização dos ácidos nucléicos ou proteínas ou metabólitos químicos

identificados. A quantificação pode incluir medição dos níveis de transcrição de um ou mais genes. Tipicamente o fundo geral de ácidos nucléicos identificados é, por exemplo, um em que a concentração dos ácidos nucléicos identificados (transcrições de pré-mRNA, transcrições de mRNA ou ácidos nucléicos derivados das transcrições de RNA) é proporcional aos níveis de expressão de genes que codificam aqueles ácidos nucléicos identificados.

Para análise de transcriptoma, o fundo geral de ácidos nucléicos pode ser marcado antes, durante ou após hibridização, embora os ácidos nucléicos sejam tipicamente marcados antes da hibridização. Marcações de fluorescência são tipicamente usadas, freqüentemente com um fluoróforo simples, e, onde marcação de fluorescência for usada, quantificação dos ácidos nucléicos hibridados pode ser por quantificação de fluorescência do ácido nucléico hibridado fluorescentemente marcado. Tal quantificação é facilitada pelo uso de um escâner a laser confocal ou microscópio de fluorescência, como um microscópio de fluorescência confocal que pode ser equipado com um estágio automatizado para permitir varredura automática do arranjo e que pode ser equipado com um sistema de aquisição de dados para registro de medição automatizado e processamento subsequente da informação de intensidade de fluorescência. Dispositivos para ler tais arranjos incluem os módulos de CloneTracker®, ImaGene®, GeneSight® e a base de dados de GeneDirector®, disponível de Biodiscovery, Inc., El Segundo, Calif., ou a leitora GeneChip®, disponível de Affymetrix, Santa Clara Inc., Calif. Em uma modalidade, hibridização ocorre em severidade baixa (por exemplo cerca de 20°C a cerca de 50°C, ou cerca de 30°C a cerca de 40°C, ou cerca de 25 37°C). Hibridização pode incluir lavagens subsequentes em severidade progressivamente crescente até um nível desejado de especificidade de hibridização ser alcançado.

Quantificação do sinal de hibridização pode ser por quaisquer dispositivos conhecidos a alguém versado na técnica. Porém, em uma modalidade, quantificação é alcançada pelo uso de um escâner de fluorescência confocal. Dados são tipicamente avaliados calculando a diferença em intensidade de sinal de hibridização entre cada sonda de oligonucleotídeo e sua

sonda de controle de disparidade correspondente. Tipicamente, esta diferença pode ser calculada e avaliada para cada gene. Certos processos analíticos são fornecidos aqui.

Técnicas foram desenvolvidas para preparar sondas de hibridização bacterianas apropriadas (ver, por exemplo Choi et al. (2003) *App. Envir. Microbic.* 69:4737-4742). Por exemplo, as células podem ser armazenadas em um agente estabilizante de RNA como RNAlater (Ambion, Austin, TX). RNA é em geral purificada em três etapas: (1) isolamento do RNA total, (2) remoção do DNA contaminante e (3) limpeza total do RNA total. RNA total pode ser isolado e depois misturado com iniciadores de hexâmero aleatórios e transcriptase reversa para fazer o cDNA. Tipicamente pelo menos uma sonda fluorescente é incorporada no cDNA. Em uma modalidade, uma sonda fluorescente é incorporada, em outra modalidade mais de uma sonda, por exemplo 2, 3, 4, 5 ou mais sondas fluorescentes são incorporadas nas mesmas amostras de cDNA ou diferentes. Em um hospedeiro eucariótico, o fundo geral dos ácidos nucléicos identificados pode também ser o mRNA de poliA+ total isolado de uma amostra biológica, ou cDNA feito por transcrição reversa do RNA ou cDNA de segundo filamento ou RNA transcrita do intermediário de cDNA bifilamentar.

Corantes fluorescentes são tipicamente incorporados em moléculas de cDNA durante a reação de transcrição reversa. Devido à estrutura de mRNA diferente entre procariotes (bactérias) e eucariotes (levedura, células mamíferas, etc.), iniciadores diferentes podem ser usados, porém iniciadores aleatórios podem ser usados em ambos os casos, e iniciadores de oligo-dT podem ser usados em eucariotes que tem caudas de poliA. Um processo alternativo é marcação de amino-alila para aumentar a intensidade de sinal. Este processo incorpora análogos de nucleotídeo caracterizando um grupo quimicamente reativo ao qual um corante fluorescente pode ser ligado após a reação de transcrição reversa (Manduchi, E., et al. (2002) Comparison of different labeling processes for two-channel high-density micro array experiments. *Physiol Genomics* 10:169-79).

O fundo geral de ácidos nucléicos identificados pode ser tratado

para reduzir a complexidade da amostra e assim reduzir o sinal de base obtido na hibridização. Os termos "base" ou "sinal de base" referem-se aos sinais de hibridização resultantes da ligação não-específica, ou outras interações, entre os ácidos nucléicos identificados marcados e os componentes do arranjo de oligonucleotídeos (por exemplo, as sondas de oligonucleotídeo, sondas de controle, o substrato de arranjo, etc.). Em um método, um fundo geral de mRNAs, derivado de uma amostra biológica, é hibridado com um fundo geral de oligonucleotídeos compreendendo as sondas de oligonucleotídeo presentes no arranjo. O fundo geral de ácidos nucléicos hibridado é depois tratado com RNase A que digere as regiões unifilamentares. Os complexos de hibridização bifilamentares restantes são depois desnaturados e as sondas de oligonucleotídeo são removidas, deixando um fundo geral de mRNAs intensificado para aqueles mRNAs complementares às sondas de oligonucleotídeo no arranjo.

Em outro método para redução de base, um fundo geral de mRNAs derivado de uma amostra biológica é hibridado com oligonucleotídeos identificados emparelhados específicos onde os oligonucleotídeos identificados emparelhados específicos são complementares às regiões flanqueando as subseqüências dos mRNAs complementares às sondas de oligonucleotídeo no arranjo. O fundo geral de ácidos nucléicos hibridado é tratado com RNase H que digere as seqüências de ácido nucléico hibridadas (bifilamentares). As seqüências de ácido nucléico unifilamentares restantes que têm um comprimento cerca de equivalente à região flanqueada pelos oligonucleotídeos identificado emparelhados específicos são depois isoladas (por exemplo através de eletroforese) e usadas como o fundo geral de ácidos nucléicos para monitorar a expressão de gene.

Um terceiro método para redução de base envolve eliminar ou reduzir a representação no fundo geral de mensagens de mRNA identificadas pré-selecionadas particulares (por exemplo, mensagens que são sobre-expressadas characteristicamente na amostra). Este processo envolve hibridar com uma sonda de oligonucleotídeo que é complementar à mensagem pré-selecionada identificada de mRNA ao fundo geral de mRNAs de poliA+

derivados de uma amostra biológica. A sonda de oligonucleotídeo hibrida com o mRNA de poliA+ pré-selecionado particular ao qual é complementar. O fundo geral de ácidos nucléicos hibridado é tratado com RNase H que digere a região bifilamentar (hibridada) assim separando a mensagem de sua cauda de poliA+. Isolamento ou amplificação (por exemplo, usando uma coluna de oligo dT) do mRNA de poliA+ no fundo geral depois fornece um fundo geral que tem uma representação reduzida ou nenhuma da mensagem de mRNA pré-selecionada identificada.

#### ANÁLISE

O gene identificado é tipicamente identificado comparando um perfil genético da célula hospedeira que expressa a proteína ou peptídeo recombinante a um perfil genético da célula hospedeira que não expressa a proteína ou peptídeo recombinante. Em modalidades iterativas, o gene identificado a ser modificado é identificado comparando um perfil genético da célula que será modificada (a segunda célula) à célula da qual foi modificada (a primeira célula). O gene identificado é identificado comparando um perfil genético da segunda célula a um perfil genético da primeira célula e identificando um ou mais genes da expressão da qual é aumentado na segunda célula.

Microarranjos de cDNA medem a abundância de mRNA relativa entre duas amostras. Uma série de amostra de ponto de tempo de pós-indução pode ser comparada à amostra de pré-indução para a mesma cepa (perfil de expressão temporal), ou as amostras de pós-indução podem ser comparadas com as cepas diferentes no mesmo ponto de tempo. A comparação pode ser através do uso de um programa de computação, como GeneSight®. Por exemplo, quando usar um microarranjo usando um marcador fluorescente, uma intensidade de mancha pode ser medida para cada amostra ligada ao arranjo (por exemplo uma seqüência de DNA). A intensidade de mancha pode depois ser corrigida para base e a razão da intensidade para as amostras do hospedeiro versus o hospedeiro expressando a proteína ou peptídeo recombinante, ou para o hospedeiro expressando a proteína ou peptídeo recombinante comparado ao hospedeiro modificado que expressa a

proteína ou peptídeo recombinante, pode ser medido. A razão fornece uma medida para identificar os genes que são sobre-regulados ou a expressão da qual é aumentado sob expressão da proteína ou peptídeo recombinante, ou sob modificação da célula hospedeira para permitir identificação de um gene identificado.

Para identificar se um gene é sobre-regulado, uma razão padrão ou "corte" é estabelecida. A razão de corte pode ser projetada para superar os efeitos de ruído de base associados a um ensaio particular. Em geral, qualquer razão de mais que 1 entre as medições pode designar um gene sobre-regulado. Porém, variação entre os ensaios pode requerer uma razão mais alta que 1, por exemplo 1,5, ou mais que 2, ou mais que 2,5, ou mais que 3, ou mais que 3,5 ou mais que 4 ou mais que 4,5, ou mais que 5 ou mais que 6, ou mais que 7, ou mais que 8, ou mais que 9 ou mais que 10. O padrão pode ser estabelecido antes do processo, contando com padrão conhecidos na técnica, ou pode ser estabelecido durante as medições comparando as razões de níveis de genes ou produtos de gene de controle, como genes de manutenção.

ETAPA III: EXPRESSÃO VARIÁVEL DO GENE OU PRODUTO DE GENE COMPENSATÓRIO IDENTIFICADO GENETICAMENTE MODIFICANDO A CÉLULA PARA FORNECER UMA CÉLULA RECOMBINANTE MODIFICADA QUE ALCANÇA UM AUMENTO NA EXPRESSÃO DE PROTEÍNA RECOMBINANTE, ATIVIDADE OU SOLUBILIDADE.

GENES COMPENSATÓRIOS IDENTIFICADOS

Os genes ou produtos de gene compensatórios que são identificados na etapa ii), ou análogos homólogos, co-fatores ou subunidades destes, são usados para projetar estratégias para geneticamente modificar a célula ou aumentar, diminuir, knock-in ou knock-out a expressão de um ou mais genes identificados. As seqüências de gene identificadas podem ser usadas em bases de dados públicos para projetar estratégias, particularmente projetar constructos para modular expressão de um gene através das técnicas descritas acima. Tais técnicas são bem conhecidas.

Em uma modalidade, o gene ou genes identificados são pelo

menos uma protease putativa, uma proteína similar à protease, um co-fator ou subunidade de uma protease. Em outras modalidades, o gene ou genes identificados são pelo menos um modulador de dobramento, modulador de dobramento putativo, co-fator ou subunidade de um modulador de dobramento. Em certas modalidades, um gene identificado é uma subunidade de uma protease. Em uma modalidade, o gene ou genes identificados podem ser uma serina, treonina, cisteína, aspártico ou metalo peptidase. Em uma modalidade, o gene ou genes identificados podem ser selecionados de *hsIV*, *hsIU*, *c/pA*, *c/pB* e *c/pX*. O gene identificado pode também ser um co-fator de uma protease. Em outra modalidade, o gene ou genes identificados são um modulador de dobramento. Em algumas modalidades, o gene ou genes identificados podem ser selecionados de uma proteína de chaperona, uma foldase, uma peptidil prolil isomerase e uma isomerase de ligação de dissulfeto. Em algumas modalidades, o gene ou genes identificados podem ser selecionados de *htpG*, *cbpA*, *dnaJ*, *dnaK* e *fkbP*.

Genes bacterianos são organizados em operons que são agrupamentos de gene que codificam as proteínas necessárias para executar a função coordenada como biossíntese de um aminoácido dado. Portanto, em uma modalidade, o gene identificado faz parte de um operon. Em uma modalidade particular, o gene identificado é um operon que codifica para uma ou mais proteínas com atividade de protease sozinho ou em combinação, ou é um operon que codifica para uma ou mais proteínas com atividade moduladora de dobramento, incluindo foldases, chaperonas, e isomerases.

### PROTEASES

Em uma modalidade da invenção, a célula hospedeira é modificada reduzindo a expressão, inibindo ou removendo pelo menos uma protease do genoma. A modificação pode também ser mais de uma protease em algumas modalidades. Em uma modalidade relacionada, a célula é modificada reduzindo a expressão de um co-fator de protease ou proteína de protease. Em outra modalidade, a célula hospedeira é modificada por inibição de um promotor para uma protease ou proteína relacionada que pode ser um promotor nativo. A modificação de gene pode ser para modular uma proteína

homóloga ao gene identificado.

Na base de dados de MEROPS, peptidases são agrupadas em clãs e famílias. As famílias são grupos de peptidases funcionalmente similares estreitamente relacionadas. Famílias são agrupadas por seu tipo catalítico: S, serina; T, treonina; C, cisteína; A, aspártico; M, metalo e U, desconhecido.

5 Mais de 20 famílias (denotadas S1 - S27) de serina protease foram identificadas, estas são agrupadas em 6 clãs (SA, SB, SC, SE, SF e SG) em base de similaridade estrutural e outra evidência funcional. Estruturas são conhecidas por quatro dos clãs (SA, SB, SC e SE). Treonina peptidases são caracterizadas por um nucleófilo de treonina no término N da enzima madura. O exemplo de tipo para este clã é o componente beta de proteasoma arqueano de *Thermoplasma acidophilum*. Cisteína peptidases têm topologias moleculares características e são peptidases em que o nucleófilo é o grupo sulfidrila de um resíduo de cisteína. Cisteína proteases são divididas em clãs (proteínas que são evolutivas relacionadas), e também sub-divididas em famílias, 10 em base da arquitetura de seu díade ou triade catalítico:

15

Clã CA contém as famílias de papaína (C1), calpaína (C2), estreptopaína (C10) e as peptidases ubiquitina-específicas (C12, C19), como também muitas famílias de endopeptidases de cisteína viral.

20 Clã CD contém as famílias de clostriplaína (C11), gingipaína R (C25), legumaína (C13), caspase-1 (C14) e separina (C50). Estas enzimas têm especificidades dominadas pelas interações do sub-sítio 51.

Clã CE contém as famílias de adenaína (C5) de adenovírus, a Ulp1 protease eucariótica (C48) e as YopJ proteases bacterianas (C55).

25 Clã CF contém apenas piroglutamil peptidase I (C15).

Clã PA contém as picornaínas (C3), que provavelmente evoluíram das serina peptidases e que formam a maior parte das enzimas neste clã.

Clãs PB e CH contêm as cisteína peptidases autolíticas.

Endopeptidases aspárticas de origem vertebrada, fúngica e reproviral foram caracterizadas. Aspartato peptidases são assim nomeadas porque os resíduos de Asp são os ligantes da molécula de água ativada em todos os exemplos onde os resíduos catalíticos foram identificados, embora

pelo menos uma enzima viral seja acreditada ter como Asp e uma Asn como sua diade catalítica. Todas ou a maioria das aspartato peptidases são endopeptidases. Estas enzimas foram atribuídas em clãs (proteínas que são evolutivas relacionadas), e também sub-divididas em famílias, em grande parte em base de sua estrutura terciária.

Metaloproteases são as mais diversas dos quatro tipos principais de protease, com mais de 30 famílias identificadas até agora. Nestas enzimas, um cátion divalente, usualmente zinco, ativa a molécula de água. O íon de metal é mantido no lugar por ligantes de aminoácido, usualmente três em número. Os ligantes de metal conhecidos são His, Glu, Asp ou Lys e pelo menos um outro resíduo é requerido para catálise que pode representar um papel eletrofílico. Das metaloproteases conhecidas, quase a metade contém um motivo de HEXXH que foi mostrado em estudos cristalográficos fazer parte do sítio de ligação de metal. O motivo de HEXXH é relativamente comum, mas pode ser mais estritamente definidos para metaloproteases como abXHEbbHbc onde 'a' é mais freqüentemente valina ou treonina e faz parte do sub-sítio do S1 em termolisina e neprilisina, 'b' é um resíduo descarregado, e 'c' um resíduo hidrofóbico. Prolina nunca é encontrada neste sítio, possivelmente porque quebraria a estrutura helicoidal adotada por este motivo em metaloproteases.

As peptidases associadas ao clã U têm um mecanismo catalítico desconhecido como a dobra de proteína do domínio de sítio ativo e os resíduos de sítio ativo não foram relatados.

Certas proteases (por exemplo OmpT) podem adsorver na superfície dos corpos de inclusão e podem degradar a proteína desejada enquanto estiver sendo redobrada. Portanto, certas proteínas identificadas podem ser proteases ou proteínas de protease que aderem aos corpos de inclusão e estes podem ser modificados, por exemplo, para reduzir a ligação.

Proteases ou proteínas de protease podem também ser classificadas como Aminopeptidases; Dipeptidases; Dipeptidil-peptidases e tripeptidil peptidases; Peptidil-dipeptidases; Carboxipeptidases do tipo serina-tipo; Metalocarboxipeptidases; Carboxipeptidases do tipo cisteína; Omegapepti-

dases; Serina proteinases; Cisteína proteinases; Aspártico proteinases; Metalo proteinases; ou Proteinases de mecanismo desconhecido.

Aminopeptidases incluem citosol aminopeptidase (leucil aminopeptidase), alanil aminopeptidase de membrana, cistinil aminopeptidase, 5 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-estereospecífica, 10 aminopeptidase ey. Dipeptidases incluem x-his dipeptidase, x-arg dipeptidase, x-metil-his dipeptidase, cis-gli 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. Dipeptidil-peptidases e tripeptidil peptidases incluem 15 dipeptidil-peptidase i, dipeptidil-peptidase ii, dipeptidil peptidase iii, dipeptidil-peptidase iv, dipeptidil-dipeptidase, tripeptidil-peptidase I, tripeptidil-peptidase II. Peptidil-dipeptidases incluem peptidil-dipeptidase a e peptidil-dipeptidase b. Carboxipeptidases do tipo serina incluem pro-x carboxipeptidase lisossomal, D-ala-D-ala carboxipeptidase do tipo serina, carboxipeptidase C, carboxipeptidase D. Metalocarboxipeptidases incluem carboxipeptidase a, carboxipeptidase B, lisina(arginina) carboxipeptidase, gly-X carboxipeptidase, alanina carboxipeptidase, muramolpentapeptídeo carboxipeptidase, carboxipeptidase h, glutamato carboxipeptidase, carboxipeptidase M, muramoltetrapeptídeo carboxipeptidase, zinco d-ala-d-ala carboxipeptidase, 20 carboxipeptidase A2, pro-x carboxipeptidase de membrana, tubulinil-tir carboxipeptidase, carboxipeptidase t. Omegapeptidases incluem acilaminoacil-peptidase, peptidil-glicinamidase, piroglutamil-peptidase I, beta-aspartil-peptidase, piroglutamil-peptidase II, n-formilmetionil-peptidase, pteroilpolili-[gama]-glutamato carboxipeptidase, gama-glu-X carboxipeptidase, acilmuramoil-ala peptidase. Serina proteinases incluem quimiotripsina, quimiotripsina c, metridina, tripsina, trombina, Fator Xa de coagulação, plasmina, enteropeptidase, acrosina, protease alfa-lítica, glutamila, endopeptidase, ca-

tepsina 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 de complemento c1r55, componente de complemento c1s55, convertase c3/c5 da via de complemento clássico, fator de complemento I, fator de complemento D, convertase c3/c5 da via de complemento alternativa, cerevisina, hipodermina C, lisil endopeptidase, endopeptidase 1a, gama-reni, venombin ab, leucil endopeptidase, triptase, escutelarina, quexina, subtilisina, orizina, endopeptidase k, termomicolina, termitase, endopeptidase SO, ativador de plasminogênio T, proteína C, endopeptidase pancreática E, elastase pancreática ii, serina endopeptidase IGA-específica, U-plasminogênio, ativador, venombina A, furina, mieloblastina, semenogelase, granzima A ou proteinase de linfócito T citotóxico 1, granzima B ou proteinase de linfócito T citotóxico 2, esteptogrisina A, treptogrisina B, glutamil endopeptidase II, oligopeptidase B, fator de coagulação de límulo c, fator de coagulação de límulo, enzima de coagulação de límulo, omptina, repressor lex, peptidase líder bacteriana I, togavirina, flavirina. Cisteína proteinases incluem catepsina B, papaína, ficina, quimiopapaína, asclepaína, clostriplaína, esteptopalaína, actinídeo, catepsina 1, catepsina H, calpaína, catepsina t, glicila, endopeptidase, pró-coagulante de câncer, catepsina S, picornaína 3C, picornaína 2A, caricaína, ananaína, bromelaína de talo, bromelaína de fruta, legumaína, histolisaína, enzima de conversão de interleucina 1-beta. Aspártico proteinases incluem pepsina A, pepsina B, gastricsina, quimiosina, catepsina D, neopentesina, renina, retropepsina, enzima de conversão de pró-opiomelanocortina, aspergilopepsina I, aspergilopepsina II, penicilopepsina, rizopuspepsina, endotiapepsina, mucoropepsina, candidapepsina, sacaropepsina, rodotorulapepsina, fisaropepsina, acrocilindropepsina, poliporopepsina, picnoporopepsina, escitalidopepsina a, escitalidopepsina b, xantomonapepsina, catepsina e pepsina de barreira, peptidase líder bacteriana I, pseudomonapepsina, plasmepsina. Metalo proteinases incluem atrolisina a, colagenase microbiana, leucolisina, colagenase intersticial, neprilisina, envelisina, metaloendopeptidase iga-específica, procolágeno

N-endopeptidase, timet oligopeptidase, neurolisina, estromelisina 1, meprina A, procolágeno C-endopeptidase, peptidil-lis metaloendopeptidase, astacina, estromelisina, 2, matrilisina gelatinase, aeromonolisina, pseudolisina, termolisina, bacilosina, aureolisina, cocolisina, micolisina, metaloendopeptidase beta-lítica, peptidil-asp metaloendopeptidase, neutrófilo colagenase, gelatinase B, leishmanolisina, sacarolisina, autolisina, deuterolisina, serralisina, atrolisina B, atrolisina C, atroxase, atrolisina E, atrolisina F, adamalisina, horrilisina, ruberlisina, botropasina, botrolisina, ofiolisina, trimerelisina I, trimere-lisina II, mucrolisina, pitrilisina, insulisina, O-sialoglicoproteína endopeptidase, russelisina, mitocondrial, intermediário, peptidase, dactilisina, nardilisina, magnolisina, meprina B, peptidase de processamento mitocondrial, elastase de macrófago, coriolisina, toxilisina. Proteinases de mecanismo desconhecido incluem termopsina e complexo de endopeptidase multicatalítica.

Certas proteases de *P. fluorescens* são listadas na Tabela A.

15

TABELA A

Classe	Família	RXF	Função Curada	Gene	Fisiologia
Homólogos de MEROPS					
Aspártico peptidases					
	A8 (família de peptidase II de sinal)				
		RXF05383	Peptidase de sinal de lipo-proteína (ec 3.4.23.36)		Processamento de numerosas lipoproteínas segregadas bacterianas.
	A24 (família de prepilina peptidase do tipo IV)				
		RXF05379	Prepilina peptidase do tipo 4 pilD (ec 3.4.99.-)		Este peptidase ligado à membrana cliva um peptídeo líder especializado de prepilina do tipo 4 durante sua secreção de muitas espécies bacterianas. Uma vez segregadas, as proteínas processadas são requeridas para funções incluindo

Classe	Família	RXF	Função Curada	Gene	Fisiologia
					formação de pilus do tipo 4, toxina e outra secreção de enzima, transferência de gene e formação de biofilme.
	Cisteína peptidases				
	C15 (família de piroglutamil peptidase I)				
		RXF02161	Pirrolidona carboxilato peptidase (ec 3.4.19.3)		Remoção de grupos de piroglutamila de peptídeos no catabolismo de proteína
	C40				
		RXF01968	Proteína associada à invasão, P60		
		RXF04920	Proteína associada à invasão, P60		
		RXF04923	Proteína associada à fosfatase papq		
	C56 (família de Pfpl endopeptidase)				
		RXF01816	Protease I (ec 3.4.--)		
	Metalopeptidases				
	M1				
		RXF08773	Alanina aminopeptidase de membrana (ec 3.4.11.2)		
	M3				
		RXF00561	Oligopeptidase A (ec 3.4.24.70)	pr/C	Degradação de peptídeos de sinal de lipoproteína, e outros oligopeptídeos intracelulares. Papel na maturação de bacteriófago precursor de P22 gp7.
		RXF04631	Oligopeptidases		

Classe	Família	RXF	Função Curada	Gene	Fisiologia
			Zn-dependentes		
	M4 (família de termolisina)				
		RXF05113	Precursor de metaloprotease extracelular (ec 3.4.24.-)		
	M41 (família de FtsH endopeptidase)				
		RXF05400	Proteína de divisão celular ftsH (ec 3.4.24.-)		Papel proposto no controle de qualidade proteolítica das moléculas reguladoras e proteínas de membrana, em levedura.
	M10				
		RXF04304	Serralisina (ec 3.4.24.40)		
		RXF04500	Serralisina (ec 3.4.24.40)		
		RXF01590	Serralisina (ec 3.4.24.40)		
		RXF04495	Serralisina (ec 3.4.24.40)		
		RXF02796	Serralisina (ec 3.4.24.40)		
	M14 (família de carboxipeptidase A)				
		RXF09091	Precursor de zinco-carboxipeptidase (ec 3.4.17.-)		
	M16 (família de pitrilisina)				
		RXF03441	Proteína de síntese de co-enzima pqq (ec 3.4.99.-)		
		RXF01918	Zinco protease (ec 3.4.99.-)		
		RXF01919	Zinco protease (ec 3.4.99.-)		
		RXF03699	Peptidase de		

Classe	Família	RXF	Função Curada	Gene	Fisiologia
			processamento (ec 3.4.24.64)		
	M17 (família de leucil aminopeptidase)				
		RXF00285	Citosol amino-peptidase (ec 3.4.11.1)		Contribui para nutrição bacteriana
	M18				
		RXF07879	Aspartil amino-peptidase (ec 3.4.11.21)		
	M20				
		RXF00811	Succinildiaminopimelato dessuccinilase (ec 3.5.1.18)	<i>dapE</i>	
		RXF04052	Xaa-His dipeptidase (ec 3.4.13.3)		
		RXF01822	Precursor de carboxipeptidase G2 (ec 3.4.17.11)		
		RXF04892	N-acil-L-aminoácido amidoidrolase (ec 3.5.1.14)		
	M28 (família de aminopeptidase Y)				
		RXF03488	Precursor de proteína de conversão de isozima de fosfatase alcalina (ec 3.4.11.-)		
	M42 (família de gutamil aminopeptidase)				
		RXF05615	Aminopeptidase de desbloqueio (ec 3.4.11.-)		
	M22				
		RXF05817	O-sialoglicoproteína endopeptidase (ec		

Classe	Família	RXF	Função Curada	Gene	Fisiologia
			3.4.24.57)		
		RXF03065	Família de proteína de glicoprotease		
	M23				
		RXF01291	Endopeptidase de parede celular, família M23/M37		
		RXF03916	Proteínas de membrana relacionadas às metaloendopeptidases		
		RXF09147	Endopeptidase de parede celular, família M23/M37		
	M24				
		RXF04693	Metionina aminopeptidas (ec 3.4.11.18)		Provável papel na remoção cotranslacional de metionina N-terminal
		RXF03364	Metionina aminopeptidase (ec 3.4.11.18)		Provável papel na remoção cotranslacional de metionina N-terminal.
		RXF02980	Xaa-Pro aminopeptidase (ec 3.4.11.9)		Envolvido em modificação da proteína retrocelular, na bactéria.
	M48 (família de Ste24 endopeptidase)				
		RXF05137	Proteína de choque térmico HtpX		
		RXF05081	Zinco metaloprotease (ec 3.4.24.-)		
	M50 (família de S2P protease)				
		RXF04692	Metaloprotease de membrana		

Classe	Família	RXF	Função Curada	Gene	Fisiologia
Serina peptidases					
	S1 (família de quimiotripsina)				
		RXF01250	Protease do (ec 3.4.21.)		
		RXF07210	Protease do (ec 3.4.21.)		
	S8 (família de subtilisina)				
		RXF06755	Serina protease (ec 3.4.21.-)		
		RXF08517	Serina protease (ec 3.4.21.-)		
		RXF08627	Serina protease extracelular (ec 3.4.21.-)		
		RXF06281	Precursor de serina protease extracelular (ec 3.4.21.-)		
		RXF08978	Serina protease extracelular (ec 3.4.21.-)		
		RXF06451	Serina protease (ec 3.4.21.-)		
	S9 (família de prolil oligopeptidase)				
		RXF02003	Protease ii (ec 3.4.21.83)		
		RXF00458	Hidrolase		
	S11 (família D-Ala-D-Ala carboxipeptidase A)				
		RXF04657	D-alanil-D-alanina endopeptidase (ec 3.4.99.-)		
		RXF00670	D-alanil-D-alanina carboxipeptidase (ec 3.4.16.4)		
	S13 (família D-Ala-D-Ala peptidase C)				
		RXF00133	D-alanil-meso-diaminopimelato en-		Age na síntese e remodelagem das paredes de células

Classe	Família	RXF	Função Curada	Gene	Fisiologia
			dopeptidase (ec 3.4.-.-)		bacterianas
		RXF04960	D-ala-nil-meso-diamin opimelato endopeptidase (ec 3.4.-.-)		
		S14 (família de ClpP endopeptidase)			
		RXF04567	Subunidade proteolítica de Clp protease atp-dependente (ec 3.4.21.92)	<i>clpP</i>	Suposto contribuir para eliminação de proteínas prejudiciais no calor térmico
		RXF04663	Subunidade proteolítica de Clp protease atp-dependente (ec 3.4.21.92)	<i>clpP</i>	Suposto contribuir para eliminação de proteínas prejudiciais no calor térmico
		S16 (família de Lon protease)			
		RXF04653	Protease atp-dependente La (3.4.21.53)		Suposto contribuir para eliminação de proteínas prejudiciais no calor térmico
		RXF08653	Protease atp-dependente La (3.4.21.53)		
		RXF05943	Protease atp-dependente La (3.4.21.53)		
		S24 (família de LexA)			
		RXF00449	Repressor de LexA (ec 3.4.21.88)		
		RXF03397	Repressor de LexA (ec 3.4.21.88)		
		S26 (família de peptidase de sinal I)			
		RXF01181	Peptidase de sinal I (ec 3.4.21.89)		Cliva os peptídeos de sinal das proteínas segregadas
		S33			

Classe	Família	RXF	Função Curada	Gene	Fisiologia
		RXF05236	Prolina imino-peptidase (ec 3.4.11.5)	<i>Pip3</i>	
		RXF04802	Prolina imino-peptidase (ec 3.4.11.5)	<i>Pip1</i>	
		RXF04808	Prolina imino-peptidase (ec 3.4.11.5)	<i>Pip2</i>	
	S41 (família de peptidase de processamento C-terminal)				
		RXF06586	Protease cauda-específica (ec 3.4.21.-)		
		RXF01037	Protease cauda-específica (ec 3.4.21.-)		
	S45				
		RXF07170	Penicilina Aci-lase (ec 3.5.1.11)	<i>pacB2</i>	
		RXF06399	Penicilina Aci-lase (ec 3.5.1.11)	<i>pacB1</i>	
	S49 (família de protease IV)				
		RXF06993	Possível protease sohb (ec 3.4.--)		
		RXF01418	Protease iv (ec 3.4.--)		
	S58 (família de DmpA aminopeptidase)				
		RXF06308	D-aminopeptida se (ec 3.4.11.19)		
Treonina peptidases					
	T1 (família de proteassoma)				
		RXF01961	Protease atp-dependente hslV (ec 3.4.25.-)	<i>hslV</i>	Julgado contribuir para eliminação de proteína prejudiciais no choque térmico

Classe	Família	RXF	Função Curada	Gene	Fisiologia
	T3 (família gama-glutamiltransferase)				
		RXF02342	Gama-glutamil-transpeptidase (ec 2.3.2.2)	ggt1	
		RXF04424	Gama-glutamil-transpeptidase (ec 2.3.2.2)	ggt2	
Peptidases não-classificadas					
	U32				
		RXF00428	Protease (ec 3.4.--)		
		RXF02151	Protease (ec 3.4.--)		
	U61				
		RXF04715	Muramoltetra-peptídeo carboxipeptidase (ec 3.4.17.13)		
	U62				
		RXF04971	Proteína de PmbA	pmbA	O produto do gene de PmbA ( <i>{Escherichia coli}</i> ) facilita a secreção do peptídeo antibiótico microcina B17, removendo um peptídeo líder de 26-aminoácido, N-terminal (Madison et al., 1997).
		RXF04968	Proteína de TldD		
Proteases de não-MEROPS					
		RXF00325	Proteína repressora C2		
		RXF02689	Dipeptidase microssomal (ec 3.4.13.19)		
		RXF02739	Dipeptidase membrana (ec		

Classe	Família	RXF	Função Curada	Gene	Fisiologia
			3.4.13.19)		
		RXF03329	Proteína citosólica hipotética		
		RXF02492	Xaa-Pro dipeptidase (ec 3.4.13.9)		
		RXF04047	Família de protease amino terminal de caax		
		RXF08136	Protease (proteína similar à transglutaminase)		
		RXF09487	Zinco metaloprotease (ec 3.4.24.-)		

Certas proteases de origem de *E. coli* são listadas na Tabela B.

**TABELA B**

Classe	Família	Código	Peptidase ou homólogo (subtipo)	Gene
Aspártico peptidases	A8	A08.001	Peptidase de sinal II	<i>ispA</i>
	A24	A24.001	Prepilina peptidase do tipo IV (proteína de EtpN (plasmídeo p0157))	<i>etpN</i>
		A24.001	Prepilina peptidase do tipo IV 1 (proteína de CofP)	<i>cofP</i>
		A24.001	Prepilina peptidase do tipo IV 1 (proteína de HofD)	<i>hofD/hopD/hopO</i>
		A24.003	Prepilina peptidase do tipo IV 2 (proteína de HopD)	<i>hopD/ECs4188</i>
		A24 não-designada	Peptidases da família A24A não-designada (proteína de ORF_F310)	<i>pppA/ORF_F310</i>
		A24 não-designada	Peptidases da família A24A não-designada (proteína de PilU (plasmídeo R721))	<i>pilU</i>
		A24	Peptidases da família A24A	<i>bfpP/bfpG</i>

Classe	Família	Código	Peptidase ou homólogo (subtipo)	Gene
		não-designada	não-designada (proteína de BfpP (plasmídeo pMAR2))	
		A24 não-designada	Peptidases da família A24A não-designada (proteína de PilU)	<i>PILU</i>
	A26	A26.001	Omptina	<i>ompT/ECS1663/B0565</i>
		A26.005	Proteinase SopA	<i>sopA</i>
Cisteína peptidases	C26	C26 não-designada	Peptidases da família C26 não-designada	<i>YC-JL/Z2490/ECS1875</i>
	C40	C40.004	spr g.p. (tipo <i>Escherichia</i> ) (proteína de spr)	<i>spr</i>
		C40 não-designada	Peptidases da família C40 não-designada (proteína de NlpC)	<i>n-lpC/C2104/Z3737/ECS2415</i>
		C40 não-designada	Peptidases da família C40 não-designada (proteína de YafL)	<i>YafL</i>
		C40 não-designada	Peptidases da família C40 não-designada (quitinase 3)	
		C40 não-designada	Peptidases da família C40 não-designada (proteína de YdhO)	<i>ydhO</i>
	C39	C39.005	Peptidase de processamento de colicina V (proteína de CvaB)	<i>cvaB</i>
		C39.005	Peptidase de processamento de colicina V (proteína de MtfB)	<i>mtfB</i>
		C39 não-designada	Peptidases da família C39 não-designada (proteína de secreção de H47 de microcina MchF)	<i>mchF/MCLB</i>
	C56	C56 não-designada	Peptidases da família C56 não-designada (proteína de YhbO)	<i>yhbo</i>
		C56 não-designada	Peptidases da família C56 não-designada (proteína de c4536)	<i>c4536</i>

Classe	Família	Código	Peptidase ou homólogo (subtipo)	Gene
Metalopeptidases	M1	M01.005	Alanil aminopeptidase (proteobactérias)	<i>pepN</i>
	M3A	M03.004	Oligopeptidase A	<i>prlC/opdA</i>
		M03.005	Peptidil-dipeptidase Dcp	<i>dcp/Z2160/ECS2147</i>
		M03.005	Peptidil-dipeptidase Dcp	<i>dcp</i>
	M41	M41.001	FtsH endopeptidase	<i>hflB/ftsH/ECS4057</i>
	M66	M66.001	StcE protease	<i>stcE</i>
	M15D	M15 não-designada	Peptidases da subfamília M15D não-designada (proteína de VanX)	<i>ddpX/vanX/B1488/Z2222/ECS2092</i>
	M16A	M16.001	Pitrilisina	<i>ptr/ECs3678</i>
	M16B	M16 não-designada	Peptidases da subfamília M16B não-designada (proteína de PqqL)	<i>pqqL/yddC</i>
	M17	M17.003	Aminopeptidase A (bactérias)	<i>pepA/xerB</i>
		M17.004	PepB aminopeptidase	<i>pepB/Z3790/ECS3389</i>
	M24A	M24.001	Metionil aminopeptidase 1	<i>map</i>
	M24B	M24.003	X-Pro dipeptidase (bactérias)	<i>pepQ/ECs4775</i>
		M24.004	Aminopeptidase P (bactérias)	<i>pepP</i>
		M24 não-designada	Peptidases da subfamília M24B não-designada (proteína de YqhT)	<i>yqhT/ypdF/B2385/c2924</i>
	M20A	M20.010	DapE peptidase (succinil diaminopimelato dessuccinilase)	<i>dapE/msgB/C2999</i>
		M20 não-designada	Peptidases da subfamília M20A não-designada (proteína de YgeY)	<i>ygey</i>
	M20B	M20.003	Peptidase T	<i>pepT/Z1832/ECS1572</i>
	M20C	M20.007	X-His dipeptidase	<i>pepD/pepH/ECs0264</i>

Classe	Família	Código	Peptidase ou homólogo (subtipo)	Gene
	M20D	M20 não-designada	Peptidases da família M20D não-designada (proteína de YdaJ)	<i>ydaJ/ECs1922</i>
	M28A	M28 não-designada	Peptidases da subfamília M28A não-designada (pro- teína de YfbL)	<i>yfbL</i>
	M28C	M28.005	IAP aminopeptidase	<i>iap</i>
	M42	M42 não-designada	Peptidases da família M42 não-designada (proteína de YjhO)	<i>yjhO</i>
		M42 não-designada	Peptidases da família M42 não-designada (proteína de FrvX)	<i>frvX</i>
		M42 não-designada	Peptidases da família M42 não-designada (proteína de FrvX)	<i>frvX/b2384/ypdE</i>
	M38	M38.001	Beta-aspartil dipeptidase	<i>iadA</i>
	M22	M22.001	O-sialoglicoproteína endo- peptidase	<i>ygjD</i>
		M22.002	Proteína de yeaZ	<i>ye- aZ/C2211/Z2850/ ECS2516</i>
	M23B	M23.006	YibP Peptidase (proteína de YibP)	<i>yibP</i>
		M23 não-designada	Peptidases da subfamília M23B não-designada (pro- teína de YebA)	<i>yebA</i>
	M48B	M48.002	HtpX endopeptidase	<i>HrpX</i>
		M48 não-designada	Peptidases da subfamília M48B não-designada	<i>YGGG/C3521</i>
		M48 não-designada	Peptidases da subfamília M48B não-designada	<i>YFGC/C3011</i>
		M48 não-designada	Peptidases da subfamília M48B não-designada (pro- teína de YggG)	<i>YggG/Z4280/EC S3811</i>
		M48 não-designada	Peptidases da subfamília M48B não-designada (pro- teína de YcaL)	<i>ycaL/C1047/Z125 5/ECS0992</i>
	M50A	M50.004	YaeL protease (proteína de	<i>ec-</i>

Classe	Família	Código	Peptidase ou homólogo (subtipo)	Gene
			Yael)	<i>fE/YAEL/B0176/Z</i> <i>0187/ECS0178/C</i> <i>0213</i>
	M52	M52.001	HybD endopeptidase (proteína de HybD)	<i>hybD/ECS3878</i>
		M52.002	HyaD endopeptidase (proteína de HyaD)	<i>hyaD</i>
		M52.003	Hycl endopeptidase (proteína de Hycl)	<i>hycl/C3277</i>
Serina peptidases	S1B	S01.260	B1598 endopeptidase	<i>B1598</i>
	S1C	S01.273	Protease Do	<i>htrA/degP</i>
		S01.274	DegQ	<i>hho-AI/degQ/ECS410</i> <i>7/Z4593</i>
		S01.275	DegS	<i>hhoB/degS</i>
	S6	S06.002	EspP g.p. ( <i>Escherichia coli</i> )	<i>espP/pssA</i>
		S06.003	Tsh peptidase ( <i>Escherichia coli</i> ) (proteína de Tsh)	<i>tsh/hbp</i>
		S06.003	Tsh peptidase ( <i>Escherichia coli</i> )	<i>C0393</i>
		S06.004	Pet endopeptidase	<i>Sat</i>
		S06.004	Pet endopeptidase	
		S06.005	Pic endopeptidase ( <i>Shigella flexneri</i> )	<i>she/pic</i>
		S6 não-designada	Peptidases da família S6 não-designada (proteína de eatA)	<i>eatA</i>
		S6 não-designada	Peptidases da família S6 não-designada (proteína de c0350)	<i>C0350</i>
		S6 não-designada	Peptidases da família S6 não-designada (proteína de EspC)	<i>espC</i>
		S6 não-designada	Peptidases da família S6 não-designada (proteína de epeA)	<i>epeA</i>
		S6 não-designada	Peptidases da família S6 não-designada	

Classe	Família	Código	Peptidase ou homólogo (subtipo)	Gene
	S8A	S8 não-designada	Peptidases da subfamília S8A não-designada	
	S9A	S09.010	Oligopeptidase B	<i>ptrB</i>
		S09.010	Oligopeptidase B	<i>ptrB/C2255</i>
	S9X	S9 não-designada	Peptidases da família S9 não-designada	<i>YF-HR/C3060/b2534/Z3802</i>
	S11	S11.002	Murein-DD-endopeptidase	<i>pbpG</i>
		S11.003	Proteína de ligação de penicilina 6	<i>dacC/Z1066/ECS0919</i>
		S11.003	Proteína de ligação de penicilina 6 (proteína de ligação de penicilina pbp-6B)	<i>DacD/phsE/ECs2812</i>
		S11.003	Proteína de ligação de penicilina 6	<i>dacA</i>
	S12	S12 não-designada	Peptidases da família S12 não-designada (proteína de c2452)	<i>C2452</i>
		S12 não-designada	Peptidases da família S12 não-designada (proteína de YaiH)	<i>yaiH/C0480</i>
	S13	S13.001	D-Ala-D-Ala peptidase C	<i>dacB/ECS4061</i>
	S14	S14.001	Endopeptidase Clp (tipo 1)	<i>clpP/lopP/ECS0491</i>
		S14 não-designada	Peptidases da família S14 não-designada (proteína de ECS0829)	<i>Z0967/ECS0829</i>
		S14 não-designada	Peptidases da família S14 não-designada (proteína de ECS2960)	<i>H0022/Z2112/EC S2960/L34</i>
	S16	S16.001	Lon protease	<i>lon/deg/ECS0493</i>
		S16 não-designada	Peptidases da família S16 não-designada (proteína de ECS1039)	<i>lonB/Z1305/ECS1039</i>
		S16 não-designada	Peptidases da família S16 não-designada (proteína de c1091)	<i>C1091</i>
	S24	S24.001	Repressor LexA (proteína de LexA)	<i>lexA/exrA</i>

Classe	Família	Código	Peptidase ou homólogo (subtipo)	Gene
		S24.003	Proteína de UmuD	
		S24.003	Proteína de UmuD	<i>umuD/C1631</i>
	S26	S26A	S26.001	<i>Peptidase de sinal I</i>
		S26.014	Proteína de transferência de plasmídeo de traF (proteína de TraF)	<i>traF</i>
	S33	S33 não-designada	Peptidases da família S33 não-designada (proteína de BioH)	<i>bi-oH/C4189//Z4767 /ECS4255</i>
	S41A	S41.001	Protease de processamento C-terminal 1	<i>prc/tsp/ECS2540/ Z2877//C2239</i>
	S45	S45.001	Precursor de penicilina G acilase	<i>pac</i>
	S49	S49.001	Protease IV	<i>sp-pA/ECs2472//C2 170</i>
		S49.002	sohB endopeptidase	<i>sohB/ECS1844/Z 2538//C1737</i>
	S51	S51.001	Dipeptidase E	<i>pepE</i>
	S54	S54 não-designada	Peptidases da família S54 não-designada (proteína de c0741)	<i>C0741</i>
		S54 não-designada	Peptidases da família S54 não-designada (glicerofosfato desidrogenase)	<i>glpG/C4201//Z47 84/ECS4267</i>
Treonina peptidases	T1B	T01.006	Componente de Hs1V de Hs1UV peptidase	<i>hs1V</i>
	T2	T02.002	Asparaginase	<i>ybiK/Z1051m/C0 913</i>
	T3	T03.001	Gama-glutamiltransferase 1 (bacteriana)	<i>ggt/C4246</i>
	S41A	S41.001	Protease de processamento C-terminal 1	<i>prc/tsp/ECS2540/ Z2877//C2239</i>
Peptidases não-classificadas	U6	U06.001	Mureína endopeptidase	<i>me-pA/ECs3212//C2 874</i>
		U32	U32 não-designada	<i>Peptidases da família U32</i>

Classe	Família	Código	Peptidase ou homólogo (subtipo)	Gene
				<i>não-designada (proteína de YdcP)</i>
		U32 não-designada	Peptidases da família U32 não-designada (proteína de YegQ)	<i>yegQ/C2611</i>
		U32 não-designada	Peptidases da família U32 não-designada (proteína de YhbU)	<i>YH-BU/C3911/Z4519 /ECS4039</i>
		U35	U35 não-designada	<i>Peptidases da família U35 não-designada</i>
		U35 não-designada	Peptidases da família U35 não-designada (proteína de ECs4973)	<i>ECs4973</i>
		U49	U49.001	<i>Lit protease (<i>Escherichia coli</i>)</i>
		U61	U61.001	<i>Muramoyil-tetrapeptídeo carboxipeptidase</i>
		U61 não-designada	Peptidases da família U61 não-designada (proteína de MccF)	<i>mccF</i>
		U62	U62.001	<i>Peptidase de processamento de microcina 1</i>
		U62.002	Peptidase de processamento de microcina 2	<i>tldD/ECs4117</i>
		M9G.035	Endopeptidase ECP 32 ( <i>Escherichia coli</i> )	

Certas proteases de origem de *S. Cerevisiae* são listadas na Tabela C.

TABELA C

Classe	Família	Código	Peptidase ou homólogo (subtipo)	Gene
Aspártico peptidases	A1	A01.015	Pepsina de barreira	<i>Bar1</i>
		A01.018	Sacaropepsina	<i>pep4/pho9</i>
		A01.030	lapsina 1	<i>iap3</i>
		A01.031	lapsina 2	<i>mkc7</i>
		A01.035	lapsina 3	<i>YPS3</i>
		A01.UPW	Peptidases da família A1 não-designada	<i>YPS7/D9476.8/YDR349C</i>
		A01.UPW	Peptidases da família A1 não-designada (proteína de YIR039C)	<i>YIR039C</i>
	A2D	A02.022	Transposon de Ty3 ( <i>Saccharomyces cerevisiae</i> ) endopeptidase (retrotransposon Ty3-1)	<i>POL3/TY3-2</i> <i>orfB/TY3B</i>
	A11B	A11.003	Transposon de Ty1 ( <i>Saccharomyces cerevisiae</i> ) endopeptidase (proteína B de transposon Ty1-17)	<i>Ty1B</i>
		A11.003	Transposon de Ty1 ( <i>Saccharomyces cerevisiae</i> ) endopeptidase (proteína B de transposon Ty1)	<i>Ty1B</i>
		A11.003	Transposon de Ty1 ( <i>Saccharomyces cerevisiae</i> ) endopeptidase (proteína B de transposon Ty1)	<i>Ty1B</i>
	A11X	A11.UPW	Peptidases da família A11 não-designada (retrotransposon Ty4)	
	A22B	A22.008	Proteína de YKL100c ( <i>Saccharomyces cerevisiae</i> )	<i>YKL100c</i>
Cisteína peptidases	C1B	C01.085	Bleomicina hidrolase (levadura)	<i>GAL6/YCP1/LAP3</i>

Classe	Família	Código	Peptidase ou homólogo (subtipo)	Gene
	C2	C02.008	Calpaína-7	Y- <i>MR154C/Cpl1/Rim13</i>
	C12	C12.002	Ubiquitinil hidrolase YUH1	<i>yuh1</i>
	C13	C13.005	Glicosilfosfatidilinositol:proteína transamidase	<i>d9798.2</i>
	C19	C19.002	Ubp1 ubiquitina peptidase	<i>ubp1</i>
		C19.003	Ubp2 ubiquitina peptidase	<i>ubp2</i>
		C19.004	Ubp3 ubiquitina peptidase	<i>ubp3</i>
		C19.005	Doa4 ubiquitina peptidase	<i>DOA4</i>
		C19.006	Ubp5 ubiquitina peptidase	<i>ubp5</i>
		C19.079	UBP6 ( <i>Saccharomyces cerevisiae</i> ) (proteína de YFRO 10W)	<i>yfr010w</i>
		C19.UPW	Peptidases da família C19 não-designada (proteína de YNL186W)	<i>YNL186W</i>
		C19.UPW	Peptidases da família C19 não-designada (UBP9)	<i>ubp9</i>
		C19.UPW	Peptidases da família C19 não-designada (proteína de YBL067C)	<i>YBL067C</i>
		C19.UPW	Peptidases da família C19 não-designada (proteína de YBRO58C)	<i>UBP12/YBR058C</i>
		C19.UPW	Peptidases da família C19 não-designada (ubiquitina hidrolase 16 carbóxi-terminal)	<i>UBP16/YPL072W/LPF12W</i>
		C19.UPW	Peptidases da família C19 não-designada (proteína de YMR304W)	Y- <i>MR304W/ym995.2.06</i>
		C19.UPW	Peptidases da família C19 não-designada (proteína de YMR223W)	Y- <i>MR223W/ym995.9.05</i>
		C19.UPW	Peptidases da família C19 não-designada (UBP7)	<i>ubp7</i>
		C19.UPW	Peptidases da família C19 não-designada (UBP13)	<i>ubp13</i>

Classe	Família	Código	Peptidase ou homólogo (subtipo)	Gene
	C44	C44.971	Glucosami-na-frutose-6-fosfato amino-transferase	
		C44.971	Glicosami-na-frutose-6-fosfato amino-transferase (glicosami-na-frutose-6-fosfato amino-transferase)	<i>gfa1</i>
	C48	C48.001	Ulp1 endopeptidase	<i>YPL020c</i>
		C48.005	Ulp2 endopeptidase (proteína de Smt4p)	<i>SMT4</i>
	C50	C50.001	Separase	<i>ESP1/YGR098C</i>
	C54	C54.001	ATG4 peptidase ( <i>Saccharomyces cerevisiae</i> )	<i>Apg4/Aut2</i>
	C56	C56.004	YDR533C g.p. ( <i>Saccharomyces cerevisiae</i> )	<i>Y-DR533C/D9719.36</i>
		C56.UPW	Peptidases da família C56 não-designada (proteína de YPL280W)	<i>YPL280W</i>
		C56.UPW	Peptidases da família C56 não-designada (proteína de YOR391C)	<i>YOR391C</i>
	I34	I34.001	Inibidor de sacaropepsina	<i>PAI3/YMR174C/YM8010</i>
Metalopeptidases	M1	M01.006	Ape2 aminopeptidase	<i>lap1/ape2</i>
		M01.007	Aap1' aminopeptidase	<i>AAP1</i>
		M01.007	Aap1' aminopeptidase	
		M01.017	Yin7 g.p. ( <i>Saccharomyces cerevisiae</i> )	<i>Yil137C</i>
		M01.UPW	Peptidases da família M1 não-designada (proteína de ynl045w)	<i>inl045w</i>
	M3A	M03.003	Sacarolisina	<i>prd1</i>
		M03.006	Peptidase intermediária mitocondrial	<i>MIP1</i>
	M16A	M16.007	Axl1 peptidase	<i>axl1</i>
		M16.008	Ste23 peptidase	<i>ste23</i>

Classe	Família	Código	Peptidase ou homólogo (subtipo)	Gene
		M16.UPA	Peptidases da subfamília M16A não-designada (proteína de orf1)	<i>Orf1</i>
	M16B	M16.003	Beta-subunidade de peptidase de processamento mitocondrial (beta)	<i>mas1/mif1</i>
	M16C	M16.UPC	Peptidases da subfamília M16C não-designada (proteína de YDR430C)	<i>YDR430C</i>
		M16.UPC	Peptidases da subfamília M16C não-designada (proteína de YOL098C)	<i>YOL098C</i>
	M16X	M16.971	Subunidade alfa de não peptidase de peptidase de processamento mitocondrial (alfa)	<i>mas2/mif2</i>
		M16.974	UCR2_HUMANO (proteína de núcleo de ubiquinol citochrome c reductase 2)	<i>ucr2/cor2/qcr2</i>
	M18	M18.001	Aminopeptidase I	<i>ape1/lap4</i>
		M18.UPW	Peptidases da família M18 não-designada (proteína de YHR113W)	<i>YHR113W</i>
	M20A	M20.005	Dipeptidase não-específica citosólica	<i>YFR044C</i>
	M20E	M20.002	Gly-X Carboxipeptidase	<i>Cps1/cps</i>
		M20.002	Gly-X Carboxipeptidase (pseudogene; deduzido da seqüência de nucleotídeo por MEROPS)	<i>AOE110, AO-E264, AOE130</i>
	M22	M22.003	Mername-AA017 peptidase (proteína de YKR038C)	<i>YKR038C</i>
		M22.UPW	Peptidases da família M22 não-designada (proteína de QR17)	<i>QR17</i>
	M24A	M24.001	Metionil aminopeptidase 1	<i>map1</i>
		M24.002	Metionil aminopeptidase 2	<i>ybl091c</i>
	M24B	M24.009	Aminopeptidase P1	<i>YLL029W</i>
		M24.026	Homólogo de aminopeptidase	<i>YER078C</i>

Classe	Família	Código	Peptidase ou homólogo (subtipo)	Gene
			dase P (proteína de YER078C)	
		M24.UPB	Peptidases da subfamília M24B não-designada (proteína de YFR006W)	<i>yfr006w</i>
	M28A	M28.001	Aminopeptidase Y	<i>ape3</i>
	M28E	M28.006	Mername-AA063 peptidase (proteína de YDR415c)	<i>YDR415c</i>
	M28X	M28.974	Glutaminil ciclase	<i>YFR018C</i>
		M28.UPW	Peptidases da família M28 não-designada (proteína de YBR074W)	<i>YBR074W</i>
	M41	M41.002	Afg3 g.p. ( <i>Saccharomyces cerevisiae</i> ) (proteína de AGF3)	<i>agf3/yta10</i>
		M41.003	m-AAA protease (proteína de RCA1)	<i>rca1/yta12</i>
		M41.004	i-AAA protease	<i>yme1/yta1/osd1</i>
	M48A	M48.001	Ste24 endopeptidase	<i>STE24</i>
	M48B	M48.018	Oma1 endopeptidase ( <i>Saccharomyces cerevisiae</i> ) (proteína de YKR087C)	<i>Y-KR087C/YKR407</i>
	M49	M49.001	Dipeptidil-peptidase III	<i>YOL057W</i>
		M49.UPW	Peptidases da família M49 não-designada	
	M67A	M67.001	Poh1 peptidase	<i>RPN11/MPR1/YFR004W</i>
		M67.002	Metaloenzima de domínio Jab1/MPN	<i>YDL216c/D0888</i>
		M67.973	Subunidade reguladora de não-ATPase de proteassoma 26S	<i>RPN8/YOR261C</i>
Serina peptidases	SIC	S01.434	Nma111 endopeptidase ( <i>Saccharomyces cerevisiae</i> ) (proteína YNL123W)	<i>ynl123w</i>
	S8A	S08.052	Cerevisina	<i>prb1</i>
		S08.UPA	Peptidases da subfamília S8A não-designada (proteína de YSP3)	<i>YSP3</i>

Classe	Família	Código	Peptidase ou homólogo (subtipo)	Gene
		S08.UPA	Peptidases da subfamília S8A não-designada (proteína de YCR54C)	<i>YCR54C</i>
	S8B	S08.070	Cexina	<i>kex2</i>
	S9B	S09.005	Dipeptidil aminopeptidase A	<i>ste13/yci1</i>
		S09.006	Dipeptidil aminopeptidase B (fungo)	<i>dap2</i>
	S9X	S09.UPW	Peptidases da família S9 não-designada (proteína de Ynl320w)	<i>YNL320W</i>
	S10	S10.001	Carboxipeptidase Y	<i>prc1</i>
		S10.007	Kex carboxipeptidase	<i>kex1</i>
		S10.UPW	Peptidases da família S10 não-designada (proteína de YBR139W)	<i>ybr139W</i>
	S16	S16.002	PIM1 endopeptidase	<i>lon/pim1</i>
	S26A	S26.002	Protease de membrana interna mitocondrial 1 (1)	<i>imp1</i>
		S26.012	Protease de membrana interna mitocondrial 2 (2)	<i>imp2</i>
	S26B	S26.010	Componente de 21 kDa de signalase (eucarioto)	<i>sec11</i>
		S33.UPW	Peptidases da família S33 não-designada	<i>ECM18/YDR125C</i>
		S33.UPW	Peptidases da família S33 não-designada	<i>ECM18/YDR125C</i>
	S54	S54.007	Proteína de Pcp1 ( <i>Saccharomyces cereviseae</i> ) (proteína de YGR101W)	<i>YGR101W</i>
	S59	S59.001	Nucleoporina 145	<i>Nup145</i>
Treonina peptidases	T1A	T01.010	Subunidade catalítica de proteassoma 1	<i>pre3</i>
		T01.011	Subunidade catalítica de proteassoma 2	<i>pup1</i>
		T01.012	Subunidade catalítica de proteassoma 3	<i>pre2/prg1</i>
		T01.983	Subunidade de proteassoma beta 3	<i>pup3</i>

Classe	Família	Código	Peptidase ou homólogo (subtipo)	Gene
		T01.984	Subunidade de proteassoma beta 2	<i>pre1</i>
		T01.986	Subunidade de proteassoma beta 1	<i>pre7/prs3</i>
		T01.987	Subunidade de proteassoma beta 4	<i>pre4</i>
	T1X	T01.971	Subunidade de proteassoma alfa 6	<i>prs2/prc2</i>
		T01.972	Subunidade de proteassoma alfa 2	<i>pre8/prs4</i>
		T01.973	Subunidade de proteassoma alfa 4	<i>pre9/prs5</i>
		T01.974	Subunidade de proteassoma alfa 7	<i>pre6</i>
		T01.975	Subunidade de proteassoma alfa 5	<i>pup2</i>
		T01.976	Subunidade de proteassoma alfa 1	<i>pre5</i>
		T01.977	Subunidade de proteassoma alfa 3	<i>pre10/prs1/prc1</i>
	T3	T03.012	Gama-glutamiltransferase (Saccharomyces) (proteína de YLR299w)	<i>L8003.4</i>
	T5	T05.002	Precursor de ornitina acetil-transferase	<i>arg7/emc40/YMR062C</i>
Peptidases não-classificadas	U48	U48.001	Prenil protease 2	<i>rce1</i>

### MODULADORES DE DOBRAMENTOS

Os genes ou produtos de gene sobre-regulados identificados podem ser um ou mais moduladores de dobramento. Moduladores de dobramento podem por exemplo ser proteínas de HSP70, proteínas de

- 5 HSP110/SSE, proteínas de HSP40 (relacionadas a DNAJ), proteínas similares a GRPE, proteínas de HSP90, proteínas de CPN60 e de CPN10, chaperoninas citosólicas, proteínas de HSP100, HSPs Pequena, Calnexina e calreticulina, PDI e proteínas relacionadas a tiorredoxina, Peptidil-prolil isome-

rases, Ciclofilina PPIases, proteínas de ligação de FK-506, Parvulina PPIases, chaperoninas individuais, chaperonas proteína-específicas ou chaperonas intramoleculares. Moduladores de dobramentos são em geral descritos em "Guidebook to Molecular Chaperones and Protein-Folding Catalysts" (1997) 5 ed. M. Gething, Universidade de Melbourne, Austrália.

As chaperonas molecular mais bem caracterizadas no citoplasma de *E. coli* são sistemas dependentes de DnaK-DnaJ-GrpE ATP- e de GroEL-GroES. Com base em estudos *in vitro* e considerações de homologia, várias proteínas citoplásmicas adicionais foram propostas para funcionarem 10 como chaperonas moleculares em *E. coli*. Estas incluem ClpB, HtpG e IbpA/B que, como DnaK-DnaJ-GrpE e GroEL-GroES, são proteínas de choque térmico (Hsps) pertencendo ao regulon de tensão. A conformação *trans* de ligações Pro é energeticamente favorecida em cadeias de proteína nascentes; porém, ~5% de todas as ligações de prolil peptídeo são encontrados em uma 15 conformação *cis* nas proteínas nativas. A isomerização de *trans* para *cis* de ligações de X-Pro é limitativa em taxa no dobramento de muitos polipeptídeos e é catalisada *in vivo* através de peptidil prolil *cis/trans* isomerases (PPIases). Três PPIases citoplásmicas, SlyD, SlpA e fator desencadeador (TF), foram 20 identificadas até agora em *E. coli*. TF, uma proteína de 48 kDa associada às subunidades ribossômicas de 50 S que foram postuladas cooperar com as chaperonas em *E. coli* para garantir dobramento apropriado das proteínas recentemente sintetizadas. Pelo menos cinco proteínas (tiorredoxinas 1 e 2, e glutarredoxinas 1, 2 e 3, os produtos dos genes de *trxA*, *trxC*, *grxA*, *grxB* e *grxC*, respectivamente) estão envolvidos na redução de ligação em pontes de 25 dissulfeto que transientemente surgem nas enzimas citoplásmicas. Desse modo, genes identificados podem ser proteínas ou chaperonas de formação de ligação de dissulfeto que permitem formação de ligação de dissulfeto apropriada.

Certos moduladores de dobramento em *O. fluorescens* são listados na Tabela D. 30

Tabela D

RXF	Gene	Função	Família
<b>GroES/EL</b>			
rxf02095	<i>groES</i>	chaperona	Hsp10
rxf06767:: rxf02090	<i>groEL</i>	chaperona	Hsp60
RXF01748	<i>ibpA</i>	Proteína de choque térmico pequena (sHSP) IbpA PA3126; Age como um retentor para dobramento de GroESL	Hsp20
RXF03385	<i>hscB</i>	proteína de chaperona hscB	Hsp20
<b>Hsp70 (DnaK/J)</b>			
rxf05399	<i>dnaK</i>	chaperona	Hsp70
RXF06954	<i>dnaK</i>	chaperona	Hsp70
RXF03376	<i>hscA</i>	chaperona	Hsp70
RXF03987	<i>cbpA</i>	Proteína de ligação de DNA curvada, atividade similar a dnaJ	Hsp40
RXF05406	<i>dnaJ</i>	proteína de chaperona dnaJ	Hsp40
RXF03346	<i>dnaJ</i>	chaperonas moleculares (família de DnaJ)	Hsp40
<b>Hsp100 (Clp/Hsl)</b>			
RXF04587	<i>clpA</i>	subunidade de ligação de atp de clp protease atp-dependente ClpA	Hsp100
RXF08347	<i>clpB</i>	proteína de ClpB	Hsp100
RXF04654	<i>clpX</i>	subunidade de ligação de atp de clp protease atp-dependente clpX	Hsp100
RXF01957	<i>hslU</i>	subunidade de ligação de atp de hsl protease atp-dependente hslU	Hsp100
RXF01961	<i>hslV</i>	subunidade de ligação de atp de hsl protease atp-dependente hslV	Hsp100
<b>Hsp33</b>			
RXF04254	<i>yrfI</i>	chaperonina de 33 kDa (homólogo de proteína de choque térmico 33) (HSP33)	Hsp33
<b>Hsp90</b>			
RXF05455	<i>htpG</i>	proteína de chaperona htpG	Hsp90
<b>SecB</b>			

RXF	Gene	Função	Família
RXF022311	<i>secB</i>	chaperona específica de secreção SecB	SecB
<b>Isomerases de ligação de dissulfeto</b>			
RXF07017	<i>dsbA</i>	dissulfeto isomerase	DSBA oxidor-reductase
RXF08657	<i>ds-bA/dsbC/dsbG/fernA</i>	dissulfeto isomerase	DSBA oxidor-reductase
rxf01002	<i>ds-bA/dsbC</i>	dissulfeto isomerase	DSBA oxidor-reductase/ Ti-orredoxina
rxf03307	<i>dsbC</i>	dissulfeto isomerase	glutarredoxina/ Tiorredoxina
Rxf04890	<i>dsbG</i>	dissulfeto isomerase	glutarredoxina/ Tiorredoxina
<b>Peptidil-prolil cis-trans isomerases</b>			
RXF03768	<i>ppiA</i>	Peptidil-prolil cis-trans isomerase A (ec 5.2.1.8)	PPIase: tipo ciclofilina
RXF05345	<i>ppiB</i>	Peptidil-prolil cis-trans isomerase B	PPIase: tipo ciclofilina
RXF06034	<i>fklB</i>	Peptidil-prolil cis-trans isomerase FklB	PPIase: tipo FKBP
RXF06591	<i>fklB/fkbP</i>	Peptidil-prolil cis-trans isomerase de proteína de ligação de fk506 (EC 5.2.1.8)	PPIase: tipo FKBP
RXF05753	<i>fklB/fkbP</i>	Peptidil-prolil cis-trans isomerase (ec 5.2.1.8)	PPIase: tipo FKBP
RXF01833	<i>slyD</i>	Peptidil-prolil cis-trans isomerase SlyD	PPIase: tipo FKBP
RXF04655	<i>tig</i>	fator desencadeador, ppiase (ec 5.2.1.8)	PPIase: tipo FKBP
RXF05385	<i>yaad</i>	peptidil-prolil cis-trans isomerase de 16 kDa do tipo FKBP provável (EC 5.2.1.8) (PPIase) (Rotamase)	PPIase: tipo FKBP
RXF00271	<i>yaad</i>	Peptidil-prolil cis-trans isomerase (ec 5.2.1.8)	PPIase: tipo FKBP

RXF	Gene	Função	Família
Chaperonas de montagem de pili (parecidas com papD)			
RXF06068	<i>cup</i>	Proteína de chaperona cup	papD de montagem de pili
RXF05719	<i>ecpD</i>	Proteína de chaperona ecpD	papD de montagem de pili
RXF03406	<i>ecpD;csuC</i>	Proteína de chaperona ecpD	papD de montagem de pili
RXF04296	<i>ecpD;cup</i>	Proteína de chaperona ecpD	papD de montagem de pili
RXF04553	<i>ecpD;cup</i>	Proteína de chaperona ecpD	papD de montagem de pili
RXF04554	<i>ecpD;cup</i>	Proteína de chaperona ecpD	papD de montagem de pili
RXF05310	<i>ecpD;cup</i>	Proteína de chaperona ecpD	papD de montagem de pili
RXF05304	<i>ecpD;cup</i>	Proteína de chaperona ecpD	papD de montagem de pili
RXF05073	<i>gltF</i>	função periplásica de chaperona de montagem de pili gram-negativa	papD de montagem de pili

Certos moduladores de dobramento em *E. coli* são listados na Tabela E.

Tabela E

Acesso de Uniprot	Uniprot ID	Anotação	Família
GroES/EL			
P05380	CH10_ECOLI	chaperonin de 10 kDa	Hsp10
P06139	CH60_ECOLI	chaperonin de 60 kDa	Hsp60
Hsp70(DnaK/J)			
P04475	DNAK_ECOLI	Proteína de chaperona dnaK	Hsp70
P77319	HSCC_ECOLI	Proteína de chaperona hscC	Hsp70
P36659	CBPA_ECOLI	Proteína de ligação de DNA curvada CbpA	Hsp40
P31680	DJLA_ECOLI	Proteína similar a DnaJ, rscG	Hsp40
P08622	DNAJ_ECOLI	Proteína de chaperona dnaJ	Hsp40

Acesso de Uniprot	Uniprot ID	Anotação	Família
P29131	FTSN_ECOLI	Proteína de divisão celular ftsN	Hsp40
P09372	GRPE_ECOLI	Proteína de GrpE	GrpE
P31658	HCHA_ECOLI	Proteína de chaperona hchA	Hsp31
<b>Hsp100 (Clp/Hsl)</b>			
P15716	CLPA_ECOLI	Subunidade de ligação de ATP de Clp proteases ATP-dependente	Hsp100
P03815	CLPB_ECOLI	Proteína de ClpB	Hsp100
P33138	CLPX_ECOLI	Subunidade de ligação de ATP de hsl proteases ATP-dependente clpX	Hsp100
P32168	HSLU_ECOLI	Subunidade de ligação de ATP de Clp protease ATP-dependente, clpY	Hsp100
<b>Proteína de choque térmico pequenas</b>			
P29209	IBPA_ECOLI	Proteína de choque térmico de 16 kDa A	Hsp16
P29210	IBPB_ECOLI	Proteína de choque térmico de 16 kDa B	Hsp16
<b>Proteína de choque térmico de 16 kDa B</b>			
P36662	TORD_ECOLI	Proteína de chaperona torD	TorD
P15040	SECB_ECOLI	Proteína de exportação de proteína secB	SecB
P45803	HSLO_ECOLI	Chaperonina de 33 kDa	Hsp33
P10413	HTPG_ECOLI	Proteína de chaperona htpG	Hsp90
HscAB			
P36541	HSCA_ECOLI	Proteína de chaperona hscA	Hsp66
P36540	HSCB_ECOLI	Proteína de Co-chaperona hscB	Hsp20
<b>Proteína veículo de lipoproteína</b>			
P61316	LOLA_ECOLI	Precursor de proteína veículo de lipoproteína de membrana externa	LolA
P61320	LOLB_ECOLI	Precursor de lolB de lipopro-	LolB

Acesso de Uniprot	Uniprot ID	Anotação	Família
		teína de membrana externa	
<b>Isomerases de ligação de dissulfeto</b>			
P24991	DSBA_ECOLI	Tiol: precursor de dsbA de proteína de interpermuta de dissulfeto	
P300018	DSBB_ECOLI	proteína de formação de ligação de dissulfeto	oxidoreductase de ligação de dissulfeto
P21892	DSBC_ECOLI	Tiol: precursor de dsbC de proteína de interpermuta de dissulfeto	
P36655	DSBD_ECOLI	Tiol: precursor de dsbD de proteína de interpermuta de dissulfeto (EC 1.8.1.8) (Proteína-dissulfeto reductase)	
P33926	DSBE_ECOLI	Tiol: precursor de dsbE de proteína de interpermuta de dissulfeto (proteína da biogênese de citocromo c ccmG)	
P77202	DSBG_ECOLI	Tiol: precursor de dsbG de proteína de interpermuta de dissulfeto	oxidoreductase de ligação de dissulfeto
<b>Peptidil-prolil cis-trans isomerases</b>			
P22257	TIG_ECOLI	fator desencadeador	PPIase: tipo FKBP
P45523	FKBA_ECOLI	precursor de fkpa de peptidil-prolil cis-trans isomerase do tipo FKBP	PPIase: tipo FKBP
P39311	FKBB_ECOLI	peptidil-prolil cis-trans isomerase de 22 kDa do tipo FKBP	PPIase: tipo FKBP
P22563	FKBX_ECOLI	peptidil-prolil cis-trans isomerase de 16 kDa do tipo FKBP	PPIase: tipo FKBP
P30856	SLYD_ECOLI	peptidil-prolil cis-trans isomerase do tipo FKBP slyD	PPIase: tipo FKBP
P20752	PPIA_ECOLI	precursor de peptidil-prolil cis-trans isomerase A	PPIase: tipo ciclofilina

Acesso de Uniprot	Uniprot ID	Anotação	Família
P23869	PPIB_ECOLI	peptidil-prolil cis-trans isomerase B -	PPIase: tipo ciclofilina
P39159	PPIC_ECOLI	peptidil-prolil cis-trans isomerase C	PPIase: tipo PPIC
P77241	PPID_ECOLI	peptidil-prolil cis-trans isomerase D	PPIase: tipo PPIC
P21202	SURA_ECOLI	precursor de surA de proteína de sobrevivência	PPIase: tipo parvulina
<b>Chaperonas de montagem de pili (similares a papD)</b>			
P53516	AFAB_ECOLI	precursor de afaB de proteína de chaperona	PapD de montagem de pili
P33128	ECPD_ECOLI	precursor de ecpD de proteína de chaperona	PapD de montagem de pili
P31697	FIMC_ECOLI	precursor de fimC de proteína de chaperona	PapD de montagem de pili
P77249	SFMC_ECOLI	precursor de sfmC de proteína de chaperona	PapD de montagem de pili
P75749	YBGP_ECOLI	precursor de ybgP de chaperona fimbrial hipotética	PapD de montagem de pili
P40876	YCBF_ECOLI	precursor de ycbF de chaperona fimbrial hipotética	PapD de montagem de pili
P75856	YCBR_ECOLI	precursor de ycbR de chaperona fimbrial hipotética	PapD de montagem de pili
P33342	YEHC_ECOLI	precursor de yehC de chaperona fimbrial hipotética	PapD de montagem de pili
P77599	YFCS_ECOLI	precursor de yfcS de chaperona fimbrial hipotética	PapD de montagem de pili
P28722	YHCA_ECOLI	precursor de yhcA de chaperona fimbrial hipotética	PapD de montagem de pili
P77616	YQIH_ECOLI	precursor de yqiH de chaperona fimbrial hipotética	PapD de montagem de pili
P42914	YRAI_ECOLI	precursor de yral de chaperona fimbrial hipotética	PapD de montagem de pili

Certos moduladores de dobramento de *S. cervisiae* são mostrados na tabela F.

Tabela F

Acesso Uniprot GroES/EL	Uniprot ID	Fonte GO	Anotação	Família
P19882	HS60_LEVE-DURA	GOA:interpro	proteína de choque térmico 60, precursor mitocondrial	Hsp60
P38228	TC62_LEVE-DURA	GOA:interpro	chaperona mitocondrial TCM62	Hsp60
P38910	CH10_LEVE-DURA	GOA:interpro	proteína de choque térmico de 10 kDa mitocondrial	Hsp10
<b>Hsp70 (DnaK/J)</b>				
P25491	MAS5_LEVE-DURA	GOA:interpro	proteína de importação de proteína mitocondrial MAS5, Ydj1	Hsp40
P10591	HS71_LEVE-DURA	PMID:9789005	proteína de choque térmico SSA1	Hsp70
P10592	HS72_LEVE-DURA	PMID:9448096	proteína de choque térmico SSA2	Hsp70
P11484	HS75_LEVE-DURA		proteína de choque térmico SSB1	Hsp70
P40150	HS76_LEVE-DURA		proteína de choque térmico SSB2	Hsp70
P09435	HS73_LEVE-DURA	PMID:7867784	proteína de choque térmico SSA3	Hsp70
P22202	HS74_LEVE-DURA		proteína de choque térmico SSA4	Hsp70
P25294	SIS1_LEVE-DURA	GOA:interpro	proteína de SIS1	Hsp40
P32527	ZUO1_LEVE-DURA	GO:0003754	Zuotina	Hsp40
P35191	MDJ1_LEVE-DURA	GOA:interpro	proteína de MDJ1, precursor mitocondrial	Hsp40
P12398	HS77_LEVE-DURA	PMID:8654364	proteína de choque térmico SSC1, precursor mitocondrial	Hsp70
P38523	GRPE_LEVEDURA	GOA:interpro	homólogo de proteína de GrpE, precursor mitocondrial, MGE1	GrpE
P14906	SC63_LEVE-	GOA:spkw	proteína de translocação	Hsp40

Acesso Uniprot GroES/EL	Uniprot ID	Fonte GO	Anotação	Família
	DURA		SEC63	
P16474	GR78_LEVE-DURA		GRP 78, BIP, Kar2	Hsp70
P25303	SCJ1_LEVE-DURA	GOA:interpro	proteína relacionada a DnaJ SCJ1	Hsp40
P39101	CAJ1_LEVE-DURA	GOA:interpro	proteína de CAJ1	Hsp40
P48353	HLJ1_LEVE-DURA	GOA:interpro	Proteína de HLJ1	Hsp40
P39102	XDJ1_LEVE-DURA	GOA:interpro	Proteína de XDJ1	Hsp40
P52868	YGM8_LEVEDURA	GOA:interpro	proteína de 41,0 kDa hipotética em região intergênica de CEG1-SOH1	Hsp40
P53940	YNH7_LEVE-DURA	GOA:interpro	proteína de 58,9 kDa hipotética em região intergênica de TPM1-MKS1	Hsp40
P38353	SSH1_LEVE-DURA		homólogo de proteína 61 de Séc	Hsp70
P36016	LHS1_LEVE-DURA	GOA:spkw	homólogo de proteína 70 de choque térmico LHS1, SSI1	Hsp70
P38788	YHMA_LEVEDURA	PMID:11054575	homólogo de proteína 70 de choque térmico YHR064C	Hsp70
<b>Hsp110/Sse</b>				
P32589	HS78_LEVE-DURA	PMID:10480867	homólogo de proteína de choque térmico SSE1	SSE
P32590	HS79_LEVE-DURA		homólogo de proteína de choque térmico SSE2	SSE
<b>Hsp100 (Clp/Hsl)</b>				
P31539	H104_LEVE-DURA	GOA:interpro	proteína de choque térmico 104	Hsp100
P33416	HSP7_LEVE-DURA		proteína de choque térmico 78, precursor mitocondrial	Hsp100
P38323	MCX1_LEVEDURA	GOA:interpro	chaperona similar a clpX mitocondrial MCX1	Hsp100
<b>Pequena proteína de choque térmico</b>				
P15992	HS26_LEVE-DURA	PMID:10581247	proteína de choque térmico 26	Hsp pequeno

Acesso Uniprot GroES/EL	Uniprot ID	Fonte GO	Anotação	Família
<b>Prefoldina</b>				
P48363	PFD3_LEVE-DURA	GOA:interpro	Subunidade de prefoldina provável 3	prefoldina
Q04493	PFD5_LEVE-DURA	GOA:interpro	subunidade de prefoldina 5	prefoldina
P43573	YFC3_LEVE-DURA	GOA:interpro	proteína de 91,4 kDa hipotética na região intergênica de STE2-FRS2	prefoldina
P46988	PFD1_LEVE-DURA	GOA:spkw	subunidade de prefoldina 1	KE2
P40005	PFD2_LEVE-DURA	GOA:spkw	subunidade de prefoldina 2	KE2
P53900	PFD4_LEVE-DURA	GOA:spkw	subunidade de prefoldina 4	KE2
P52553	PFD6_LEVE-DURA	GOA:spkw	subunidade de prefoldina 6	KE2
<b>Hsp 90</b>				
P02829	HS82_LEVE-DURA	GOA:interpro	proteína de choque térmico HSP82	Hsp90
P15108	HS83_LEVE-DURA	GOA:interpro	proteína de cognato de choque térmico HSC82	Hsp90
P06101	CC37_LEVE-DURA	GOA:spkw	co-chaperona de Hsp90 Cdc37	Cdc37
P33313	CNS1_LEVE-DURA	GOA:spkw	supressor 7 de ciclofinila 1	CNS1
P15705	STI1_LEVE-DURA	PMID:8972212	proteína de choque térmico STI1	
<b>Calnexina</b>				
P27825	CALX_LEVE-DURA	GOA:spkw	precursor de homólogo de calnexina	calnexina
<b>Complexo T de chaperoninas citosólicas</b>				
P12612	TCPA_LEVE-DURA	GOA:interpro	Proteína de complexo T 1, subunidade alfa	TCP-1, Hsp60
P39076	TCPB_LEVE-DURA	GOA:interpro	Proteína de complexo T 1, subunidade beta	TCP-1, Hsp60
P39078	TCPD_LEVE-DURA	GOA:interpro	Proteína de complexo T 1, subunidade delta	TCP-1, Hsp60
P40413	TCPE_LEVE-	GOA:interpro	Proteína de complexo T 1,	TCP-1,

Acesso Uniprot GroES/EL	Uniprot ID	Fonte GO	Anotação	Família
	DURA		subunidade epsilon	Hsp60
P39077	TCPG_LEVEDURA	GOA:interpro	Proteína de complexo T 1, subunidade gama	TCP-1, Hsp60
P42943	TCPH_LEVEDURA	GOA:interpro	Proteína de complexo T 1, subunidade eta	TCP-1, Hsp60
P47079	TCPQ_LEVEDURA	GOA:interpro	proteína de complexo T 1, subunidade teta	TCP-1, Hsp60
P39079	TCPZ_LEVEDURA	GOA:interpro	Proteína de complexo T 1, subunidade zeta	TCP-1, Hsp60
<b>Proteína específica</b>				
P48606	TBCA_LEVEDURA	GOA:spkw	Chaperona tubulina-específica A	Proteína específica
P53904	TBCB_LEVEDURA	GOA:spkw	Chaperona tubulina-específica B	Proteína específica
P46670	CIN2_LEVEDURA	GOA:spkw	co-fator de dobragem de tubulina C Cin2	Proteína específica
P40987	CIN1_LEVEDURA		co-fator de dobragem de tubulina D Cin1	Proteína específica
P39937	PAC2_LEVEDURA	GOA:spkw	co-fator de dobragem de tubulina E PAC2	Proteína específica
P21560	CBP3_LEVEDURA	GOA:spkw	proteína de CBP3, precursor mitocondrial	Proteína específica
Q12287	COXS_LEVEDURA	GOA:spkw	chaperona de cobre de oxidase de citocromo c	Proteína específica
P40202	LYS7_LEVEDURA	GOA:interpro	chaperona de cobre de superóxido dismutase 1	
Q02774	SHR3_LEVEDURA	PMID:10564255	Proteína de componente secretório SHR3	Proteína específica
P38293	UMP1_LEVEDURA	GOA:spkw	Fator de maturação de proteassoma UMP1	Proteína específica
P38784	VM22_LEVEDURA	PMID:7673216	proteína de montagem de ATPase vacuolar VMA22	Proteína específica
P38072	SCO2_LEVEDURA	GOA:spkw	Proteína de SCO2, precursor mitocondrial	proteína específica
P53266	SHY1_LEVEDURA	PMID:11389896	Proteína de SHY1	Proteína específica
P40046	VTC1_LEVEDURA	GOA:spkw	Chaperona de transportador vacuolar 1	Proteína específica

Acesso Uniprot GroES/EL	Uniprot ID	Fonte GO	Anotação	Família
P38958	PT00_LEVE-DURA	PMID:11498004	proteína de PET100, precursor mitocondrial	Proteína específica
<b>Isomerases de ligação de dissulfeto</b>				
P17967	PDI_LEVE-DURA	PMID:11157982	precursor de proteína dissulfeto isomerase	oxidorreductase de ligação de dissulfeto
P32474	EUG1_LEVE-DURA	PMID:11157982	precursor de EUG1 de proteína dissulfeto isomerase	oxidorreductase de ligação de dissulfeto
Q12404	MPD1_LEVEDURA	PMID:11157982	precursor de MPD1 de dissulfeto isomerase	oxidorreductase de ligação de dissulfeto
Q99316	MPD2_LEVEDURA	PMID:11157982	precursor de proteína dissulfeto isomerase MPD2 (EC 5.3.4.1)	oxidorreductase de ligação de dissulfeto
Q03103	ERO1_LEVE-DURA	PMID:9659913	precursor de oxidoreductina endoplasmica 1 (EC 1.8.4.-) (Proteína de oxidoreductase endoplasmica 1)	oxidorreductase de ligação de dissulfeto
P388866	FMO1_LEVEDURA	PMID:10077572	Monooxigenase tiol-específica (EC 1.14.13.-) (Monooxigenase flavina-dependente)	oxidorreductase de ligação de dissulfeto
<b>Peptidil-prolil cis-trans isomerases</b>				
P14832	CYPH_LEVEDURA	GOA:interpro	ciclofilina de peptidil-prolil cis-trans isomerase A/Cpr1/Cyp1/CPH1/Scc1	PPIase: tipo ciclofilina
P23285	CYPB_LEVE-DURA	GOA:interpro	Ciclofilina de peptidil-prolil cis-trans isomerase B/Cpr2/Cyp2	PPIase: tipo ciclofilina
P25719	CYPC_LEVEDURA	GOA:interpro	Peptidil-prolil cis-trans isomerase C/CYP3/CPR3, mitocondrial	PPIase: tipo ciclofilina
P25334	CYPR_LEVEDURA	GOA:interpro	Peptidil-prolil cis-trans isomerase CPR4/Scc3	
P35176	CYPD_LE-	GOA:interpro	Peptidil-prolil cis-trans iso-	PPIase:

Acesso Uniprot GroES/EL	Uniprot ID	Fonte GO	Anotação	Família
	VEDURA		merase D CypD/Cpr5	tipo ciclofílina
P53691	CYP6_LEVE-DURA	PMID:1094276 7	Peptidil-prolil cis-trans isomerase CPR6	PPlase: tipo ciclofílina
P47103	CYP7_LEVE-DURA	PMID:1094276 7	Peptidil-prolil cis-trans isomerase CYP7	PPlase: tipo ciclofílina
P53728	CYP8_LEVE-DURA	GOA:interpro	Peptidil-prolil cis-trans isomerase CYP8	PPlase: tipo ciclofílina
Q02770	Q02770	GOA:interpro	Yp1064cp	PPlase: tipo ciclofílina
P20081	FKBP_LEVE-DURA	GOA:interpro	proteína de ligação de FK506 FKB1/RBP1	PPlase: tipo FKBP
P32472	FKB2_LEVE-DURA	GOA:interpro	proteína de ligação de FK506 FKBP-13/FKB-15/FKB2, FPR2	PPlase: tipo FKBP
P38911	FKB3_LEVE-DURA	GOA:interpro	proteína nuclear de ligação de FK506 FKB-70/Npi46/Fpr3	PPlase: tipo FKBP
Q06205	FKB4_LEVE-DURA	GOA:interpro	proteína de ligação de FK506 FPR4	PPlase: tipo FKBP
P22696	ES-SI_LEVEDURA	GOA:spkw	proteína de ESS1	PPlase: tipo parvulina
Mista fracamente caracterizada				
P27697	ABC1_LEVE-DURA	GOA:spkw	proteína de ABC1, precursor mitocondrial	ABC1
P53193	YGB8_LEVE-DURA	GOA:interpro	proteína de 21,8 kDa hipotética em região intergênica de CKB1-ATE1	Hsp20
P28707	YKL7_LEVE-DURA	PMID:9632755	proteína de 24,1 kDa em região intergênica de VMA12-APN1	P23/wos2
P38932	VP45_LEVE-DURA	PMID:1143282 6	proteína associada à classificação de proteína vacuolar 45	similar a SEC1
Q12019	MDN1_LEVEDURA	GOA:spkw	midasina	

## MANIPULAÇÃO GENÉTICA

Na etapa iii), o processo inclui alterar expressão do gene ou produto de gene compensatório identificado na célula recombinante através de modificação genética para fornecer uma célula recombinante modificada.

- 5 Após identificação de um ou mais genes sobre-regulados, proteínas ou processos metabólicos, o genoma do hospedeiro pode ser modificado. Certos genes ou produtos de gene, embora identificados como sobre-regulados, podem não estar disponíveis para modulação porque eles são essenciais à célula ou são conhecidos afetar outros processos que podem ser essenciais à 10 célula ou organismo.

O genoma pode ser modificado incluindo um gene exógeno ou elemento de promotor no genoma ou no hospedeiro com um vetor de expressão, intensificando a capacidade de um gene identificado para produzir mRNA ou proteína, ou deletando ou rompendo um gene ou elemento de

- 15 promotor, ou reduzindo a capacidade de um gene para produzir mRNA ou proteína. O código genético pode ser alterado, assim afetando a transcrição e/ou translação de um gene, por exemplo através de técnicas de substituição, deleção ("knock-out"), co-expressão ou de inserção ("knock-in"). Genes adicionais para uma proteína desejada ou seqüência reguladora que modula a 20 transcrição de uma seqüência existente podem também ser inseridos.

## RECOMBINAÇÃO

O genoma da célula hospedeira que expressa proteína ou peptídeo recombinante pode ser modificado por meio de um evento alvejante genético que pode ser por inserção ou recombinação por exemplo recombi-

- 25 nação homóloga. Recombinação homóloga refere-se ao processo de recombinação de DNA com base em homologia de seqüência. Recombinação homóloga permite modificações sítio-específicas em genes endógenos e desse modo novas alterações podem ser criadas em um genoma. Uma etapa em recombinação homóloga é permuta de filamento de DNA, que envolve um 30 emparelhamento de um duplex de DNA de pelo menos um filamento de DNA que contém uma seqüência complementar para formar uma estrutura de recombinação intermediária contendo DNA de heteroduplex (ver, por exemplo

Radding, C. M. (1982) *Ann. Rev. Genet.* 16: 405; Pat. U.S. Nº 4.888.274). O DNA de heterodúplex pode ter várias formas, incluindo um filamento de três DNA contendo forma triplex em que um filamento complementar simples invade o dúplex de DNA (Hsieh, et al., *Genes and Development* 4: 1951 (1990); 5 Rao, et al., (1991) *PNAS* 88:2984)) e, quando dois filamentos de DNA complementares emparelham com um dúplex de DNA, uma junção de recombinação de Holliday clássica ou estrutura qui (Holliday, R., *Genet. Res.* 5: 282 (1964)) pode formar, ou uma alça D dobrada ("Diagnostic Applications of Double-D Loop Formation" U.S. Nº 07/755.462, depositada em 4 de set, 1991).

10 Uma vez formada, uma estrutura de heterodúplex pode ser solucionada através de rompimento e permuta de filamento, de forma que toda ou uma porção de um filamento de DNA invasor é encaixada em um dúplex de DNA recipiente, adicionando ou substituindo um segmento do dúplex de DNA recipiente. Alternativamente, uma estrutura de heterodúplex pode resultar em 15 conversão do gene, em que uma seqüência de um filamento invasor é transferida para um dúplex de DNA recipiente mediante reparo das bases desemparelhadas usando o filamento invasor como um modelo (*Genes*, 3<sup>a</sup> Ed. (1987) Lewina, B., John Wiley, Nova Iorque, N. I.; Lopez, et al., *Nucleic Acids Res.* 15: 5643(1987)). Seja pelo mecanismo de rompimento e rejunção ou 20 pelo(s) mecanismo(s) de conversão de gene, formação de DNA de heterodúplex nas juntas homologamente emparelhadas pode servir para transferir informação de seqüência genética de uma molécula de DNA para outra.

Em recombinação homóloga, o DNA entrante interage e integra em um sítio no genoma que contém uma seqüência de DNA substancialmente 25 homóloga. Em integração não-homóloga ("aleatória" ou "ilícita"), o DNA entrante não integra com uma seqüência homóloga no genoma mas em outro lugar, em uma de um número grande de localizações potenciais. Vários documentos descrevem o uso de recombinação homóloga em células mamíferas.

30 Vários constructos podem ser preparados para recombinação homóloga em um lócus identificado. Usualmente, a construção pode incluir pelo menos 10 bp, 20 bp, 30 bp, 40 bp, 50 bp, 70 bp, 100 bp, 500 bp, 1kbp, 2

kbp, 4 kbp, 5 kbp, 10 kbp, 15 kbp, 20 kbp, ou 50 kbp de seqüência homólogo com o lócus identificado. Podem ser envolvidas várias considerações determinando a extensão de homologia de seqüências de DNA identificadas, como, por exemplo, o tamanho do lócus identificado, disponibilidade de seqüências, 5 eficiência relativa dos eventos de dupla troca no lócus identificado e a similaridade da seqüência identificada com outras seqüências.

O DNA alvo pode incluir uma seqüência em que o DNA substancialmente isogênico flanqueia as modificações de seqüência desejadas com uma seqüência identificada correspondente no genoma a ser modificado. A 10 seqüência substancialmente isogênica pode ser pelo menos cerca de 95%, 97-98%, 99,0-99,5%, 99,6-99,9% ou 100% idêntica à seqüência identificada correspondente (com exceção das modificações de seqüência desejadas). O DNA alvo e o DNA identificado podem compartilhar extensões de DNA de pelo menos cerca de 10, 20, 30, 50, 75, 150 ou 500 pares de base que são 100% 15 idênticos.

Os constructos de DNA podem ser projetados para modificar o produto de gene endógeno, identificado. A seqüência homóloga para identificar o constructo pode ter uma ou mais deleções, inserções, substituições ou combinações destas, projetadas para romper a função do produto de gene 20 resultante. Em uma modalidade, a alteração pode ser a inserção de um gene marcador selecionável fundido na estrutura de leitura com a seqüência a montante do gene identificado.

O genoma pode também ser modificado usando deleção insercional. Nesta modalidade, o genoma é modificado recombinando uma seqüência no gene que inibe a formação do produto de gene. Esta inserção ou pode romper o gene inserindo um elemento separado, ou remover uma porção essencial do gene. Em uma modalidade, a deleção insercional inclui inserção de um gene que codifica para resistência a um estressor particular, como um antibiótico, ou para crescimento em um meio particular, por exemplo 25 para a produção de um aminoácido essencial.

O genoma pode também ser modificado pelo uso de transposons, que são elementos genéticos capazes de inserir em sítios em genomas

procariotes por mecanismos independentes de recombinação homóloga. Transposons podem incluir, por exemplo, Tn7 em *E. coli*, Tn554 em *S. aureus*, IS900 em *M. paratuberculosis*, IS492 de *Pseudomonas atlantica*, IS116 de *Streptomyces* e IS900 de *M. paratuberculosis*. Etapas acreditadas estar envolvidas em transposição incluem clivagem da terminação do transponson para render 3' OH; transferência do filamento, em que transposase reúne a terminação de 3'OH exposta do transponson e a seqüência identificada; e uma reação de transesterificação de etapa simples para render uma ligação covalente do transponson ao DNA identificado. A reação fundamental executada por transposase é em geral julgada ser corte ou permuta de filamento, o resto do processo é feito através de enzimas hospedeiro.

Em uma modalidade, um processo é fornecido para aumentar o nível de um gene identificado ou homólogo deste incorporando uma seqüência genética que codifica o gene ou homólogo no genoma através de recombinação. Em outra modalidade, um promotor é inserido no genoma para intensificar a expressão do gene identificado ou homólogo. Em uma modalidade separada, um processo é fornecido para diminuir a expressão de um gene identificado ou homólogo deste por recombinação com um gene inativo. Em outra modalidade, uma seqüência que codifica um gene diferente, que pode ter uma função separada na célula ou pode ser um gene repórter como um marcador de resistência ou um gene marcador do contrário detectável pode ser inserido em um genoma através de recombinação. Em ainda outra modalidade, uma cópia de pelo menos uma porção do gene identificado que foi mutado para uma ou mais localizações é inserida no genoma através de recombinação. A versão mutada do gene identificado não pode codificar uma proteína, ou a proteína codificada pelo gene mutado pode ser dada inativa, a atividade pode ser modulada (aumentada ou diminuída), ou a proteína mutante pode ter uma atividade diferente quando comparada à proteína nativa.

Há estratégias para knock-out genes em bactérias, que foram em geral exemplificadas em *E. coli*. Uma via é clonar um fragmento de DNA interno no gene em um vetor contendo um gene de resistência antibiótico (por exemplo ampicilina). Antes de as células serem transformadas por meio de

transferência conjugativa, transformação química ou eletroporação (Puehler, et al. (1984) *Advanced Molecular Genetics* Nova Iorque, Heidelberg, Berlim, Tóquio, Springer Verlag), uma origem de replicação, como a replicação de plasmídeo vegetativo (o lócus de *oriV*), é excisada e o fragmento de DNA restante é re-ligado e purificado (Sambrook, et al. (2000) *Molecular cloning: A laboratory manual*, terceira edição, Cold Spring Harbor, Nova Iorque, Cold Spring Harbor Laboratory Press). Alternativamente, plasmídeos resistentes a antibióticos tendo uma origem de replicação de DNA podem ser usados. Após transformação, as células foram banhadas por exemplo em placas de ágar LB contendo os antibióticos apropriados (por exemplo 200 µg/ml de ampicilina). As colônias que crescem nas placas contendo os antibióticos presumivelmente sofreram um evento de recombinação simples (Snyder, L., W. Champness, et al. (1997) *Molecular Genetics of Bacteria* Washington DC, ASM Press) que leva à integração do fragmento de DNA inteiro no genoma no lócus homólogo. Análise adicional das células resistentes a antibióticos para verificar que o knock-out de gene desejado ocorreu no lócus desejado é por exemplo por PCR diagnóstica (McPherson, M. J., P. Quirke, et al. (1991) *PCR: A Practical Approach* Nova Iorque, Oxford University Press). Aqui, pelo menos dois iniciadores de PCR são projetados: um que hibrida fora da região de DNA que foi usado para a construção do knock-out do gene; e um que hibridar dentro da cadeia principal de plasmídeo restante. Amplificação de PCR com sucesso do fragmento de DNA com o tamanho correto seguido por análise de seqüência de DNA verificará que o knock-out do gene ocorreu na localização correta no cromossomo bacteriano. O fenótipo da cepa mutante recentemente construída pode depois ser analisado por exemplo eletroforese em gel de poliacrilamida de SDS (Simpson, R., J. (2003) *Proteins and Proteomics - A Laboratory Manual*, Cold Spring Harbor, Nova Iorque, Cold Spring Harbor Laboratory Press).

Uma via alternativa para gerar um knock-out de gene é pelo uso de um replicon sensível à temperatura, como o replicon de pSC101 para facilitar a substituição do gene (Hamilton, et al. (1989) New process for generation deletions and gene replacements in *Escherichia coli*. *Journal of Bac-*

teriology 171(9): 4617-22). O processo prossegue através de recombinação homóloga entre um gene em um cromossomo e seqüências homólogas realizadas em uma temperatura de plasmídeo sensível à replicação de DNA. Transformação do plasmídeo no hospedeiro apropriado foi possível selecionar para integração do plasmídeo no cromossomo a 44°C. Crescimento subsequente destes cointegrates a 30°C leva a um segundo evento de recombinação, resultando em sua resolução. Dependendo de onde o segundo evento de recombinação ocorre, o cromossomo ou terá sofrido uma substituição de gene ou terá retido a cópia original do gene.

Outras estratégias para inibir expressão de produtos de gene particulares foram desenvolvidas. Por exemplo, interferência de RNA (RNAi), particularmente usando interferência de RNA pequena (siRNA), foi desenvolvida para extensivamente reduzir ou mesmo eliminar a expressão de um produto de gene particular. siRNAs são moléculas de RNA curtas, bi-filamentares que podem alvejar mRNAs complementares para degradação. RNAi é o fenômeno em que a introdução de um RNA bifilamentar suprime a expressão do gene homólogo. Moléculas de dsRNA são reduzidas *in vivo* para siRNAs de 21-23 nt que são os mediadores do efeito de RNAi. Sob introdução, os RNAs bifilamentares são processados em siRNAs de 20-25 nucleotídeos por uma enzima similar à RNase III chamada Dicer (etapa de iniciação). Depois, os siRNAs rearranjam-se em complexos contendo endorribonuclease conhecidos como complexos de silenciamento induzido por RNA (RISCs), desenrolando no processo. Os filamentos de siRNA subsequentemente guiam os RISCs para as moléculas de RNA complementares onde eles clivam e destroem o RNA cognato (etapa efetadora). Clivagem de RNA cognato ocorre próxima do meio da região ligada pelo filamento de siRNA. RNAi foi de forma bem sucedida usado para reduzir a expressão de gene em uma variedade de organismos incluindo peixe-zebra, nematódeos (*C. elegans*), insetos (*Drosophila melanogaster*), planária, cnidária, tripanossomas, camundongos e células mamíferas.

### MUTAÇÃO

O genoma pode também ser modificado mediante mutação de um

ou mais nucleotídeos em uma estrutura de leitura aberta que codifica um gene identificado, particularmente uma protease identificada. Técnicas para mutação genética, por exemplo mutagênese loco-dirigida, são bem conhecidas na técnica. Alguns métodos focalizam na geração de mutações aleatórias em DNA cromossômico como aqueles induzidos por raios X e químicas. Mutagênese alvejada para uma região definida de DNA inclui muitas técnicas, mais popular que outras. Métodos *in vitro* para mutagênese dirigida podem ser em geral agrupados em três categorias: i) processos que reestruturam os fragmentos de DNA, como mutagênese de cassette; ii) mutagênese aleatória localizada; e iii) mutagênese oligonucleotídeo-direcionada.

Mutagênese oligonucleotídeo-direcionada é com base no conceito que um oligonucleotídeo que codifica uma(s) mutação(ões) desejada(s) é anelado em um filamento do DNA de interesse e serve como um iniciador para iniciação da síntese de DNA. Desta maneira, o oligonucleotídeo mutagênico é incorporado no filamento recentemente sintetizado. Oligonucleotídeos mutagênicos incorporam pelo menos uma alteração de base mas podem ser projetados para gerar múltiplas substituições, inserções ou deleções. Exemplos incluem processos com base em PCR e praticamente todos os processos não baseados em PCR em uso hoje. Estas técnicas incluem seleção antibiótica positiva (Lewis, M. K. e Thompson, D. V. (1990) *Nucl. Acids Res.* 18, 3439; Bohnsack, R. N. (1996) *Meth. Mol. Biol.* 57, 1; Vavra, S. e Brondyk, W. H. (1996) *Promega Notes* 58, 30; *Altered Sites® II in vitro Mutagenesis Systems Technical Manual #TM001*, Promega Corporation), seleção de sítio de restrição único (Deng, W. P. e Nickoloff, J. A. (1992) *Anal. Biochem.* 200, 81), incorporação de uracila (Kunkel, T. A. (1985) *Proc. Natl. Acad. Sci. USA* 82, 488; Kunkel, T. A., Roberts, J. D. e Zakour, R. A. (1987) *Meth. Enzymol.* 154, 367) e incorporação de fosforotioato (Taylor, J. W., Ott, J. e Ecksteina, F., (1985) *Nucl. Acids Res.* 13, 8764; Nakamaye, K. e Ecksteina, F., (1986) *Nucl. Acids Res.* 14, 9679). Oligonucleotídeos podem também codificar uma biblioteca de mutações randomizando a composição de base nos sítios durante a síntese química resultando em oligonucleotídeos degenerados ou "dopados". A habilidade para localizar e especificar mutações é

grandemente intensificada pelo uso de oligonucleotídeos sintéticos hibridados com vetor de plasmídeo contendo inserção de DNA.

- O formato geral para mutagênese dirigida é: desnaturação de plasmídeo DNA que contém o modelo de interesse (cDNA, promotor, etc.)
- 5 para produzir regiões unifilamentares; anelamento de um oligonucleotídeo mutante sintético no filamento identificado; síntese de um filamento complementar novo usando, por exemplo, T4 DNA polimerase; e vedação do corte resultante entre a terminação do filamento novo e do oligonucleotídeo, por exemplo usando T4 DNA Ligase. O heterodúplex resultante é propagado
- 10 através de transformação, por exemplo, em *E. coli*. Processos de seleção e enriquecimento foram incorporados nos processos de mutagênese para grandemente melhorar a eficiência de recuperação do filamento mutante e taxas que chegam a 80-90% são possíveis. Numerosos processos existem para gerar tipos diferentes de mutações e intensificar a seleção do mutante.
- 15 Exemplos de processos para intensificar a seleção do mutante incluem seleção antibiótica positiva do filamento mutante, uso de um filamento de DNA contendo uracila, que pode ser seletivamente degradado *in vivo* e incorporação de análogo de dNTP, que pode dar um filamento de DNA de heterodúplex impérvio para digestão. Alguns métodos podem ser combinados, como
- 20 mutagênese de cassete e o uso de oligonucleotídeos "dopados" para criar uma biblioteca de mutações aleatórias em uma região pequena, definida.

Uma extensão dos assim-chamados processos "padrão" de mutagênese dirigida inclui aqueles que contam com amplificação de DNA, especificamente a reação em cadeia de polimerase (PCR). O ponto comum principal em mutagênese dirigida é o uso de um oligonucleotídeo mutagênico. O oligonucleotídeo mutagênico deveria hibridar eficazmente com o modelo. Para hibridização eficiente, pode haver, por exemplo, 100% de emparelhamento de base em qualquer terminação da seqüência identificada sem formação de estrutura secundária, mas pode também ser menor que 100% de identidade, como 98%, 95%, 92%, 90%, 85%, 80%, 70% ou apenas uma porção da seqüência pode ser idêntica. Para substituições pequenas, 10-15 bases que hibridam com qualquer um dos lados da disparidade são usual-

mente suficientes. A composição da terminação 3' do iniciador é particularmente importante visto que as polimerases tipicamente não se estendem de uma terminação 3' desemparelhada ou fracamente hibridada.

A base para mutagênese dirigida através de seleção antibiótica positiva é que um oligonucleotídeo ou oligonucleotídeos de seleção são anelados simultaneamente, com o oligonucleotídeo mutagênico, para reparar um gene de resistência a antibióticos (10-13). Seleção pelo filamento mutante é permitida mediante resistência a antibióticos do DNA mutado e sensibilidade do filamento não-mutado. Este método oferece um meio muito eficiente para gerar um número indefinido das mutações desejadas com pouco tempo em mãos.

Mutagênese sítio-direcionada pelo uso de um sítio de restrição único é com base nos processos de Deng e Nickoloff (Deng, W. P. e Nickoloff, J. A. (1992) *Anal. Biochem.* 200, 81). Neste método, um oligonucleotídeo de seleção que contém uma seqüênciia mutada para um sítio de restrição único é anelado simultaneamente com o oligonucleotídeo mutagênico. O oligonucleotídeo de seleção transmite o sítio não-essencial imune para restrição pela enzima correspondente. Seleção pelo filamento mutante é intensificada digerindo o fundo geral resultante de plasmídeos com a enzima de restrição única. A digestão lineariza o plasmídeo parental assim eficazmente diminuindo sua habilidade para transformar bactérias.

Mutagênese sítio-direcionada através de incorporação de deoxiuridina contam com a habilidade de uma cepa hospedeira para degradar o DNA modelo contendo uracila (U) no lugar de timidina (T). Um número pequeno de dUTPs está incorporado no filamento modelo no lugar de dTTP em um hospedeiro que carece de atividades de dUTPase (*dut-*) e de uracil N-deglicosidase (*ung-*). (Uracila por si não é mutagênica e seus pares de base com adenina.) Normalmente, dUTPase degrada deoxiuridina e uracil N-deglicosidase remove qualquer uracila incorporada. Replicação de pós-mutação em uma cepa de *dut+* *ung+* é usada depois para degradar o filamento de DNA não-identificado. Este método requer que o DNA unifilamentado seja usado de forma que apenas um filamento contenha os Us que

são suscetíveis à degradação.

O método de incorporação de fosforotioato para mutagênese dirigida conta com a habilidade de um análogo de dNTP que contém um grupo tiol de dar DNA de heterodúplex resistente à digestão da enzima de restrição.

- 5 O filamento de mutante é estendido do oligonucleotídeo mutagênico e sintetizado na presença de dCTPalphaS. DNA modelo não-usado é removido através de digestão com uma exonuclease. Teoricamente, apenas DNA circular, heterodúplex permanece. O heterodúplex é depois entalhado, mas não cortado, no(s) sítio(s) de restrição. Exonuclease III é usada para digerir o
- 10 filamento entalhado e o fragmento restante depois age como um iniciador para repolimerização, criando um homodúplex mutante.

No método com base em reação em cadeia de polimerase (PCR) para gerar uma mutação em DNA, um modelo é amplificado usando um conjunto de iniciadores de oligonucleotídeo gene-específicos exceto que um

- 15 oligonucleotídeo, ou mais nos protocolos que usam amplificações múltiplas, contenha a mutação desejada. Variações incluem alterar o sítio de hibridização dos oligonucleotídeos para produzir múltiplos fragmentos de PCR de sobreposição com a mutação na sobreposição e o método de "mega-iniciador", que usa três oligonucleotídeos e duas rodadas de amplificação
- 20 em que um filamento do produto da primeira amplificação serve como um iniciador na segunda amplificação.

- 25 No método de extensão de sobreposição, iniciadores de oligodeoxirribonucleotídeo complementares (oligo) são usados e a reação em cadeia de polimerase para gerar dois fragmentos de DNA tendo terminações de sobreposição. Estes fragmentos são combinados em uma reação de 'fusão' subsequente em que as terminações de sobreposição anelam-se, permitindo a sobreposição de 3' de cada filamento servir como um iniciador para a extensão de 3' do filamento complementar. O produto de fusão resultante é amplificado também por PCR. Alterações específicas na seqüência de nucleotídeo (nt) podem ser introduzidas incorporando alterações de nucleotídeo nos iniciadores de oligo de sobreposição.
- 30

## CONSTRUCTOS DE VETOR

Em uma modalidade separada, a célula hospedeira é modificada incluindo um ou mais vetores que codificam um gene identificado, tipicamente um modulador de dobramento ou um co-fator de um modulador de dobramento. Em outra modalidade, a célula hospedeira é modificada intensificando 5 um promotor para um modulador de dobramento ou um co-fator para um modulador de dobramento, incluindo acrescentar um promotor exógeno ao genoma de célula hospedeira.

Em outra modalidade, a célula hospedeira é modificada incluindo 10 um ou mais vetores que codificam um inibidor de um gene compensatório identificado, como um inibidor de protease. Um tal inibidor pode ser uma molécula anti-sentido que limita a expressão do gene compensatório identificado, um co-fator do gene identificado ou um homólogo do gene identificado. Anti-sentido é em geral usado para referir-se a uma molécula de ácido nu- 15cléico com uma seqüência complementar para pelo menos uma porção do gene identificado. Além disso, o inibidor pode ser um RNA de interferência ou um gene que codifica um RNA de interferência. Em organismos eucarióticos, um tal RNA de interferência pode ser um RNA de interferência pequena ou uma ribozima, como descrito, por exemplo, em Burn, A. et al. ( 1998) *Nature* 20 391:806-11, Elbashir et al. (2001) *Genes & Development* 15(2):188-200, El- bashir et al. (2001) *Nature* 411(6836):494-8, Patentes U.S. Nºs 6.506.559 de Carnegie Institute, 6.573.099 de Benitec, Pedidos de Patente U.S. Nºs 2003/0108923 de Whitehead Inst. e 2003/0114409, Publicações do PCT Nºs WO03/006477, WO03/012052, WO03/023015, WO 03/056022, WO 25 03/064621 e WO 03/070966. O inibidor pode também ser outra proteína ou peptídeo. O inibidor pode, por exemplo, ser um peptídeo com uma seqüência de consenso para a protease ou proteína de protease. O inibidor pode tam- bém ser uma proteína ou peptídeo que pode produzir uma molécula inibidora direta ou indireta para a protease ou proteína de protease no hospedeiro. 30 Inibidores de protease podem incluir Amastatina, E-64, Antipaína, Elastatinal, APMSF, Leupeptina, Bestatina, Pepstatina, Benzamidina, 1,10-Fenantrolina, Quimiostatina, Fosforamidon, 3,4-dicloroisocumarina, TLCK, DFP, TPCK.

Mais de 100 inibidores de protease de proteína de ocorrência natural foram identificados até agora. Eles foram isolados em uma variedade de organismos de bactérias para animais e plantas. Eles se comportam como inibidores de ligação firme reversíveis ou pseudo-irreversíveis de proteases impedindo

5 acesso do substrato ao sítio ativo através de obstáculo estérico. Seu tamanho é também extremamente variável de 50 resíduos (por exemplo BPTI: Inibidor de Tripsina Pancreático Bovino) até 400 resíduos (por exemplo alfa-1PI: Inibidor de Proteinase alfa-1). Eles são estritamente classe-específicos exceto

10 as proteínas da família de alfa-macroglobulina (por exemplo alfa-2 macroglobulina) que liga e inibe a maioria das proteases através de um mecanismo de captura molecular.

Um constructo vetor ou DNA exógena pode ser transfeccionado ou transformado na célula hospedeira. Técnicas para transfeccionar e transformar células eucarióticas e procarióticas respectivamente com ácidos nucleicos exógenos são bem conhecidas na técnica. Estas podem incluir absorção mediada por vesícula de lipídio, transfecção mediada por fosfato de cálcio (co-precipitação de fosfato de cálcio/DNA), infecção viral, particularmente usando vírus modificados como, por exemplo, adenovírus modificados, microinjeção e eletroporação. Para transformação procariótica, as técnicas

15 podem incluir absorção mediada por choque térmico, fusão de protoplasto bacteriano com células intactas, microinjeção e eletroporação. Técnicas para transformação de planta incluem transferência mediada por *Agrobacterium*, como por *A. tumefaciens*, micropojéteis de tungstênio ou ouro rapidamente impelidos, eletroporação, microinjeção e absorção mediada por polietileno

20 glicol. O DNA pode ser DNA uni ou bifilamentar, linear ou circular, relaxado ou super-enrolado. Para várias técnicas para transfeccionar células mamíferas, ver, por exemplo, Keown et al. (1990) *Processes in Enzymology* Vol. 185, págs. 527-537.

Para eventos de recombinação, os constructos podem incluir uma

30 ou mais seqüências de inserção que podem inserir ou transpor uma ou mais seqüência de ácido nucléico em uma seqüência diferente. Porém, a construção pode ser projetada para expressão exógena de um gene compensa-

tório identificado ou homólogo deste sem incorporação no DNA/genoma celular existente.

Os constructos podem conter um, ou mais que um, sítio de entrada de ribossoma interno (IRES). O constructo pode também conter um promotor operavelmente ligado à seqüência de ácido nucléico que codifica pelo menos uma porção do gene identificado, ou um co-fator do gene identificado, uma versão mutante de pelo menos uma porção do gene compensatório identificado, ou no caso de proteases, um inibidor do gene identificado. Alternativamente, o constructo pode ser sem promotor. Em casos em que o constructo não é projetado para incorporar no DNA/genoma celular, o vetor tipicamente contém pelo menos um elemento promotor. Além das seqüências de ácido nucléico, o vetor de expressão pode conter seqüências de marcador selecionáveis. Os constructos de expressão pode também conter sítios para iniciação, terminação de transcrição e/ou sítios de ligação de ribossoma. Os constructos identificados podem ser inseridos e expressados em qualquer célula procariótica ou eucariótica, incluindo, mas não limitada às células bacterianas, como *P. fluorescens* ou *E. coli*, células de levedura, células mamíferas, como células de CHO, ou células de planta.

Vetores de clonagem podem incluir por exemplo plasmídeo p-BR322 (Bolivar, Rodriguez et al. 1977), a série de pUC dos plasmídeos (Vieira e Messing 1982), pBluescript (Short, Fernandez et al. 1988), pACYC177 e pACYC184 (Chang e Cohen 1978). Promotores exógenos para o uso em tais constructos, incluem, mas não são limitados a, o promotor de PL de fago lambda, promotores lac de *E. coli*, trp de *E. coli*, phoA de *E. coli*, tac de *E. coli*, SV40 tardia, LTRs de SV40 tardia, retroviral, genes de PGK1, GAL1, de GAL10, CYC1, PH05, TRP1, ADH1, ADH2, forglimaldeídeo fosfato desidrogenase, hexocinase, piruvato descarboxilase, fosfofructocinase, triose fosfato isomerase, fosfoglucose isomerase, feromônio de fator de emparelhamento de glucocinase alfa, promotor de PRBI, GUT2, de GPDI, promotor de metalotioneína e/ou promotores virais mamíferos, como aqueles derivados de adenovírus e vírus de vaccínia. Outros promotores serão conhecidos a alguém versado na técnica.

Promotores para vetores exógenos, ou promotores exógenos projetados para ser inseridos no genoma pode ser com base em elementos de resposta específicos em uma célula. Por exemplo, os promotores podem ser responsivos aos compostos químicos, por exemplo a antranilato ou benzoato,

5 como descritos na Publicação do PCT Nº WO 2004/005221. Os constructos podem incluir um ou mais promotores. Estes podem ser independentes, ou podem estar em tandem. Por exemplo os promotores podem ser projetados de forma que um gene compensatório identificado seja sobre ou sub-regulado em um prazo particular com a proteína ou peptídeo recombinante. Por exemplo, em um caso em que o gene identificado é um modulador de dobramento, o modulador ou co-fator de dobramento pode ser induzido logo antes da indução da proteína ou peptídeo recombinante. Promotores podem incluir, mas não são limitados aos seguintes:

10

Promotor	Fonte	Regulação	Indução
lac	<i>E. coli</i>	lacI, lacI <sup>q</sup>	IPTG
lacUV5	<i>E. coli</i>	lacI, lacI <sup>q</sup>	IPTG
tac (híbrido)	<i>E. coli</i>	lacI, lacI <sup>q</sup>	IPTG
trc (híbrido)	<i>E. coli</i>	lacI, lacI <sup>q</sup>	IPTG
P <sub>syn</sub> (sintético)	<i>E. coli</i>	lacI, lacI <sup>q</sup>	IPTG
trp	<i>E. coli</i>		inanição de triptofano
araBAD	<i>E. coli</i>	araC	1-arabinose
lpp <sup>a</sup>	<i>E. coli</i>	IPTG	lactose
1pp-lac (híbrido)	<i>E. coli</i>	lacI	IPTG
phoA	<i>E. coli</i>	phoB (positivo) phoR (negativo)	inanição de fosfato
recA	<i>E. coli</i>	lexA	ácido nalidíxico
proU	<i>E. coli</i>		osmolaridade
cst-1	<i>E. coli</i>		inanição de glicose
tetA	<i>E. coli</i>		tetraciclina
cadA	<i>E. coli</i>	cadR	pH
nar	<i>E. coli</i>	fnr	condições anaeróbicas
P <sub>L</sub>	λ	λ clts857	térmico (alteração para

Promotor	Fonte	Regulação	Indução
			42°C)
cspA	<i>E. coli</i>		térmico (alteração para abaixo de 20°C)
T7	T7	$\lambda$ clts857	térmico
Operador de T7-lac	T7	lacI <sup>q</sup>	IPTG
Operador de T3-lac	T3	lacI <sup>q</sup>	IPTG
Operador de T5-lac	T5	lacI, lacI <sup>q</sup>	IPTG
Gene de T4 32	T4		Infecção de T4
Operador de n-prM-lac	<i>Bacillus</i>	lacI <sup>q</sup>	IPTG
VHb	<i>Vitreoscilla</i>		oxigênio
Proteína A	<i>S. aureus</i>		

Constructos podem incluir marcadores de seleção para identificar células modificadas. Genes marcadores selecionáveis adequados incluem, mas não são limitados a: genes que conferem a habilidade para crescer em certos substratos de meios, como o gene tk (timidina cinase) ou o gene hprt (hipoxantina fosforibosiltransferase) que confere a habilidade para crescer em meio HAT (hipoxantina, aminopterina e timidina); o gene gpt bacteriano (guanine/xantina fosforibosiltransferase) que permite crescimento em meio MAX (ácido micofenólico, adenina e xantina). Ver, por exemplo, Song, K-Y., et al. (1987) *Proc. Nat'l Acad. Sci. U. S. A.* 84:6820-6824; Sambrook, J. et al. (1989) *Molecular Cloning - A Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y., Capítulo 16. Outros exemplos de marcadores selecionáveis incluem: genes que conferem resistência a compostos como antibióticos, genes que conferem a habilidade para crescer em substratos selecionados, genes que codificam proteínas que produzem sinais detectáveis como luminescência, como proteína fluorescente verde, proteína fluorescente verde intensificada (eGFP). Uma ampla variedade de tais marcadores é conhecida e disponível, incluindo, por exemplo, genes de resistência

a antibióticos como o gene de resistência à neomicina (neo) (Southern, P. e P. Berg, (1982) *J. Mol. Appl. Genet.* 1:327-341); e o gene de resistência à higromicina (hyg) ((1983) *Nucleic Acids Research* 11:6895-6911, e Te Riele, H., et al. (1990) *Nature* 348:649-651). Outros genes marcadores selecionáveis incluem: ácido acetoidróxi sintase (AHAS), fosfatase alcalina (AP), galactosidase beta (LacZ), glucoronidase beta (GUS), cloranfenicol acetiltransferase (CAT), proteína fluorescente verde (GFP), proteína fluorescente vermelha (RFP), proteína fluorescente amarela (YFP), proteína fluorescente ciana (CFP), peroxidase de raiz-forte (HRP), luciferase (Luc), nopalina sintase (NOS), octopina sintase (OCS) e derivados destes. Marcadores selecionáveis múltiplos estão disponíveis que confere resistência a ampicilina, bleomicina, cloranfenicol, gentamicina, higromicina, canamicina, lincomicina, metotrexato, fosfinotricina, puromicina e tetraciclina. Genes marcadores selecionáveis adicionais úteis nesta invenção, por exemplo, são descritos nas patentes U. S. Nos: 6.319.669; 6.316.181; 6.303.373; 6.291.177; 6.284.519; 6.284.496; 6.280.934; 6.274.354; 6.270.958; 6.268.201; 6.265.548; 6.261.760; 6.255.558; 6.255.071; 6.251.677; 6.251.602; 6.251.582; 6.251.384; 6.248.558; 6.248.550; 6.248.543; 6.232.107; 6.228.639; 6.225.082; 6.221.612; 6.218.185; 6.214.567; 6.214.563; 6.210.922; 6.210.910; 6.203.986; 6.197.928; 6.180.343; 6.172.188; 6.153.409; 6.150.176; 6.146.826; 6.140.132; 6.136.539; 6.136.538; 6.133.429; 6.130.313; 6.124.128; 6.110.711; 6.096.865; 6.096.717; 6.093.808; 6.090.919; 6.083.690; 6.077.707; 6.066.476; 6.060.247; 6.054.321; 6.037.133; 6.027.881; 6.025.192; 6.020.192; 6.013.447; 6.001.557; 5.994.077; 5.994.071; 5.993.778; 5.989.808; 5.985.577; 5.968.773; 5.968.738; 5.958.713; 5.952.236; 5.948.889; 5.948.681; 5.942.387; 5.932.435; 5.922.576; 5.919.445; e 5.914.233.

Deleções podem ser pelo menos cerca de 5 bp, 10 bp, 20 bp, 30 bp, 40 bp ou 50 bp, comumente pelo menos cerca de 100 bp, e em geral não mais que cerca de 20 kbp, onde a deleção normalmente pode incluir pelo menos uma porção da região de codificação incluindo uma porção de ou um ou mais éxons, uma porção de ou um ou mais íntrons, e pode ou não pode incluir uma porção das regiões de não-codificação de flanqueamento, particularmente a região de não-codificação de 5' (região reguladora transcri-

cional). Desse modo, a região homóloga pode estender-se além da região de codificação na região de não-codificação de 5' ou alternativamente na região de não-codificação de 3'. Inserções podem em geral não exceder a 10 kbp, usualmente não exceder a 5 kbp, em geral sendo pelo menos 50 bp, mais 5 usualmente pelo menos 200 bp.

A(s) região(ões) de homologia pode(m) incluir mutações onde as mutações podem também inativar o gene identificado, fornecendo um deslocamento de estrutura, ou alterando um aminoácido fundamental, ou a mutação pode corrigir um alelo disfuncional, etc. Usualmente, a mutação pode 10 ser uma alteração sutil, não excedendo cerca de 5% das seqüências de flanqueamento homólogas.

O constructo pode ser preparado de acordo com os processos conhecidos na técnica, vários fragmentos podem ser reunidos, introduzidos em vetores apropriados, clonados, analisados e depois manipulados também 15 até que o constructo desejado tenha sido alcançado (ver, por exemplo Figuras 5-11). Várias modificações podem ser feitas à seqüência, para permitir análise de restrição, excisão, identificação de sondas, etc. Mutações silenciosas podem ser introduzidas, como desejado. Em vários estágios, análise de restrição, seqüenciação, amplificação com a reação em cadeia de polimerase, 20 reparo de iniciador, mutagênese *in vitro*, etc. podem ser empregados. Processos para a incorporação de genes de resistência a antibióticos e fatores de seleção negativos são familiarizados àqueles de habilidade usual na técnica (ver, por exemplo, WO 99/15650; patente U.S. Nº 6.080.576; patente U.S. Nº 25 6.136.566; Niwa, et al., *J. Biochem.* 113:343-349 (1993); e Yoshida, et al., *Transgenic Research*, 4:277-287 (1995)).

O constructo pode ser preparado usando um vetor bacteriano, incluindo um sistema de replicação procariótico, por exemplo uma origem reconhecível por uma célula procariótica como *P. fluorescens* ou *E. coli*. Um marcador, igual ou diferente do marcador a ser usado para inserção, pode ser 30 empregado que pode ser removido antes da introdução na célula identificada. Uma vez o vetor contendo o constructo foi completado, ele pode ser também manipulado, como por deleção de certas seqüências, linearização, ou intro-

duzindo mutações, deleções ou outras seqüências na seqüência homóloga. Após manipulação final, o constructo pode ser introduzido na célula.

O processo pode ser iterativo. Em uma modalidade, após modificação do hospedeiro e da expressão da proteína recombinante no hospedeiro modificado, um perfil genético da célula hospedeira modificada é analisado para identificar um ou mais outros genes identificados na expressão que é alterada na célula hospedeira modificada. Em particular, genes compensatórios podem ser aqueles que mostram expressão aumentada no hospedeiro modificado expressando proteína recombinante quando comparada a uma célula hospedeira modificada que não expressa a proteína ou peptídeo recombinante, ou quando comparada a uma célula hospedeira inalterada. O processo também inclui alterar a expressão do gene ou outros genes identificados e expressar a proteína ou peptídeo na célula duplamente modificada. Estas etapas podem ser iteradas para melhorar a expressão de proteína e podem ser repetidas uma, duas, três, quatro, cinco, seis, sete, oito, nove, ou pelo menos dez vezes.

#### PRODUÇÃO DE PROTEÍNA

O processo da invenção otimamente leva à produção aumentada de proteína ou peptídeo recombinante em uma célula hospedeira. A produção aumentada pode incluir uma quantidade aumentada de proteína por grama de proteína do hospedeiro em uma quantidade dada de tempo, ou pode incluir um aumento no comprimento de tempo que a célula hospedeira está produzindo proteína ou peptídeo recombinante. A produção aumentada pode também incluir uma melhoria nos requerimentos para crescimento da célula hospedeira recombinante. A produção aumentada pode ser uma produção aumentada de proteína ou peptídeo de comprimento total. Se a melhoria for em níveis aumentados de proteína, a proteína ou peptídeo pode ser produzido em um ou mais corpos de inclusão em uma célula hospedeira.

A produção aumentada pode ser alternativamente um nível aumentado de proteína ou peptídeo ativo por grama de proteína produzida, ou por grama de proteína do hospedeiro. A produção aumentada pode também ser um nível aumentado de proteína ou peptídeo recuperável, como proteína

solúvel, produzida por grama de recombinante ou por grama de proteína de célula hospedeira. A produção aumentada pode também ser qualquer combinação de nível total aumentado e nível de ativo ou solúvel aumentado de proteína.

5           Produção aumentada é tipicamente medida comparando o nível de produção após um certo período de indução em uma célula modificada para a mesma indução na célula inalterada.

#### SOLÚVEL/INSOLÚVEL

10          A expressão melhorada de proteína recombinante pode ser um aumento na solubilidade da proteína. A proteína ou peptídeo recombinante pode ser produzido e restabelecido do citoplasma, periplasma ou meio extracelular da célula hospedeira. A proteína ou peptídeo pode ser insolúvel ou solúvel. A proteína ou peptídeo pode incluir uma ou mais seqüências alvos ou seqüências para ajudar a purificação.

15          Em certas modalidades, a invenção fornece um processo para melhorar a solubilidade de uma proteína ou peptídeo recombinante em uma célula hospedeira. O termo "solúvel" como aqui usado significa que a proteína não é precipitada através de centrifugação entre cerca de 5.000 e 20.000 x gravidade quando girado durante 10-30 minutos em um tampão sob condições fisiológicas. Proteínas solúveis, ativas são capazes de exibir a função, e não fazem parte de um corpo de inclusão ou outra massa precipitada.

20          A invenção pode também melhorar a recuperação das proteínas ou peptídeos recombinantes ativos. Por exemplo, a interação entre um polipeptídeo identificado e um de origem, polipeptídeo variante, polipeptídeo segmento-substituído e/ou polipeptídeo resíduo-substituído pode ser medida por qualquer ensaio conveniente *in vitro* ou *in vivo*. Desse modo, ensaios *in vitro* podem ser usados para determinar qualquer interação detectável entre um polipeptídeo identificado e, por exemplo entre enzima e substrato, entre hormônio e receptor de hormônio, entre anticorpo e antígeno, etc. Tal detecção pode incluir a medição das alterações colorimétricas, alterações na radioatividade, alterações na solubilidade, alterações no peso molecular quando medido por eletroforese em gel e/ou processos de exclusão em gel;

etc. Ensaios *in vivo* incluem, mas não são limitados a, ensaios para detectar efeitos fisiológicos, por exemplo ganho de peso, alteração no equilíbrio de eletrólito, alteração no tempo de coágulo sanguíneo, alterações na dissolução de coágulo e a indução de resposta antigênica. Em geral, qualquer ensaio *in vivo* pode ser usado desde que um parâmetro variável exista para detectar uma alteração na interação entre o polipeptídeo identificado e o de interesse. Ver, por exemplo, patente US Nº 5.834.250.

#### CITOPLÁSMICO/PERIPLÁSMICO/SECRETADO

Em certas modalidades, a proteína pode também ser segregada no periplasma se fundida em uma seqüência de secreção de sinal apropriada. Em uma modalidade, a seqüência sinal pode ser uma proteína de ligação de fosfato, um peptídeo sinal de secreção de proteína de ligação de Lys-Arg-Orn (LAObp ou KRObp), um peptídeo de sinal de secreção de Porina de Membrana Externa E (OprE), um peptídeo de sinal de secreção de azurina, um peptídeo de sinal de secreção de proteína de ligação de ferro (III) [Fe(III)bp] ou um peptídeo de sinal de secreção de lipoproteína B (LprB).

Em uma modalidade, nenhuma condição ou agente de promoção de ligação de dissulfeto adicional é requerido para restabelecer o polipeptídeo contendo ligação de dissulfeto identificado na forma ativa, solúvel da célula hospedeira modificada ou célula duplamente ou múltipla modificada. Em uma modalidade, o peptídeo, polipeptídeo, proteína transgênica(o) ou fragmento deste tem uma conformação intramolecular dobrada em seu estado ativo. Foi observado que as proteínas mamíferas complexas solúveis no citoplasma podem configurar-se apropriadamente ao posicionamento apropriado dos grupos tiol para formação de ligação de dissulfeto posterior no periplasma. Em uma modalidade, o peptídeo, polipeptídeo, proteína transgênica(o) ou fragmento contém pelo menos uma ligação de dissulfeto intramolecular em seu estado ativo; e talvez até 2, 4, 6, 8, 10, 12, 14, 16, 18 ou 20 ou mais ligações de dissulfeto.

Em uma modalidade, mais que 50% do peptídeo, polipeptídeo, proteína transgênica(o) ou fragmento deste expresso produzido será produzido como peptídeos, polipeptídeos, proteínas simples funcionais, ou frag-

mentos destes em forma solúvel, ativa ou forma insolúvel facilmente renaturada no citoplasma ou periplasma. Em outra modalidade cerca de 60%, 70%, 75%, 80%, 85%, 90%, 95% da proteína expressa são obtidos ou facilmente renaturados em forma ativa.

5

### EXEMPLOS

As cepas bacterianas usadas no estudo atual estão listadas na TABELA 1. Cepas de *P. fluorescens* foram crescidas em frascos de agitação a 30°C. OD<sub>575</sub> foi registrada para cada cepa em vários pontos de tempo.

**TABELA 1. VISÃO GERAL DAS CEPAS BACTERIANAS**

Cepa	Genótipo de Cepa Relevante	Plasmídeo	Proteína recombinante
MB214	cepa hospedeira de <i>P. Fluorescens</i>		
MB206	<i>pyrF</i>		
DC240	<i>pyrF</i>	pDOW2415	Nitrilase
DC271	<i>pyrF</i>	pDOW1323	pbp:hGH
DC280	<i>pyrF</i>	pDOW1339	plasmídeo apenas de vetor
DC369	<i>pyrF</i>	pDOW1426	hGH
DC462	<i>pyrF</i>	pDOW3501	GrpE, DnaKJ
DC463	<i>pyrF</i>	pDOW3501, pDOW1426	GrpE, DnaKJ, hGH
HJ104	<i>pyrF</i>	pDOW1349	hGH-COP
DC370	<i>pyrF, hslU</i>		
DC372	<i>pyrF, hslU</i>	pDOW1426	hGH
DC373	<i>pyrF, hslU</i>	pDOW1323	pbp:hGH
HJ105	<i>pyrF, hslU</i>	pDOW1349	hGH-COP
DC417	<i>pyrF, hslUV</i>		
HJ115	<i>pyrF, hslUV</i>	pDOW1426	hGH
HJ117	<i>pyrF, hslUV</i>	pDOW1349	hGH-COP

10

Os plasmídeos usados nos experimentos a seguir estão listados na Tabela 2.

**TABELA 2: VISÃO GERAL DOS PLASMÍDEOS**

Plasmídeos	Relevância
pDOW2236	vetor de clonagem
pDOW2240	Ptac <i>grpE-dnaKJ, pyrF</i> <sup>+</sup>
pDOW2247	Pmtl nenhum gene recombinante; vetor vazio
pDOW3501	Pmti <i>grpE-dnaKJ, pyrF</i> <sup>+</sup>
pDOW1349	<i>pyrF</i> <sup>+</sup> , <i>hGH::COP</i>
pDOW1426	<i>pyrF</i> <sup>+</sup> , hGH
pDOW1261-2	vetor suicida, <i>pyrF</i> <sup>+</sup>
pDOW2050	usado para construção das cepas de deleção de <i>hsfUV</i>

**COLETÂNEA DE AMOSTRA E ISOLAMENTO DE RNA**

Todas as amostras foram colhidas de uns experimentos de frascos de agitação de 250 ml padrão. As amostras foram tiradas em pontos de tempo diferentes como indicados nas figuras. A cada ponto de tempo, 10 ml de cultura de célula dos frascos de agitação foram colhidos e misturados com 10 ml de reagente *RNAlater* (Ambion, Austin, TX) para estabilizar o RNA.

**HIBRIDIZAÇÃO DE MICROARRANJO E ANÁLISE DE DADOS**

Para cada amostra de RNA, os nucleotídeos fluorescentes Cy3-dUTP ou Cy5-dUTP (Amersham Pharmacia, Piscataway, NJ) foram incorporados no cDNA em uma reação de transcrição reversa (RT) usando iniciador de hexâmero aleatório (Amersham). Os dois fundos gerais de cDNA marcados foram combinados e aplicados a uma lâmina de microarranjo. As lâminas de microarranjo contêm oligodeoxirribonucleotídeos amino-modificados de 50 mer (oligos) representando cada um por ORF de *P. fluorescens*. Cada oligo foi impresso duas vezes para manchas em duplicata em localização diferente usando o robô SDDC-2 (Virtek, Toronto, Canadá - agora distribuído através de Bio-Rad Laboratories, Hercules, CA) e pinos de SMP3 (TeleChem Internacional Inc., Sunnyvale, CA). As lâminas de microscópio usadas foram revestidas com uma resina epóxi positivamente carregada para ligação de DNA eficiente (MWG Inc, Alameda, CA). Após imprimir, as lâminas foram pós-processadas de acordo com as especificações de MWG. Um pa-

cote de software de BioDiscovery Inc. (El Segundo, CA) foi usado para facilitar a análise de dados. Este pacote consiste em módulos CloneTracker®, ImaGene®, GeneSight® e a base de dados de GeneDirector®. Cada lâmina hibridada foi varrida usando ScanArray 5000 (Packard BioScience, Billerica, MA)

5 para medir a fluorescência do cDNA Cy3- e Cy5-marcado ligado ao microarranjo. As imagens adquiridas foram quantificadas em ImaGene® e os dados brutos foram processados em GeneSight®. Durante a preparação dos dados, a intensidade das manchas para cada gene foi corrigida com a base; o sinal para o canal de Cy5 foi normalizado ao canal de Cy3 usando a intensidade de

10 sinal total para o arranjo inteiro; a razão normalizada de Cy5 para Cy3 para cada gene foi transformada de log2, e as réplicas foram combinadas.

#### ANÁLISE DA EXPRESSÃO DE PROTEÍNA ATRAVÉS DE SDS-PAGE

Alíquotas de cultura foram colhidas em vários pontos de tempo após indução de IPTG, normalizada para OD<sub>600</sub> de 10. Os lisados de célula

15 foram separados em frações solúveis e insolúveis através de centrifugação a 11000 g durante 5 minutos. As alíquotas de 2,5 ul foram combinadas com 5 ul de tampão de amostra de 2X NuPAGE LDS (Invitrogen, San Diego, CA), 50 uM de DTT e H<sub>2</sub>O para 10 µl, depois aquecidas para 95°C durante 5 minutos. As proteínas foram separadas e visualizadas em 12% de géis de Nupage

20 tingidos com Coomassie Blue usando Simply Blue Safestain (Invitrogen, San Diego, CA).

#### MEDIÇÃO DA ATIVIDADE DE FLUORESCÊNCIA

Rendimento da proteína foi também medido por atividade de fluorescência da fusão de proteína de fluorescência verde (COP) e hormônio de crescimento humano (hGH). O constructo de fusão de hgh::COP foi transformado nas cepas do tipo selvagem ou mutantes de *hs/U* e selecionadas na placa de ágar de glicose M9 sem uracila. A cultura de célula induzida por IPTG foi normalizada para OD<sub>600</sub> de cinco. Atividade de fluorescência relativa (RF) foi medida usando o espectrofluorímetro de microplaca de Spectramax Gemini (Molecular Devices, Sunnyvale, CA) sob o ajuste apropriado (Ex485, filtro de passagem de banda Em538530).

EXEMPLO 1: ANÁLISE DE EXPRESSÃO DE GENE DE CEPAS QUE PRODUZEM PROTEÍNAS CITOPLÁSMICAS E PERIPLÁSMICAS - COMPARAÇÃO DE DIFERENTES PONTOS DE TEMPO

Para estudo dos FMs e expressão de gene de protease durante a produção de proteína heteróloga, cepas de *P. fluorescens* DC206, 280, 240 e 271 foram usadas nos experimentos de microarranjo iniciais. DC206 é a cepa hospedeira e foi usada como um controle para crescimento de célula; DC280 tem um plasmídeo apenas de vetor e foi usado como um controle para o experimento de microarranjo; DC240 é DC206 com uma enzima citoplásica de nitrilase de codificação de plasmídeo que é solúvel; DC271 é DC206 com um plasmídeo que codifica o hormônio de crescimento humano periplásico (pbp::hGH) que é parcialmente insolúvel. Cepas foram crescidas em 200 ml de meio de frasco de agitação e crescimento das células foi monitorado medindo a OD<sub>575</sub>. Indução de IPTG foi executada 24 h após inoculação. Todas as cepas cresceram similarmente e as amostras de cultura foram tiradas logo antes (0 h) e 4 h após a indução para isolamento de RNA e perfilção transcripcional (TxP) usando microarranjo de DNA (Figura 1).

Os perfis genéticos, isto é, perfis transcripcionais foram com base na comparação da amostra de ponto de tempo de indução de 4 h com aquele da amostra de 0 h, as duas amostras foram rotuladas com corantes fluorescentes, Cy3-dUTP ou Cy5-dUTP, e co-hibridadas na mesma lâmina para cada cepa. Cada hibridização foi duplicada com experimentos de troca de corante (isto é, amostras foram marcadas com Cy3-dUTP ou Cy5-dUTP) (Tabela 3, lâminas 1 a 6). As lâminas hibridadas foram varridas usando um escâner a laser confocal. Intensidade de sinal para cada gene foi determinada e processada usando o pacote de software de microarranjo de Biodiscovery (El Segundo, CA). A razão de expressão dos dois pontos de tempo para cada gene foi calculada e as razões para todos os genes ao longo das cepas foram agrupadas com base no valor da razão e tendência entre as três cepas (DC280, DC240 e DC271) (Figura 2).

**TABELA 3.** Sumário dos experimentos de microarranjo executados nos Exemplos 1-3

Experi- mento	Lâmina	Cy3	Cy5	Experi- mento	Lâmina	Cy3	Cy5
DC280	1	amostra de 4 h	amostra de 0 h	0 h	7	DC240	DC271
	2	amostra de 0 h	amostra de 4 h		8	DC271	DC240
DC240	3	amostra de 4 h	amostra de 0 h	4 h	9	DC240	DC271
	4	amostra de 0 h	amostra de 4 h		10	DC271	DC240
DC271	5	amostra de 4 h	amostra de 0 h	DC369	11	amostra de 4 h	amostra de 0 h
	6	amostra de 0 h	amostra de 4 h		12	amostra de 0 h	amostra de 4 h

Para focalizar em FM e expressão de gene de protease em *P. fluorescens* sob a tensão imposta pela produção de proteína recombinante de nível alto, uma lista de FM e genes de protease foi comparada à análise de agrupamento. Após análise de agrupamento hierárquica de todos os genes de DC280, DC240 e DC271, FMs e proteases foram identificadas em dois agrupamentos (linhas nos agrupamentos 6 e 7; Figura 2).

Quatro genes no agrupamento 7 mostram expressão mais alta significativa em DC271 que expressa principalmente hormônio de crescimento humano periplásrmico insolúvel quando comparado a DC240 que produz nitrilase citoplásrmica solúvel ou DC280 que não superproduz nenhuma proteína. Os quatro genes são HslV de codificação de rxf01961, HslU de codificação de rxf01957, CbpA de codificação de rxf03987 e HtpG de codificação de rxf05455. O HslV de *E. coli* (ClpQ) e HslU (ClpY) juntos formam uma protease citoplásrmica. A subunidade pequena, HslV, é uma peptidase relacionada às α-subunidades proteassomais de eucariotes. A subunidade grande, HslU, é um ATPase com homologia para outra família de ATPases de Clp como ClpA e ClpX. CbpA de *E. coli* é um análogo da co-chaperona bem caracterizada DnaJ como julgado de não só sua estrutura mas também sua

função. O fenótipo de lesões em DnaJ, como sensibilidade de temperatura para crescimento, é restabelecido sob introdução do gene de *cbpA* em um plasmídeo de multicópias. HtpG de *E. coli* funciona como uma chaperona molecular ATP-independente *in vitro*. Ela reconhece e transientemente liga intermediários de dobramentos estranhos, reduzindo sua concentração livre na solução e desse modo impedindo a agregação não específica.

Os genes foram agrupados no agrupamento 6 da Figura 2 novamente usando agrupamento hierárquico para identificar efeitos menos pronunciados. Figura 3 mostra que FMs e proteases foram identificados em dois agrupamentos principais (linhas no agrupamento 6 e 8). Os dois FMs no agrupamento 8 são DnaK e DnaJ, duas chaperonas principais que são bem conhecidas trabalhar juntas para dobrar numerosas proteínas. Outra análise dos valores de expressão de genes do agrupamento 6 identificou um FM adicional, ClpX que é mais alta expressa em DC271 produzindo pbp::hGH quando comparada a DC240 produzindo nitrilase ou DC280 que não superproduz nenhuma proteína. A proteína de choque térmico de ClpX de *E. coli* é homóloga aos membros da família de ATPases de HSP100/Clp procarióticos e eucarióticos. ClpX de *E. coli* foi isolado como um componente específico das proteases de Clp ATP-dependentes, que mantêm certos polipeptídeos em uma forma competente para proteólise pela subunidade de protease de ClpP. ClpX pode agir como uma chaperona molecular, na ausência de ClpP, ativando as proteínas de iniciação envolvidas na replicação de DNA. FMs e proteases identificados importantes para produção de hGH periplásmico estão listadas na TABELA 4.

**TABELA 4.** Lista de FM e genes de protease cujos níveis de razão de mRNA de estado constante são mais altos em DC271 quando comparado a DC240 e DC280. Os valores listados são a razão de 4 h após indução de IPTG por 0 h.

Gene ID	DC280 (4 h vs. 0 h)	DC240 (4 h vs. 0 h)	DC271 (4 h vs. 0 h)	Gene	Função
RXF05455_1	0,8	0,6	5,3	<i>htpG</i>	Proteína de chaperona HtpG
RXF03987_1	1,0	0,5	5,2	<i>cbpA</i>	Proteína de ligação de DNA curvada

Gene ID	DC280 (4 h vs. 0 h)	DC240 (4 h vs. 0 h)	DC271 (4 h vs. 0 h)	Gene	Função
RXF01961_1	0,9	0,4	5,0	<i>hsIV</i>	HslV Protease ATP-dependente (ec 3.4.25.-)
RXF01957_1	1,0		4,8	<i>hsIU</i>	Hsl Protease ATP-dependente, subunidade de ligação de ATP HsIU
RXF05399	1,0	0,6	3,3	<i>dnaK*</i>	Proteína de chaperona DnaK
RXF05399_1	1,3	0,6	3,0	<i>dnaK*</i>	Proteína de chaperona DnaK
RXF05406_1	1,2	0,7	3,0	<i>dnaJ</i>	Proteína de chaperona DnaJ
RXF04654_1	1,1	0,9	2,0	<i>cpx</i>	Cpx protease ATP-dependente, subunidade de ligação de ATP Cpx

\* Para *dnaK*, duas sondas estão presentes na fatia de microarranjo e desse modo são fornecidos dois valores de expressão de gene.

#### EXEMPLO 2 - ANÁLISE DE EXPRESSÃO DE GENE DE CEPAS QUE PRODUZEM PROTEÍNAS CITOPLÁSMICAS E PERIPLÁSMICAS - COM-

##### 5 PARADA DIRETA DE CEPAS DIFERENTES

Para confirmar os resultados obtidos acima, experimentos de microarranjo adicionais foram executados por comparação direta das duas cepas DC271 e DC240 (lâminas 7 a 10 na Tabela 3). A comparação das duas cepas em 4 h após ponto de tempo de indução confirmou que um conjunto

- 10 quase idêntico de FM e genes de protease foi sobre-regulado em células que expressam pbp::hGH parcialmente solúveis (Tabela 5). Todos os genes listados na Tabela 5 são significativamente (isto é > 2 vezes) mais alto expressos nas cepas que produzem hGH parcialmente insolúvel quando comparado às células que produzem nitrilase completamente solúvel. Na comparação direta de DC271 a DC240, algumas proteínas adicionais foram identificadas quando comparadas à comparação de ponto de tempo (ver Ta-
- 15

bela 4) que mostrou valores de expressão de gene significativamente mais altos durante a produção de hGH parcialmente insolúvel. Aqueles genes incluíram ClpB de codificação de rxf08347, ClpA de codificação de rxf04587 e FkbP de codificação de rxf05753. O homólogo de ClpB de *E. coli* está envolvido na reativação dos corpos de inclusão juntos com DnaKJ-GrpE. ClpA de *E. coli* tem uma função de chaperona ou, quando junto com ClpP, degrada as proteínas. Em *E. coli*, FkbP funciona como uma peptidil-prolil isomerase.

**TABELA 5.** Lista de FM e genes de protease cujos níveis de mRNA de estado constante são mais altos em DC271 quando comparados a DC240. Os valores listados são a razão de DC271 para DC240 em 4 h após indução de IPTG.

Gene ID	DC271 vs. DC240 a 0 h	DC271 vs. DC240 a 4 h	Gene	Função
RXF03987_1	0,8	10,8	<i>cpxA</i>	Proteína de ligação de DNA curvada
RXF01957_1	0,9	10,0	<i>hsfU</i>	Hsf protease ATP-dependente, subunidade de ligação de ATP HsfU
RXF01961_1	0,7	10,0	<i>hsfV</i>	Protease ATP-dependente HsfV (ec 3.4.25.-)
RXF05455_1	0,7	7,8	<i>htpG</i>	Proteína de chaperona HtpG
RXF05406_1	1,0	4,7	<i>dnaJ</i>	Proteína de chaperona DnaJ
RXF08347_1	0,6	3,8	<i>clpB</i>	Proteína de ClpB
RXF05399_1	1,0	3,7	<i>dnaK*</i>	Proteína de chaperona DnaK
RXF05399	0,9	2,9	<i>dnaK*</i>	Proteína de chaperona DnaK
RXF04587_1	0,9	2,8	<i>clpA</i>	Clp protease ATP-dependente, subunidade de ligação de ATP ClpA

Gene ID	DC271 vs. DC240 a 0 h	DC271 vs. DC240 a 4 h	Gene	Função
RXF05753_1	1,1	2,1	<i>fkbP</i>	Peptidil-prolil cis-trans isomerase (ec 5.2.1.8)
RXF04654_1	1,2	2,0	<i>clpX</i>	Clp protease ATP-dependente, subunidade de ligação de ATP ClpX

\*Para *dnaK*, duas sondas estão presentes na fatia de microarranjo e desse modo dois valores de expressão de gene são fornecidos.

#### EXEMPLO 3: ANÁLISE DE EXPRESSÃO DE GENE DE UMA CEPA QUE PRODUZ UMA PROTEÍNA CITOPLÁSMICA INSOLÚVEL

Uma vez que DC271 parcialmente expressa hormônio de crescimento humano periplásrmico (*pbp::hGH*), ele foi investigado se FMs e genes de protease similares ou diferentes fossem sobre-regulados em uma cepa que principalmente expressa hGH citoplásrmico insolúvel. DC369 foi usado neste experimento. As 4 h após indução, a amostra foi comparada à amostra de ponto de tempo 0 h, e experimentos de microarranjo foram executados como mostrado na Tabela 3 (lâminas 11 e 12). Novamente, FM e genes de protease similares foram observados ser sobre-regulados indicando que os genes identificados estão envolvidos em dobramento citoplásrmico ao invés periplásrmico e degradação de proteína (Tabela 6). Um resumo de quais genes que foram identificados em qual experimento junto com a sobre-regulação de dobramento é mostrado no diagrama de Venn da Figura 4.

TABELA 6. Lista de FM e genes de protease cujos níveis de mRNA de estado constante foram mais altos em DC369 em 4 h após indução quando comparados a zero no tempo. Os valores listados são a razão de 4 h após indução de IPTG para 0 h (logo antes da indução).

Gene ID	DC369 (4 h vs. 0 h)	Gene	Função
RXF01961_1	4,8	<i>hsIV</i>	Protease ATP-dependente HsIV (ec 3.4.25.-)
RXF01957_1	4,3	<i>hsIU</i>	HsI protease ATP-dependente, subunidade de ligação de ATP HsIU

Gene ID	DC369 (4 h vs. 0 h)	Gene	Função
RXF03987_1	4,1	<i>cbpA</i>	Proteína de ligação de DNA curvada
RXF05455_1	3,3	<i>htpG</i>	Proteína de chaperona HtpG
RXF05406_1	2,3	<i>dnaJ</i>	Proteína de chaperona DnaJ
RXF08347_1	2,2	<i>clpB</i>	Proteína de ClpB
RXF05399_1	2,1	<i>dnaK*</i>	Proteína de chaperona DnaK
RXF02095_1	2,0	<i>groES</i>	Chaperonina de 10 kDa GroES
RXF06767_1	2,0	<i>groEL</i>	Chaperonina de 10 kDa GroEL
RXF05399	1,8	<i>dnaK*</i>	Proteína de chaperona DnaK
RXF04587_1	1,7	<i>clpA</i>	Clp protease ATP-dependente, subunidade de ligação de ATP ClpA

\* Para *dnaK*, duas sondas estão presentes na fatia de microarranjo e desse modo são fornecidos dois valores de expressão de gene.

**EXEMPLO 4: GERAÇÃO DE UMA CEPA MUTANTE DE *hslU* EM DC206 DE *P. fluorescens***

5 O dois genes *hslU* foram encontrados estar entre os genes identificados mais altamente sobre-regulados. HslU é uma ATPase citoplasmica. A proteína homóloga em *E. coli* pode agir em combinação com uma segunda proteína para promover degradação de proteína energia-dependente em *E. coli*. HslU interage com HslV, uma proteína com homologia para as α-subunidades de proteasoma. Os homólogos de HslVU de *E. coli* foram relatados estar envolvidos em proteólise geral das proteínas desdobradas em Missiakas, D., et al. (1996) Identification and characterization of HslV HslU (ClpQ ClpY) proteins involved in overall proteolysis of misfolded proteins in *Escherichia coli*. *Embo J*15:6899-909. análise de seqüência de DNA sugeriu 10 que os genes de *hslVU* de *P. fluorescens* são prováveis de fazer parte de um operon bicistrônico (Figura 5).

15

20 Para verificar se HslVU estão de fato envolvidos na degradação de hGH, um cepa knock-out de *hslU* foi construída. Uma tal cepa foi gerada por inativação insercional de *hslU* (Figura 6). Um fragmento de DNA interno de cerca de 550 bp para *hslU* foi clonado no vetor de pCR2.1-TOPO resistente

à canamicina. Uma vez que este vetor tem uma origem de replicação (ColE1) que é funcional em *E. coli* mas não em *P. fluorescens*, os plasmídeos construídos integrarão no cromossomo de DC206 através de recombinação homóloga para conferir resistência à canamicina. O sítio de inserção correto para as colônias resistentes à canamicina foi confirmado através de PCR de diagnóstico de colônia usando iniciadores que hibridam fora da região originalmente amplificada e dentro da cadeia principal de plasmídeo (Tabela 3). A cepa mutante de *hsfU* construída foi designada DC370.

Os iniciadores foram projetados que amplificariam uma região interna de ~550 bp do gene de *hsfU* (Tabela 7). O fragmento interno foi amplificado usando Taq Polimerase (Promega), purificado e clonado em vetor de pCR2.1-TOPO (Invitrogen, San Diego, CA). Os plasmídeos foram transformados em DC206 de *P. fluorescens* competente e selecionados nas placas de ágar de glicose de M9 suplementada com 250 µg/ml de uracila e 50 µg/ml de canamicina.

**TABELA 7. INICIADORES**

Iniciadores	Seqüência	Propósito
hsfU_sens	accgaagtccgtatgtggg (SEQ ID NO: 1)	usado na amplificação de PCR do fragmento de <i>hsfU</i> interno
HsfU_antis	aatcgcgctgcacgcctcg (SEQ ID NO: 2)	
hsf_F2	ttcatcaaggcgtcgaaagcg (SEQ ID NO: 3)	
hsf_R2	tcaatgttgcaccatgcc (SEQ ID NO: 4)	
M13_R	caggaaacagactatgac (SEQ ID NO: 5)	usado em PCR de diagnóstico
M13_F	taaaaacgacgcccag (SEQ ID NO: 6)	
hsf_UP	gtggcagccaccaaggctgc (SEQ ID NO: 7)	
hsf_middleUp	ccccacattggatggaggcttacaaggggagagtctccacg (SEQ ID NO: 8)	usado em PCR de SOE o fragmento de DNA acima e abaixo de <i>hsfUV</i>
hsf_middleDown	cgtggagactctcccttgtaagcctcactcaatgtggg (SEQ ID NO: 9)	
hsf_down	ggccaatggttggccacgcg (SEQ ID NO: 10)	
hsf_UpUp	tgccgacgcccacagggtgc (SEQ ID NO: 11)	

Iniciadores	Seqüência	Propósito
hsI_DownDown	gcctggtaactgcgactcg (SEQ ID NO: 12)	usado em PCR de diagnóstico
RC199	atatactagtaggaggtaacttatggctgacgaaca-gacgca (SEQ ID NO: 13)	usado na clonagem de <i>gr-pE-DnaKJ</i>
RC200	atattcttagattacagggtcgccgaagaagc (SEQ ID NO: 14)	

### COMPARAÇÃO DE EXPRESSÃO DA PROTEÍNA POR ANÁLISE DE SDS-PAGE

Para estudar o efeito do knockout do gene de *hsI/U*, duas expressões de proteína exógena foram comparadas entre a cepa de origem DC206 e a cepa mutante recentemente construída DC370. Os plasmídeos abrigando o gene codificando pbp::hGH (pDOW1323) e hGH (pDOW1426) foram cada um transformados em células de DC370 competentes e resultaram em cepas DC373 e DC372, respectivamente. Os experimentos de crescimento padrão em frasco de agitação foram executados com as quatro cepas.

Figura 7 mostra que as cepas do tipo selvagem e mutantes têm taxas de crescimento similares. As amostras foram operadas em géis de SDS-PAGE (Figura 8 e 9). Os resultados sugerem que o mutante produziu quantidades mais altas de proteínas devido à deleção da subunidade de protease HsI.U.

### COMPARAÇÃO DE EXPRESSÃO DA PROTEÍNA POR ATIVIDADE DE FLUORESCÊNCIA

Uma vez que o efeito observado da falta de HsI.U no rendimento de hGH é difícil para quantificar usando análises de SDS-PAGE, o perfil temporal de produção de proteína foi monitorado pela fluorescência de uma proteína de fusão entre proteína fluorescente verde COP e hGH. Um plasmídeo contendo uma fusão de hGH::COP foi construído e transformado na cepa de origem DC206 e na cepa de deleção de gene de *hsI/U* DC370 resultando em cepas HJ104 e HJ105 (Tabela 1). Os experimentos de frasco de agitação padrão foram executados e as amostras tiradas em vários pontos de tempo para medições da fluorescência (Figura 10). As leituras do fluorímetro claramente mostraram que a cepa mutante de protease de *hsI/U* teve níveis de

expressão de proteína significativamente mais altos comparados aos da cepa parental (Figura 11). Esta descoberta confirma os resultados obtidos por análise de SDS-PAGE. Comparando à cepa de tipo selvagem, o mutante de *hslU* aumentou 33,05% da fluorescência relativa em 24 h após indução (ver inserção na Figura 11).

#### EXEMPLO 5: CONSTRUÇÃO DE UMA CEPA KNOCKOUT LIMPA DE *hslUV*

A protease de Hsl consiste em duas subunidades: uma subunidade de ligação de ATP codificada por *hslU* e uma subunidade de protease codificada por *hslV*. A cepa knock-out de protease de Hsl previamente construída é uma inativação insercional do gene de *hslU*. Para remover a preocupação que *HslV* poderia ainda funcionar como uma protease podendo acoplar a uma subunidade de ligação de ATP de outra protease, uma cepa de deleção foi construída que teve os genes de *hslU* e *hslV* removidos do cromossomo.

Como mostrado na Figura 13, plasmídeo pDOW2050 foi construído por amplificação de PCR de dois fragmentos de DNA flanqueando a região de *hslUV*, os dois fragmentos foram subsequentemente fundidos usando o método de PCR de Encaixe através de Extensão de Sobreposição (SOE) (ver Ho, S. N. (1991) Method for gene splicing by overlap extension using the polymerase chain reaction. Pedido: Patente US 89-3920955023171). Os fragmentos de DNA fundidos foram depois ligados no sítio de *SrfI* de vetor pDOW1261-2. O plasmídeo de deleção foi nomeado pDOW2050 após a inserção ter sido confirmada através de seqüenciamento de DNA.

Plasmídeo pDOW2050 foi eletroporado em DC206 e banhado sobre placas de ágar de M9 suplementadas com 1% de glicose e 15 pg/ml de tetraciclina. Resistência à tetraciclina é devido a um evento de integração que recombina o plasmídeo inteiro no cromossomo com uma das duas regiões homólogas dentro do genoma (Figura 13). Para selecionar as células que têm uma deleção dos genes de *hslUV*, que resulta de uma segunda recombinação homóloga entre o plasmídeo integrado e a região de DNA homóloga no cromossomo, as colônias resistentes à tetraciclina foram crescidas para fase estacionária em meio de LB suplementado com 250 µg/ml de uracila. As cé-

Iulas são depois banhadas sobre placas de ágar de LB suplementadas com 500 pg/ml de ácido 5-fluoroorótico (5-FOA). As células que perderam o plasmídeo integrado por um segundo evento de recombinação também perderam o gene de *pyrF* e desse modo são resistentes a 5-FOA, resultando na 5 cepa de deleção de *hsf/UV* cromossômico desejada, chamada DC417.

#### ANÁLISE FENOTÍPICA DA CEPA DE DELEÇÃO DE *hsf/UV*

Análise de SDS-PAGE da cepa de deleção de *hsf/UV* expressando a proteína de hGH (cepa HJ115) mostrou rendimento de proteína muito mais alto que a cepa do tipo selvagem DC369, similar ao que foi observado mais 10 cedo usando a cepa mutante insercional de *hsf/U* DC372 (dados não mostrados).

Rendimento da proteína foi também medido por atividade de fluorescência da fusão de hGH::COP usando o mesmo método descrito mais cedo. Plasmídeo pDOW1349 contendo a fusão de hGH::COP foi transformado nas cepas do tipo selvagem e mutante que resultam em cepas HJ104 e HJ117, respectivamente. Experimentos de frasco de agitação padrão foram executados e as amostras foram tiradas em vários pontos de tempo para medições de fluorescência relativa (Figura 14). As leituras do fluorímetro indicaram que a cepa de deleção de protease de *hsf/UV* teve níveis de expressão de proteínas significativamente mais altos (cerca de 50% de aumento de rendimento) quando comparada aos da cepa do tipo selvagem. Este resultado é similar ao que foi previamente observado com a cepa knock-out de *hsf/U*. 20

#### EXEMPLO 6: IDENTIFICAÇÃO DE ALVO ITERATIVO USANDO TECNOLOGIA DE MICROARRANJO DE DNA

Para investigar se um conjunto novo de proteases é sobre-regulado na cepa de deleção de protease de *hsf/UV*, experimentos de microarranjo de DNA foram conduzidos. Experimentos de frasco de agitação padrão foram executados usando a cepa do tipo selvagem (DC369) e mutante (HJ115) para expressar hGH. Para cada cepa, as 4 h após as amostras de indução term sido comparadas à amostra de ponto de tempo 0 h (logo antes da indução de proteína heteróloga) os experimentos de microarranjo de DNA 30

foram executados. Comparando a razão de dois pontos de tempo entre as cepas do tipo selvagem e mutante, uma lista nova de genes de protease que são sobre-regulados na cepa de deleção de protesase de *hs/UV* foi identificada (Tabela 8). Estes genes recentemente identificados codificando proteases podem agora ser os alvos para uma segunda rodada de eventos de deleção de gene para também melhorar o rendimento de produção de proteína heteróloga.

**TABELA 8:** Genes de protease cujos níveis de mRNA de estado constante são mais altos na cepa de deleção de protease de *hsfUV* (HJ115) quando comparados aos da cepa do tipo selvagem (DC369), com base na razão de 4 h após indução de IPTG para 0 h (logo antes da indução).

ORFID	Função Curada	Seqüênci a
rxf00133	D-alanil-me- so-diaminop- imelato en- dopeptidase (ec 3.4.-.-)	atgcacttggaaaatgtttcacaccagcaccctgtggcggcttgatgttgtgtctggggcggt ctgcgcgcacgcgtctccaaacctccaccccgcaaccctggataagctgttgtgacgcaccccg cgctgcacaggcgcaccgtctcgatgttgtgtatgtccgcacaggcaccacgtgtat cagcacaacccacgcacccggctgtgtgcgcgcgtccaacctcaagctgtgaccacggcgg cagccatggatgtatggggccgcagttaccgcgtccacgcacactgtgtggcaatggctac gccaggcgcacccggctgactggcaacctgtaccgtgtggcttggcgcacccggatgttac tttgcgcactatcaggcgcgtccgcgcacattggccaggcaggcgtgcgcagggtgcagg gtgaccctgtgttcgcacgcacacttgttgcgtatgcgcaggcggctggcgtggactgtcccatg atgtgaaccacccatctacgcgcgcacgttgcgcgtgcacgttgcgcacccataccgt tttgcgtggcgcgcgtgttgcaccgcacccggcgcgtgcacgtgcgtgcgcggcgcgc gtggagatctacccgcacccgactacctgcacactgaataaccgcgcgcgcgcgc gtaacagctatggatcaaccgcgtgcgcaccaacctgtgcacgcgcgcgcgc gcccgc acccgttgttgcacgcacccggcgcacccggcgcacccggcgcacccggcgcacccggc ggggcaagtctgtgggtgacgggtgtggccgagcacaatggccgcgtgcaggagc tgatcgtgcgcgtgtcaagctgcacacaacgcgcacgtccgcgcgcgcgcgcgc gcgcgcacacggccacgcgcgcacggcggcggcggcggcgcgcgcgcgc ctcaagcgcgcacccggcgcgcgcgcgcgcgcgcgcgcgcgcgcgc ggcgtiaaccgtgtgtgcacgcacccgcgtgtgcacccggcgcacccggcgc tgggtcgcacgcgcgtgtgcacacgcgcgtccgcgcgcgcgcgc cagccctgggttaccgcgtgcgcgcacggcgtgcggaaaataaccctgcacgc ccatggccggcgtgtgcacatgtggcgtgtgcacccggcgcacccggc cgcgttgtgaccaacaactatgtggcgtgtgcgcgcacccggcgc tgcgcgttgttgcacgcgcacccggcgcacccggcgcacccggc tgcgcgttgttgcacgcgcacccggcgcacccggcgcacccggc





EXEMPLO 7: CO-SOBRE-EXPRESSÃO DE MODULADORES DE DOBRAMENTO AUMENTA A SOLUBILIDADE DA PROTEÍNA ALVO hGH

Com base nos dados de perfilção transcripcional mostrados na Figura 4, expressão de moduladores de dobramentos (FMs) DnaK e DnaJ foi aumentada em cepas que produzem proteína recombinante comparadas às cepas de controle (ver Tabelas 4 e 5). Uma cepa que co-superproduziu GrpE, DnaK e DnaJ junto com hGH foi produzida e testada para identificar se isto resultou na acumulação de níveis aumentados de hGH solúvel.

CONSTRUCTO DE PLASMÍDEO CONTENDO grpE-dnaKJ PARA CO-SOBRE-EXPRESSÃO COM hGH

Os genes de *grpE-dnaKJ* de *P. fluorescens* foram amplificados usando o DNA cromossômico isolado de MB214 (DNeasy; Qiagen, Valença, CA) como um modelo, RC199 (5'-ATATACTAGTAGGAGGTAACTTATGGC TGACGAACAGACGCA-3') e RC200 (5'-ATATTCTAGATTACAGGTCGCCGA AGAACG-3') como iniciadores, *Pfu Turbo* (Stratagene, La Jolla, CA) foi usado seguindo as recomendações do fabricante. O produto de PCR resultante (4 kb) foi digerido com *Spel* e *Xba*I (sítios de restrição sublinhados acima nos iniciadores) e ligado a pDOW2236 para criar pDOW2240 contendo os genes de *grpE-dnaKJ* sob o controle do promotor de tac. Plasmídeo pDOW2240 foi digerido com *Spel* e *Hind*III e o fragmento de 4,0 kb contendo *grpE-dnaKJ* resultante foi purificado em gel usando Qiaquick (Qiagen) e ligado a pDOW2247 também digerido com *Spel* e *Hind*III. O plasmídeo resultante, pDOW3501, contendo *grpE-dnaKJ* sob o controle do promotor de manitol, foi transformado em DC388 mediante seleção em placas de glicose de M9 suplementadas com 250 µg/ml de uracila. Por fim, pDOW1426 foi eletroporado na cepa acima (DC462) e selecionado em placas de glicose de M9, resultando em cepa DC463 com dois plasmídeos induzíveis: 1) pDOW1426 carregando  $P_{tac}$  hGH e 2) pDOW3501 carregando  $P_{mts}$  *grpE-dnaKJ*.

FERMENTAÇÃO EM FRASCO DE AGITAÇÃO, COLETA DE AMOSTRA E ANÁLISE

Culturas em duplicata de DC463 foram crescidas em frascos de agitação. Indução de proteína foi realizada pela adição de 0,1 mM de IPTG

para hGH e 0,5% de manitol para GrpE-DnaKJ em 24 hs após inoculação. As amostras foram colhidas a 0, 4, 8, 24 e 48 horas após indução. Em cada ponto de tempo, 20 células de OD<sub>600</sub> normalizada em 1 ml foram colhidas, lisadas usando Easylyse® (Epicentro, Madison, WI) e separadas em frações solúveis e insolúveis mediante centrifugação a 14000 rpm durante 30 minutos. Volumes iguais das amostras foram combinados com BioRad (Hercules, CA) 2x tampão de Laemmli, aquecidos para 95°C durante 5 minutos com 30 µL carregados em um gel de Criterion a 15% de Tris HCl de BioRad usando tampão de operação de SDS de 1x Tris Glicina corrente (BioRad). As proteínas foram visualizadas com Simply Blue Safestain (Invitrogen, Carlsbad, CA) como mostrado na Figura 15. Os géis tingidos com Coomassie resultantes foram varridos usando um Densitômetro Pessoal de Molecular Devices (Molecular Devices, Sunnyvale, CA) com análises executadas usando Image-Quant e Excel. Como mostrado na Figura 15, a co-sobre-expressão de GrpE, 15 DnaKJ significativamente aumentou a solubilidade de hGH, convertendo quase 100% da proteína alvo na fração solúvel, embora a um rendimento de proteína total inferior. Experimentos adicionais repetindo crescimento e indução de DC463 usando a adição simultânea de IPTG e manitol de forma bem próxima imitaram os resultados mostrados aqui, embora com um grau 20 variado de solubilidade de hGH (entre 50-100%; dados não mostrados), quando GrpE DnaKJ foram co-superproduzidos. Estas descobertas também demonstram que engenharia de cepa alvejada com base em perfilamento transcripcional pode levar a um projeto de cepa racional para aumentar a solubilidade e/ou rendimento de uma proteína recombinante.

25 A invenção foi descrita com referência a certas modalidades e exemplos não-limitativos. Estará claro a alguém de habilidade na técnica que outras modalidades da invenção são também possíveis.

## REIVINDICAÇÕES

1. Processo para melhorar a expressão de uma proteína ou peptídeo recombinante em uma célula ou organismo hospedeiro, compreendendo:

5           i) expressar a proteína ou peptídeo recombinante na célula ou organismo recombinante hospedeiro;

10          ii) analisar um perfil genético da célula recombinante para identificar um ou mais genes ou produtos de gene compensatórios que são expressos a um nível mais alto na célula recombinante que em ou uma célula hospedeira que não foi modificada para expressar a proteína recombinante ou

15          uma célula recombinante que não está expressando a proteína recombinante; e

15          iii) alterar a expressão do gene ou produto de gene compensatório identificado na célula recombinante através de modificação genética para

15          fornecer uma célula recombinante modificada que alcança um aumento em expressão, atividade ou solubilidade da proteína recombinante.

20          2. Processo de acordo com a reivindicação 1, ainda compreendendo expressar a proteína ou peptídeo na célula recombinante modificada.

20          3. Processo de acordo com a reivindicação 1, ainda compreendendo:

20           (a) expressar a proteína ou peptídeo recombinante na célula recombinante modificada;

25           (b) analisar um perfil genético da célula recombinante modificada para identificar pelo menos um/uns segundo(s) gene(s) ou produto(s) de gene que é/são diferencialmente expresso(s) na célula recombinante modificada;

25           (c) alterar a expressão do segundo produto de gene identificado na célula recombinante modificada para fornecer uma célula duplamente modificada; e

30           (d) expressar a proteína ou peptídeo na célula recombinante duplamente modificada.

30          4. Processo de acordo com a reivindicação 3, ainda compreendendo repetir as etapas a) a d).

5. Processo de acordo com a reivindicação 4, compreendendo repetir as etapas a) a d) até a viabilidade da célula ser afetada alterando a expressão do(s) gene(s) ou produto(s) de gene modificado(s).
6. Processo de acordo com a reivindicação 4, compreendendo  
5 repetir as etapas a) a d) até a expressão da proteína ou peptídeo recombinante alcançar um ponto terminal alvejado.
7. Processo de acordo com a reivindicação 1, em que um perfil genético é analisado comparando um perfil genético da célula recombinante a um segundo perfil genético da célula hospedeira.  
10
8. Processo de acordo com a reivindicação 1, em que o perfil genético é um perfil de transcriptoma.
9. Processo de acordo com a reivindicação 8, em que o perfil de transcriptoma é determinado por um microarranjo.
10. Processo de acordo com a reivindicação 1, em que o perfil  
15 genético é um perfil de proteoma.
11. Processo de acordo com a reivindicação 10, em que o perfil de proteoma é determinado por eletroforese em gel bidimensional, ICAT ou LC/MS.
12. Processo de acordo com a reivindicação 10, em que o perfil de  
20 proteoma é determinado por um arranjo de peptídeo.
13. Processo de acordo com a reivindicação 12, em que o arranjo de peptídeo é um arranjo de anticorpo.
14. Processo de acordo com a reivindicação 1, em que o produto de gene identificado é uma protease, uma subunidade de uma protease, um  
25 co-fator de uma protease, um modulador celular ou um genético que afeta a expressão de uma protease.
15. Processo de acordo com a reivindicação 14, em que o produto de gene identificado é uma protease.
16. Processo de acordo com a reivindicação 14, em que o produto  
30 de gene identificado é uma subunidade de uma protease.
17. Processo de acordo com a reivindicação 14, em que o produto de gene identificado é um co-fator de uma protease.

18. Processo de acordo com a reivindicação 14, em que o produto de gene identificado é um modulador celular ou genético que afeta a expressão de uma protease.

19. Processo de acordo com a reivindicação 14, em que o produto de gene identificado é selecionado do grupo que consiste em D-alanil-meso-diaminopimelato endopeptidase, zinco protease, dipeptidase microssomal, precursor de metaloprotease extracelular, proteína de divisão de célula ftsH e produtos de gene derivados dos genes *hsIV*, *hsIU*, *clpX*, *clpA* e *clpB*.

20. Processo de acordo com a reivindicação 14, em que o nível de rRNA do produto de gene identificado é sobre-regulado quando a proteína ou peptídeo recombinante for expressado na célula hospedeira.

21. Processo de acordo com a reivindicação 14, em que o produto de gene identificado é removido de um genoma de célula hospedeira.

22. Processo de acordo com a reivindicação 21, em que o produto de gene identificado é removido através de recombinação homóloga.

23. Processo de acordo com a reivindicação 1, em que o produto de gene identificado é um modulador de dobra, uma subunidade de um modulador de dobra, um co-fator de um modulador de dobra ou um modulador celular ou genético que afeta a expressão de um modulador de dobra.

24. Processo de acordo com a reivindicação 23, em que o produto de gene identificado é um modulador de dobra.

25. Processo de acordo com a reivindicação 23, em que o produto de gene identificado é uma subunidade de um modulador de dobra.

26. Processo de acordo com a reivindicação 23, em que o produto de gene identificado é um co-fator de um modulador de dobra.

27. Processo de acordo com a reivindicação 23, em que o produto de gene identificado é um modulador celular ou genético que afeta a expressão de um modulador de dobra.

28. Processo de acordo com a reivindicação 23, em que o modulador de dobra é uma proteína de chaperona.

29. Processo de acordo com a reivindicação 23, em que o modulador de dobra é selecionado do grupo que consiste em produtos de gene

dos genes *cbpA*, *htpG*, *dnak*, *dnaJ*, *fkbP2*, *groES* e *groEL*.

30. Processo de acordo com a reivindicação 23, em que a expressão do produto de gene identificado é alterada aumentando a expressão do gene identificado, um co-fator de um gene identificado ou um modulador 5 celular ou genético do gene identificado.

31. Processo de acordo com a reivindicação 30, em que a expressão aumentada é por inclusão de um DNA que codifica o produto de gene identificado.

10 32. Processo de acordo com a reivindicação 30, em que a expressão aumentada é por inserção de um promotor em um genoma da célula hospedeira.

33. Processo de acordo com a reivindicação 30, em que a expressão aumentada é por inclusão de um vetor exógeno na célula hospedeira.

15 34. Processo de acordo com a reivindicação 1, em que a célula hospedeira é uma célula microbiana.

35. Processo de acordo com a reivindicação 1, em que a célula hospedeira é uma Pseudomonada.

36. Processo de acordo com a reivindicação 1, em que a célula hospedeira é uma célula de *P. fluorescens*.

20 37. Processo de acordo com a reivindicação 1, em que a célula hospedeira é uma célula de *E. coli*.

38. Processo de acordo com a reivindicação 1, em que a célula hospedeira é selecionada do grupo que consiste em uma célula de inseto, uma célula mamífera, uma célula de levedura, uma célula fúngica e uma cé- 25 lula de planta.

39. Processo de acordo com a reivindicação 9, em que o micro-arranjo compreende amostras de pares de ligação para pelo menos 50% de um genoma da célula hospedeira.

30 40. Processo de acordo com a reivindicação 9, em que a técnica de microarranjo compreende amostras de pares de ligação para pelo menos 80% de um genoma da célula hospedeira.

41. Processo de acordo com a reivindicação 9, em que o micro-

arranjo compreende amostras de pares de ligação para pelo menos 90% de um genoma da célula hospedeira.

42. Processo de acordo com a reivindicação 9, em que o micro-arranjo compreende amostras de pares de ligação para pelo menos 95% de  
5 um genoma da célula hospedeira.

43. Processo de acordo com a reivindicação 1, em que a expressão melhorada é um aumento na quantidade de proteína ou peptídeo recombinante.

10 44. Processo de acordo com a reivindicação 1, em que a expressão melhorada é uma solubilidade aumentada da proteína ou peptídeo recombinante.

45. Processo de acordo com a reivindicação 1, em que a expressão melhorada é uma atividade aumentada da proteína ou peptídeo recombinante.

15 46. Processo de acordo com a reivindicação 1, em que o perfil genético é um perfil de genes em uma família de gene.

47. Processo de acordo com a reivindicação 1, em que o perfil compreende proteases e moduladores de dobra.

20 48. Processo de acordo com a reivindicação 46, em que o perfil consiste essencialmente em proteases.

49. Célula ou organismo hospedeiro, em que expressa uma proteína recombinante que foi geneticamente modificada para reduzir a expressão de pelo menos duas proteases.

25 50. Célula ou organismo hospedeiro, em que expressa uma proteína recombinante que foi geneticamente modificada para reduzir a expressão de pelo menos uma protease selecionada do grupo que consiste em D-alanil-meso-diaminopimelato endopeptidase, zinco protease, dipeptidase microssomal, precursor de metaloprotease extracelular, proteína de divisão de célula *ftsH* e produtos de gene derivados dos genes *hsIV*, *hsIU*, *clpX*, *clpA*  
30 e *clpB*.

51. Célula ou organismo hospedeiro, em que expressa uma proteína derivada mamífera recombinante que foi geneticamente modificada para

reduzir a expressão de pelo menos uma protease.

52. Célula ou organismo hospedeiro de acordo com a reivindicação 52, em que a proteína recombinante é hormônio de crescimento humano.

53. Célula ou organismo hospedeiro, em que expressa uma proteína recombinante que foi geneticamente modificada para aumentar a expressão de pelo menos dois moduladores de dobra que não sejam subunidades de modulador de dobra.

FIG. 1

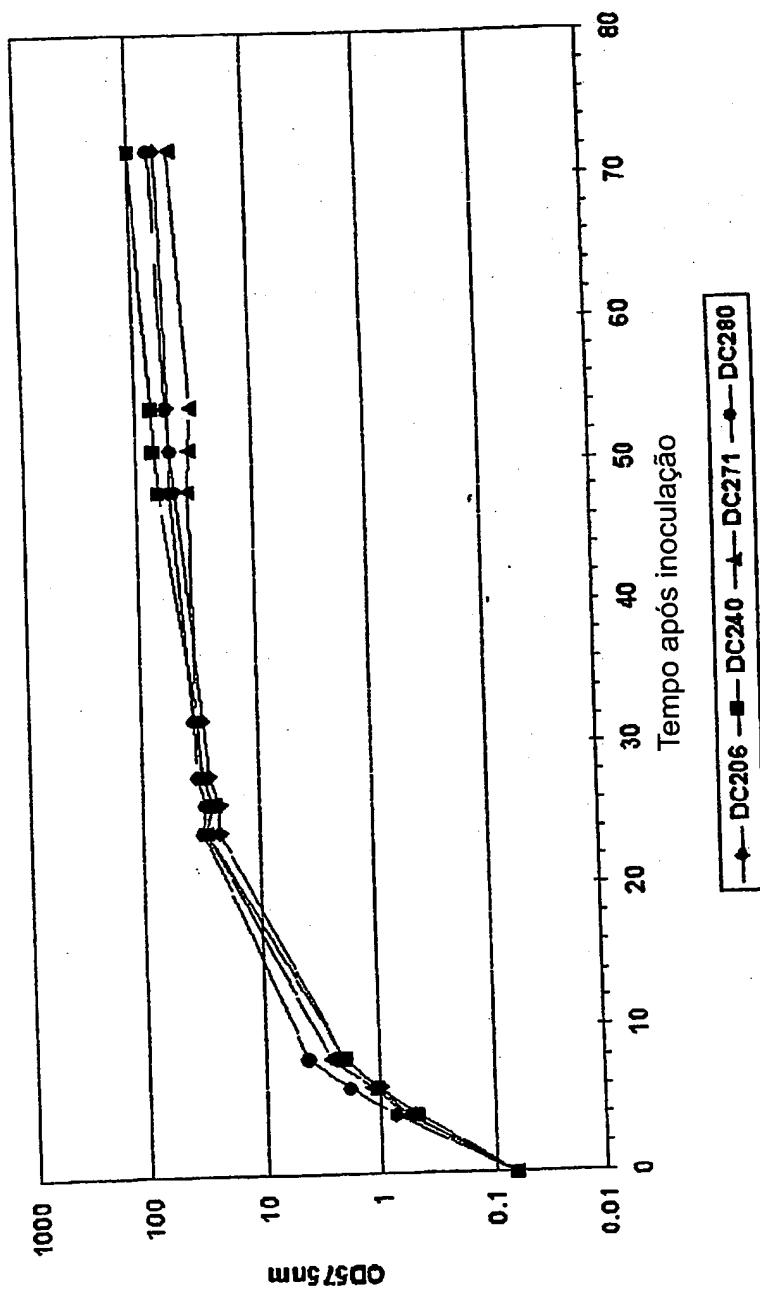


FIG. 2

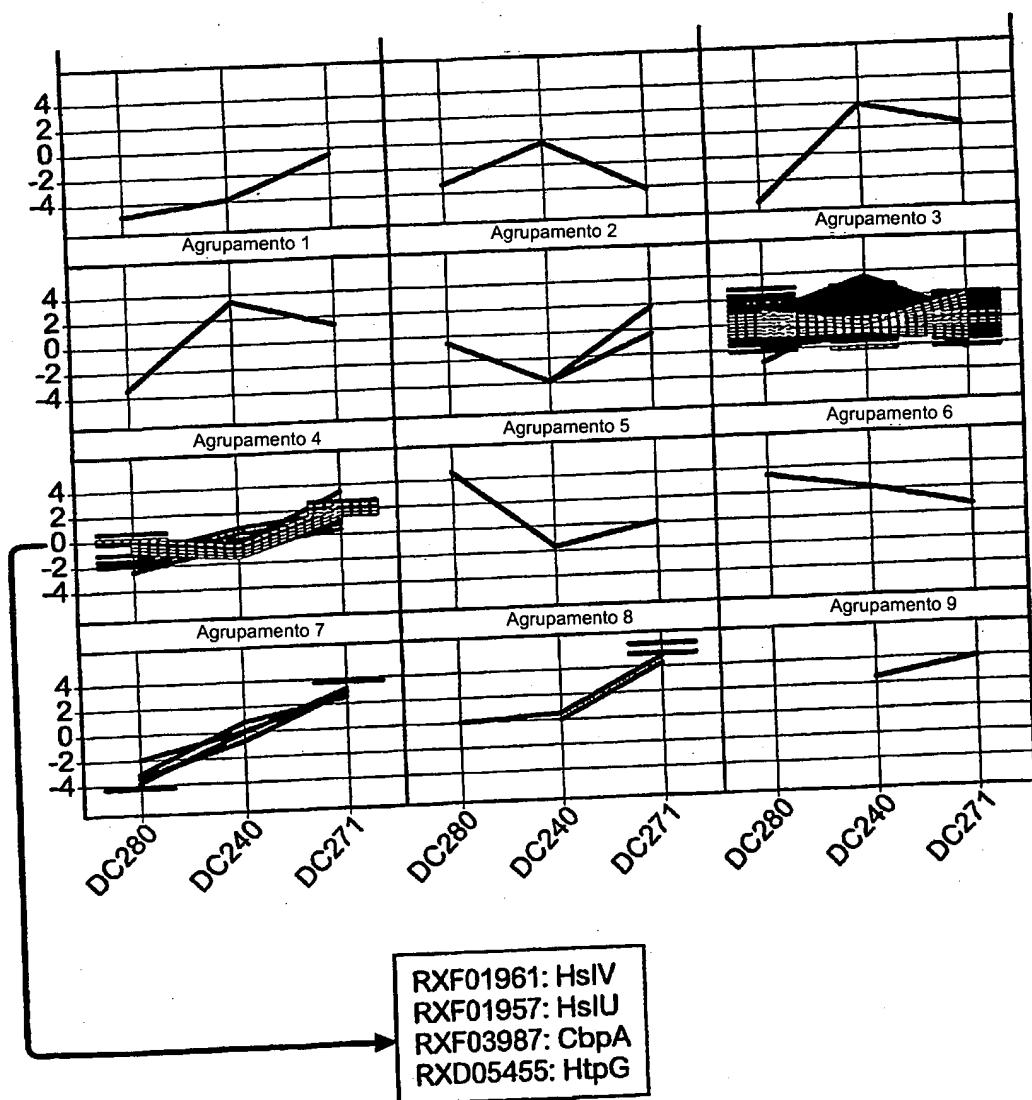


FIG. 3

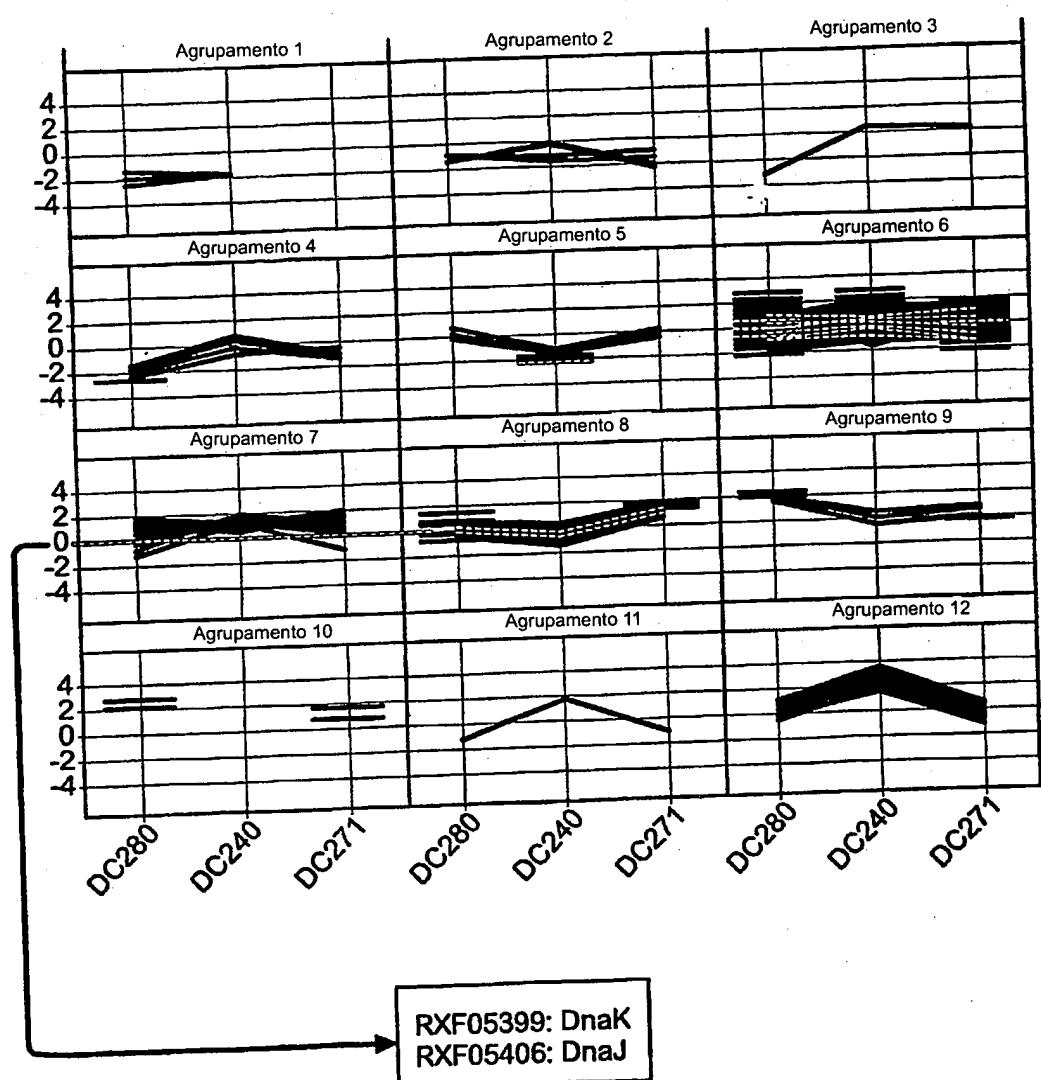
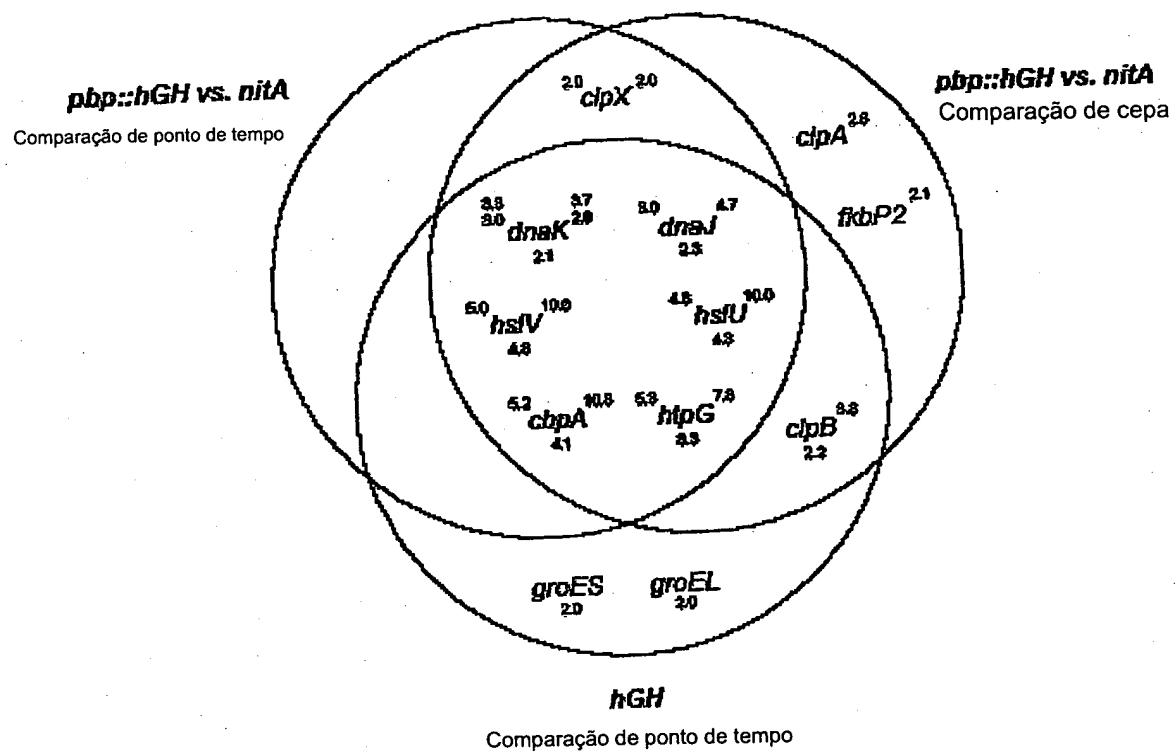


FIG. 4



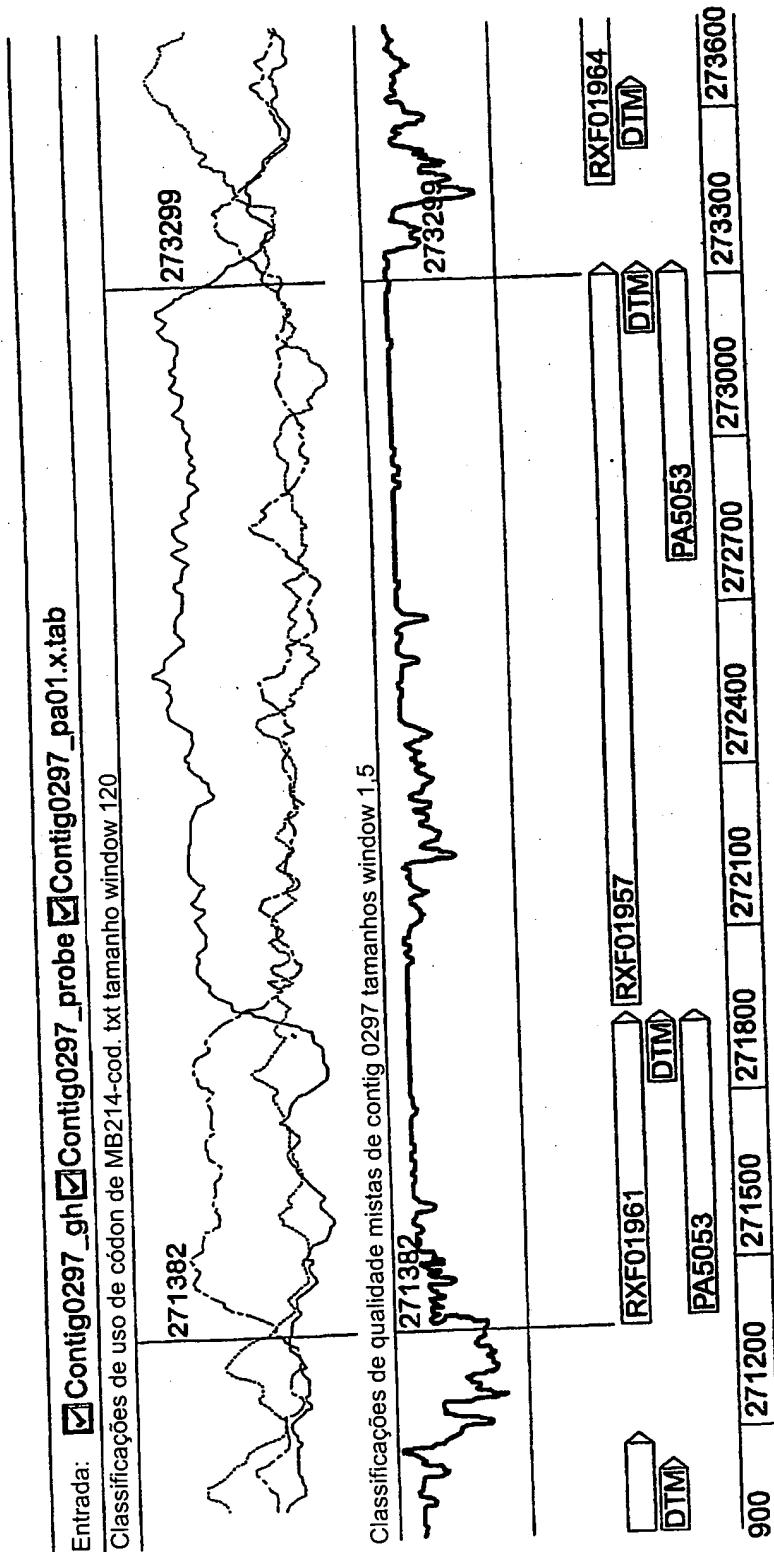


FIG. 5

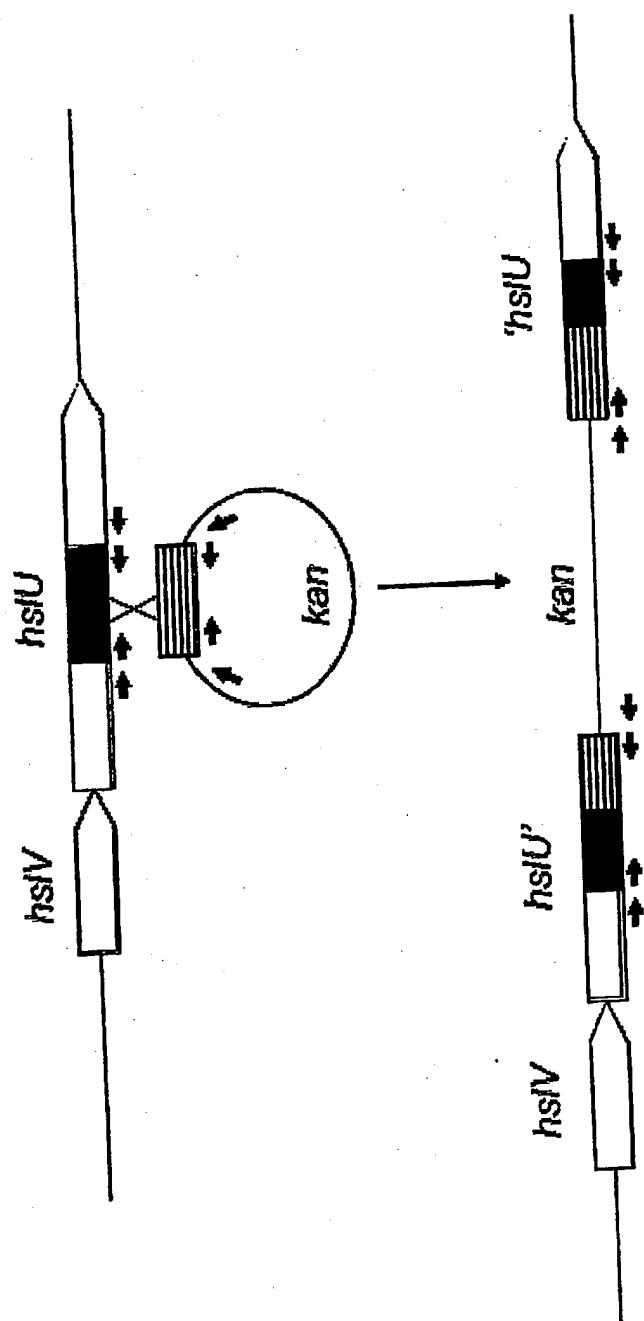
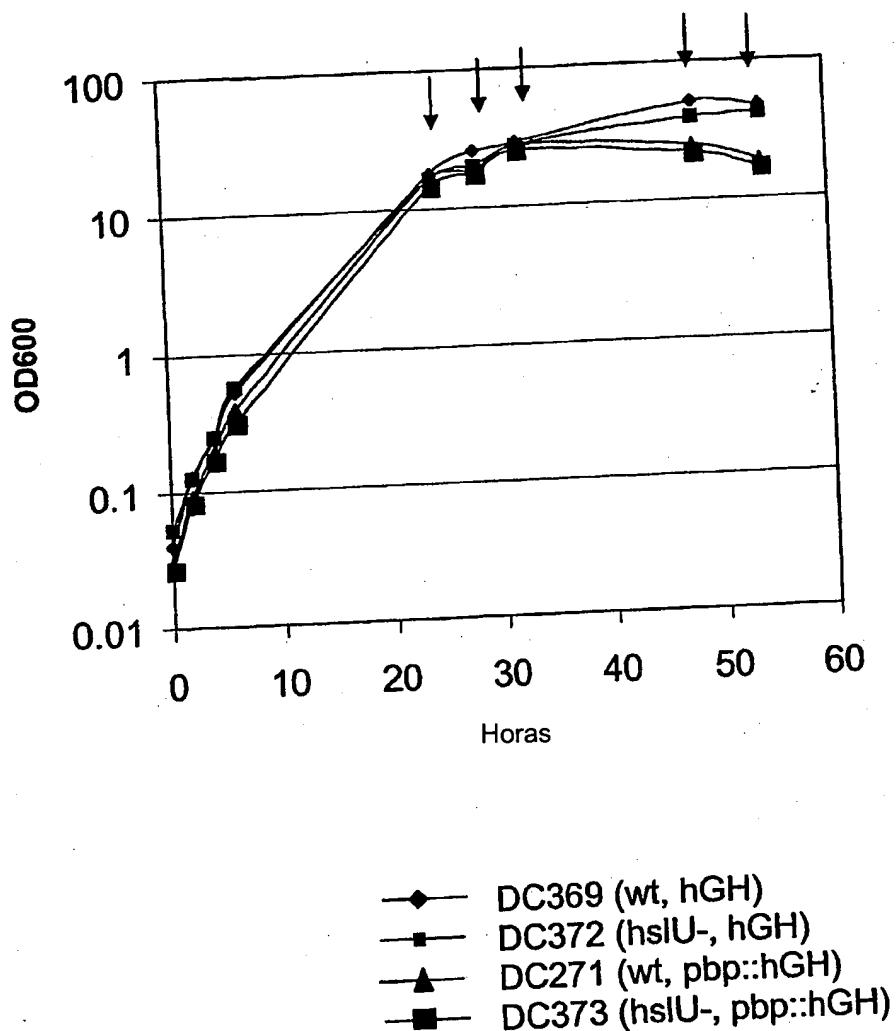


FIG. 6

7/15

FIG. 7



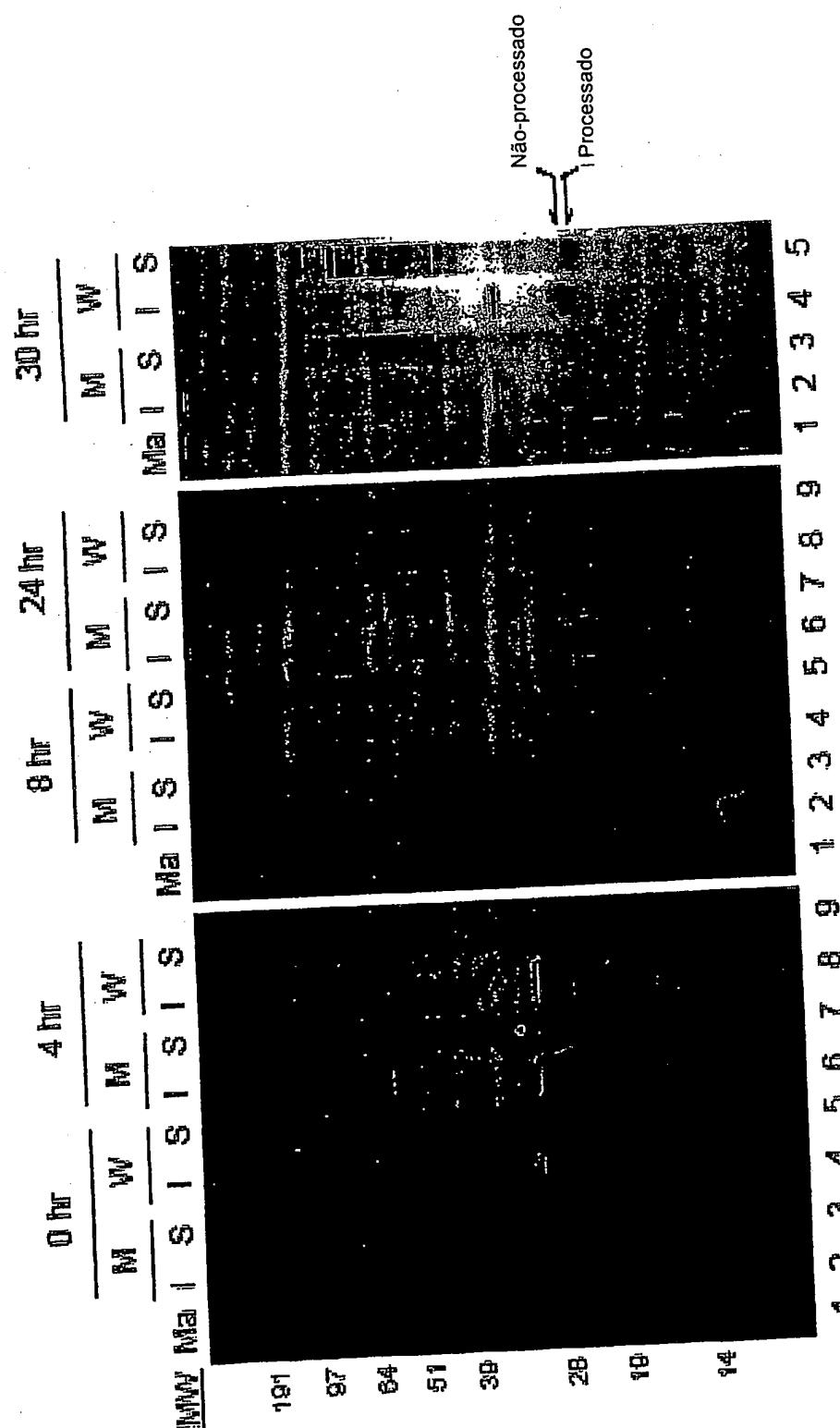


FIG. 8

9/15

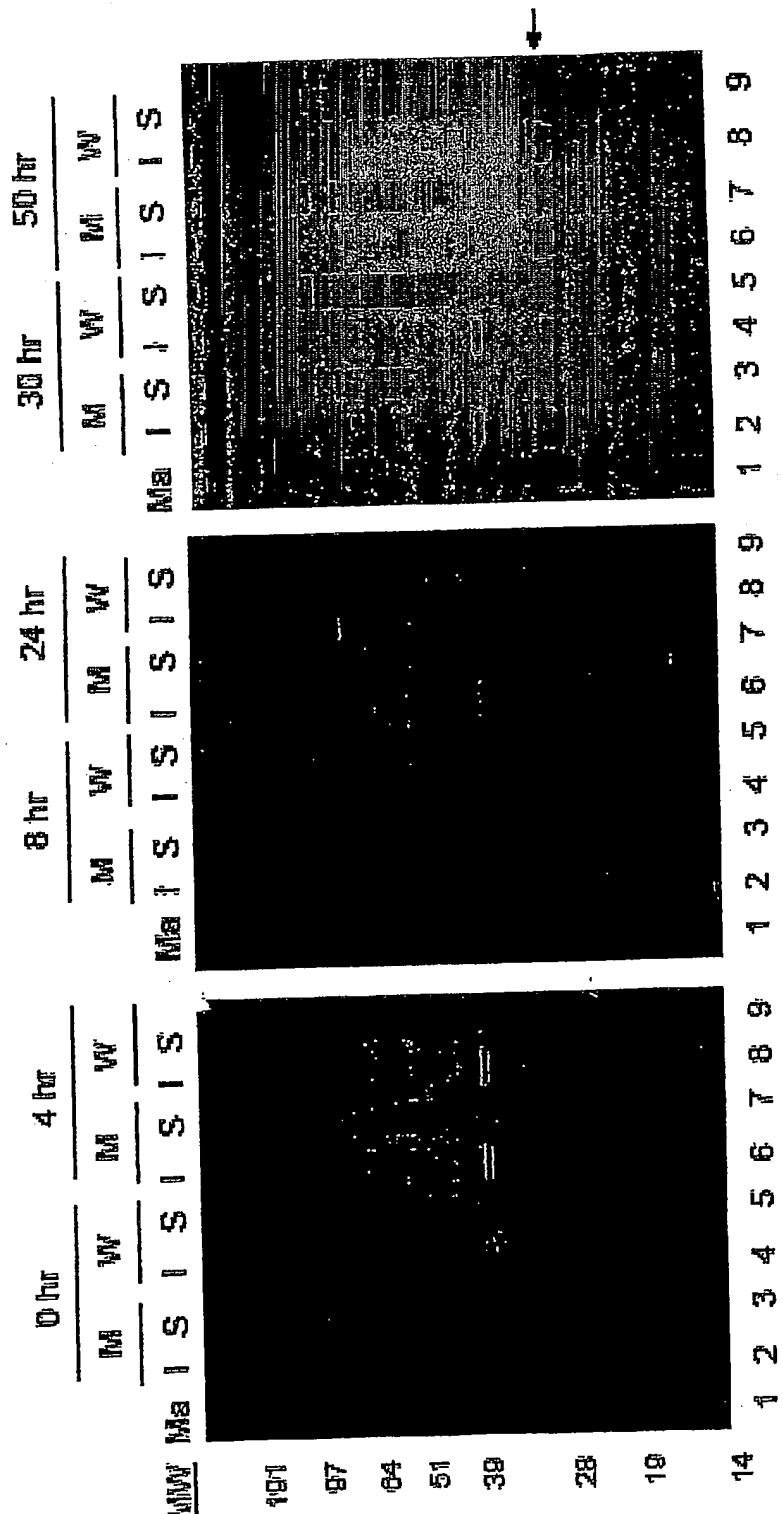


FIG. 9

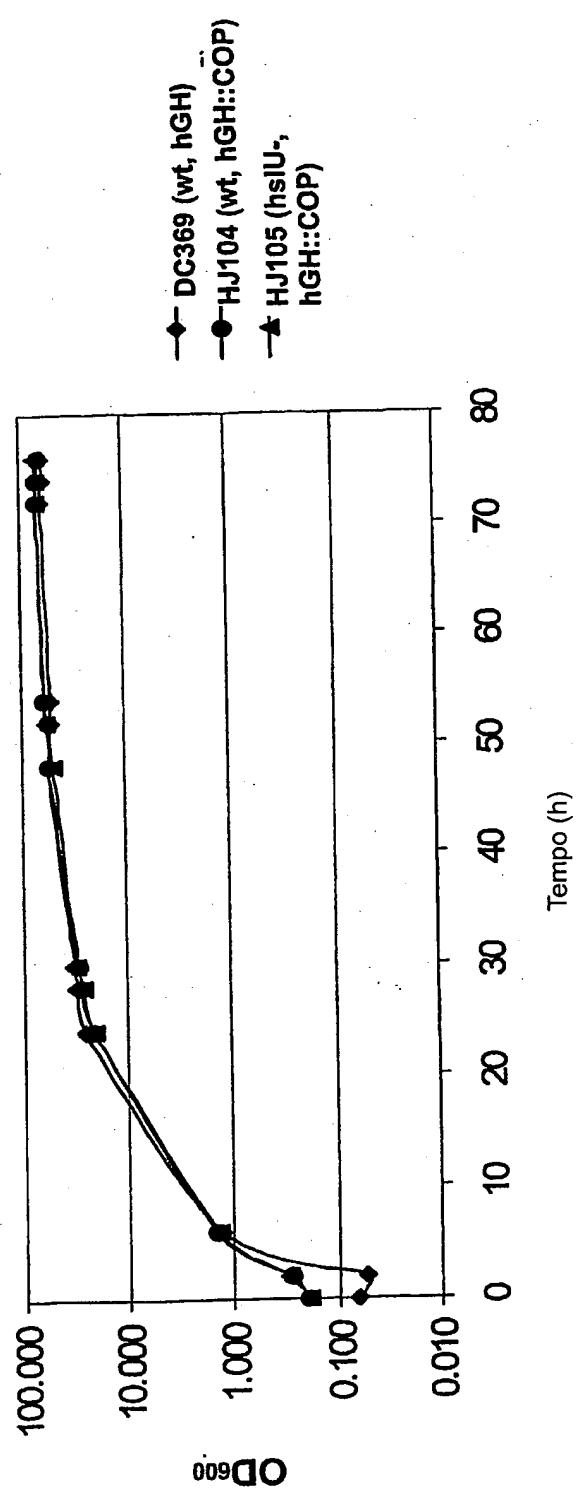


FIG. 10

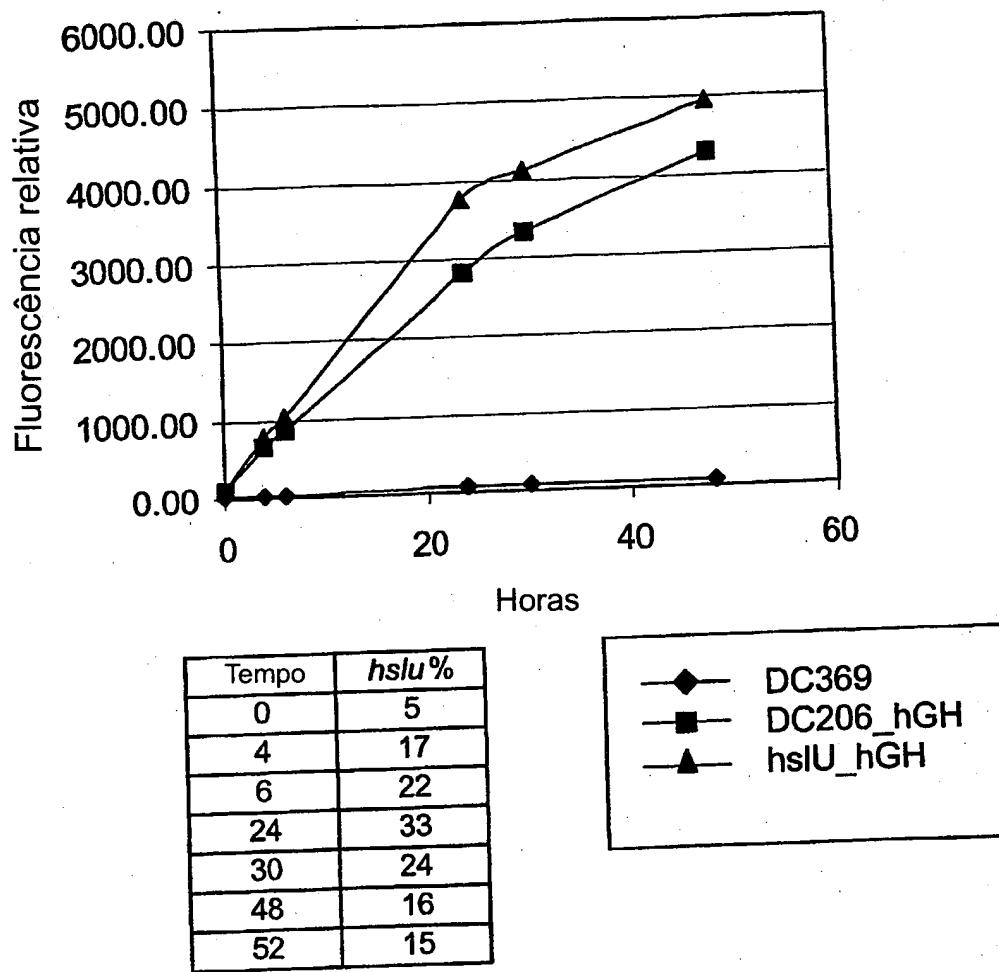


FIG. 11

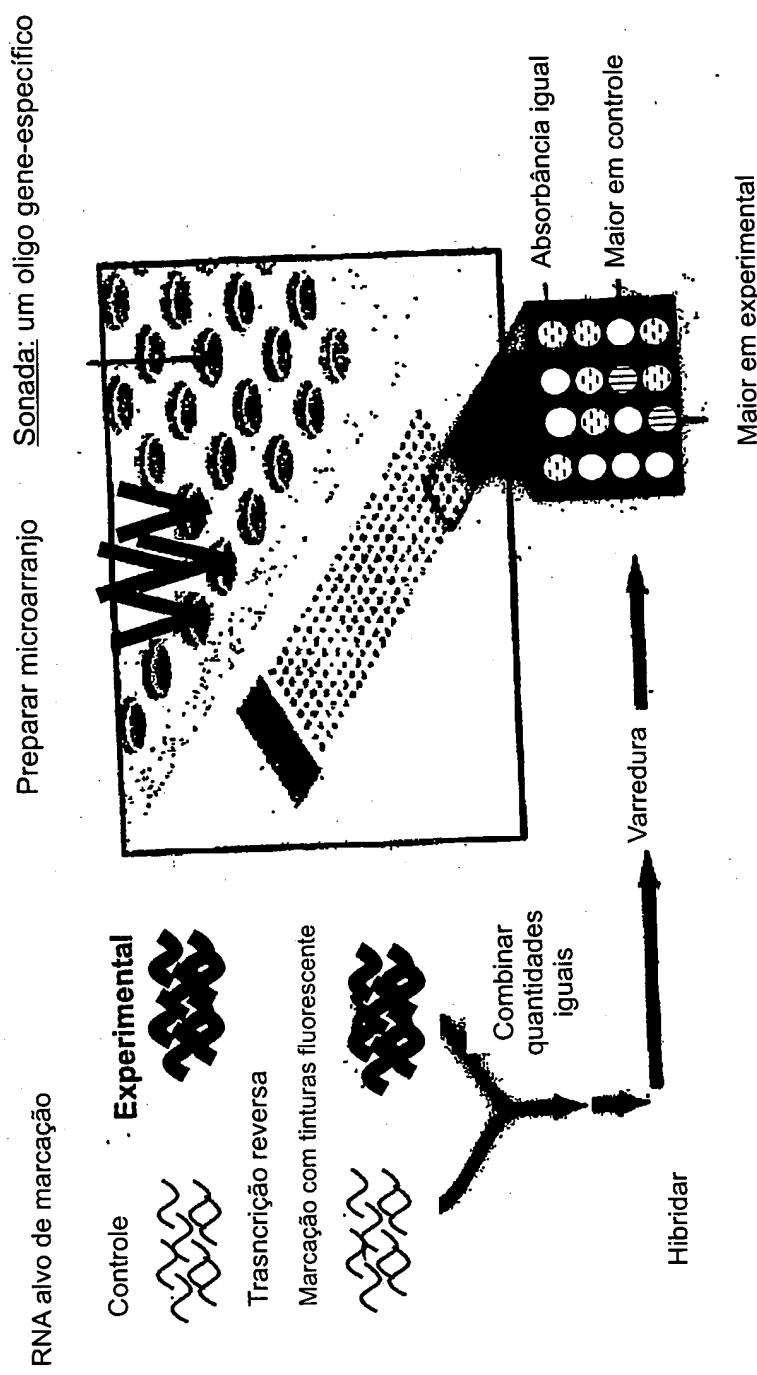


FIG. 12

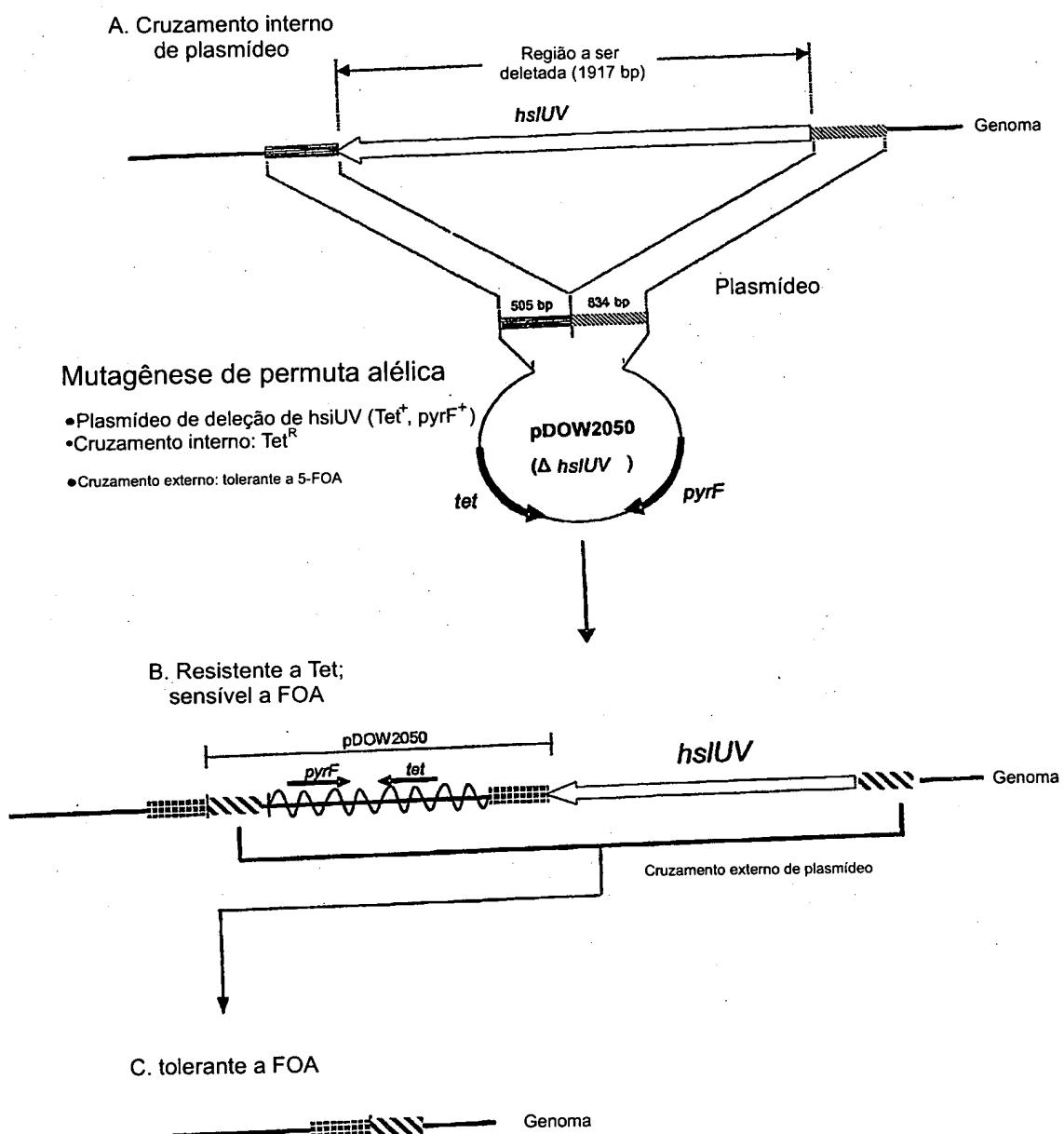


FIG. 13

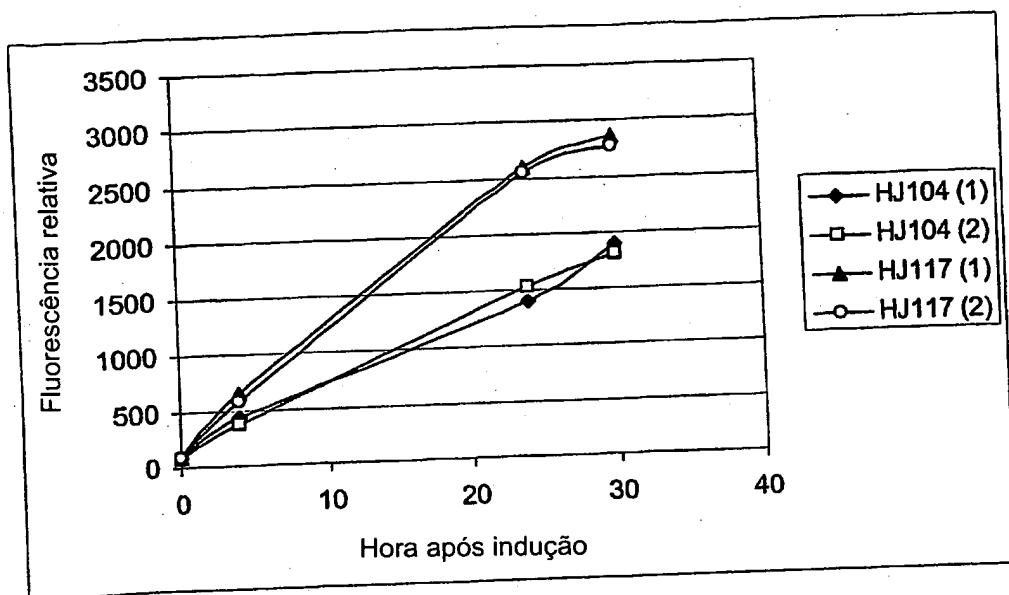


FIG. 14

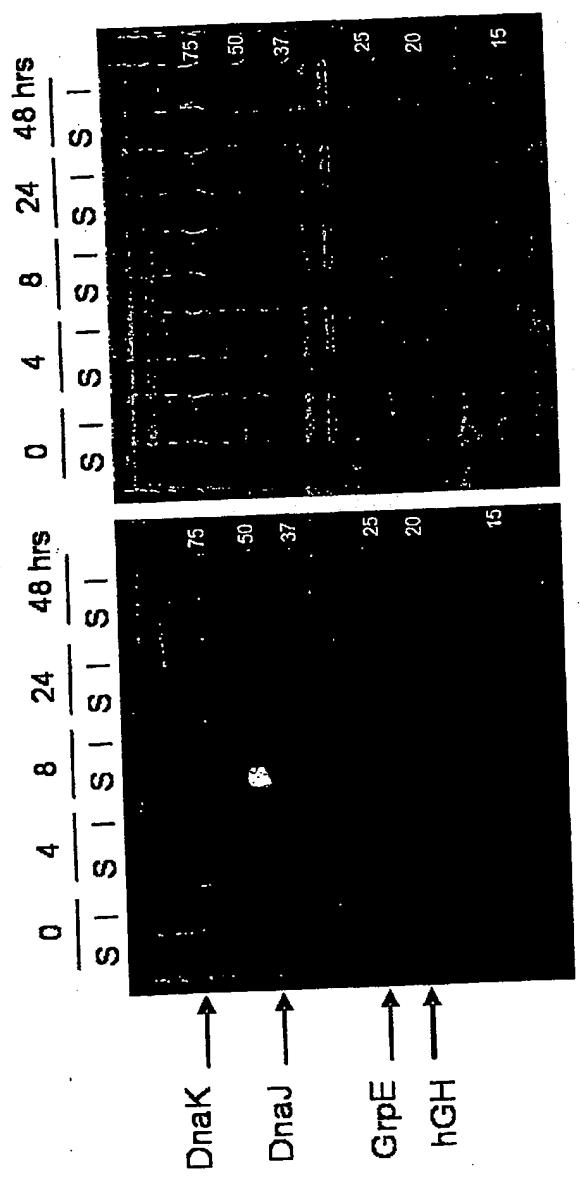


FIG. 15

Sem moduladores de dobramento

DnaKJ-GrpE

## RESUMO

### **Patente de Invenção: "PROCESSO PARA EXPRESSÃO DE PROTEÍNA MELHORADA ATRAVÉS DE ENGENHARIA DE CEPA".**

A presente invenção refere-se a um processo para melhorar os níveis de produção de proteínas ou peptídeos recombinantes ou melhorar o nível de proteínas ou peptídeos recombinantes ativos expressos em células hospedeiras. A invenção é um processo de comparar dois perfis genéticos de uma célula que expressa uma proteína recombinante e modificar a célula para alterar a expressão de um produto de gene que é sobre-regulado em resposta à expressão da proteína recombinante. O processo pode melhorar a produção de proteína ou pode melhorar a qualidade da proteína, por exemplo, aumentando a solubilidade de uma proteína recombinante.

Descritivo da Patente de Invenção para "PROCESSO PARA EXPRESSÃO DE PROTEÍNA MELHORADA ATRAVÉS DE ENGENHARIA DE CEPA".

REFERÊNCIA CRUZADA AO PEDIDO DE PATENTE RELACIONADO

5 Este pedido de patente reivindica prioridade ao Pedido Provisório US Nº 60/591.489, depositado em 26 de julho de 2004.

CAMPO DA INVENÇÃO

A presente invenção refere-se ao campo de produção de proteína, e em particular é um processo para melhorar os níveis de produção de proteínas ou peptídeos recombinantes ou melhorar o nível de proteínas ou peptídeos recombinantes ativos expressos em células hospedeiras.

ANTECEDENTES

Mais que 155 proteínas e peptídeos recombinantemente produzidos foram aprovados pela U.S. Food and Drug Administration (FDA) para o uso como fármacos e vacinas de biotecnologia, com outros 370 em experimentações clínicas.

Diferente das terapêuticas de molécula pequena que são produzidas através de síntese química, as proteínas e peptídeos são mais eficazmente produzidos em células vivas. Em muitos casos, a célula ou organismo foi geneticamente modificado para produzir ou aumentar a produção da proteína.

Quando uma célula é modificada para produzir quantidades grandes de uma proteína alvo, a célula é colocada sob tensão e freqüentemente reage induzindo ou suprimindo outras proteínas. A tensão que uma célula hospedeira sofre durante a produção de proteínas recombinantes pode aumentar a expressão, por exemplo, de proteínas específicas ou co-fatores para causar degradação da proteína recombinante sobre-expressada. A expressão aumentada de proteínas compensatórias pode ser contra-produtiva à meta de expressar níveis altos de proteína recombinante ativa, de comprometimento total. Expressão diminuída ou carência de expressão adequada de

outras proteínas pode causar desdobramento e agregação da proteína recombinante. Embora seja conhecido que uma célula sob tensão alterará seu perfil de expressão de proteína, não é conhecido em qualquer exemplo dado

que proteínas específicas serão sobre-reguladas ou sub-reguladas.

### MICROARRANJOS

Tecnologia de microarranjo pode ser usada para identificar a presença e nível de expressão de um número grande de polinucleotídeos em um ensaio simples. Ver por exemplo patente U.S. Nº 6.040.138, depositada em 15 de setembro de 1995, patente U.S. Nº 6.344.316, depositada em 25 de junho de 1997, patente U.S. Nº 6.261.776, depositada em 15 de abril de 1999, patente U.S. Nº 6.403.957, depositada em 16 de outubro de 2000, patente U.S. Nº 6.451.536, depositada em 27 de setembro de 2000, patente U.S. Nº 10 6.532.462, depositada em 27 de agosto de 2001, patente U.S. Nº 6.551.784, depositada em 9 de maio de 2001, patente U.S. Nº 6.420.108, depositada em 9 de fevereiro de 1998, patente U.S. Nº 6.410.229, depositada em 14 de dezembro de 1998, patente U.S. Nº 6.576.424, depositada em 25 de janeiro de 2001, patente U.S. Nº 6.687.692, depositada em 2 de novembro de 2000, 15 patente U.S. Nº 6.600.031, depositada em 21 de abril de 1998 e patente U.S. Nº 6.567.540, depositada em 16 de abril de 2001, todas atribuídas a Affymetrix, Inc.

Patente U.S. Nº 6.607.885 de E. I. duPont de Nemours e Co. descreve métodos para perfilar e identificar as alterações de expressão de gene após submeter uma célula bacteriana a expressão alterando as condições mediante comparação de uma primeira e segunda medição de microarranjo.

Wei et al. usou uma análise de microarranjo para investigar perfis de expressão de gene de *E. coli* com indução de gene de lac (Wei Y., et al. 25 (2001) High-density microarray-mediated gene expression profiling of *Escherichia coli*. *J Bacteriol.* 183(2):545-56). Outros grupos também investigaram os perfis transcripcionais regulados após mutação de genes endógenos ou deleção de genes reguladores (Sabina, J., et al (2003) Interfering with Different Steps of Protein Synthesis Explored by Transcriptional Profiling of 30 *Escherichia coli* K-12 *J Bacteriol.* 185:6158-6170; Lee JH (2003) Global analyses of transcriptomes and proteomas of a parent strain and an L-threonine-overproducing mutant strain. *J Bacteriol.* 185(18):5442-51; Kabir

MM, et al. (2003) Gene expression patterns for metabolic pathway in pgi knockout *Escherichia coli* with and without phb genes based on RT-PCR *J Biotechnol.* 105(1-2):11-31; Eymann C., et al. (2002) *Bacillus subtilis* functional genomics: global characterization of the stringent response by proteome and transcriptome analysis. *JBacteriol.* 184(9):2500-20.

Gill et al. descreve o uso de tecnologia de microarranjo para identificar alterações na expressão de genes relacionados à tensão em *E. coli* após expressão de proteínas de fusão recombinantes de cloranfenicol acetil-transferase (Gill et al. (2001) Genomic Analysis of High-Cell-Density Recombinant *Escherichia coli* Fermentation and "Cell Conditioning" for Improved Recombinant Protein Yield *Biotech. Bioengin.* 72:85-95). O perfil de transcrição de gene de tensão, compreendendo apenas 16% do genoma total, em densidade de célula alta foi usado para avaliar estratégias de "condicionamento de célula" para alterar os níveis das chaperonas, proteases e outras proteínas intracelulares antes da sobre-expressão da proteína recombinante. As estratégias para "condicionamento" envolveram manipulação farmacológica das células, incluindo através de tratamentos de ditiotreitol e etanol.

Asai et al. descreve o uso de análise de microarranjo para identificar genes alvos ativados por sobre-expressão de certos fatores sigma que são tipicamente induzidos após tensões de célula (Asai K., et al. (2003) DNA microarray analysis of *Bacillus subtilis* sigma factors of extracytoplasmic function family. *FEMS Microbiol. Lett.* 220(1):155-60). Células que sobre-expressam fatores sigma como também genes repórteres ligados aos promotores de fator sigma foram usadas para mostrar indução de gene regulada por tensão.

Choi et al. descreveu a análise e sobre-regulação de genes metabólicos que são sub-regulados em culturas de batelada de alta densidade de *E. coli* que expressa proteína de fusão de fator de crescimento similar à insulina humana (IGF-I<sub>f</sub>) (Choi et al. (2003) Enhanced Production of Insulin-Like Growth Factor I Fusion Protein in *Escherichia coli* by Coexpression of the Down-Regulated Genes Identified by Transcriptome Profiling App. *Envir. Microbio.* 69:4737-4742). O foco deste trabalho estava nas alterações me-

tabólicas que ocorrem durante as condições de alta densidade após indução de proteína. Genes que foram sub-regulados após indução de produção de proteína recombinante durante condições de crescimento de densidade alta foram identificados e genes metabólicos específicos que tinham sido 5 sub-regulados foram expressados em células produzindo IGF- $I_f$  recombinante. O trabalho mostrou que aumentando a produção metabólica de certas bases de nucleotídeo e aminoácidos poderia aumentar a produção de proteína e aquelas taxas de crescimento poderiam ser modificadas aumentando a expressão de uma molécula transportadora metabólica sub-regulada. Estas 10 estratégias foram projetadas para alterar o ambiente celular para reduzir tensões metabólicas associadas à produção de proteína em geral ou com cultura de densidade alta.

### DEGRADAÇÃO DE PROTEÍNA

Degradação indesejada de proteína recombinante apresenta um 15 obstáculo ao uso eficiente de certos sistemas de expressão. A expressão de proteínas exógenas freqüentemente induz respostas de tensão em células hospedeiras que podem ser, por exemplo, defesas naturais a uma fonte de carbono limitada. Todas as células contêm um número grande de genes capazes de produzir proteínas degradativas. Não é possível prognosticar quais 20 proteases serão reguladas por um hospedeiro dado em resposta à expressão de uma proteína recombinante particular. Por exemplo, as bactérias *P. fluorescens* contém até 200 proteases e proteínas relacionadas à protease.

No citoplasma de *E. coli*, a proteólise é em geral realizada por um grupo de proteases e moléculas de co-fator. Etapas de degradação mais 25 precoces são realizadas por cinco Hsps ATP-dependentes: Lon/La FtsH/HflB, ClpAP, ClpXP e ClpYQ/HslUV (Gottesman S (1996) Proteases and their targets in *Escherichia coli*. *Annu. Rev. Genet.* 30:465-506). Junto com FtsH (uma protease associada à membrana interna, o sítio ativo desta opõem-se ao citoplasma), ClpAP e ClpXP são responsáveis pela degradação de proteínas 30 modificadas em seus términos carboxila pela adição da cauda de desestabilização não-polar AANDENYALAA (Gottesman S, et al. (1998) The ClpXP and ClpAP proteases degrade proteins with carboxil-terminal peptide tails added

by the SsrA-tagging system. *Genes Dev.* 12:1338-1347; Herman C, et al. (1998) Degradation of carbóxi-terminal-tagged cytoplasmic proteins by the *Escherichia coli* protease HflB (FtsH). *Genes Dev.* 12:1348-1355).

Vários métodos foram tomados para evitar degradação durante a produção de proteína recombinante. Um método é produzir cepas hospedeiras que carregam mutações em um gene de protease. Baneyx e Georgiou, por exemplo, utilizaram uma cepa deficiente de protease para melhorar o rendimento de uma proteína de fusão de proteína A-β-lactamase (Baneyx F, Georgiou G. (1991) Construction and characterization of *Escherichia coli* strains deficient in multiple secreted proteases: protease III degrades high-molecular-weight substrates in vivo. *J Bacteriol* 173: 2696-2703). Park et al. usou um método mutacional similar para melhorar a atividade da proteína recombinante 30% comparada com a cepa de origem de *E. coli* (Park S. et al. (1999) Secretory production of recombinant protein by a high cell density culture of a protease negative mutant *Escherichia coli* strain. *Biotechnol. Progr.* 15:164-167). Patentes U.S. Nºs 5.264.365 e 5.264.365 descrevem a construção de *E. coli* deficiente de protease, particularmente múltiplas cepas deficientes de protease, para produzir polipeptídeos proteoliticamente sensíveis. Publicação de PCT Nº WO 90/03438 descreve a produção de cepas de *E. coli* que incluem cepas deficientes de protease ou cepas incluindo um inibidor de protease. Similarmente, Publicação de PCT Nº WO 02/48376 descreve cepas de *E. coli* deficientes em proteases DegP e Prc.

#### DOBRAMENTO DE PROTEÍNA

Outro obstáculo principal na produção de proteínas recombinantes em células hospedeiras é que a célula freqüentemente não é adequadamente equipada para produzir proteína solúvel ou ativa. Embora a estrutura primária de uma proteína seja definida por sua seqüência de aminoácido, a estrutura secundária é definida pela presença de hélices alfas ou folhas beta, e a estrutura terciária por ligações covalentes entre extensões de proteína adjacentes, como ligações de dissulfeto. Ao expressar proteínas recombinantes, particularmente em grande escala, a estrutura secundária e terciária da própria proteína é de importância crítica. Qualquer alteração significativa

em estrutura de proteína pode render uma molécula funcionalmente inativa, ou uma proteína com atividade biológica significativamente reduzida. Em muitos casos, uma célula hospedeira expressa moduladores de dobramento (FMs) que são necessários para produção apropriada de proteína recombinante ativa. Porém, nos níveis altos de expressão em geral requeridos para produzir produtos de biotecnologia utilizáveis, economicamente satisfatórios, uma célula freqüentemente não pode produzir modulador ou moduladores de dobramento nativo suficiente para processar a proteína recombinante.

Em certos sistemas de expressão, superprodução de proteínas exógenas pode ser acompanhada por seu desdobramento e segregação em agregados insolúveis. Em células bacterianas, estes agregados são conhecidos como corpos de inclusão. Em *E. coli*, a rede de moduladores/ chaperonas de dobramento inclui a família de Hsp70. A chaperona de Hsp70 principal, DnaK, impede agregação da proteína e suporta eficazmente o redobramento das proteínas estragadas. A incorporação de proteínas de choque térmico em agregados de proteína pode facilitar a desagregação. Porém, proteínas processadas para corpos de inclusão podem, em certos casos, ser restabelecidas através de processamento adicional da fração insolúvel. Proteínas encontradas em corpos de inclusão tipicamente têm que ser purificadas através de etapas múltiplas, incluindo desnaturação e renaturação. Processos de renaturação típicos para proteínas alvejadas por corpo de inclusão envolvem tentativas de dissolver o agregado em desnaturante concentrado e remoção subsequente do desnaturante por diluição. Agregados são freqüentemente formados novamente neste estágio. O processamento adicional soma custo, não há nenhuma garantia que o redobramento *in vitro* renderá produto biologicamente ativo, e as proteínas restabelecidas podem incluir quantidades grandes de impurezas de fragmento.

Um método para reduzir agregação de proteína é através de engenharia de fermentação, mais comumente reduzindo a temperatura de cultivo (ver Baneyx F (1999) *In vivo folding of recombinant proteins in Escherichia coli*. In *Manual of Industrial Microbiology and Biotechnology*, Ed. Davies et al. Washington, DC: American Society for Microbiology ed. 2:551-565 e

referências nele). A mais recente realização que dobramento de proteína *in vivo* é assistida por chaperonas moleculares, que promove a isomerização apropriada e alvejamento celular de outros polipeptídeos transientemente interagindo com intermediários de dobramentos, e por foldases, que acelera as etapas limitativas de taxa ao longo da via de dobramento, forneceu métodos adicionais para combater o problema de formação de corpo de inclusão (ver por exemplo Thomas JG et al. (1997). Molecular chaperones, folding catalysts and the recovery of active recombinant proteins from *E. coli*: to fold or to refold. *Appl Biochem Biotechnol*, 66:197-238).

Em certos casos, a sobre-expressão de chaperonas foi observada aumentar os rendimentos solúveis de proteínas propensas à agregação (ver Baneyx, F. (1999) Recombinant Protein Expression in *E. coli* *Curr. Opin. Biotech.* 10:411-421 e referências nele). O processo não parece envolver dissolução de corpos de inclusão recombinantes pré-formados mas está relacionado ao dobramento melhorado das cadeias de proteína recentemente sintetizadas. Por exemplo, Nishihara et al. co-expressou *groESL* e *dnaJK/grpE* no citoplasma para melhorar a estabilidade e acumulação de Cryj2 recombinante (um alergênio de pólen de cedro japonês) (Nishihara K, Kanemori M, Kitagawa M, Yanagi H, Yura T. 1998. Chaperone coexpression plasmids: differential and synergistic roles of DnaK-DnaJ-GrpE and GroEL-GroES in assisting folding of an allergen of Japanese cedar pollen, Cryj2, in *Escherichia coli*. *Appl. Environ. Microbiol.* 64:1694). Lee e Olins também co-expressaram GroESL e DnaK e aumentaram a acumulação de procolagenase humana em dez vezes (Lee S, Olins P. 1992. Effect of overproduction of heat shock chaperones GroESL and DnaK on human procollagenase production in *Escherichia coli*. *JBC* 267:2849-2852). O efeito benéfico associado a um aumento na concentração intracelular destas chaperonas parece altamente dependente da natureza da proteína superproduzida, e o sucesso não é de forma alguma garantido.

Uma necessidade existe por processos para desenvolvimento de cepas hospedeiras que mostram produção, atividade ou solubilidade melhoradas de proteína ou peptídeo recombinante para reduzir os custos de fa-

bricação e aumentar o rendimento dos produtos ativos.

É, portanto, um objetivo da invenção fornecer processos para melhorar expressão de proteína recombinante em um hospedeiro.

É um outro objetivo da invenção fornecer processos que aumentam os níveis de expressão em células hospedeiras expressando proteínas ou peptídeos recombinantes.

É outro objetivo da invenção fornecer processos para aumentar os níveis de proteína solúvel feita em sistemas de expressão recombinante.

É ainda outro objetivo da invenção fornecer processos para aumentar os níveis de proteína ativa feita em sistemas de expressão recombinante.

#### SUMÁRIO

Um processo é fornecido para melhorar a expressão de uma proteína ou peptídeo recombinante compreendendo:

- 15            i) expressar a proteína ou peptídeo recombinante em uma célula hospedeira;
- ii) analisar um perfil genético da célula e identificar um ou mais produtos de gene endógeno que são sobre-regulados na expressão ou sobre-expressão da proteína ou peptídeo recombinante; e
- 20            iii) alterar a expressão de um ou mais produtos de gene endógeno identificados geneticamente modificando a célula.

O processo pode fornecer expressão melhorada como medida por rendimentos melhorados de proteína, ou pode melhorar a recuperação da proteína ativa, por exemplo aumentando a solubilidade da proteína recombinante expressa, ou uma proteína ou peptídeo relacionado.

Usando este processo, pode ser determinado quais das muitas proteínas celulares são "escolhidas" pela célula para compensar a expressão da proteína recombinante estranha, e esta informação pode levar ao desenvolvimento de sistemas de expressão de proteína mais eficazes. Por exemplo, 30 é conhecido que, tipicamente, uma célula seletivamente sobre-regulará uma ou mais proteases para degradar uma proteína recombinante sobre-expres-sada. Porém, pode não ser prognosticado com antecedência qual(is) prote-

ase(s) a célula sobre-regulará para compensar a tensão causada por qualquer proteína recombinante dada. Análise do perfil genético da célula por microarranjo ou tecnologia equivalente pode identificar quais proteases são sobre-reguladas em uma célula dada em resposta à produção de proteína 5 exógena. Esta informação é depois usada para geneticamente modificar a célula para diminuir a expressão destas proteases particulares, poupando outras proteínas que são úteis ou até mesmo necessárias para homeostase de célula.

Como outro exemplo, uma célula pode seletivamente sobre-regular um ou mais moduladores ou co-fatores de dobramentos para aumentar a capacidade de dobramento ou solubilidade da proteína recombinante. Novamente, não pode ser prognosticado com antecedência quais moduladores ou co-fatores de dobramentos serão selecionados em um sistema dado para ajudar no processamento de uma proteína recombinante 15 específica. Analisando o perfil genético através de microarranjo ou tecnologia equivalente permite a identificação dos moduladores ou co-fatores de dobramentos que foram sobre-regulados. Com base nesta informação, a célula é geneticamente modificada para aumentar a expressão dos moduladores ou co-fatores de dobramentos selecionados preferidos pela célula para a proteína recombinante dada. Esta modificação pode aumentar o percentual de 20 proteína ativa se restabelecida, enquanto minimizando o impacto prejudicial em homeostase de célula.

Portanto, o rendimento e/ou atividade e/ou solubilidade da proteína recombinante pode(m) ser aumentado(s) ou modificando o organismo 25 hospedeiro ou aumentando ou diminuindo a expressão de uma proteína compensatória (isto é uma proteína que é sobre-regulada em resposta à tensão de célula dada) de uma maneira que é seletiva e que deixa outros mecanismos benéficos inteiros da célula.

O processo pode ser usado iterativamente até a expressão de 30 proteína recombinante ativa ser otimizado. Por exemplo, usando o processo descrito acima, a célula ou organismo hospedeiro é geneticamente modificado para sobre-regular, sub-regular, knock-in ou knock-out uma ou mais

proteínas compensatórias identificadas. A célula ou organismo hospedeiro assim modificado pode depois ser cultivado para expressar a proteína recombinante, ou uma proteína ou peptídeo relacionado, e proteínas compensatórias adicionais identificadas por meio de microarranjo ou análise equivalente. A célula ou organismo hospedeiro modificado é depois de novo geneticamente modificado para sobre-regular, sub-regular, knock-in ou knock-out as proteínas compensatórias adicionais selecionadas. Este processo pode ser iterado até uma célula ou organismo hospedeiro ser obtido que exibe expressão máxima de ativo e/ou proteína solúvel sem enfraquecimento 5 devido do organismo ou célula hospedeira. Estas etapas por exemplo podem ser repetidas por exemplo, uma, duas, três, quatro, cinco, seis, sete, oito, 10 nove, ou dez ou mais vezes.

Em outra modalidade, o processo também compreende: iv) expressar a proteína ou peptídeo recombinante em uma célula geneticamente 15 modificada. Em ainda outra modalidade, o processo também compreende: v) analisar um segundo perfil genético da célula geneticamente modificada expressando proteína ou peptídeo recombinante e identificando um ou mais produtos de gene adicionais que são diferencialmente expressos na célula modificada que expressa a proteína ou peptídeo recombinante. Em uma outra 20 modalidade, o processo adicionalmente compreende: vi) alterar a expressão de um ou mais produtos de gene adicionais identificados para fornecer uma célula modificada dupla. Opcionalmente, a proteína ou peptídeo recombinante, ou uma proteína ou peptídeo relacionado, pode ser expressado na célula modificada dupla. Os produtos de gene diferencialmente regulados identificados 25 na célula modificada podem ser sobre ou sub-regulados quando comparados à célula hospedeira ou quando comparados à célula modificada que não expressa proteína ou peptídeo recombinante.

Em ainda outra modalidade, o processo também compreende: iv) analisar um segundo perfil genético de uma célula geneticamente modificada 30 expressando a proteína ou peptídeo recombinante e identificar um ou mais produtos de gene adicionais que são diferencialmente expressos na célula modificada que não está expressando a proteína ou peptídeo recombinante.

Em uma outra modalidade, o processo adicionalmente compreende: v) alterar a expressão de um ou mais produtos de gene adicionais identificados na célula modificada para fornecer uma célula modificada dupla. Os produtos de gene diferencialmente regulados identificados na célula modificada podem 5 ser sobre ou sub-regulados quando comparados à célula ou organismo hospedeiro ou quando comparados à célula modificada que não expressa a proteína ou peptídeo recombinante.

Em uma modalidade específica, um processo é fornecido para melhorar a expressão de uma proteína ou peptídeo recombinante compreendendo: i) expressar a proteína ou peptídeo recombinante em uma célula hospedeira; ii) analisar um perfil genético da célula e identificar pelo menos uma protease que é sobre-regulada quando a proteína ou peptídeo recombinante é expressado; e iii) alterar a expressão de uma protease identificada geneticamente modificando a célula ou organismo hospedeiro para reduzir a 10 expressão da protease sobre-regulada. Em uma outra modalidade, o processo compreende alterar a expressão de pelo menos uma segunda protease identificada na célula modificada para fornecer uma célula modificada de protease dupla. Em outra modalidade, o processo também compreende: iv) 15 expressar a proteína ou peptídeo recombinante, ou uma proteína ou peptídeo relacionado, em uma célula modificada de protease. Em outra modalidade, o processo também compreende analisar um segundo perfil genético da célula modificada de protease para identificar um ou mais produtos de gene adicionais que são diferencialmente expressos na célula modificada.

Em outra modalidade, um processo é fornecido para melhorar a 20 expressão de uma proteína ou peptídeo recombinante compreendendo: i) expressar a proteína ou peptídeo recombinante em uma célula hospedeira; ii) analisar um perfil genético da célula e identificar pelo menos um modulador de dobramento sobre-regulado (FM) que é sobre-regulado após sobre-expressão da proteína ou peptídeo recombinante; e iii) alterar a expressão de pelo 25 menos um modulador de dobramento identificado geneticamente modificando a célula para fornecer uma célula modificada de FM. Em uma outra modalidade, o processo compreende alterar a expressão de pelo menos um se-

gundo modulador de dobramento identificado na célula modificada para fornecer uma célula modificada de FM duplo. Em outra modalidade, o processo também compreende: iv) expressar a proteína ou peptídeo recombinante, ou uma proteína ou peptídeo relacionado, em uma célula modificada de FM. Em 5 outra modalidade, o processo também compreende analisar um segundo perfil genético da célula modificada de FM para identificar um ou mais produtos de gene adicionais que são diferencialmente expressos na célula modificada.

O termo "perfil genético" como aqui usado é significado para incluir uma análise de genes em um genoma, mRNA transscrito de genes no genoma (ou o cDNA equivalente), produtos de transcrição que foram modificados por uma célula como variantes de encaixe de genes em sistemas eucarióticos, ou proteínas ou peptídeos translados de genes em um genoma, incluindo proteínas que são modificadas pela célula ou transladada de variantes de encaixe de mRNA transladado do genoma. Um perfil genético é significado incluir mais de um gene ou produto de gene, e tipicamente inclui um grupo de pelo menos 5, 10, 50, 100 ou mais genes ou produtos de gene que são analisados.

Em uma modalidade, o perfil genético analisado pode ser um 20 perfil de transcriptoma, isto é um perfil dos produtos de transcrição de genes do genoma. O processo pode incluir analisar o perfil de transcriptoma usando um microarranjo ou tecnologia equivalente. Nesta modalidade, o microarranjo pode incluir os pares de ligação para pelo menos uma porção do transcriptoma da célula hospedeira, e tipicamente inclui amostras de pares de ligação 25 para os produtos de gene de pelo menos 50% do genoma do organismo. Mais tipicamente, o microarranjo inclui amostras de pelo menos 80%, 90%, 95%, 98%, 99% ou 100% dos pares de ligação para produtos de gene no genoma da célula hospedeira.

Em uma modalidade separada, o microarranjo pode incluir um 30 subconjunto selecionado de pares de ligação para genes ou produtos de gene que representam classes de produtos que são afetadas pela expressão de proteína recombinante. Exemplos não-limitativos incluem proteases putativas

ou conhecidas, co-fatores de proteases ou proteínas similares à protease; moduladores de dobramentos, moduladores co-fatores de dobramentos ou proteínas que podem melhorar o dobramento ou solubilidade da proteína; fatores de transcrição; proteínas envolvidas em estabilidade de ácido nucléico ou iniciação translacional; cinases; receptores extracelulares ou intracelulares; enzimas metabólicas; co-fatores metabólicos; proteínas de envelope; fatores de sigma; proteínas ligadas à membrana; proteínas de transmembrana; proteínas associadas à membrana e genes de manutenção. O perfil genético pode ser analisado medindo a ligação dos genes expressos da célula hospedeira que expressa a proteína ou peptídeo recombinante para o microarranjo. O perfil de transcriptoma pode também ser analisado usando ensaios de não-microarranjo como ensaios de mancha, incluindo ensaios de northern blot, ou colunas revestidas com pares de ligação.

Em outra modalidade, o perfil genético analisado pode ser um perfil de proteoma, isto é um perfil das proteínas produzidas de genes em um organismo dado. O processo pode incluir analisar o perfil de proteoma usando, por exemplo, eletroforese bidimensional. Técnicas como espectrometria de massa em combinação com ferramentas de separação como eletroforese em gel bidimensional ou cromatografia líquida multidimensional, podem também ser usadas no processo. Em eletroforese bidimensional, as proteínas separadas podem incluir proteínas de pelo menos 10% do proteoma do organismo. Mais tipicamente, proteínas de pelo menos 20%, 30%, 40%, 60%, 80% ou 90% das proteínas no proteoma da célula hospedeira são separados e analisados através de técnicas como tingimento de proteínas e/ou espectrometria de massa.

Em modalidade adicional, o perfil de proteoma é analisado usando espectrometria de massa. Há várias técnicas relacionadas que usam cromatografia líquida (LC) acoplada à espectrometria de massa (MS) e espectrometria de massa de tandem (MS/MS) para identificar proteínas e medir sua abundância relativa. Freqüentemente, uma amostra é marcada com um marcador de isótopo pesado que permite a comparação com a outra amostra sem alterar as propriedades químicas. Por exemplo, em uma amostra a cis-

teína de aminoácido pode ser marcada com um marcador que contém oito átomos de hidrogênio. A outra amostra é marcada com um marcador que contém oito átomos de deutério ("pesado") no lugar de (+8 Dáltons). Os dados de MS podem ser usados para encontrar pares de peptídeos de 8 Dáltons separadamente e quantificar a diferença. Dados de MS/MS dos mesmos peptídeos fornecem uma aproximação de seqüência primária, e a proteína ID. Outros experimentos identificam as proteínas *in vivo* mediante cultivo das células com aminoácidos "pesados". Estes tipos de técnicas podem ser usados para identificar milhares de proteínas em um experimento simples e estimar a abundância relativa se presente em ambas as amostras (ver Goodlett DR e Aebersold RH (2001). Mass Spectrometry in Proteomics. Chem Rev 101:269-295). ICAT é um tipo de MS/MS, representa Marcadores de Afinação Codificados de Isótopo (ver Gygi SP, Rist B, Gerber SA, Turecek F, Gelb MH, e Aebersold RH (1999). Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. Nat Biotech 17:994-999).

Em outra modalidade, o processo pode incluir analisar o perfil de proteoma usando, por exemplo, um microarranjo. Nesta modalidade, o arranjo pode incluir os pares de ligação para pelo menos uma porção das proteínas expressas pela célula hospedeira sob condições de crescimento apropriadas, e tipicamente inclui os pares de ligação para proteínas de pelo menos 10% do proteoma do organismo. Mais tipicamente, o microarranjo inclui os pares de ligação para proteínas de pelo menos 20%, 30%, 40%, 60%, 80% ou 90% das proteínas no proteoma da célula hospedeira. Os pares de ligação podem ser anticorpos que podem ser fragmentos de anticorpo como fragmentos de anticorpo de cadeia simples. Em uma modalidade separada, o microarranjo pode incluir os pares de ligação para um subconjunto selecionado de proteínas do proteoma, incluindo, por exemplo, proteínas de protease putativa ou moduladores de dobramentos putativos. O microarranjo pode tipicamente também incluir um conjunto de pares de ligação para proteínas que são usadas como controles. O perfil genético pode ser analisado medindo a ligação das proteínas da célula hospedeira que expressa a proteína ou peptídeo recombinante aos pares de ligação no microarranjo. O perfil de proteoma pode

também ser analisado em um formato de ensaio padrão, como um ensaio de Elisa ou um ensaio de western blot padrão.

As amostras no perfil genético podem ser analisadas individualmente ou agrupadas em agrupamentos. Os agrupamentos podem tipicamente ser agrupados por similaridade em expressão de gene. Em modalidades particulares, os agrupamentos podem ser agrupados como genes que são sobre-regulados a uma extensão similar ou genes que são sub-regulados a uma extensão similar.

O gene sobre-regulado identificado é tipicamente identificado comparando um perfil genético da célula hospedeira que expressa a proteína ou peptídeo recombinante a um perfil genético da célula hospedeira que não expressa a proteína ou peptídeo recombinante. Em uma outra modalidade, uma célula hospedeira que expressa uma proteína homóloga é analisada para a primeira proteína recombinante.

O genoma da célula hospedeira que expressa a proteína ou peptídeo recombinante pode ser modificado através de recombinação, por exemplo recombinação homóloga ou recombinação heteróloga. O genoma pode também ser modificado por mutação de um ou mais nucleotídeos em uma estrutura de leitura aberta que codifica um gene, particularmente uma protease identificada. Em outra modalidade, a célula hospedeira é modificada incluindo um ou mais vetores que codificam um inibidor de um gene ou produto de gene identificado, como um inibidor de protease. Em outra modalidade, a célula hospedeira é modificada por inibição de um promotor que pode ser um promotor nativo. Em uma modalidade separada, a célula hospedeira é modificada incluindo um ou mais vetores que codificam um gene, tipicamente um modulador de dobramento ou um co-fator de um modulador de dobramento. Em outra modalidade, a célula hospedeira é modificada intensificando um promotor para um modulador de dobramento identificado ou um co-fator para um modulador de dobramento, incluindo adicionando um promotor e-xógeno ao genoma da célula hospedeira.

A célula hospedeira pode ser qualquer célula capaz de produzir proteína ou peptídeo recombinante. Em uma modalidade, a célula hospedeira

é um procarioto, como uma célula bacteriana incluindo, mas não limitada a uma espécie de Escherichia ou uma de Pseudonaonas. A célula hospedeira pode ser uma célula de Pseudomonas como uma célula de *P. fluorescens*. Em outras modalidades, a célula hospedeira é uma célula de *E. coli*. Em outra 5 modalidade a célula hospedeira é uma célula eucariótica, por exemplo uma célula de inseto, incluindo mas não limitado a uma célula de uma espécie de *Spodoptera*, *Trichoplusia*, *Drosophila* ou *Estignaene*, ou uma célula mamífera, incluindo mas não limitada a uma célula murina, uma célula de hamster, uma de macaco, uma de primata ou uma célula humana. Em outra modalidade, a 10 célula hospedeira é uma célula de planta, incluindo, mas não limitada a, uma célula de tabaco, milho, uma célula de uma espécie de *Arabidopsis*, célula de batata ou arroz. Em outra modalidade, um organismo inteiro é analisado no processo, incluindo mas não limitado a um organismo transgênico.

Em uma modalidade, os genes ou produtos de gene compensatórios sobre-regulados identificados são uma ou mais proteases e/ou um ou mais moduladores de dobramentos. Em certas modalidades, um gene ou produto de gene identificado pode também ser uma subunidade de uma protease ou um modulador de dobramento ou um co-fator de uma protease ou um co-fator de um modulador de dobramento. Em uma modalidade, o gene 15 identificado pode ser selecionado de uma peptidase serina, treonina, cisteína, aspártico ou metalo. Em certas outras modalidades, o gene ou produto de gene identificado pode ser selecionado de *hsIV*, *hsIU*, *clpA*, *clpB* e *clpX*. O gene ou produto de gene identificado pode também ser um co-fator de uma protease. Em outra modalidade, o gene ou produto de gene identificado é um 20 modulador de dobramento. Em certas modalidades, o gene ou produto de gene identificado pode ser selecionado de uma proteína de chaperona, uma foldase, uma peptidil prolil isomerase e uma isomerase de ligação de dissulfeto. Em uma modalidade, o gene ou produto de gene identificado pode ser selecionado de *htpG*, *cbpA*, *dnaJ*, *dnaK* e *fkbP*. Em uma modalidade, um gene 25 ou produto de gene homólogo ao gene sobre-regulado identificado é modificado no genoma do hospedeiro.

O processo pode levar à produção aumentada de proteína ou

peptídeo recombinante em uma célula hospedeira, por exemplo, aumentando a quantidade de proteína por grama de proteína de hospedeiro (proteína de célula total) em uma quantidade dada de tempo, ou aumentando a quantidade de comprimento de tempo durante o qual a célula ou organismo está produzindo a proteína recombinante. A produção aumentada pode otimizar a eficiência da célula ou organismo, por exemplo, diminuindo a despesa de energia, aumentando o uso de recursos disponíveis ou diminuindo os requerimentos para suplementos de crescimento em meios de crescimento. A produção aumentada pode também resultar em um nível aumentado de proteína recuperável ou peptídeo, como proteína solúvel, produzido por grama de recombinante ou por grama de proteína de célula hospedeira.

A invenção também inclui uma célula hospedeira recombinante melhorada que é produzida pelo processo reivindicado.

#### BREVE DESCRIÇÃO DOS DESENHOS

Figura 1 é um gráfico de uma comparação de crescimento (densidade óptica com o passar do tempo) de cepas diferentes de *P. fluorescens*. As células foram induzidas com 0,3 M de IPTG em 24 h após inoculação. As cepas são: DC280 que abriga o vetor vazio pDOW1339, DC240 que produz a enzima de nitrilase citoplasmática solúvel e DC271 que produz a hGH periplasmática parcialmente insolúvel. DC206, a cepa parental de DC280, DC240 e DC271, foi incluída como um controle. As amostras foram tiradas em 0 e 4 h de indução pós-IPTG para isolamento de RNA e profilação da expressão de gene, como indicado pelas setas.

Figura 2 é um gráfico de agrupamento hierárquico de todos os genes de cepas de *P. fluorescens* DC280, DC240 e DC271 em 12 agrupamentos em 4 h após IPTG quando comparados em 0 h IPTG (indicado no fundo da figura). Com base no valor e tendência, os genes foram aglomerados e agrupados usando o algoritmo de agrupamento hierárquico de Spotfire DecisionSite. Linhas pontilhadas indicam pontos de dados que foram filtrados devido à qualidade de manchas fraca ou baixo nível de expressão. O eixo geométrico x representa a comparação de cada cepa; o eixo geométrico y representa o valor de expressão com relação 4 h atrás para antes da indução

de IPTG. Todos os FMs identificados são realçados. Agrupamento 7 mostra 2 genes de FM e 2 de subunidade de protease que são altamente expressos em cepa DC271, que superproduzem a proteína de hGH periplásica. Os genes de FM restantes são agrupados no agrupamento 6.

5           Figura 3 é uma análise de agrupamento hierárquico do agrupamento 6 da Figura 2. No agrupamento 8 novo, dois moduladores de dobramentos, DnaK e DnaJ, foram identificados ambos destes mostraram níveis de expressão mais altos para produção de proteína recombinante periplásica similar ao HslVU, CbpA e HtpG previamente identificado. Agrupamento 6  
10 mostra onde o resto dos FMs estão agrupados.

Figura 4 é um diagrama de Venn mostrando a protease sobre-regulada e FMs dos três conjuntos de experimentos na Tabela 5, 6 e 7. Como resumido na Tabela 5, 6 e 7, a lista de genes foi organizada em esquema de Venn para realçar a sobreposição da lista de gene entre os três  
15 conjuntos de experimentos indicados no canto. Para cada gene, a razão de cada experimento foi exibida com 2 como um corte.

Figura 5 é um gráfico da análise de seqüência dos genes *hsIV* (RXF01961) e *hsIU* (RXF01957) de *P. fluorescens* gerados por Artemis. O diagrama de uso de códon (painel do topo) indica que o limite de gene está correto. Este é confirmado pelos melhores homólogos das seqüências de proteína de HslV e HslU de *P. aeruginosa* como indicado em baixo dos genes de RXF01961 e RXF01957. O diagrama de contagem de qualidade de Phrap mostra que a qualidade de seqüência é boa, isto é a linha de contagem está acima da linha horizontal indicando uma qualidade melhor que 1 erro em 10  
20 kb (painel mediano). As caixas apontadas brancas abertas abaixo dos genes mostram a localização das sondas geradas para uso nos experimentos de microarranjo de DNA.

Figura 6 é uma ilustração esquemática de uma construção de mutante de *hsIU* onde um produto de PCR de aproximadamente 550 bp de  
30 *hsIU* (caixa azul claro) foi ligado no vetor de clonagem TOPO TA2.1 (círculo). O plasmídeo resultante foi transformado em células de *P. fluorescens* competentes e colônias resistentes à canamicina (kan) foram analisadas em PCR

diagnóstica para confirmar a construção de uma mutação de inserção no gene de *hslU*.

Figura 7 é um gráfico de uns ensaios de curva de crescimento comparando cepa do tipo selvagem com de mutante de *hslU* que superproduz hGH ou pbp::hGH em meio de produção de frasco de agitação. As setas indicam pontos de tempo onde foram tiradas as amostras.

Figura 8 é uma imagem de análise de SDS-PAGE das cepas DC271 e DC373 que expressam pbp::hGH. As amostras de DC271 (tipo selvagem, W) e DC373 (mutante de *hslU*, M) foram tiradas logo antes da indução da proteína (0 h) e depois 4 h, 8 h, 24 h, e 30 h após adição de IPTG. Frações solúveis (S) e insolúveis (I) foram preparadas para cada amostra analisada. A produção de hGH não processado e processado é indicada pelas setas. O marcador (Ma) de peso molecular (MW) é mostrado à direita dos géis.

Figura 9 é uma imagem da análise de SDS-PAGE das cepas DC369 e DC372 que expressam hGH no citoplasma. As amostras foram tiradas de DC369 (tipo selvagem, W) e DC372 (mutante de *hslU*, M) logo antes da indução da proteína (0 h) e depois 4 h, 8 h, 24 h, 30 h e 50 h após adição de IPTG. Frações solúveis (S) e insolúveis (I) foram preparadas para cada amostra analisada. A produção de hGH é indicada por uma seta. O marcador (Ma) de peso molecular (MW) é mostrado à direita dos géis.

Figura 10 é um gráfico de curvas de crescimento de cepas que expressam a proteína de fusão de hGH::COP. As cepas incluem: DC369 que expressa hGH apenas (não fundido com COP) como um controle negativo; HJ104, o tipo selvagem que expressa hGH::COP; HJ105, o mutante de *hslU* que expressa hGH::COP.

Figura 11 é um gráfico das medições da atividade de fluorescência verde para cepas que expressam a proteína de fusão de hGH::COP usando um fluorímetro. Cinco OD600 de cultura de célula foram ensaiadas para cada cepa abrigando hGH ou hGH::COP em pontos de tempo diferentes após indução de IPTG. As cepas testadas incluem: DC369 que expressa hGH apenas (não fundido com COP) como um controle negativo; HJ104, o tipo

selvagem que expressa hGH::COP; HJ105, o mutante de *hslU* que expressa hGH::COP. A tabela inserida mostra o percentual de aumento da fluorescência relativa no mutante de *hslU* comparado ao tipo selvagem em pontos de tempo diferentes após indução de IPTG.

5 Figura 12 é uma representação pictórica do processo de medir abundância relativa de mRNA entre duas amostras.

Figura 13 é uma representação da construção de deleção cromossômica de gene de *hslUV* na cepa *pyrF*-negativa. A. Plasmídeo PDOW2050 contém fragmentos de DNA de 505 bp e 634 bp flanqueando o 10 gene de *hslUV*. Uma vez que plasmídeo de suicídio pDOW2050 não pode replicar em *P. fluorescens*, as células resistentes à tetraciclina apenas serão geradas após um evento de recombinação simples em uma das regiões homólogas que integram o plasmídeo inteiro no genoma. B. Células resistentes à tetraciclina contêm o plasmídeo inteiro integrado no genoma. Estas 15 células também contêm o gene de *pyrF* codificado do plasmídeo. Seleção para células que têm o segundo evento recombinante ocorreu através de células de colocação em placa em placas de ágar suplementadas com FOA, que em cepas *pyrF*-positivas, são convertidas em um composto tóxico. C. A cepa de deleção cromossômica foi confirmada por análise de seqüenciação.

20 Figura 14 é um gráfico de fluorescência relativa com o passar do tempo para medições de atividade de fluorescência verde para as cepas que expressam a proteína de fusão de hGH::COP usando um fluorímetro. Duplicatas foram usadas tanto para cepa de deleção do tipo selvagem (HJ104) como de *hslUV* (HJ117).

25 Figura 15 são imagens de géis de SDS-PAGE de cepas que expressam hGH com ou sem moduladores de dobramentos GrpE-DnakJ. As amostras foram removidas em vários tempos após indução por IPTG (0, 4, 8, 24 e 48 h), normalizadas em OD600 de 20 e lisadas usando EasyLyse. As frações solúveis (S) insolúveis (I) foram separadas em um Critério de BioRad 30 15% de Tris HCl gel de SDS-PAGE e tingido com Coomassie.

#### DESCRÍÇÃO DETALHADA

Um processo é fornecido para melhorar a expressão de uma

proteína ou peptídeo recombinante compreendendo i) expressar a proteína ou peptídeo recombinante em uma célula hospedeira; ii) analisar um perfil genético da célula e identificar um ou mais produtos de gene sobre-regulados endógenos, incluindo uma ou mais proteases ou moduladores de dobramentos que são sobre-regulados na expressão da proteína ou peptídeo recombinante; e iii) alterar a expressão de um ou mais produtos de gene identificados geneticamente modificando a célula. Em outra modalidade, o processo também compreende expressar a proteína ou peptídeo recombinante em uma célula geneticamente modificada. Em outra modalidade, o processo 5 também compreende analisar um segundo perfil genético da célula geneticamente modificada para identificar um ou mais produtos de gene adicionais que são diferencialmente expressos na célula modificada. Em uma outra modalidade, o processo compreende alterar a expressão de pelo menos um segundo produto de gene identificado na célula modificada para fornecer uma 10 célula modificada dupla. O processo pode fornecer expressão melhorada como medido por rendimentos melhorados de proteína, ou pode melhorar a recuperação de proteína ativa, por exemplo aumentando solubilidade da 15 proteína recombinante expressa.

Mais em geral, a invenção inclui um processo para melhorar a expressão de uma proteína ou peptídeo recombinante em uma célula ou organismo hospedeiro compreendendo:

- i) expressar a proteína ou peptídeo recombinante na célula ou organismo hospedeiro recombinante;
- ii) analisar um perfil genético da célula recombinante para identificar 25 um gene compensatório ou produto de gene que é expressado em um nível mais alto na célula recombinante que em um de uma célula hospedeira que não foi modificada para expressar a proteína recombinante ou uma célula recombinante que não está expressando a proteína recombinante; e
- iii) alterar a expressão do gene compensatório ou produto de gene 30 identificado na célula recombinante através de modificação genética para fornecer uma célula recombinante modificada que alcança um aumento na expressão, atividade ou solubilidade de proteína recombinante.

Ao longo do relatório descritivo, quando uma faixa for fornecida, deveria ser entendido que os componentes são significados ser independentes. Por exemplo, uma faixa de 1-6 significa independentemente 1, 2, 3, 4, 5 ou 6.

5 As etapas do processo são descritas em mais detalhe abaixo.

ETAPA I: MODIFICAÇÃO GENÉTICA DE CÉLULA OU ORGANISMO HOSPEDEIRO PARA EXPRESSAR UMA PROTEÍNA OU PEPTÍDEO RE-COMBINANTE EM UMA CÉLULA HOSPEDEIRA

Na primeira etapa do processo, uma célula hospedeira é modificada para ter a capacidade para expressar uma proteína ou peptídeo recombinante. A célula hospedeira pode ser modificada usando qualquer técnica conhecida na técnica. Por exemplo, a proteína recombinante pode ser expressada de um vetor de expressão que é exógeno ao genoma da célula e que é transfeccionado ou transformado na célula. A construção de vetores de expressão como também técnicas para transfecção ou transformação é descrita abaixo. A célula hospedeira pode também ser modificada para expressar uma proteína ou peptídeo recombinante de uma inserção genômica como descrito abaixo. Um gene que codifica a proteína ou peptídeo recombinante pode ser inserido no genoma da célula ou organismo hospedeiro através de técnicas como recombinação homóloga ou heteróloga. Estas técnicas são descritas abaixo.

A proteína ou peptídeo recombinante pode ser expressado sob o controle de um elemento que requer manipulação adicional da célula. Por exemplo, tratamento químico da célula pode ser requerido para iniciar ou intensificar a expressão de proteína ou de peptídeo. Os elementos de promotor e de repressor que governam a expressão de proteínas ou peptídeos recombinantes nas células hospedeiras são descritos abaixo e são bem conhecidos na técnica. Estes podem incluir elementos de promotor com base no promotor "lac" responsivo a IPTG.

30 SELEÇÃO DE UMA CÉLULA OU ORGANISMO HOSPEDEIRO

O processo da invenção pode ser usado em qualquer sistema hospedeiro dado, incluindo de origem eucariótica ou procariótica. O processo

é em geral limitado apenas pela disponibilidade de informação bastante genética para análise de um perfil genético para identificar um gene identificado. Embora seja em geral típico que as seqüências representativas de uma porcentagem grande do genoma estão disponíveis, por exemplo pelo menos 5 50%, 60%, 70%, 80%, 90%, 95%, 98%, 99% ou 100% das seqüências expressas ou encontradas no genoma, transcriptoma ou proteoma, a invenção pode ser praticada usando apenas uma porção das seqüências no genoma, transcriptoma ou proteoma. Em particular, em circunstâncias quando a informação disponível incluir informação sobre um grupo de seqüências relacionadas, como um grupo metabolicamente ligado, apenas uma porção pequena das seqüências representativas do genoma pode ser usada para o processo da invenção. O processo é também não limitado às proteínas recombinantes particulares sendo expressadas, como um aspecto fundamental 10 do processo a capacidade é racional e iterativamente para sistemas de expressão de projeto com base em técnicas para identificar alterações celulares que ocorrem em uma célula hospedeira na expressão de proteínas ou peptídeos recombinantes e para modular a célula hospedeira usando procedimentos conhecidos na técnica.

A célula hospedeira pode ser qualquer célula capaz de produzir 20 proteína ou peptídeo recombinante. Em uma modalidade, a célula hospedeira é uma célula microbiana, isto é, uma célula de uma bactéria, fungo, levedura ou outros eucariotas unicelulares, procariotos e vírus. Os sistemas comumente usados para produzir proteínas ou peptídeos recombinantes incluem certas células bacterianas, particularmente *E. coli*, por causa de seus 25 requerimentos de crescimento relativamente baratos e capacidade de potencial para produzir proteína em culturas de batelada grandes. Levedura é também usada para expressar proteínas e peptídeos biologicamente relevantes, particularmente para propósitos de pesquisa. Sistemas incluem *Saccharomyces cerevisiae* ou *Pichia pastoris*. Estes sistemas são bem caracterizados, fornecem em geral níveis aceitáveis de expressão de proteína total e são 30 comparativamente rápidos e baratos. Sistemas de expressão de célula de inseto também emergiram como uma alternativa para expressar proteínas

recombinantes em forma biologicamente ativa. Em alguns casos, proteínas corretamente dobradas que são pós-translacionalmente modificadas podem ser produzidas. Sistemas de expressão de célula mamífera, como células de ovário de hamster chinês, também foram usados para a expressão de proteínas recombinantes. Em uma escala pequena, estes sistemas de expressão são freqüentemente eficazes. Certos biológicos podem ser derivados de proteínas mamíferas, particularmente em aplicações de saúde animal ou humana. Em outra modalidade, a célula hospedeira é uma célula de planta, incluindo, mas não limitada a, uma célula de tabaco, milho, uma célula de uma espécie de *Arabidopsis*, célula de batata ou de arroz. Em outra modalidade, um organismo multicelular é analisado ou modificado no processo, incluindo mas não limitado a um organismo transgênico. Técnicas para analisar e/ou modificar um organismo multicelular são em geral com base nas técnicas descritas para modificar as células descritas abaixo.

Em uma modalidade, a célula hospedeira pode ser um procarionto como uma célula bacteriana incluindo, mas não limitada a uma espécie de *Escherichia* ou *Pseudomonas*. Células bacterianas típicas são descritas, por exemplo, em "Biological Diversity: Bacteria and Archaeans", a chapter of the On-Line Biology Book, fornecido por Dr MJ Farabee do Estrella Mountain Community College, Arizona, USA em URL: [http://www.emc.maricopa.edu/faculty/farabee/BIOBK/BioBookDiversity\\_2.html](http://www.emc.maricopa.edu/faculty/farabee/BIOBK/BioBookDiversity_2.html).

Em certas modalidades, a célula hospedeira pode ser uma célula de *Pseudomonas*, e pode tipicamente ser uma célula de *P. fluorescens*. Em outras modalidades, a célula hospedeira pode também ser uma célula de *E. coli*. Em outra modalidade a célula hospedeira pode ser uma célula eucariótica, por exemplo uma célula de inseto, incluindo mas não limitada a uma célula de uma espécies de *Spodoptera*, *Trichoplusia*, *Drosophila* ou *Estigmene*, ou uma célula mamífera, incluindo mas não limitada a uma célula murina, uma célula de hamster, um macaco, um primata ou uma célula humana.

Em certas modalidades, a célula hospedeira é uma célula de *Pseudomonad*, e pode ser por exemplo um organismo de *P. fluorescens*.

Em uma modalidade, a célula hospedeira pode ser um membro

de qualquer da taxa bacteriana. A célula pode, por exemplo, ser um membro de qualquer espécie de eubactéria. O hospedeiro pode ser um membro qualquer um da taxa: *Acidobacteria*, *Actinobacteria*, *Aquificae*, *Bacteroidetes*, *Chlorobi*, *Chlamydiae*, *Choroflexi*, *Chrysogenetes*, *Cyanobacteria*, *Deferribacteres*, *Deinococcus*, *Dictyoglomi*, *Fibrobacteres*, *Firmicutes*, *Fusobacteria*, *Gemmatimonadetes*, *Lentisphaerae*, *Nitrospirae*, *Planctomycetes*, *Proteobacteria*, *Spirochaetes*, *Thermodesulfobacteria*, *Thermomicrobia*, *Thermotogae*, *Thermus (Thermales)* ou *Verrucomicrobia*. Em uma modalidade de uma célula hospedeira eubacteriana, a célula pode ser um membro de qualquer espécie de eubactéria, excluindo *Cyanobacteria*.

O hospedeiro bacteriano pode também ser um membro de qualquer espécie de *Proteobacteria*. Uma célula hospedeira proteobacteriana pode ser um membro de qualquer um da taxa *Alphaproteobacteria*, *Betaproteobacteria*, *Gammaproteobacteria*, *Deltaproteobacteria* ou *Epsilonproteobacteria*. Além disso, o hospedeiro pode ser um membro de qualquer um da taxa *Alphaproteobacteria*, *Betaproteobacteria*, or *Gammaproteobacteria*, e um membro de qualquer espécie de *Gammaproteobacteria*.

Em uma modalidade de um hospedeiro *Gammaproteobacteriano*, o hospedeiro será o membro de qualquer um da taxa *Aeromonadales*, *Alteromonadales*, *Enterobacteriales*, *Pseudomonadales* ou *Xanthomonadales*; ou um membro de qualquer espécie dos *Enterobacteriales* ou *Pseudomonadales*. Em uma modalidade, a célula hospedeira pode ser da ordem *Enterobacteriales*, a célula hospedeira será um membro da família *Enterobacteriaceae*, ou um membro de qualquer um dos gêneros *Erwinia*, *Escherichia* ou *Serratia*; ou um membro do gênero *Escherichia*. Em uma modalidade de uma célula hospedeira da ordem *Pseudomonadales*, a célula hospedeira será um membro da família *Pseudomonadaceae*, até mesmo do gênero *Pseudomonas*. Hospedeiros de proteobacterianos gama incluem os membros das espécies *Escherichia coli* e membros das espécies *Pseudomonas fluorescens*.

Outros organismos de *Pseudomonas* podem também ser úteis. *Pseudomonads* e espécies estritamente relacionadas incluem *Proteobacteria Gram (-) Subgrupo 1*, que inclui o grupo de *Proteobacteria* que pertence às

famílias e/ou gêneros descritos como "Gram-Negative Aerobic Rods and Coccii" por R.E. Buchanan e N.E. Gibbons (eds.), Bergey's Manual of Determinative Bacteriology, págs. 217-289 (8<sup>a</sup> ed., 1974) (The Williams & Wilkins Co., Baltimore, MD, USA) (doravante "Bergey (1974)"). A tabela a seguir  
5 apresenta estes famílias e gêneros de organismos.

As famílias e gêneros listados na Parte, "Gram-Negative Aerobic Rods and Coccii" (em Bergey (1974))	
Família I. Pseudomonadaceae	<i>Gluconobacter</i> <i>Pseudomonas</i> <i>Xanthomonas</i> <i>Zoogloea</i>
Família II. Azotobacteraceae	<i>Azomonas</i> <i>Azotobacter</i> <i>Beijerinckia</i> <i>Dexria</i>
Família III. Rhizobiaceae	<i>Agrobacterium</i> <i>Rhizobium</i>
Família IV. Methylomonadaceae	<i>Methylococcus</i> <i>Methylomonas</i>
Família V. Halobacteriaceae	<i>Halobacterium</i> <i>Halococcus</i>
Outros gêneros	<i>Acetobacter</i> <i>Alcaligenes</i> <i>Bordetella</i> <i>Brucella</i> <i>Francisella</i> <i>Thermus</i>

"Proteobacteria Gram(-) Subgrupo 1" também inclui Proteobacte-  
ria que seria classificada neste título de acordo com os critérios usados na

classificação. O título também inclui grupos que foram previamente classificados nesta seção mas não são mais, como os gêneros *Acidovorax*, *Brevundimonas*, *Burkholderia*, *Hydrogenophaga*, *Oceanimonas*, *Ralstonia* e *Stenotrophomonas*, o gênero *Sphingomonas* (e o gênero *Blastomonas*, dele derivado), que foi criado reagrupando organismos pertencentes (e previamente chamados espécie de) ao gênero *Xanthomonas*, ao gênero *Acidomonas*, que foi criado reagrupando os organismos pertencentes ao gênero *Acetobacter* conforme definido em Bergey (1974). Além disso, hospedeiros podem incluir células do gênero *Pseudomonas*, *Pseudomonas enalia* (ATCC 5 14393), *Pseudomonas nigrifaciens* (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) foram reclassificados 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*. "Proteobacteria Gram(-) Subgrupo 1" também inclui Proteobacteria classificadas como pertencentes a qualquer 10 uma das famílias: Pseudomonadaceae, Azotobacteraceae (agora freqüentemente chamada pelo sinônimo, o "grupo de Azotobacter" de Pseudomonadaceae), Rhizobiaceae e Methylomonadaceae (agora freqüentemente chamada pelo sinônimo, "Methylococcaceae"). Conseqüentemente, além daqueles gêneros do contrário descritos aqui, outros gêneros de Proteobacterial caindo dentro de "Proteobacteria Gram(-) Subgrupo 1" incluem: 1) 15 bactérias do grupo de Azotobacter do gênero *Azorhizophilus*; bactérias da família Pseudomonadaceae dos gêneros *Cellvibrio*, *Oligella* e *Teredinibacter*; 2) bactérias da família Rhizobiaceae dos gêneros *Chelatobacter*, *Ensifer*, *Liberibacter* (também chamado "*Candidatus Liberibacter*") e *Sinorhizobium*; e 20 3) bactérias da família Methylococcaceae dos gêneros *Methylobacter*, *Methylocaldum*, *Methylomicrobium*, *Methylosarcina* e *Methylosphaera*.

Em outra modalidade, a célula hospedeira é selecionada de

"Proteobacteria Gram (-) Subgrupo 2". "Proteobacteria Gram (-) Subgrupo 2" é definido como o grupo de Proteobacteria dos gêneros a seguir (com os números totais das cepas depositadas publicamente disponíveis listadas no catálogo destes indicadas entre parêntese, todas depositadas na ATCC, exceto como do contrário indicado): Acidomonas (2); *Acetobacter* (93); *Gluconobacter* (37); *Brevundimonas* (23); *Beijerinckia* (13); *Dexia* (2); *Brucella* (4); *Agrobacterium* (79); *Chelatobacter* (2); *Ensifer* (3); *Rhizobium* (144); *Sinorhizobium* (24); *Blastomonas* (1); *Sphingomonas* (27); *Alcaligenes* (88); *Bordetella* (43); *Burkholderia* (73); *Ralstonia* (33); *Acidovorax* (20); *Hydrogenophaga* (9); *Zoogloea* (9); *Methylobacter* (2); *Methylocaldum* (1 em NCIMB); *Methylococcus* (2); *Methylomicrobium* (2); *Methylomonas* (9); *Methylosarcina* (1); *Methylosphaera*; *Azomonas* (9); *Azorhizophilus* (5); *Azotobacter* (64); *Cellvibrio* (3); *Oligella* (5); *Pseudomonas* (1139); *Francisella* (4); *Xanthomonas* (229); *Stenotrophomonas* (50); e *Oceanimonas* (4).

15 Espécies de célula hospedeira exemplares de "Proteobacteria Gram (-) Subgrupo 2" incluem, mas não são limitadas às bactérias a seguir (com os números de depósito de ATCC ou outros de cepa(s) exemplar(es) destas mostrados entre parêntese): *Acidomonas methanolica* (ATCC 43581); *Acetobacter aceti* (ATCC 15973); *Gluconobacter oxidans* (ATCC 19357);  
20 *Brevundimonas diminuta* (ATCC 11568); *Beijerinckia indica* (ATCC 9039 e ATCC 19361); *Dexia gumrnosa* (ATCC 15994); *Brucella melitensis* (ATCC 23456); *Brucella abortus* (ATCC 23448); *Agrobacterium tumefaciens* (ATCC 23308), *Agrobacterium radiobacter* (ATCC 19358), *Agrobacterium rhizogenes* (ATCC 11325); *Chelatobacter heintzii* (ATCC 29600); *Ensifer adhaerens*  
25 (ATCC 33212); *Rhizobium leguminosarum* (ATCC 10004); *Sinorhizobium fredii* (ATCC 35423); *Blastomonas natatoria* (ATCC 35951); *Sphingomonas paucimobilis* (ATCC 29837); *Alcaligenes faecalis* (ATCC 8750); *Bordetella pertussis* (ATCC 9797); *Burkholderia cepacia* (ATCC 25416); *Ralstonia pickettii* (ATCC 27511); *Acidovorax facilis* (ATCC 11228); *Hydrogenophaga flava*  
30 (ATCC 33667); *Zoogloea ramigera* (ATCC 19544); *Methylobacter luteus* (ATCC 49878); *Methylocaldum gracile* (NCIMB 11912); *Methylococcus capsulatus* (ATCC 19069); *Methylomicrobium agile* (ATCC 35068); *Methylomo-*

*nas methanica* (ATCC 35067); *Methylosarcina fibrata* (ATCC 700909); *Methylosphaera hansonii* (ACAM 549); *Azomonas agilis* (ATCC 7494); *Azorhizophilus paspali* (ATCC 23833); *Azotobacter chroococcum* (ATCC 9043); *Cellvibrio mixtus* (UQM 2601); *Oligella urethralis* (ATCC 17960); *Pseudomonas aeruginosa* (ATCC 10145), *Pseudomonas fluorescens* (ATCC 35858); *Francisella tularensis* (ATCC 6223); *Stenotrophomonas maltophilia* (ATCC 13637); *Xanthomonas campestris* (ATCC 33913); e *Oceanimonas doudoroffii* (ATCC 27123).

Em outra modalidade, a célula hospedeira é selecionada de "Proteobacteria Gram(-) Subgrupo 3". "Proteobacteria Gram(-) Subgrupo 3" é definida como o grupo de Proteobacteria dos seguintes gêneros: *Brevundimonas*; *Agrobacterium*; *Rhizobium*; *Sinorhizobium*; *Blastomonas*; *Sphingomonas*; *Alcaligenes*; *Burkholderia*; *Ralstonia*; *Acidovorax*; *Hydrogenophaga*; *Methylobacter*; *Methylocaldum*; *Methylococcus*; *Methylomicrobium*; *Methylomonas*; *Methylosarcina*; *Methylosphaera*; *Azomonas*; *Azorhizophilus*; *Azotobacter*; *Cellvibrio*; *Oligella*; *Pseudomonas*; *Teredinibacter*; *Francisella*; *Stenotrophomonas*; *Xanthomonas*; e *Oceanimonas*.

Em outra modalidade, a célula hospedeira é selecionada de "Proteobacteria Gram(-) Subgrupo 4". "Proteobacteria Gram(-) Subgrupo 4" é definida como o grupo de Proteobacteria dos seguintes gêneros: *Brevundimonas*; *Blastomonas*; *Sphingomonas*; *Burkholderia*; *Ralstonia*; *Acidovorax*; *Hydrogenophaga*; *Methylobacter*; *Methylocaldum*; *Methylococcus*; *Methylomicrobium*; *Methylomonas*; *Methylosarcina*; *Methylosphaera*; *Azomonas*; *Azorhizophilus*; *Azotobacter*; *Cellvibrio*; *Oligella*; *Pseudomonas*; *Teredinibacter*, *Francisella*; *Stenotrophomonas*; *Xanthomonas*; e *Oceanimonas*.

Em outra modalidade, a célula hospedeira é selecionada de "Proteobacteria Gram(-) Subgrupo 5". "Proteobacteria Gram(-) Subgrupo 5" é definida como o grupo de Proteobacteria dos seguintes gêneros: *Methylobacter*; *Methylocaldum*; *Methylococcus*; *Methylomicrobium*; *Methylomonas*; *Methylosarcina*; *Methylosphaera*; *Azomonas*; *Azorhizophilus*; *Azotobacter*; *Cellvibrio*; *Oligella*; *Pseudomonas*; *Teredinibacter*; *Francisella*; *Stenotrophomonas*; *Xanthomonas*; e *Oceanimonas*.

A célula hospedeira pode ser selecionada de "Proteobacteria Gram(-) Subgrupo 6". "Proteobacteria Gram(-) Subgrupo 6" é definida como o grupo de Proteobacteria dos seguintes gêneros: *Brevundimonas*; *Blastomonas*; *Sphingomonas*; *Burkholderia*; *Ralstonia*; *Acidovorax*; *Hydrogenophaga*; 5 *Azomonas*; *Azorhizophilus*; *Azotobacter*; *Cellvibrio*; *Oligella*; *Pseudomonas*; *Teredinibacter*; *Stenotrophomonas*; *Xanthomonas*; e *Oceanimonas*.

A célula hospedeira pode ser selecionada de "Proteobacteria Gram(-) Subgrupo 7". "Proteobacteria Gram(-) Subgrupo 7" é definida como o grupo de Proteobacteria dos seguintes gêneros: *Azomonas*; *Azorhizophilus*; 10 *Azotobacter*; *Cellvibrio*; *Oligella*; *Pseudomonas*; *Teredinibacter*; *Stenotrophomonas*; *Xanthomonas*; e *Oceanimonas*.

A célula hospedeira pode ser selecionada de "Proteobacteria Gram(-) Subgrupo 8". "Proteobacteria Gram(-) Subgrupo 8" é definida como o grupo de Proteobacteria dos seguintes gêneros: *Brevundimonas*; *Blastomonas*; *Sphingomonas*; *Burkholderia*; *Ralstonia*; *Acidovorax*; *Hydrogenophaga*; 15 *Pseudomonas*; *Stenotrophomonas*; *Xanthomonas*; e *Oceanimonas*.

A célula hospedeira pode ser selecionada de "Proteobacteria Gram(-) Subgrupo 9". "Proteobacteria Gram(-) Subgrupo 9" é definida como o grupo de Proteobacteria dos seguintes gêneros: *Brevundimonas*; *Burkholderia*; *Ralstonia*; *Acidovorax*; *Hydrogenophaga*; *Pseudomonas*; *Stenotrophomonas*; e *Oceanimonas*. 20

A célula hospedeira pode ser selecionada de "Proteobacteria Gram(-) Subgrupo 10". "Proteobacteria Gram(-) Subgrupo 10" é definida como o grupo de Proteobacteria dos seguintes gêneros: *Burkholderia*; *Ralstonia*; 25 *Pseudomonas*; *Stenotrophomonas*; e *Xanthomonas*.

Em outra modalidade, a célula hospedeira é selecionada de "Proteobacteria Gram(-) Subgrupo 11". "Proteobacteria Gram(-) Subgrupo 11" é definida como o grupo de Proteobacteria dos seguintes gêneros: *Pseudomonas*; *Stenotrophomonas*; e *Xanthomonas*. A célula hospedeira pode ser 30 selecionada de "Proteobacteria Gram(-) Subgrupo 12". "Proteobacteria Gram(-) Subgrupo 12" é definida como o grupo de Proteobacteria dos seguintes gêneros: *Burkholderia*; *Ralstonia*; *Pseudomonas*. A célula hospedeira

pode ser selecionada de "Proteobacteria Gram(-) Subgrupo 13". "Proteobacteria Gram(-) Subgrupo 13" é definida como o grupo de Proteobacteria dos seguintes gêneros: *Burkholderia*; *Ralstonia*; *Pseudomonas*; e *Xanthomonas*.

A célula hospedeira pode ser selecionada de "Proteobacteria Gram(-) Sub-

5 grupo 14". "Proteobacteria Gram(-) Subgrupo 14" é definida como o grupo de Proteobacteria dos seguintes gêneros: *Pseudomonas* e *Xanthomonas*. A célula hospedeira pode ser selecionada de "Proteobacteria Gram(-) Subgrupo 15". "Proteobacteria Gram(-) Subgrupo 15" é definida como o grupo de Proteobacteria do gênero *Pseudomonas*.

10 A célula hospedeira pode ser selecionada de "Proteobacteria Gram(-) Subgrupo 16". "Proteobacteria Gram(-) Subgrupo 16" é definida como o grupo de Proteobacteria das seguintes espécies de *Pseudomonas* (com os

números de depósito da ATCC ou outros de cepa(s) exempla(es) mostrada(s) entre parênteses): *Pseudomonas abietaniphila* (ATCC 700689); *Pseudomonas*

15 *aeruginosa* (ATCC 10145); *Pseudomonas alcaligenes* (ATCC 14909); *Pseudomonas anguilliseptica* (ATCC 33660); *Pseudomonas citronellolis* (ATCC 13674); *Pseudomonas fluorescens* (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*

*alcaliplaila*; *Pseudomonas alginovora*; *Pseudomonas andersonii*; *Pseudomonas asplenii* (ATCC 23835); *Pseudomonas azelaica* (ATCC 27162); *Pseudomonas beijerinckii* (ATCC 19372); *Pseudomonas borealis*; *Pseudomonas*

20 *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*

25 (ATCC 4683); *Pseudomonas cissicola* (ATCC 33616); *Pseudomonas coronafaciens*; *Pseudomonas diterpeniphila*; *Pseudomonas elongata* (ATCC 10144); *Pseudomonas flectens* (ATCC 12775); *Pseudomonas azotoformans*;

*Pseudomonas brenneri*; *Pseudomonas cedrella*; *Pseudomonas corrugata* (ATCC 29736); *Pseudomonas exagitaçaoientalis*; *Pseudomonas fluorescens* (ATCC 35858); *Pseudomonas gessardii*; *Pseudomonas libanensis*; *Pseudomonas mandelii* (ATCC 700871); *Pseudomonas marginalis* (ATCC 10844);  
5     *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 grimonii*; *Pseudomonas halodenitrificans*; *Pseudomonas halophila*; *Pseudomonas hibiscicola* (ATCC 19867); *Pseudomonas huttiensis* (ATCC 14670);  
10    *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);  
15    *Pseudomonas psychrophila*; *Pseudomonas fulva* (ATCC 31418); *Pseudomonas monteilii* (ATCC 700476); *Pseudomonas mosselii*; *Pseudomonas oryzihabitans* (ATCC 43272); *Pseudomonas plecoglossicida* (ATCC 700383);  
20    *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);  
25    *Pseudomonas viridiflava* (ATCC 13223); *Pseudomonas thermocarboxidovorans* (ATCC 35961); *Pseudomonas thermotolerans*; *Pseudomonas thivervalensis*; *Pseudomonas vancouverensis* (ATCC 700688); *Pseudomonas wisconsinensis*; e *Pseudomonas xiamenensis*.  
30

A célula hospedeira pode ser selecionada da "Proteobacteria Gram(-) Subgrupo 17". "Proteobacteria Gram(-) Subgrupo 17" é definida como

o grupo de Proteobacteria conhecida na técnica como as "Pseudomonads fluorescentes" incluindo aquelas pertencentes, por exemplo, às seguintes espécies de *Pseudomonas*: *Pseudomonas azotofrmans*; *Pseudomonas brenneri*; *Pseudomonas cedrella*; *Pseudomonas corrugata*; *Pseudomonas exagitaçãoorientalis*; *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*.

A célula hospedeira pode ser selecionada de "Proteobacteria Gram(-) Subgrupo 18". "Proteobacteria Gram(-) Subgrupo 18" é definida como o grupo de todas as subespécies, variedades, cepas e outras unidades sub-especiais das espécies *Pseudomonas fluorescens*, incluindo aquelas pertencentes, por exemplo, às seguintes (com os números de depósito de ATCC ou outros de cepa(s) exemplar(es) mostrados entre parênteses): *Pseudomonas fluorescens* biotipo A, também chamada biovar 1 ou biovar I (ATCC 13525); *Pseudomonas fluorescens* biotipo B, também chamada biovar 2 ou biovar II (ATCC 17816); *Pseudomonas fluorescens* biotipo C, também chamada biovar 3 ou biovar III (ATCC 17400); *Pseudomonas fluorescens* biotipo F, também chamada biovar 4 ou biovar IV (ATCC 12983); *Pseudomonas fluorescens* biotipo G, também chamada biovar 5 ou biovar V (ATCC 17518); *Pseudomonas fluorescens* biovar VI; *Pseudomonas fluorescens* Pf0-1; *Pseudomonas fluorescens* Pf-5 (ATCC BAA-477); *Pseudomonas fluorescens* SBW25; and *Pseudomonas fluorescens* subsp. *cellulosa* (NCIMB 10462).

A célula hospedeira pode ser selecionada de "Proteobacteria Gram (-) Subgrupo 19". "Proteobacteria Gram (-) Subgrupo 19" é definida como o grupo de todas as cepas de *Pseudomonas fluorescens* biotipo A. Uma cepa típica deste biotipo é a cepa de *P. fluorescens* MB101 (ver patente US Nº 5.169.760 de Wilcox), e derivados desta. Um exemplo de um derivado desta é a cepa de *P. fluorescens* MB214, construída inserindo no lócus de asd cromossômico de MB 101 (gene de aspartato desidrogenase), uma construção

nativo de *E. coli* de Placl-lacl-lacZYA (isto é em que PlacZ foi deletado).

Cepas de *P. fluorescens* adicionais que podem ser usadas na presente invenção incluem *Pseudomonas fluorescens* Migula e *Pseudomonas fluorescens* Loitokitok, enquanto tendo as designações de ATCC a seguir:

5 A célula hospedeira pode ser selecionada de "Proteobacteria Gram(-) Subgrupo 19". "Proteobacteria Gram(-) Subgrupo 19" é definida como o grupo de todas as cepas de *Pseudomonas fluorescens* biotipo A. Uma cepa típica A típica deste biotipo é cepa de *P. fluorescens* MB101 (ver Patente US Nº 5.169.760 de Wilcox), e derivados destes. Um exemplo de um derivado  
10 deste é a cepa de *P. fluorescens* MB214, construída inserindo dentro do lócus de asd cromossômico de MB 101 (gene de aspartato desidrogenase), a construção nativa de P1acl-lacl-lacZYA de *E. coli* (isto é em que PlacZ foi deletado).

Cepas adicionais de *P. Fluorescens* que podem ser usadas na  
15 presente invenção incluem *Pseudomonas fluorescens* Migula e *Pseudomonas fluorescens* Loitokitok, tendo as seguintes designações da ATCC: [NCIB 8286]; NRRL B-1244; NCIB 8865 cepa CO1; NCIB 8866 cepa CO2; 1291 [ATCC 17458]; IFO 15837; NCIB 8917; LA; NRRL B-1864; pirrolidina; PW2 [ICMP 3966; NCPPB 967; NRRL B-899]; 13475; NCTC 10038; NRRL B1603  
20 [6; IFO 15840]; 52-1C; CCEB 488-A [BU 140]; CCEB 553 [IEM 15/47]; IAM 1008 [A1-H-27]; IAM 1055 [AHH-23]; 1 [IFO 15842]; 12 [ATCC 25323; NTH 11; den Dooren de Jong 216]; 18 [IFO 15833; WRRL P-7]; 93 [TR-10]; 108 [52-22; IFO 15832]; 143 [IFO 15836; PL]; 149 [2-40-40; IFO 15838]; 182 [IFO 3081; PJ 73]; 184 [IFO 15830]; 185 [W2 LA]; 186 [IFO 15829; PJ 79]; 187 [NCPPB  
25 263]; 188 [NCPPB 316]; 189 [PJ227; 1208]; 191 [IFO 15834; PJ 236; 22/1]; 194 [Klinge R-60; PJ 253]; 196 [PJ 288]; 197 [PJ 290]; 198 [PJ 302]; 201 [PJ 368]; 202 [PJ 372]; 203 [PJ 376]; 204 [IFO 15835; PJ 682]; 205 [PJ 686]; 206 [PJ 692]; 207 [PJ 693]; 208 [PJ 722]; 212 [PJ 832]; 215 [PJ 849]; 216 [PJ 885]; 267 [B-9]; 271 [B-1612]; 401 [C71A; IFO 15831; PJ 187]; NRRL B-3178 [4; IFO  
30 15841]; KY 8521; 3081; 30-21; [IFO 3081]; N; PYR; PW; D946-B83 [BU 2183; FERM-P 3328]; P-2563 [FERM-P 2894; IFO 13658]; IAM-1126 [43F]; M-1; A506 [A5-06]; A505 [A5-05-1]; A526 [A5-26]; B69; 72; NRRL B-4290; PMW6

[NCIB 11615]; SC 12936; AI [IFO 15839]; F 1847 [CDC-EB]; F 1848 [CDC 93]; NCIB 10586; P17; F-12; AmMS 257; PRA25; 6133D02; 6519E01; N1; SC15208; BNL-WVC; NCTC 2583 [NCIB 8194]; H13; 1013 [ATCC 11251; CCEB 295]; IFO 3903; 1062; ou Pf-5.

5 Outros hospedeiros adequados incluem aqueles classificados em outras partes da referência, como Proteobactérias Gram (+). Em uma modalidade, a célula hospedeira é um *E. coli*. A seqüência de genoma para *E. coli* foi estabelecida para MG1655 de *E. coli* (Blattner, et al. (1997) The complete genome sequence of *Escherichia coli* K-12 *Science* 277(5331): 1453-74) e  
10 microarranjo de DNA estão comercialmente disponíveis para K12 de *E. coli* (MWG Inc, High Point, NC). *E. coli* pode ser cultivado em um meio rico como Luria-Bertani (LB) (10 g/L de triptona, 5 g/L de NaCl, 5 g/L de extrato de levedura) ou um meio mínimo definido como M9 (6 g/L de Na<sub>2</sub>HPO<sub>4</sub>, 3 g/L de KH<sub>2</sub>PO<sub>4</sub>, 1 g/L de NH<sub>4</sub>Cl, 0,5 g/L NaCl, pH 7,4) com uma fonte de carbono  
15 apropriada como 1% de glicose. Habitualmente, uma cultura durante a noite de células de *E. coli* é diluída e inoculadas em meio fresco rico ou mínimo em um frasco agitador ou um fermentador e crescida a 37°C.

Um hospedeiro pode também ser de origem mamífera, como uma célula derivada de um mamífero incluindo qualquer mamífero humano ou  
20 não-humano. Mamíferos podem incluir, mas não são limitados a primatas, macacos, porcino, ovino, bovino, roedores, ungulado, porcos, suínos, ovelha, carneiro, cabras, gado, cervos, mulas, cavalos, macacos, bugios, cachorros, gatos, ratos e camundongos.

Uma célula hospedeira pode também ser de origem de planta.  
25 Qualquer planta pode ser selecionada para a identificação de genes e seqüências reguladoras. Exemplos de alvos de planta adequados para o isolamento de genes e seqüências reguladoras incluiriam mas não limitariam a alfafa, maçã, abricó, Arabidopsis, alcachofra, rúcula, espargos, abacate, banana, cevada, feijões, beterraba, amora-preta, mirtilo, brócolos, couve-de-bruxelas, repolho, óleo de canola, cantalupo, cenoura, mandioca, mamona, couve-flor, aipo, cereja, chicória, coentro, cítrico, clementinas, trevo, coco, café, milho, algodão, oxicoco, pepino, abeto de Douglas, berinjela, en-

dívia, escarola, eucalipto, funcho, figos, alho, cabaço, uva, toronja, melão honeydew, jicama, kiwi, alface, alho-porró, limão, lima, pinheiro-amarelo, linhaça, manga, melão, cogumelo, nectarina, noz, aveia, óleo de palma, óleo de colza, quiabo, azeitona, cebola, laranja, uma planta ornamental, palma, 5 mamão, salsa, pastinaca, ervilha, pêssego, amendoim, pêra, pimenta, caqui, pinheiro, abacaxi, musa, ameixa, romã, álamo, batata, jerimum, marmelo, pinho, radicchio, rabanete, colza, framboesa, arroz, centeio, sorgo, pinheiro do Sul, feijão-soja, espinafre, abóbora, morango, beterraba-açucareira, cana-de-açúcar, girassol, batata-doce, liquidambar, mexerica, chá, tabaco, to-10 mate, triticale, grama, nabo, uma videira, melancia, trigo, inhames e abobrinha. Em algumas modalidades, plantas úteis no processo são *Arabidopsis*, milho, trigo, feijão-soja e algodão.

Para expressão de uma proteína ou peptídeo recombinante, ou para modulação de um gene compensatório identificado, qualquer promotor 15 de planta pode ser usado. Um promotor pode ser um promotor de RNA polimerase II de planta. Elementos incluídos em promotores de planta podem ser uma caixa de TATA ou caixa de Goldberg-Hogness, tipicamente posicionada aproximadamente 25 a 35 pares de base a montante (5') do sítio de iniciação de transcrição, e a caixa de CCAAT, localizada entre 70 e 100 pares 20 de base a montante. Em plantas, a caixa de CCAAT pode ter uma seqüência de consenso diferente que a seqüência funcionalmente análoga dos promotores mamíferos (Messing et al., Em: *Genetic Engineering of Plants*, Kosuge et al., eds., págs. 211-227, 1983). Virtualmente além disso, todos promotores incluem seqüências de ativação ou intensificadores adicionais a montante 25 (Benoist e Chambon, *Nature* 290:304-310, 1981; Gruss et al., *Proc. Nat. Acad. Sci. USA* 78:943-947, 1981; e Khoury e Gruss, *Cell* 27:313-314, 1983) estendendo-se por volta de -100 bp a -1.000 bp ou mais a montante do sítio de iniciação de transcrição.

#### EXPRESSÃO DE PROTEÍNA OU PEPTÍDEO RECOMBINANTE

30 Como descrito abaixo, uma célula ou organismo hospedeiro pode ser criado para expressar proteína ou peptídeo recombinante usando técnicas padrão. Por exemplo, proteína recombinante pode ser expressada de um

vetor ou de um gene exógeno inserido no genoma do hospedeiro. Vetores que podem ser usados para expressar proteínas exógenas são bem conhecidos na técnica e são descritos abaixo. Genes para expressar proteína ou peptídeo recombinante podem também ser inseridos no genoma usando técnicas como 5 recombinação homóloga ou heteróloga, como descritas abaixo.

A proteína ou peptídeo recombinante pode ser expressado após indução com um composto químico sob expressão de um gene endógeno ou de produto de gene. A proteína recombinante pode também ser expressa quando a célula hospedeira for colocada em um ambiente particular. Ele-10 mentos promotores específicos são descritos abaixo. Estes incluem, mas não são limitados a, promotores que podem ser induzidos sob tratamento da cé-  
lula com químicas como IPTG, benzoato ou antranilato.

#### PROTEÍNAS/PEPTÍDEOS RECOMBINANTES

A célula hospedeira foi projetada para expressar uma proteína ou peptídeo recombinante. Este pode ser de qualquer espécie e de qualquer tamanho. Porém, em certas modalidades, a proteína ou peptídeo recombinante é uma proteína ou peptídeo terapeuticamente útil. Em algumas modalidades, a proteína pode ser uma proteína mamífera, por exemplo uma proteína humana, e pode ser, por exemplo, um fator de crescimento, uma citocina, 15 um quimiocina ou uma proteína de sangue. A proteína ou peptídeo recombinante pode ser expressado primariamente em uma forma inativa na célula hospedeira. Em certas modalidades, a proteína ou peptídeo recombinante é menos que 100 kD, menos que 50 kD ou menos que 30 kD em tamanho. Em 20 modalidades de ceratina, a proteína ou peptídeo recombinante é um peptídeo de pelo menos 5, 10, 15, 20, 30, 40, 50 ou 100 aminoácidos.

Vetores de expressão existentes permitem a produção de proteína recombinante em *E. coli*. Para todos estes sistemas de expressão de proteína procedimentos de clonagem rotineiros como descritos mais cedo podem ser seguidos (Sambrook, et al. (2000) *Molecular cloning: A laboratory manual*, terceira edição, Cold Spring Harbor, Nova Iorque, Cold Spring Harbor Laboratory Press).

O sistema expressão Champion® pET fornece um nível alto de

produção de proteína. Expressão é induzida do promotor de T7lac forte. Este sistema tira proveito da atividade alta e especificidade da bacteriófago T7 RNA polimerase para transcrição de nível alto do gene de interesse. O operador de lac localizado na região de promotor fornece regulação mais firme

5 que os vetores baseados em T7 tradicionais, melhorando a estabilidade do plasmídeo e viabilidade da célula (Studier, F. W. e B. A. Moffatt (1986) Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes *Journal of Molecular Biology* 189(1): 113-30; Rosenberg, et al. (1987) Vectors for selective expression of cloned DNAs by T7 RNA poly-  
10 merase *Gene* 56(1): 125-35). O sistema de expressão de T7 usa o promotor de T7 e T7 RNA polimerase (T7 RNAP) para transcrição de nível alto do gene de interesse. Expressão de nível alto é alcançada em sistemas de expressão de T7 porque o T7 RNAP é mais processiva que *E. coli* RNAP nativa e é dedicada à transcrição do gene de interesse. Expressão do gene identificado é  
15 induzida fornecendo uma fonte de T7 RNAP na célula hospedeira. Isto é realizado usando um hospede de *E. coli* de BL21 contendo uma cópia cromosômica do gene de T7 RNAP. O gene de T7 RNAP está sob o controle do promotor lacUV 5 que pode ser induzido por IPTG. T7 RNAP é expressado em indução e transcreve o gene de interesse.

20 O sistema de expressão de pBAD permite expressão firmemente controlada, titulável da proteína recombinante através da presença de fontes de carbono específicas como glicose, glicerol e arabinose (Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter" *Journal of Bacteriology* 177(14): 412130). Os vetores de  
25 pBAD são projetados exclusivamente para dar controle preciso nos níveis de expressão. Expressão de gene heterólogo dos vetores de pBAD é iniciada no promotor de araBAD. O promotor é positiva e negativamente regulado pelo produto do gene de araC. AraC é um regulador transcripcional que forma um complexo com L-arabinose. Na ausência de L-arabinose, o dímero de AraC  
30 bloqueia a transcrição. Para ativação transcripcional máxima dois eventos são requeridos: (i.) L-arabinose ligue a AraC permitindo a transcrição começar. (ii.) O complexo de proteína de ativador de cAMP (CAP)-cAMP liga ao DNA e

estimula a ligação de AraC na localização correta da região de promotor.

O sistema de expressão de *trc* permite nível alto, expressão regulada em *E. coli* do promotor de *trc*. Os vetores de expressão de *trc* foram otimizados para expressão de genes eucarióticos em *E. coli*. O promotor de

- 5 *trc* é um promotor híbrido forte derivado dos promotores triptofano (*trp*) e lactose (*lac*). Ele é regulado pelo operador de *lacO* e o produto do gene de *lac<sup>Q</sup>* (Brosius, J. (1984) Toxicity of an overproduced foreign gene product in *Escherichia coli* and its use in plasmid vectors for the selection of transcription terminators *Gene* 27(2): 161-72).

10 A invenção também inclui a célula hospedeira recombinante melhorada que é produzida pelo processo reivindicado. Em uma modalidade, a invenção inclui uma célula produzida pelo processo descrito. Em outra modalidade, a invenção inclui uma célula ou organismo hospedeiro que ex-

pressa uma proteína recombinante que foi geneticamente modificada para 15 reduzir a expressão de pelo menos duas proteases. Em outras modalidades, a invenção inclui uma célula ou organismo hospedeiro que expressa uma proteína recombinante que foi geneticamente modificada para reduzir a ex-

pressão de pelo menos uma protease selecionada do grupo que consiste em 20 produtos de genes de *hsIV*, *hsIU*, *cipX*, *cipA* e *cipB*, e em certas submoda-

lides, a célula ou organismo foi modificado para reduzir a expressão de *HsIV* 25 ou *HsIU*. Em certas modalidades, a célula ou organismo hospedeiro modifi- cado expressa uma proteína derivada mamífera recombinante, e pode ex-

pressar proteína derivada que pode ser hormônio de crescimento humano para um humano recombinante. A célula pode ser modificada por qualquer

25 técnica conhecida na técnica, por exemplo por uma técnica em que pelo menos um gene de protease é knock-out do genoma, ou mutando pelo menos um gene de protease para reduzir a expressão de uma protease, ou alterando pelo menos um promotor de pelo menos um gene de protease para reduzir expressão de uma protease.

30 Em outra modalidade, hospedeiro ou organismo que expressam uma proteína recombinante que é apresentada que foi geneticamente modi- ficada para aumentar a expressão de pelo menos um, pelo menos dois mo-

duladores de dobramentos, ou pelo menos três moduladores de dobramentos. Em certas submodalidades, os moduladores de dobramentos não são subunidades de modulador de dobramentos. O modulador de dobramento pode ser selecionado do grupo que consiste em produtos de genes de *cbpA*, *htpG*, 5 *dnaK*, *dnaJ*, *fkbP2*, *groES* e *groEL*, e, em certas submodalidades, pode ser *htpG* ou *cbpA*. A célula ou organismo hospedeiro podem, em um exemplo não-limitativo, expressar uma proteína mamífera, como uma proteína humana. A proteína pode ser hormônio de crescimento humano. O modulador ou moduladores de dobramento podem ser aumentados, por exemplo, incluindo 10 um vetor de expressão como descrito aqui na célula. A expressão de modulador de dobramento pode também ser aumentada, por exemplo, mutando um promotor de um modulador de dobramento ou subunidade de modulador de dobramento. Uma célula ou organismo hospedeiro que expressa uma proteína recombinante pode também ser geneticamente modificado para aumentar 15 a expressão de pelo menos um modulador de dobramento e diminui a expressão de pelo menos uma protease ou proteína de protease. Organismos compreendendo uma ou mais células produzidas pelo processo descrito estão também inclusos na invenção.

ETAPA II: ANÁLISE DE UM PERFIL GENÉTICO PARA IDENTIFICAR UM  
20 GENE OU PRODUTO DE GENE COMPENSATÓRIO QUE É EXPRESSADO  
EM UM NÍVEL MAIS ALTO NA CÉLULA RECOMBINANTE

O processo da invenção inclui analisar um perfil genético da célula recombinante para identificar um gene compensatório ou produto de gene que são expressados a um nível mais alto na célula recombinante que em ou 25 uma célula hospedeira que não foi modificada para expressar a proteína recombinante ou uma célula recombinante que não estão expressando a proteína recombinante.

Um "perfil genético" como aqui usado pode incluir genes em um genoma, mRNA transcreveram de genes no genoma ou cDNA derivaram de 30 mRNA transscrito de genes no genoma. Um perfil genético pode também incluir produtos de transcrição que foram modificados por uma célula como variantes de encaixe de genes em sistemas eucarióticos, ou proteínas

transladadas dos genes em um genoma, incluindo proteínas que são modificadas pela célula ou transladas de variantes de encaixe de mRNA transladado do genoma. Um perfil genético é significado referir-se somente à análise simultânea de entidades múltiplas, como em um arranjo ou outro sistema

- 5 de multiplexação, incluindo análise de borrão simultânea múltipla ou cromatografia de coluna com pares de ligação de múltiplos ligados ao pacote. De acordo com a invenção, pelo menos 5, 10, 25, 50, 70, 80, 90 ou 100 ou mais genes ou produtos de gene são analisados simultaneamente.

### TRANSCRIPTOMA

- 10 Em uma modalidade, o perfil genético analisado é um perfil de transcriptoma. Um transcriptoma completo refere-se ao conjunto completo de transcrições de mRNA produzidas pelo genoma de uma só vez. Ao contrário do genoma, o transcriptoma é dinâmico e varia consideravelmente em circunstâncias divergentes devido aos padrão diferentes de expressão de gene.
- 15 Transcriptômicos, o estudo do transcriptoma, é um meio inclusivo de identificar padrão de expressão de gene. O transcriptoma analisado pode incluir o conjunto conhecido completo de genes transcritos, isto é o conteúdo de mRNA ou cDNA correspondente de uma célula ou organismo hospedeiro de hospedeiro. O cDNA pode ser uma cadeia de nucleotídeos, um polinucleotídeo, nucleotídeo, molécula de ácido nucléico isolados ou qualquer fragmento
- 20 ou complemento destes que originou recombinante ou sinteticamente e é bifilamentar ou unifilamentar, de codificação e/ou de não-codificação, um éxon ou um ítron de uma molécula de DNA genômica, ou combinado com carboidrato, lipídios, proteína ou elementos ou substâncias inorgânicas. A
- 25 cadeia de nucleotídeo pode ser pelo menos 5, 10, 15, 30, 40, 50, 60, 70, 80, 90 ou 100 nucleotídeos em comprimento. O transcriptoma pode também incluir apenas uma porção do conjunto conhecido de transcrições genéticas. Por exemplo, o transcriptoma pode incluir menos que 98%, 95, 90, 85, 80, 70,
- 30 60 ou 50% das transcrições conhecidas em um hospedeiro. O transcriptoma pode também ser alvejado a um conjunto específico de genes.

Em uma modalidade, o processo de triagem pode incluir triagem usando um arranjo ou um microarranjo para identificar um perfil genético. Em

outra modalidade, o perfil de transcriptoma pode ser analisado usando processos conhecidos como hibridização em ensaios de borrão como northern blots. Em outra modalidade, o processo pode incluir processos com base em PCR como RT-PCR que podem quantificar a expressão de um conjunto particular de genes. Em uma modalidade da invenção, um gene identificado, por exemplo uma proteína de modulador de dobramento (FM) ou proteína de protease, isto é uma protease, peptidase, ou polipeptídeo ou co-fator associado, é identificada por um processo de triagem de processamento alto.

O processo pode incluir analisar o perfil de transcriptoma usando 10 um microarranjo ou técnica equivalente. Nesta modalidade, o microarranjo pode incluir pelo menos uma porção do genoma transcrita da célula hospedeira, e tipicamente inclui os pares de ligação para amostras de genes de pelo menos 50% dos genes transcritos do organismo. Mais tipicamente, o microarranjo ou técnica equivalente inclui os pares de ligação para amostras de 15 pelo menos 80%, 90%, 95%, 98%, 99% ou 100% dos genes transcritos no genoma da célula hospedeira. Porém, em uma modalidade separada, o microarranjo pode incluir os pares de ligação para um subconjunto selecionado de genes do genoma, incluindo mas não limitados a, genes de protease putativos ou genes de modulador de dobramento putativos. Um microarranjo ou 20 técnica equivalente pode tipicamente também incluir os pares de ligação para um conjunto de genes que são usados como controles, como genes de manutenção. Um microarranjo ou técnica equivalente pode também incluir genes crescidos em cachos em grupos como genes que codificam para proteínas degradativas, moduladores e co-fatores de dobramentos, proteínas metabólicas como proteínas envolvidas em metabolismo de glicose ou aminoácido ou 25 síntese de nucleobase, fatores de transcrição, fatores de estabilização de ácido nucléico, genes regulados de sinal extracelular como cinases e receptores ou proteínas de andaime.

Um microarranjo é em geral formado ligando um número grande 30 de pares de ligação distintos que podem incluir polinucleotídeos, aptâmeros, químicas, anticorpos ou outras proteínas ou peptídeos a um suporte sólido como um microfatia, lâmina de vidro ou outros, em um padrão definido.

Contatando o microarranjo com uma amostra obtida de uma célula de interesse e detectando a ligação dos pares de ligação expressos na célula que hibridam com as seqüências na fatia, o padrão formado pelos polinucleotídeos de hibridização permite a identificação de genes ou agrupamentos de genes que são expressados na célula. Além disso, onde cada membro ligado ao suporte sólido for conhecido, a identidade dos pares de hibridização com os da amostra de ácido nucléico pode ser identificada. Uma tecnologia de resistência de microarranjo é que permite a identificação da expressão de gene de diferencial simplesmente comparando os padrão de hibridização.

Exemplos de processos de triagem de processamento alto incluem hibridização de mRNA de célula hospedeira ou cDNA substancialmente correspondente, para um/uns arranjo(s) hidrolizável(is) ou microarranjo(s). O arranjo ou microarranjo pode ser um ou mais arranjo(s) de oligômeros ou polímeros de ácido nucléico ou de análogos de ácido nucléico. Em uma modalidade, o(s) arranjo(s) ou microarranjo(s) será(ão) independente ou coletivamente um/uns arranjo(s) ou microarranjo(s) amplo(s) em genoma de célula hospedeira, contendo uma população de oligômeros ou polímeros de ácido nucléico ou de análogos de ácido nucléico cujas seqüências de nucleotídeo são hidrolizáveis em porções representativas de todos os genes conhecidos codificar ou prognosticados como codificando FMs na cepa de célula hospedeira ou todos os genes conhecidos codificar ou prognosticados codificar proteases ou proteínas de protease na cepa de célula hospedeira. Um microarranjo amplo em genoma inclui seqüências que ligam a uma porção representativa de todas das seqüências de estrutura de leitura aberta (ORF) conhecidas ou prognosticadas, como de mRNA ou cDNA correspondente do hospedeiro.

As seqüências de oligonucleotídeo ou análogos no arranjo tipicamente hibridam com as seqüências de mRNA ou de cDNA correspondentes da célula hospedeira e tipicamente compreendem uma seqüência de nucleotídeo complementar a pelo menos uma porção de uma seqüência de mRNA ou cDNA hospedeira, ou uma seqüência homóloga à seqüência de mRNA ou de cDNA hospedeira. Filamentos de DNA simples com seqüências comple-

mentares podem emparelhar entre si e formarem moléculas bifilamentares. Microarranjo em geral aplicam-se ao princípio de hibridização em um formato altamente paralelo. Em vez de um identificado, milhares de identificados potenciais diferentes podem ser arranjados em um suporte sólido de miniatura.

- 5 Em vez de uma única sonda de DNA marcado, uma mistura complexa de moléculas de DNA marcado é usada, preparada do RNA de um tipo de célula ou tecido particular. As abundâncias de moléculas de DNA marcado individuais nesta sonda complexa tipicamente refletem os níveis de expressão dos genes correspondentes. Em um processo simplificado, quando hibridar com o  
10 arranjo, as seqüências abundantes gerarão sinais fortes e seqüências raras gerarão sinais fracos. A resistência do sinal pode representar o nível de expressão de gene na amostra original.

Em uma modalidade, um arranjo ou microarranjo amplo em genoma será usado. Em uma modalidade, o arranjo representa mais que 50%  
15 das estruturas de leitura aberta no genoma do hospedeiro, ou mais que 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% ou 100% das estruturas de leitura aberta conhecidas no genoma. O arranjo pode também representar pelo menos uma porção de pelo menos 50% das seqüências conhecidas codificar proteína na célula hospedeira. Em modalidades separadas, o arranjo representa mais que 50% dos genes ou genes putativos da célula hospedeira, ou mais que 55%, 60%, 65%,  
20 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% ou 100% dos genes conhecidos ou genes putativos. Em uma modalidade, mais de um oligonucleotídeo ou análogo podem ser usados para cada gene  
25 ou seqüência de gene putativo ou estrutura de leitura aberta. Em uma modalidade, estes múltiplos oligonucleotídeos ou análogos representam porções diferentes de um gene conhecido ou seqüência de gene putativo. Para cada gene ou seqüência de gene putativo, de cerca de 1 a cerca de 10000 ou de 1 a cerca de 100 ou de 1 a cerca de 50, 45, 40, 35, 30, 25, 20, 15, 10 ou menos  
30 oligonucleotídeos ou análogos podem estar presentes no arranjo.

Um microarranjo ou um arranjo ou microarranjo amplo em genoma completo pode ser preparado de acordo com qualquer processo co-

nhecido na técnica, com base no conhecimento das(s) seqüênci(a)s do genoma de célula hospedeira, ou das seqüências de codificação propostas no genoma, ou com base no conhecimento das seqüências de mRNA expressas na célula hospedeira ou organismo hospedeiro.

5 Para tipos diferentes de células hospedeiras, o mesmo tipo de microarranjo pode ser aplicado. Os tipos de microarranjo incluem microarranjo de DNA complementar (cDNA) (Schena, M., et al. (1995) Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* 270:467-70) e microarranjo de oligonucleotídeo (Lockhart, et al.  
10 (1996) Expression monitoring by hybridization to high-density oligonucleotide arrays. *Nat Biotechnol* 14:1675-80). Para microarranjo de cDNA, o fragmento de DNA de uma estrutura de leitura aberta parcial ou inteira é impresso nas lâminas. As características de hibridização podem ser diferentes ao longo da lâmina porque porções diferentes das moléculas podem ser impressas em localizações diferentes. Para os arranjos de oligonucleotídeo, oligos 2080 mer  
15 podem ser sintetizados ou *in situ* (na fatia) ou por síntese convencional seguida por imobilização na fatia, porém em geral todas as sondas são projetadas para ser similares com respeito à temperatura de hibridização e afinidade de ligação (Butte, A. (2002) The use and analysis of microarray data. *Nat Rev Drug Discov* 1:951-60).

Na análise do perfil de transcriptoma, os oligômeros ou polímeros de ácido nucléico ou de análogos de ácido nucléico podem ser RNA, DNA, ou um análogo de RNA ou DNA. Tais análogos de ácido nucléico são conhecidos na técnica e incluem, por exemplo: ácidos nucléicos de peptídeo (PNA); ácidos nucléicos de arabinose; ácidos nucléicos de altritol; ácidos nucléicos ligados (BNA), por exemplo, ácidos nucléicos ligados a 2'-0,4'-C-etileno e ácidos nucléicos ligados a 2'-0,4'-C-metileno; ácidos nucléicos de cicloexenila; ácidos nucléicos com base em nucleotídeo 2',5'-ligado; ácidos nucléicos de morfolino (unidades de morfolino nucleobase-substituídas conectadas, por exemplo, através de ligações de fosforodiamidato); análogos de ácido nucléico substituídos na cadeia principal, por exemplo, ácidos nucléicos 2'-substituídos, em que pelo menos um dos átomos de carbono 2' de um ácido

nucléico do tipo oligo ou polissacarídeo ou análogo é independentemente substituído com, por exemplo, qualquer um de um grupo halo, tio, amino, alifático, oxialifático, tioalifático ou aminoalifático (em que alifático é tipicamente C<sub>1</sub>-C<sub>10</sub> alifático).

5 Oligonucleotídeos ou análogos de oligonucleotídeo no arranjo podem ser de tamanho uniforme e, em uma modalidade, podem ser cerca de 10 a cerca de 1.000 nucleotídeos, cerca de 20 a cerca de 1.000, 20 a cerca de 500, 20 a cerca de 100, cerca de 20, cerca de 25, cerca de 30, cerca de 40, cerca de 50, cerca de 60, cerca de 70, cerca de 80, cerca de 90 ou cerca de  
10 100 nucleotídeos de comprimento.

O arranjo de sondas de oligonucleotídeo pode ser um arranjo de densidade alta compreendendo mais que cerca de 100, ou mais que cerca de 1.000 ou mais sondas de oligonucleotídeo diferentes. Tais arranjos de densidade alta podem compreender uma densidade de sonda de mais que cerca  
15 de 60, mais em geral maior que cerca de 100, mais em geral maior que cerca de 600, freqüentemente maior que cerca de 1.000, mais freqüentemente maior que cerca de 5.000, o mais freqüentemente maior que cerca de 10.000, tipicamente maior que cerca de 40.000 mais tipicamente maior que cerca de 100.000, e em certas circunstâncias é maior que cerca de 400.000 sondas de  
20 oligonucleotídeo diferentes por cm<sup>2</sup> (onde oligonucleotídeos diferentes referem-se aos oligonucleotídeos que têm seqüências diferentes). As sondas de oligonucleotídeo variam de cerca de 5 a cerca de 500, ou cerca de 5 a 50, ou de cerca de 5 a cerca de 45 nucleotídeos, ou de cerca de 10 a cerca de 40 nucleotídeos e o mais tipicamente de cerca de 15 a cerca de 40 nucleotídeos  
25 em comprimento. Arranjos particulares contêm sondas que variam de cerca de 20 a cerca de 25 oligonucleotídeos em comprimento. O arranjo pode compreender mais que 10, ou mais que 50, ou mais que 100, e tipicamente mais de 1.000 sondas de oligonucleotídeo específicas para cada gene identificado. Em uma modalidade, o arranjo compreende pelo menos 10 sondas de  
30 oligonucleotídeo diferentes para cada gene. Em outra modalidade, o arranjo tem 20 ou menos oligonucleotídeos complementares a cada gene. Embora uma superfície de arranjo planar seja típica, o arranjo pode ser fabricado em

uma superfície de virtualmente qualquer forma ou até mesmo em superfícies múltiplas.

O arranjo pode também compreender sondas de controle de disparidade. Onde tais controles de disparidade estão presentes, a etapa de quantificação pode compreender calcular a diferença em intensidade do sinal de hibridização entre cada uma das sondas de oligonucleotídeo e sua sonda de controle de disparidade correspondente. A quantificação pode também compreender calcular a diferença média em intensidade do sinal de hibridização entre cada uma das sondas de oligonucleotídeo e sua sonda de controle de disparidade correspondente para cada gene.

Em alguns formatos de ensaio, a sonda de oligonucleotídeo pode ser presa, isto é, através de ligação covalente, a um suporte sólido. Arranjos de oligonucleotídeo podem ser quimicamente sintetizados através de processos de síntese de polímero imobilizado paralela ou por processos de síntese de polímero de direcionada por luz, por exemplo em substratos de poli-L-lisina como lâminas. Arranjos quimicamente sintetizados são vantajosos em que a preparação da sonda não requer clonagem, uma etapa de amplificação de ácido nucléico ou síntese enzimática. O arranjo inclui sondas de teste que são sondas de oligonucleotídeo cada uma destas tem uma seqüência que é complementar a uma subseqüência de um dos genes (ou o mRNA ou o cRNA antissense correspondente) cuja expressão é para ser detectada. Além disso, o arranjo pode conter controles de normalização, controles de disparidade e controles do nível de expressão como descritos aqui.

Um arranjo pode ser projetado para incluir um oligonucleotídeo de hibridização por gene conhecido em um genoma. Os oligonucleotídeos ou pares de ligação equivalentes podem ser 5'-amino modificados para suportar ligação covalente às lâminas revestidas com epóxi. Os oligonucleotídeos podem ser projetados para reduzir hibridização cruzada, por exemplo reduzindo a identidade de seqüência para menos que 25% entre os oligonucleotídeos. Em geral, a temperatura de fundição dos oligonucleotídeos é analisada antes de o projeto do arranjo assegurar teor de GC consistente e Tm, e a

estrutura secundária dos pares de ligação de oligonucleotídeo é otimizada. Para perfilação de transcriptoma, a estrutura secundária é tipicamente minimizada. Em uma modalidade, cada oligonucleotídeo é impresso em pelo menos duas localizações diferentes na lâmina para aumentar a precisão.

- 5 Oligonucleotídeos de controle podem também ser projetados com base nas seqüências de diferentes espécies que a célula ou organismo hospedeiro para mostrar ligação de base.

As amostras no perfil genético podem ser analisadas individualmente ou agrupadas em agrupamentos. Os agrupamentos podem tipicamente ser agrupados por similaridade em expressão de gene. Em uma modalidade, os agrupamentos podem ser individualmente agrupados como genes que são regulados a uma extensão similar em uma célula hospedeira. Os agrupamentos podem também incluir grupos de genes que são regulados a uma extensão similar em uma célula hospedeira recombinante, por exemplo 10 genes que são sobre-regulados ou sub-regulados a uma extensão similar comparada a uma célula hospedeira ou uma célula modificada ou uma inalterada. Os agrupamentos podem também incluir grupos relacionados por estrutura de gene ou de proteína, função ou, no caso de um arranjo de transcriptoma, por colocação ou agrupamento dos pares de ligação aos genes 15 no genoma do hospedeiro. Grupos de pares de ligação ou grupos de genes ou proteínas analisados podem incluir genes selecionados de, mas não limitados a: genes que codificam para proteases putativas ou conhecidas, co-fatores de proteases ou proteínas similares à protease; moduladores de dobramentos, co-fatores de moduladores de dobramentos ou proteínas que poderiam melhorar o dobramento ou solubilidade da proteína; fatores de transcrição; proteínas envolvidas em estabilidade de ácido nucléico ou iniciação translacional; 20 cinases; receptores extracelulares ou intracelulares; enzimas metabólicas; co-fatores metabólicos; proteínas de envelope; fatores de sigma; proteínas ligadas à membrana; proteínas de transmembrana; proteínas associadas à 25 membrana e genes de manutenção.

### PROTEOMA

Em outra modalidade, o perfil genético analisado é um perfil de

proteoma. O proteoma de um hospedeiro é o conjunto completo de proteínas produzidas pelo genoma de uma só vez. O proteoma é em geral muito mais complexo que o genoma ou o transcriptoma porque cada proteína pode ser quimicamente modificada após a síntese. Muitas proteínas são clivadas durante a produção, fosforiladas, acetiladas, metiladas, ou tem grupos carboidrato acrescentados a elas, dependendo da célula hospedeira. O proteoma é também muito dinâmico. Proteômicos, o estudo do proteoma, pode abranger vários aspectos diferentes da estrutura de proteína, expressão de proteína e função. As técnicas para análise de proteoma não são tão diretas quanto às usadas em transcriptônicos. Porém, uma vantagem dos proteômicos é que as moléculas funcionais da célula estão sendo estudadas.

O processo pode incluir técnicas que medem os níveis de expressão de proteína, interações de proteína-proteína, interações de proteína-molécula pequena ou atividades enzimáticas. Em uma modalidade, o proteoma é analisado usando um processo de triagem que inclui medição de tamanho de certas proteínas, tipicamente usando espectrometria de massa. Em uma modalidade, a técnica para analisar o perfil de proteoma inclui hibridização de um anticorpo com uma proteína de interesse. Por exemplo, o processo pode incluir processos de western blot como conhecido na técnica ou pode incluir cromatografia de coluna. O processo pode também incluir processos padrão como triagem de Elisa conhecida na técnica. O processo pode também incluir ligação de pares de ligação modificados de ácido nucléico que podem ser aptâmeros ou podem ser proteína ou os pares de ligação química para proteínas ou fragmentos de peptídeo no proteoma e um processo de triagem pode incluir amplificação dos ácidos nucléicos. O processo pode também incluir compostos químicos que ligam às proteínas ou fragmentos de proteínas em um proteoma e o processo pode incluir medição da ligação através de meios químicos. A medição pode também incluir medição dos produtos de reação em uma reação química, ou por ativação de um fluoróforo. Técnicas como espectrometria de massa em combinação com ferramentas de separação como eletroforese em cromatografia líquida em gel bidimensional ou multidimensional, pode também ser usado no processo.

Tipicamente, o processo inclui uma técnica de triagem de processamento alto.

O processo da invenção pode incluir analisar o perfil de proteoma usando, por exemplo, eletroforese bidimensional. Este é um método para a separação e identificação das proteínas em uma amostra através de deslocamento em duas dimensões orientadas em ângulos retos um ao outro. Isto permite a amostra separar em uma área maior, aumentando a resolução de cada componente. A primeira dimensão é tipicamente com base na carga de uma molécula particular enquanto a segunda dimensão pode ser com base no tamanho de uma molécula. Na primeira dimensão, as proteínas são resolvidas de acordo com seus pontos isoelétricos usando eletroforese de gradiente de pH imobilizado (IPGE), focalização isoelétrica (IEF) ou eletroforese de gradiente de pH de não-equilíbrio. Sob condições padrão de temperatura e concentração de uréia, os pontos de focalização observados da grande maioria das proteínas estritamente aproxima-se dos pontos isoelétricos prognosticados calculados das composições de aminoácido das proteínas. Em geral, a primeira etapa após a preparação de uma amostra do hospedeiro inclui operar a amostra junto de um gradiente de pH, um processo conhecido como focalização isoelétrico. Os gradientes de pH podem ser gerados acrescentando anfólitos a um gel de acrilamida. Esta é uma mistura de espécies anfotéricas com uma faixa de valores de pl. Os gradientes de pH podem também ser gerados adicionando Immobilines que é similar aos anfólitos mas foram imobilizados dentro do gel de poliacrilamida produzindo um gradiente de pH imobilizado que não necessita ser pré-focalizado.

A segunda dimensão em eletroforese bidimensional pode ser separação por tamanho de proteínas. As proteínas podem ser separadas de acordo com seu peso molecular aproximado usando poli-acrilamida-eletroforese de dodecil sulfato de sódio (SDS-PAGE). A técnica é extensamente usada e conhecida na técnica. A idéia básica é revestir as proteínas com um detergente (SDS) que reveste todas as proteínas em uma amostra e negativamente as carrega. As proteínas são depois submetidas à eletroforese em gel. Os géis podem tipicamente ser géis de acrilamida e pode estar em um gradiente de densidade. A carga colocada no gel empurra as proteínas a-

través do gel com base no tamanho. Em eletroforese bidimensional, as proteínas separadas podem incluir proteínas de pelo menos 10% do proteoma do organismo. Mais tipicamente, as proteínas de pelo menos 20%, 30%, 40%, 60%, 80% ou 90% das proteínas no proteoma da célula hospedeira são separadas e analisadas através de técnicas como tingimento de proteínas e/ou espectrometria de massa.

O processo da invenção pode também incluir analisar o perfil de proteoma usando um microarranjo. Nesta modalidade, o microarranjo pode incluir os pares de ligação para pelo menos uma porção das proteínas expressas pela célula hospedeira sob condições de crescimento apropriadas, e tipicamente inclui os pares de ligação para proteínas de pelo menos 5% do proteoma do organismo. Mais tipicamente, o microarranjo inclui os pares de ligação para proteínas de pelo menos 10%, 20%, 30%, 40%, 60%, 80% ou 90% das proteínas no proteoma da célula hospedeira. Os pares de ligação podem ser anticorpos que podem ser fragmentos de anticorpo como fragmentos de anticorpo de cadeia simples. Os pares de ligação podem também incluir aptâmeros que são moléculas incluindo ácidos nucléicos que ligam às proteínas específicas ou porções de proteínas. Em uma modalidade separada, o microarranjo pode incluir pares de ligação para um subconjunto selecionado de proteínas do proteoma, incluindo, por exemplo, proteínas de protease putativas ou moduladores de dobramento putativos. O microarranjo pode tipicamente também incluir um conjunto de pares de ligação para proteínas que são usadas como controles. O perfil genético pode ser analisado medindo a ligação das proteínas da célula hospedeira que expressa a proteína ou peptídeo recombinante aos pares de ligação no microarranjo.

O formato de arranjo de proteína mais simples em geral consiste em um número grande de reagentes de captura de proteína ligados às manchas definidas em um material de suporte planar. Este arranjo é depois exposto a uma amostra de proteína complexa. A ligação das proteínas de analito específicas às manchas individuais pode depois ser monitorada usando métodos diferentes. Em casos onde os analitos foram pré-marcados com uma tintura fluorescente, a ligação pode ser diretamente monitorada

usando um escâner de fluorescência. Freqüentemente o anticorpo no formato do tipo sanduíche clássico é usado em que dois reagentes de ligação de proteína simultaneamente ligam ao mesmo antígeno: um anticorpo é imobilizado sobre a superfície, e o outro é fluorescentemente marcado ou conjugado com uma enzima que pode produzir um produto fluorescente, lúminescente ou colorido quando fornecido com o substrato apropriado.

Anticorpos monoclonais ou seus fragmentos de ligação de antígeno são correntemente uma escolha para agentes de captura devido a sua especificidade alta, afinidade e estabilidade. Eles foram usados em uma variedade de ensaios de perfilação de proteína de analito simples clássicos como ensaios imunoabsorventes ligados à enzima (ELISA) desde os anos setenta. Adicionalmente, bibliotecas de exibição de fago dos fragmentos de anticorpo oferecem o potencial para produção de anticorpo em escalas proteômicas. Estas bibliotecas podem ser usadas para isolar os agentes de ligação de afinidade alta contra proteína identificada em um prazo significativamente mais curto que é possível com processos com base em imunização. Exibição de ribossoma e exibição de mRNA são processos adicionais, completamente *in vitro*, que contam ligação física das proteínas de biblioteca a suas seqüências de mRNA de codificação. Tais processos têm sido usados com sucesso para selecionar reagentes de ligação de afinidade alta para proteínas identificadas (Wilson, DS, et al. (2001) *The use of mRNA display to select high-affinity protein-binding peptides Proc Natl Acad Sci USA* 98:3750-3755). Vários grupos tomaram um método diferente para desenvolver reagentes de captura de proteína de afinidade alta para biofatias de proteína. Por exemplo, os aptâmeros foram usados, que são moléculas de RNA ou DNA unifilamentares que originam de experimentos de seleção *in vitro* (denominados SELEX: evolução sistemática de ligandos por enriquecimento exponencial) com afinidades altas para proteínas. Um outro desenvolvimento em tecnologias de aptâmero é assim-chamado fotoaptâmeros. Estas moléculas têm um atributo adicional que intensifica sua utilidade como reagentes de captura de proteína. Elas carregam o grupo de reticulação fototativável 5'-bromodeoxiuridina que, quando ativado por luz UV, pode causar

reticulação covalente com as proteínas identificadas ligadas (Petach, H & Gold, L (2002) Dimensionality is the issue: use of photoaptamers in protein microarrays *Curr Opin Biotechnol* 13:309-314). O evento de foto-reticulação fornece uma segunda dimensão de especificidade similar à ligação de um anticorpo de detecção secundário em um imunoensaio em sanduíche.

Uma ampla variedade de substratos de superfície e químicas de ligação foi avaliada pela imobilização de agentes de captura em microarranjo de proteína. Um modo para imobilizar as proteínas em um suporte sólido conta com interações não-covalentes com base em hidrofóbico ou interações de van der Waals, ligação de hidrogênio ou forças eletrostáticas. Exemplos de imobilização eletrostática incluem o uso de materiais como lâminas de vidro revestidas com nitrocelulose e polilisina ou aminopropil silano. Microarranjos de proteína foram também fabricados por meio de adsorção física sobre as superfícies de plástico de placas de 96 cavidades. Um exemplo de ligação covalente de proteínas à superfície foi descrito por MacBeath e Schreiber (MacBeath, G & Schreiber, SL(2000) Printing proteins as microarrays for high-throughput function determination *Science* 289:1760-1763). Devido à afinidade muito alta de estreptavidina para biotina, a imobilização de proteínas biotiniladas sobre as superfícies de estreptavidina pode ser considerada quase covalente (Peluso, P et al. (2003) Optimizing antibody immobilization strategies for the construction of protein microarrays *Anal Biochem* 312:113-124). Outras estratégias foram descritas (X-ray photoelectron spectroscopy and radiometry studies of biotin-derivatized poly(L-lysine)-grafted-poly(ethylene glycol) monolayers on metal oxides (*Langmuir*) 7313-7322; Ruiz-Taylor, LA et al. (2001) Monolayers of derivatized poly(L-lysine)-grafted poly(ethylene glycol) on metal oxides as a class of biomolecular interfaces *Proc Natl Acad Sci U S A* 2001, 98:852-857; Espejo A, Bedford MT. (2004) Protein-domain microarrays *Processes Mol Biol.* 264:173-81; Zhu, H. et al. (2001) Global analysis of protein activities using proteome chips. *Science Express*).

As amostras no perfil genético podem ser analisadas individualmente ou agrupados em agrupamentos. Os agrupamentos podem tipica-

mente ser agrupados por similaridade em expressão de gene. Em uma modalidade, os agrupamentos podem ser individualmente agrupados como proteínas que são reguladas a uma extensão similar em uma célula hospedeira. Os agrupamentos podem também incluir grupos de proteínas que são

5 reguladas a uma extensão similar em uma célula hospedeira recombinante, por exemplo, que é sobre-regulado ou sub-regulado a uma extensão similar comparada a uma célula hospedeira ou uma célula modificada ou uma inalterada. Os agrupamentos podem também incluir grupos relacionados por estrutura de proteína, função ou processamento. Grupos de pares de ligação

10 de proteína em um arranjo, ou grupos de proteínas analisados em um ensaio diferente como eletroforese bidimensional, podem ser selecionados de, mas não são limitados a: proteases putativas ou conhecidas, co-fatores de proteases ou proteínas similares à protease; moduladores de dobramentos, co-fatores de moduladores de dobramentos ou proteínas que poderiam melhorar o dobramento ou solubilidade da proteína; fatores de transcrição; proteínas envolvida na estabilidade de ácido nucléico ou iniciação translacional; cinases; receptores extracelulares ou intracelulares; enzimas metabólicas; co-fatores metabólicos; proteínas de envelope; e genes de manutenção.

15

### METABOLOMA

20 Processos de análise proteômica permitem determinar a abundância e distribuição de muitas proteínas simultaneamente. Porém, as consequências funcionais de alterações para o proteoma são relatadas apenas indiretamente. Outro método é medir os níveis destas moléculas pequenas, ou metabólitos. Um perfil genético analisado no processo da invenção pode

25 desse modo incluir um perfil metabolômico. Processos para analisar o metaboloma de um hospedeiro específico incluem cromatografia de gás, cromatografia líquida de pressão alta e eletroforese capilar para separar metabólitos de acordo com várias propriedades química e físicas. As moléculas podem depois ser identificadas usando processos como espectrometria de

30 massa.

### DETECÇÃO/ANÁLISE

O processo inclui analisar um perfil genético para identificar um

gene ou produto de gene compensatório que é expressado a um nível mais alto na célula recombinante. Em geral, esta etapa inclui o monitoramento da expressão (por exemplo detecção e ou quantificação da expressão) de uma multidão de genes ou produtos de gene. A expressão é em geral monitorada detectando a ligação dos produtos de gene de célula hospedeira para um perfil de transcriptoma, proteoma ou de metaboloma como descrito acima. A análise da ligação pode envolver uma comparação da ligação entre uma célula hospedeira recombinante que expressa proteína ou peptídeo recombinante e uma célula hospedeira pura ou uma célula hospedeira recombinante que não expressa a proteína ou peptídeo.

#### DETECÇÃO

Esta etapa inclui o monitoramento da expressão (por exemplo detecção e ou quantificação da expressão) de uma multidão de genes ou produtos de gene. A expressão é em geral monitorada detectando a ligação dos produtos de gene de célula hospedeira a um perfil de transcriptoma, proteoma ou metaboloma como descrito acima. Tipicamente, pelo menos cerca de 10 genes, ou pelo menos cerca de 100, ou pelo menos cerca de 1000 e ou pelo menos cerca de 10.000 genes diferentes podem ser ensaiados de uma vez. O processo pode envolver fornecer um fundo geral de ácidos nucléicos identificados compreendendo transcrições de RNA de um ou mais dos ditos genes, ou ácidos nucléicos derivados das transcrições de RNA; hibridar o fundo geral de ácidos nucléicos com um arranjo de sondas de oligonucleotídeo imobilizadas em uma superfície onde o arranjo compreende mais de 100 oligonucleotídeos diferentes e cada oligonucleotídeo diferente está localizado em uma região predeterminada da dita superfície, cada oligonucleotídeo diferente é ligado à superfície através de pelo menos uma ligação covalente, e as sondas de oligonucleotídeo são complementares às transcrições de RNA ou ácidos nucléicos derivados das transcrições de RNA; e quantificar os ácidos nucléicos hibridados no arranjo. Uma representação pictórica de uma técnica para monitorar a expressão de um produto de gene entre duas amostras é descrita na Figura 12.

O processo pode também envolver fornecer um fundo geral de

proteínas celulares. Estes podem ser derivados de lisados celulares que são feitos através de células de lise usando detergentes ou tensoativos; usar lise osmótica; usar alterações térmicas, como ciclos de congelamento-descongelamento; usar dispositivos mecânicos ou usar alterações de pressão. Tipicamente as químicas são incluídas no processo de lise de uma célula ou sistema de células que inibem certas proteínas, como proteases, particularmente proteases não-específicas, para limitar a degradação das proteínas. Além disso, lisados de célula são tipicamente mantidos em ou abaixo de 4°C, e podem ser mantidos em ou abaixo de 0°C ou em ou abaixo de 20°C durante o processamento. Os lisados de célula podem ser separados antes de outro processamento, por exemplo por cromatografia de exclusão de tamanho, permuta de íons ou cromatografia de matriz de afinidade como usando HPLC.

Tipicamente, o produto genético identificado, mRNA, cDNA, proteína ou metabólito é marcado com um marcador ou sonda detectável. O marcador ou sonda pode ser uma ou mais moléculas fluorescentes ou fluoróforos. Estes podem incluir moléculas comercialmente disponíveis como Cy3 e Cy5 ligados, por exemplo, a nucleotídeos particulares que podem ser incorporados em um cDNA transcritos reverso para fornecer moléculas detectáveis para triagem. Em uma modalidade, uns primeiros fluoróforos são incorporados em uma amostra do hospedeiro e um segundo fluoróforo é incorporado em uma amostra de um hospedeiro expressando proteína ou peptídeo recombinante. Em uma modalidade, o primeiro fluoróforo e segundo fluoróforo emitem diferentes comprimentos de onda de luz. Nesta modalidade, a ligação das amostras do hospedeiro e do hospedeiro expressando proteína recombinante pode ser monitorada no mesmo ensaio. Em outra modalidade, os fluoróforos são excitados com diferentes comprimentos de onda de luz. Em outra modalidade, o primeiro e segundo fluoróforo é excitado ou emite luz no mesmo comprimento de onda. Nesta modalidade, as amostras do hospedeiro e da proteína recombinante expressando hospedeiro são tipicamente monitoradas em ensaios diferentes.

O processo pode adicionalmente incluir uma etapa de quantificar a hibridização dos ácidos nucléicos ou proteínas ou metabólitos químicos

identificados. A quantificação pode incluir medição dos níveis de transcrição de um ou mais genes. Tipicamente o fundo geral de ácidos nucléicos identificados é, por exemplo, um em que a concentração dos ácidos nucléicos identificados (transcrições de pré-mRNA, transcrições de mRNA ou ácidos nucléicos derivados das transcrições de RNA) é proporcional aos níveis de expressão de genes que codificam aqueles ácidos nucléicos identificados.

Para análise de transcriptoma, o fundo geral de ácidos nucléicos pode ser marcado antes, durante ou após hibridização, embora os ácidos nucléicos sejam tipicamente marcados antes da hibridização. Marcações de fluorescência são tipicamente usadas, freqüentemente com um fluoróforo simples, e, onde marcação de fluorescência for usada, quantificação dos ácidos nucléicos hibridados pode ser por quantificação de fluorescência do ácido nucléico hibridado fluorescentemente marcado. Tal quantificação é facilitada pelo uso de um escâner a laser confocal ou microscópio de fluorescência, como um microscópio de fluorescência confocal que pode ser equipado com um estágio automatizado para permitir varredura automática do arranjo e que pode ser equipado com um sistema de aquisição de dados para registro de medição automatizado e processamento subsequente da informação de intensidade de fluorescência. Dispositivos para ler tais arranjos incluem os módulos de CloneTracker®, ImaGene®, GeneSight® e a base de dados de GeneDirector®, disponível de Biodiscovery, Inc., El Segundo, Calif., ou a leitora GeneChip®, disponível de Affymetrix, Santa Clara Inc., Calif. Em uma modalidade, hibridização ocorre em severidade baixa (por exemplo cerca de 20°C a cerca de 50°C, ou cerca de 30°C a cerca de 40°C, ou cerca de 37°C). Hibridização pode incluir lavagens subsequentes em severidade progressivamente crescente até um nível desejado de especificidade de hibridização ser alcançado.

Quantificação do sinal de hibridização pode ser por quaisquer dispositivos conhecidos a alguém versado na técnica. Porém, em uma modalidade, quantificação é alcançada pelo uso de um escâner de fluorescência confocal. Dados são tipicamente avaliados calculando a diferença em intensidade de sinal de hibridização entre cada sonda de oligonucleotídeo e sua

sonda de controle de disparidade correspondente. Tipicamente, esta diferença pode ser calculada e avaliada para cada gene. Certos processos analíticos são fornecidos aqui.

Técnicas foram desenvolvidas para preparar sondas de hibridização bacterianas apropriadas (ver, por exemplo Choi et al. (2003) *App. Environ. Microbiol.* 69:4737-4742). Por exemplo, as células podem ser armazenadas em um agente estabilizante de RNA como RNAlater (Ambion, Austin, TX). RNA é em geral purificada em três etapas: (1) isolamento do RNA total, (2) remoção do DNA contaminante e (3) limpeza total do RNA total. RNA total pode ser isolado e depois misturado com iniciadores de hexâmero aleatórios e transcriptase reversa para fazer o cDNA. Tipicamente pelo menos uma sonda fluorescente é incorporada no cDNA. Em uma modalidade, uma sonda fluorescente é incorporada, em outra modalidade mais de uma sonda, por exemplo 2, 3, 4, 5 ou mais sondas fluorescentes são incorporadas nas mesmas amostras de cDNA ou diferentes. Em um hospedeiro eucariótico, o fundo geral dos ácidos nucléicos identificados pode também ser o mRNA de poliA+ total isolado de uma amostra biológica, ou cDNA feito por transcrição reversa do RNA ou cDNA de segundo filamento ou RNA transcrita do intermediário de cDNA bifilamentar.

Corantes fluorescentes são tipicamente incorporados em moléculas de cDNA durante a reação de transcrição reversa. Devido à estrutura de mRNA diferente entre procariotes (bactérias) e eucariotes (levedura, células mamíferas, etc.), iniciadores diferentes podem ser usados, porém iniciadores aleatórios podem ser usados em ambos os casos, e iniciadores de oligo-dT podem ser usados em eucariotes que tem caudas de poliA. Um processo alternativo é marcação de amino-alila para aumentar a intensidade de sinal. Este processo incorpora análogos de nucleotídeo caracterizando um grupo quimicamente reativo ao qual um corante fluorescente pode ser ligado após a reação de transcrição reversa (Manduchi, E., et al. (2002) Comparison of different labeling processes for two-channel high-density micro array experiments. *Physiol Genomics* 10:169-79).

O fundo geral de ácidos nucléicos identificados pode ser tratado

para reduzir a complexidade da amostra e assim reduzir o sinal de base obtido na hibridização. Os termos "base" ou "sinal de base" referem-se aos sinais de hibridização resultantes da ligação não-específica, ou outras interações, entre os ácidos nucléicos identificados marcados e os componentes do arranjo de oligonucleotídeos (por exemplo, as sondas de oligonucleotídeo, sondas de controle, o substrato de arranjo, etc.). Em um método, um fundo geral de mRNAs, derivado de uma amostra biológica, é hibridado com um fundo geral de oligonucleotídeos compreendendo as sondas de oligonucleotídeo presentes no arranjo. O fundo geral de ácidos nucléicos hibridado é depois tratado com RNase A que digere as regiões unifilamentares. Os complexos de hibridização bifilamentares restantes são depois desnaturados e as sondas de oligonucleotídeo são removidas, deixando um fundo geral de mRNAs intensificado para aqueles mRNAs complementares às sondas de oligonucleotídeo no arranjo.

Em outro método para redução de base, um fundo geral de mRNAs derivado de uma amostra biológica é hibridado com oligonucleotídeos identificados emparelhados específicos onde os oligonucleotídeos identificados emparelhados específicos são complementares às regiões flanqueando as subseqüências dos mRNAs complementares às sondas de oligonucleotídeo no arranjo. O fundo geral de ácidos nucléicos hibridado é tratado com RNase H que digere as seqüências de ácido nucléico hibridadas (bifilamentares). As seqüências de ácido nucléico unifilamentares restantes que têm um comprimento cerca de equivalente à região flanqueada pelos oligonucleotídeos identificado emparelhados específicos são depois isoladas (por exemplo através de eletroforese) e usadas como o fundo geral de ácidos nucléicos para monitorar a expressão de gene.

Um terceiro método para redução de base envolve eliminar ou reduzir a representação no fundo geral de mensagens de mRNA identificadas pré-selecionadas particulares (por exemplo, mensagens que são sobre-expressadas caracteristicamente na amostra). Este processo envolve hibridar com uma sonda de oligonucleotídeo que é complementar à mensagem pré-selecionada identificada de mRNA ao fundo geral de mRNAs de poliA+

derivados de uma amostra biológica. A sonda de oligonucleotídeo hibrida com o mRNA de poliA+ pré-selecionado particular ao qual é complementar. O fundo geral de ácidos nucléicos hibridado é tratado com RNase H que digere a região bifilamentar (hibridada) assim separando a mensagem de sua cauda de poliA+. Isolamento ou amplificação (por exemplo, usando uma coluna de oligo dT) do mRNA de poliA+ no fundo geral depois fornece um fundo geral que tem uma representação reduzida ou nenhuma da mensagem de mRNA pré-selecionada identificada.

#### ANÁLISE

O gene identificado é tipicamente identificado comparando um perfil genético da célula hospedeira que expressa a proteína ou peptídeo recombinante a um perfil genético da célula hospedeira que não expressa a proteína ou peptídeo recombinante. Em modalidades iterativas, o gene identificado a ser modificado é identificado comparando um perfil genético da célula que será modificada (a segunda célula) à célula da qual foi modificada (a primeira célula). O gene identificado é identificado comparando um perfil genético da segunda célula a um perfil genético da primeira célula e identificando um ou mais genes da expressão da qual é aumentado na segunda célula.

Microarranjos de cDNA medem a abundância de mRNA relativa entre duas amostras. Uma série de amostra de ponto de tempo de pós-indução pode ser comparada à amostra de pré-indução para a mesma cepa (perfil de expressão temporal), ou as amostras de pós-indução podem ser comparadas com as cepas diferentes no mesmo ponto de tempo. A comparação pode ser através do uso de um programa de computação, como GeneSight®. Por exemplo, quando usar um microarranjo usando um marcador fluorescente, uma intensidade de mancha pode ser medida para cada amostra ligada ao arranjo (por exemplo uma seqüência de DNA). A intensidade de mancha pode depois ser corrigida para base e a razão da intensidade para as amostras do hospedeiro versus o hospedeiro expressando a proteína ou peptídeo recombinante, ou para o hospedeiro expressando a proteína ou peptídeo recombinante comparado ao hospedeiro modificado que expressa a

proteína ou peptídeo recombinante, pode ser medido. A razão fornece uma medida para identificar os genes que são sobre-regulados ou a expressão da qual é aumentado sob expressão da proteína ou peptídeo recombinante, ou sob modificação da célula hospedeira para permitir identificação de um gene

5 identificado.

Para identificar se um gene é sobre-regulado, uma razão padrão ou "corte" é estabelecida. A razão de corte pode ser projetada para superar os efeitos de ruído de base associados a um ensaio particular. Em geral, qualquer razão de mais que 1 entre as medições pode designar um gene sobre-regulado. Porém, variação entre os ensaios pode requerer uma razão mais alta que 1, por exemplo 1,5, ou mais que 2, ou mais que 2,5, ou mais que 3, ou mais que 3,5 ou mais que 4 ou mais que 4,5, ou mais que 5 ou mais que 6, ou mais que 7, ou mais que 8, ou mais que 9 ou mais que 10. O padrão pode ser estabelecido antes do processo, contando com padrão conhecidos na técnica, ou pode ser estabelecido durante as medições comparando as razões de níveis de genes ou produtos de gene de controle, como genes de manutenção.

ETAPA III: EXPRESSÃO VARIÁVEL DO GENE OU PRODUTO DE GENE COMPENSATÓRIO IDENTIFICADO GENETICAMENTE MODIFICANDO A

20 CÉLULA PARA FORNECER UMA CÉLULA RECOMBINANTE MODIFICADA QUE ALCANÇA UM AUMENTO NA EXPRESSÃO DE PROTEÍNA RECOMBINANTE, ATIVIDADE OU SOLUBILIDADE.

GENES COMPENSATÓRIOS IDENTIFICADOS

Os genes ou produtos de gene compensatórios que são identificados na etapa ii), ou análogos homólogos, co-fatores ou subunidades destes, são usados para projetar estratégias para geneticamente modificar a célula ou aumentar, diminuir, knock-in ou knock-out a expressão de um ou mais genes identificados. As seqüências de gene identificadas podem ser usadas em bases de dados públicos para projetar estratégias, particularmente projetar constructos para modular expressão de um gene através das técnicas descritas acima. Tais técnicas são bem conhecidas.

Em uma modalidade, o gene ou genes identificados são pelo

menos uma protease putativa, uma proteína similar à protease, um co-fator ou subunidade de uma protease. Em outras modalidades, o gene ou genes identificados são pelo menos um modulador de dobramento, modulador de dobramento putativo, co-fator ou subunidade de um modulador de dobramento. Em certas modalidades, um gene identificado é uma subunidade de uma protease. Em uma modalidade, o gene ou genes identificados podem ser uma serina, treonina, cisteína, aspártico ou metalo peptidase. Em uma modalidade, o gene ou genes identificados podem ser selecionados de *hsIV*, *hsIU*, *clpA*, *clpB* e *clpX*. O gene identificado pode também ser um co-fator de uma protease. Em outra modalidade, o gene ou genes identificados são um modulador de dobramento. Em algumas modalidades, o gene ou genes identificados podem ser selecionados de uma proteína de chaperona, uma foldase, uma peptidil prolil isomerase e uma isomerase de ligação de dissulfeto. Em algumas modalidades, o gene ou genes identificados podem ser selecionados de *htpG*, *cbpA*, *dnaJ*, *dnaK* e *fkbP*.

Genes bacterianos são organizados em operons que são agrupamentos de gene que codificam as proteínas necessárias para executar a função coordenada como biossíntese de um aminoácido dado. Portanto, em uma modalidade, o gene identificado faz parte de um operon. Em uma modalidade particular, o gene identificado é um operon que codifica para uma ou mais proteínas com atividade de protease sozinho ou em combinação, ou é um operon que codifica para uma ou mais proteínas com atividade moduladora de dobramento, incluindo foldases, chaperonas, e isomerases.

### PROTEASES

Em uma modalidade da invenção, a célula hospedeira é modificada reduzindo a expressão, inibindo ou removendo pelo menos uma protease do genoma. A modificação pode também ser mais de uma protease em algumas modalidades. Em uma modalidade relacionada, a célula é modificada reduzindo a expressão de um co-fator de protease ou proteína de protease. Em outra modalidade, a célula hospedeira é modificada por inibição de um promotor para uma protease ou proteína relacionada que pode ser um promotor nativo. A modificação de gene pode ser para modular uma proteína

homóloga ao gene identificado.

Na base de dados de MEROPS, peptidases são agrupadas em clãs e famílias. As famílias são grupos de peptidases funcionalmente similares estreitamente relacionadas. Famílias são agrupadas por seu tipo catalítico: S,

5 serina; T, treonina; C, cisteína; A, aspártico; M, metalo e U, desconhecido.

Mais de 20 famílias (denotadas S1 - S27) de serina protease foram identificadas, estas são agrupadas em 6 clãs (SA, SB, SC, SE, SF e SG) em base de similaridade estrutural e outra evidência funcional. Estruturas são conhecidas por quatro dos clãs (SA, SB, SC e SE). Treonina peptidases são caracterizadas por um nucleófilo de treonina no término N da enzima madura. O exemplo de tipo para este clã é o componente beta de proteasoma arqueano de *Thermoplasma acidophilum*. Cisteína peptidases têm topologias moleculares características e são peptidases em que o nucleófilo é o grupo sulfidrila de um resíduo de cisteína. Cisteína proteases são divididas em clãs (proteínas que são evolutivas relacionadas), e também sub-divididas em famílias, em base da arquitetura de seu díade ou tríade catalítico:

Clã CA contém as famílias de papaína (C1), calpaína (C2), estreptopaína (C10) e as peptidases ubiquitina-específicas (C12, C19), como também muitas famílias de endopeptidases de cisteína viral.

20 Clã CD contém as famílias de clostriplaína (C11), gingipaína R (C25), legumaína (C13), caspase-1 (C14) e separina (C50). Estas enzimas têm especificidades dominadas pelas interações do sub-sítio 51.

Clã CE contém as famílias de adenaína (C5) de adenovírus, a Ulp1 protease eucariótica (C48) e as YopJ proteases bacterianas (C55).

25 Clã CF contém apenas piroglutamil peptidase I (C15).

Clã PA contém as picornaínas (C3), que provavelmente evoluíram das serina peptidases e que formam a maior parte das enzimas neste clã.

Clãs PB e CH contêm as cisteína peptidases autolíticas.

Endopeptidases aspárticas de origem vertebrada, fúngica e re-  
30 troviral foram caracterizadas. Aspartato peptidases são assim nomeadas porque os resíduos de Asp são os ligantes da molécula de água ativada em todos os exemplos onde os resíduos catalíticos foram identificados, embora

pelo menos uma enzima viral seja acreditada ter como Asp e uma Asn como sua diade catalítica. Todas ou a maioria das aspartato peptidases são endopeptidases. Estas enzimas foram atribuídas em clãs (proteínas que são evolutivas relacionadas), e também sub-divididas em famílias, em grande parte em base de sua estrutura terciária.

Metaloproteases são as mais diversas dos quatro tipos principais de protease, com mais de 30 famílias identificadas até agora. Nestas enzimas, um cátion divalente, usualmente zinco, ativa a molécula de água. O íon de metal é mantido no lugar por ligantes de aminoácido, usualmente três em número. Os ligantes de metal conhecidos são His, Glu, Asp ou Lys e pelo menos um outro resíduo é requerido para catálise que pode representar um papel eletrofílico. Das metaloproteases conhecidas, quase a metade contém um motivo de HEXXH que foi mostrado em estudos cristalográficos fazer parte do sítio de ligação de metal. O motivo de HEXXH é relativamente comum, mas pode ser mais estritamente definidos para metaloproteases como abXHEbbHbc onde 'a' é mais freqüentemente valina ou treonina e faz parte do sub-sítio do S1 em termolisina e neprilisina, 'b' é um resíduo descarregado, e 'c' um resíduo hidrofóbico. Prolina nunca é encontrada neste sítio, possivelmente porque quebraria a estrutura helicoidal adotada por este motivo em metaloproteases.

As peptidases associadas ao clã U têm um mecanismo catalítico desconhecido como a dobra de proteína do domínio de sítio ativo e os resíduos de sítio ativo não foram relatados.

Certas proteases (por exemplo OmpT) podem adsorver na superfície dos corpos de inclusão e podem degradar a proteína desejada enquanto estiver sendo redobrada. Portanto, certas proteínas identificadas podem ser proteases ou proteínas de protease que aderem aos corpos de inclusão e estes podem ser modificados, por exemplo, para reduzir a ligação.

Proteases ou proteínas de protease podem também ser classificadas como Aminopeptidases; Dipeptidases; Dipeptidil-peptidases e tripeptidil peptidases; Peptidil-dipeptidases; Carboxipeptidases do tipo serina-tipo; Metalocarboxipeptidases; Carboxipeptidases do tipo cisteína; Omegapepti-

dases; Serina proteinases; Cisteína proteinases; Aspártico proteinases; Metalo proteinases; ou Proteinases de mecanismo desconhecido.

Aminopeptidases incluem citosol aminopeptidase (leucil aminopeptidase), alanil aminopeptidase de membrana, cistinil aminopeptidase, 5 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-estereospecífica, 10 aminopeptidase ey. Dipeptidases incluem x-his dipeptidase, x-arg dipeptidase, x-metil-his dipeptidase, cis-gli 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. Dipeptidil-peptidases e tripeptidil peptidases incluem 15 dipeptidil-peptidase i, dipeptidil-peptidase ii, dipeptidil peptidase iii, dipeptidil-peptidase iv, dipeptidil-dipeptidase, tripeptidil-peptidase I, tripeptidil-peptidase II. Peptidil-dipeptidases incluem peptidil-dipeptidase a e peptidil-dipeptidase b. Carboxipeptidases do tipo serina incluem pro-x carboxipeptidase lisossomal, D-ala-D-ala carboxipeptidase do tipo serina, carboxipeptidase C, carboxipeptidase D. Metalocarboxipeptidases incluem carboxipeptidase a, carboxipeptidase B, lisina(arginina) carboxipeptidase, gly-X carboxipeptidase, alanina carboxipeptidase, muramolpentapeptídeo carboxipeptidase, carboxipeptidase h, glutamato carboxipeptidase, carboxipeptidase M, muramoltetrapeptídeo carboxipeptidase, zinco d-ala-d-ala carboxipeptidase, 20 carboxipeptidase A2, pro-x carboxipeptidase de membrana, tubulinil-tir carboxipeptidase, carboxipeptidase t. 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. Serina proteinases incluem quimiotripsina, quimiotripsina c, metridina, tripsina, trombina, Fator Xa de coagulação, plasmina, enteropeptidase, acrosina, protease alfa-lítica, glutamila, endopeptidase, ca-

tepsina 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 de complemento c1r55, componente de complemento c1s55, convertase c3/c5 da via de complemento clássico, fator de complemento I, fator de complemento D, convertase c3/c5 da via de complemento alternativa, cerevisina, hipodermina C, lisil endopeptidase, endopeptidase 1a, gama-reni, venombin ab, leucil endopeptidase, triptase, escutelarina, quexina, subtilisina, orizina, endopeptidase k, termomicolina, termitase, endopeptidase SO, ativador de plasminogênio T, proteína C, endopeptidase pancreática E, elastase pancreática ii, serina endopeptidase IGA-específica, U-plasminogênio, ativador, venombina A, furina, mieloblastina, semenogelase, granzima A ou proteinase de linfócito T citotóxico 1, granzima B ou proteinase de linfócito T citotóxico 2, esteptogrisina A, treptogrisina B, glutamil endopeptidase II, oligopeptidase B, fator de coagulação de límulo c, fator de coagulação de límulo, enzima de coagulação de límulo, omptina, repressor lex, peptidase líder bacteriana I, togavirina, flavirina. Cisteína proteinases incluem catepsina B, papaína, ficina, quimiopapaína, asclepaína, clostriplaína, esteptopalaína, actinídeo, catepsina 1, catepsina H, calpaína, catepsina t, glicila, endopeptidase, pró-coagulante de câncer, catepsina S, picornaína 3C, picornaína 2A, caricaína, ananaína, bromelaína de talo, bromelaína de fruta, legumaína, histolisaína, enzima de conversão de interleucina 1-beta. Aspártico proteinases incluem pepsina A, pepsina B, gastricsina, quimiosina, catepsina D, neopentesina, renina, retropepsina, enzima de conversão de pró-opiomelanocortina, aspergilopepsina I, aspergilopepsina II, penicilopepsina, rizopuspepsina, endotiapepsina, mucoropepsina, candidapepsina, sacaropepsina, rodotorulapepsina, fisaropepsina, acrocilindropepsina, poliporopepsina, picnoporopepsina, escitalidopepsina a, escitalidopepsina b, xantomonapepsina, catepsina e pepsina de barreira, peptidase líder bacteriana I, pseudomonapepsina, plasmepsina. Metalo proteinases incluem atrolisina a, colagenase microbiana, leucolisina, colagenase intersticial, neprilisina, envelisina, metaloendopeptidase iga-específica, procolágeno

- N-endopeptidase, timet oligopeptidase, neurolisina, estromelisina 1, meprina A, procolágeno C-endopeptidase, peptidil-lis metaloendopeptidase, astacina, estromelisina, 2, matrilisina gelatinase, aeromonolisina, pseudolisina, termolisina, bacilosina, aureolisina, cocolisina, micosina, metaloendopeptidase beta-lítica, peptidil-asp metaloendopeptidase, neutrófilo colagenase, gelatinase B, leishmanolisina, sacarolisina, autolisina, deuterolisina, serralisina, atrolisina B, atrolisina C, atroxase, atrolisina E, atrolisina F, adamalisina, horrulisina, ruberlisina, botropasina, botrolisina, ofiolisina, trimerelisina I, trimere-lisina II, mucrolisina, pitrilisina, insulisina, O-sialoglicoproteína endopeptidase, russelisina, mitocondrial, intermediário, peptidase, dactilisina, nardilisina, magnolisina, meprina B, peptidase de processamento mitocondrial, elastase de macrófago, coriolisina, toxilisina. Proteinases de mecanismo desconhecido incluem termopsina e complexo de endopeptidase multicatalítica.

Certas proteases de *P. fluorescens* são listadas na Tabela A.

15

TABELA A

Classe	Família	RXF	Função Curada	Gene	Fisiologia
	<b>Homólogos de MEROPS</b>				
	<b>Aspártico peptidases</b>				
	A8 (família de peptidase II de sinal)				
		RXF05383	Peptidase de sinal de lipo-proteína (ec 3.4.23.36)		Processamento de numerosas lipoproteínas segregadas bacterianas.
	A24 (família de prepilina peptidase do tipo IV)				
		RXF05379	Prepilina peptidase do tipo 4 pilD (ec 3.4.99.-)		Este peptidase ligado à membrana cliva um peptídeo líder especializado de prepilina do tipo 4 durante sua secreção de muitas espécies bacterianas. Uma vez segregadas, as proteínas processadas são requeridas para funções incluindo

Classe	Família	RXF	Função Curada	Gene	Fisiologia
					formação de pilus do tipo 4, toxina e outra secreção de enzima, transferência de gene e formação de biofilme.
	Cisteína peptidases				
		C15 (família de piroglutamil peptidase I)			
		RXF02161	Pirrolidona carboxilato peptidase (ec 3.4.19.3)		Remoção de grupos de piroglutamila de peptídeos no catabolismo de proteína
	C40				
		RXF01968	Proteína associada à invasão, P60		
		RXF04920	Proteína associada à invasão, P60		
		RXF04923	Proteína associada à fosfatase papq		
	C56 (família de Pfpl endopeptidase)				
		RXF01816	Protease I (ec 3.4.--)		
	Metalopeptidases				
	MI				
		RXF08773	Alanina aminopeptidase de membrana (ec 3.4.11.2)		
	M3				
		RXF00561	Oligopeptidase A (ec 3.4.24.70)	prC	Degradação de peptídeos de sinal de lipoproteína, e outros oligopeptídeos intracelulares. Papel na maturação de bacteriófago precursor de P22 gp7.
		RXF04631	Oligopeptidases		

Classe	Família	RXF	Função Curada	Gene	Fisiologia
			Zn-dependentes		
	M4 (família de termolisina)				
		RXF05113	Precursor de metaloprotease extracelular (ec 3.4.24.-)		
	M41 (família de FtsH endopeptidase)				
		RXF05400	Proteína de divisão celular ftsH (ec 3.4.24.-)		Papel proposto no controle de qualidade proteolítica das moléculas reguladoras e proteínas de membrana, em levedura.
	M10				
		RXF04304	Serralisina (ec 3.4.24.40)		
		RXF04500	Serralisina (ec 3.4.24.40)		
		RXF01590	Serralisina (ec 3.4.24.40)		
		RXF04495	Serralisina (ec 3.4.24.40)		
		RXF02796	Serralisina (ec 3.4.24.40)		
	M14 (família de carboxipeptidase A)				
		RXF09091	Precursor de zinco-carboxipeptidase (ec 3.4.17.-)		
	M16 (família de pitrilisina)				
		RXF03441	Proteína de síntese de coenzima pqq (ec 3.4.99.-)		
		RXF01918	Zinco protease (ec 3.4.99.-)		
		RXF01919	Zinco protease (ec 3.4.99.-)		
		RXF03699	Peptidase de		

Classe	Família	RXF	Função Curada	Gene	Fisiologia
			processamento (ec 3.4.24.64)		
	M17 (família de leucil aminopeptidase)				
		RXF00285	Citosol amino-peptidase (ec 3.4.11.1)		Contribui para nutrição bacteriana
	M18				
		RXF07879	Aspartil amino-peptidase (ec 3.4.11.21)		
	M20				
		RXF00811	Succinildiaminopimelato dessuccinilase (ec 3.5.1.18)	<i>dapE</i>	
		RXF04052	Xaa-His dipeptidase (ec 3.4.13.3)		
		RXF01822	Precursor de carboxipeptidase G2 (ec 3.4.17.11)		
		RXF04892	N-acil-L-aminoácido amidoidrolase (ec 3.5.1.14)		
	M28 (família de aminopeptidase Y)				
		RXF03488	Precursor de proteína de conversão de isozima de fosfatase alcalina (ec 3.4.11.-)		
	M42 (família de gutamil aminopeptidase)				
		RXF05615	Aminopeptidase de desbloqueio (ec 3.4.11.-)		
	M22				
		RXF05817	O-sialoglicoproteína endopeptidase (ec		

Classe	Família	RXF	Função Curada	Gene	Fisiologia
			3.4.24.57)		
		RXF03065	Família de proteína de glico-protease		
	M23				
		RXF01291	Endopeptidase de parede celular, família M23/M37		
		RXF03916	Proteínas de membrana relacionadas às metaloendopeptidases		
		RXF09147	Endopeptidase de parede celular, família M23/M37		
	M24				
		RXF04693	Metionina aminopeptidas (ec 3.4.11.18)		Provável papel na remoção cotranslacional de metionina N-terminal
		RXF03364	Metionina aminopeptidase (ec 3.4.11.18)		Provável papel na remoção cotranslacional de metionina N-terminal.
		RXF02980	Xaa-Pro aminopeptidase (ec 3.4.11.9)		Envolvido em modificação da proteína retrocelular, na bactéria.
	M48 (família de Ste24 endopeptidase)				
		RXF05137	Proteína de choque térmico HtpX		
		RXF05081	Zinco metaloprotease (ec 3.4.24.-)		
	M50 (família de S2P protease)				
		RXF04692	Metaloprotease de membrana		

Classe	Família	RXF	Função Curada	Gene	Fisiologia
Serina peptidases					
	S1 (família de quimiotripsina)				
		RXF01250	Protease do (ec 3.4.21.)		
		RXF07210	Protease do (ec 3.4.21.)		
	S8 (família de subtilisina)				
		RXF06755	Serina protease (ec 3.4.21.-)		
		RXF08517	Serina protease (ec 3.4.21.-)		
		RXF08627	Serina protease extracelular (ec 3.4.21.-)		
		RXF06281	Precursor de serina protease extracelular (ec 3.4.21.-)		
		RXF08978	Serina protease extracelular (ec 3.4.21.-)		
		RXF06451	Serina protease (ec 3.4.21.-)		
	S9 (família de prolil oligopeptidase)				
		RXF02003	Protease ii (ec 3.4.21.83)		
		RXF00458	Hidrolase		
	S11 (família D-Ala-D-Ala carboxipeptidase A)				
		RXF04657	D-alanil-D-alanina endopeptidase (ec 3.4.99.-)		
		RXF00670	D-alanil-D-alanina carboxipeptidase (ec 3.4.16.4)		
	S13 (família D-Ala-D-Ala peptidase C)				
		RXF00133	D-alanil-meso-diaminopimelato en-		Age na síntese e remodelagem das paredes de células

Classe	Família	RXF	Função Curada	Gene	Fisiologia
			dopeptidase (ec 3.4.-.-)		bacterianas
		RXF04960	D-ala-nil-meso-diamin opimelato endopeptidase (ec 3.4.-.-)		
		S14 (família de ClpP endopeptidase)			
		RXF04567	Subunidade proteolítica de Clp protease atp-dependente (ec 3.4.21.92)	<i>clpP</i>	Suposto contribuir para eliminação de proteínas prejudiciais no calor térmico
		RXF04663	Subunidade proteolítica de Clp protease atp-dependente (ec 3.4.21.92)	<i>clpP</i>	Suposto contribuir para eliminação de proteínas prejudiciais no calor térmico
		S16 (família de Lon protease)			
		RXF04653	Protease atp-dependente La (3.4.21.53)		Suposto contribuir para eliminação de proteínas prejudiciais no calor térmico
		RXF08653	Protease atp-dependente La (3.4.21.53)		
		RXF05943	Protease atp-dependente La (3.4.21.53)		
		S24 (família de LexA)			
		RXF00449	Repressor de LexA (ec 3.4.21.88)		
		RXF03397	Repressor de LexA (ec 3.4.21.88)		
		S26 (família de peptidase de sinal I)			
		RXF01181	Peptidase de sinal I (ec 3.4.21.89)		Cliva os peptídeos de sinal das proteínas segregadas
		S33			

Classe	Família	RXF	Função Curada	Gene	Fisiologia
		RXF05236	Prolina imino-peptidase (ec 3.4.11.5)	<i>Pip3</i>	
		RXF04802	Prolina imino-peptidase (ec 3.4.11.5)	<i>Pip1</i>	
		RXF04808	Prolina imino-peptidase (ec 3.4.11.5)	<i>Pip2</i>	
	S41 (família de peptidase de processamento C-terminal)				
		RXF06586	Protease cauda-específica (ec 3.4.21.-)		
		RXF01037	Protease cauda-específica (ec 3.4.21.-)		
	S45				
		RXF07170	Penicilina Aci-lase (ec 3.5.1.11)	<i>pacB2</i>	
		RXF06399	Penicilina Aci-lase (ec 3.5.1.11)	<i>pacB1</i>	
	S49 (família de protease IV)				
		RXF06993	Possível protease sohb (ec 3.4.--)		
		RXF01418	Protease iv (ec 3.4.--)		
	S58 (família de DmpA aminopeptidase)				
		RXF06308	D-aminopeptida se (ec 3.4.11.19)		
Treonina peptidases					
	T1 (família de proteassoma)				
		RXF01961	Protease atp-dependente hslV (ec 3.4.25.-)	<i>hslV</i>	Julgado contribuir para eliminação de proteína prejudiciais no choque térmico

Classe	Família	RXF	Função Curada	Gene	Fisiologia
	T3 (família gama-glutamiltransferase)				
		RXF02342	Gama-glutamil-transpeptidase (ec 2.3.2.2)	ggt1	
		RXF04424	Gama-glutamil-transpeptidase (ec 2.3.2.2)	ggt2	
Peptidases não-classificadas					
	U32				
		RXF00428	Protease (ec 3.4.-.-)		
		RXF02151	Protease (ec 3.4.-.-)		
	U61				
		RXF04715	Muramolitetra-peptídeo carboxipeptidase (ec 3.4.17.13)		
	U62				
		RXF04971	Proteína de PmbA	pmbA	O produto do gene de PmbA ( <i>{Escherichia coli}</i> ) facilita a secreção do peptídeo antibiótico microcina B17, removendo um peptídeo líder de 26-aminoácido, N-terminal (Madison et al., 1997).
		RXF04968	Proteína de TldD		
Proteases de não-MEROPS					
		RXF00325	Proteína repressora C2		
		RXF02689	Dipeptidase microssomal (ec 3.4.13.19)		
		RXF02739	Dipeptidase membrana (ec		

Classe	Família	RXF	Função Curada	Gene	Fisiologia
			3.4.13.19)		
		RXF03329	Proteína citosólica hipotética		
		RXF02492	Xaa-Pro dipeptidase (ec 3.4.13.9)		
		RXF04047	Família de protease amino terminal de caax		
		RXF08136	Protease (proteína similar à transglutaminase)		
		RXF09487	Zinco metaloprotease (ec 3.4.24.-)		

Certas proteases de origem de *E. coli* são listadas na Tabela B.

TABELA B

Classe	Família	Código	Peptidase ou homólogo (subtipo)	Gene
Aspártico peptidases	A8	A08.001	Peptidase de sinal II	<i>ispA</i>
	A24	A24.001	Prepilina peptidase do tipo IV (proteína de EtpN (plasmídeo p0157))	<i>etpN</i>
		A24.001	Prepilina peptidase do tipo IV 1 (proteína de CofP)	<i>cofP</i>
		A24.001	Prepilina peptidase do tipo IV 1 (proteína de HofD)	<i>hofD/hopD/hopO</i>
		A24.003	Prepilina peptidase do tipo IV 2 (proteína de HopD)	<i>hopD/ECs4188</i>
		A24 não-designada	Peptidases da família A24A não-designada (proteína de ORF_F310)	<i>pppA/ORF_F310</i>
		A24 não-designada	Peptidases da família A24A não-designada (proteína de PilU (plasmídeo R721))	<i>pilU</i>
		A24	Peptidases da família A24A	<i>bfpP/bfpG</i>

Classe	Família	Código	Peptidase ou homólogo (subtipo)	Gene
		não-designada	não-designada (proteína de BfpP (plasmídeo pMAR2))	
		A24 não-designada	Peptidases da família A24A não-designada (proteína de PilU)	<i>PILU</i>
	A26	A26.001	Omptina	<i>ompT/ECS1663/B0565</i>
		A26.005	Proteinase SopA	<i>sopA</i>
Cisteína peptidases	C26	C26 não-designada	Peptidases da família C26 não-designada	<i>YC-JL/Z2490/ECS1875</i>
	C40	C40.004	spr g.p. (tipo <i>Escherichia</i> ) (proteína de spr)	<i>spr</i>
		C40 não-designada	Peptidases da família C40 não-designada (proteína de NlpC)	<i>n-lpC/C2104//Z3737/ECS2415</i>
		C40 não-designada	Peptidases da família C40 não-designada (proteína de YafL)	<i>YafL</i>
		C40 não-designada	Peptidases da família C40 não-designada (quitinase 3)	
		C40 não-designada	Peptidases da família C40 não-designada (proteína de YdhO)	<i>ydhO</i>
	C39	C39.005	Peptidase de processamento de colicina V (proteína de CvaB)	<i>cvaB</i>
		C39.005	Peptidase de processamento de colicina V (proteína de MtfB)	<i>mtfB</i>
		C39 não-designada	Peptidases da família C39 não-designada (proteína de secreção de H47 de microcina MchF)	<i>mchF/MCLB</i>
	C56	C56 não-designada	Peptidases da família C56 não-designada (proteína de YhbO)	<i>yhbo</i>
		C56 não-designada	Peptidases da família C56 não-designada (proteína de c4536)	<i>c4536</i>

Classe	Família	Código	Peptidase ou homólogo (subtipo)	Gene
Metalopeptidases	M1	M01.005	Alanil aminopeptidase (proteobactérias)	<i>pepN</i>
	M3A	M03.004	Oligopeptidase A	<i>prlC/opdA</i>
		M03.005	Peptidil-dipeptidase Dcp	<i>dcp/Z2160/ECS2147</i>
		M03.005	Peptidil-dipeptidase Dcp	<i>dcp</i>
	M41	M41.001	FtsH endopeptidase	<i>hflB/ftsH/ECS4057</i>
	M66	M66.001	StcE protease	<i>stcE</i>
	M15D	M15 não-designada	Peptidases da subfamília M15D não-designada (proteína de VanX)	<i>ddpX/vanX/B1488/Z2222/ECS2092</i>
	M16A	M16.001	Pitrilisina	<i>ptr/ECs3678</i>
	M16B	M16 não-designada	Peptidases da subfamília M16B não-designada (proteína de PqqL)	<i>pqqL/yddC</i>
	M17	M17.003	Aminopeptidase A (bactérias)	<i>pepA/xerB</i>
		M17.004	PepB aminopeptidase	<i>pepB/Z3790/ECS3389</i>
	M24A	M24.001	Metionil aminopeptidase 1	<i>map</i>
	M24B	M24.003	X-Pro dipeptidase (bactérias)	<i>pepQ/ECs4775</i>
		M24.004	Aminopeptidase P (bactérias)	<i>pepP</i>
		M24 não-designada	Peptidases da subfamília M24B não-designada (proteína de YqhT)	<i>yqhT/ypdF/B2385/c2924</i>
	M20A	M20.010	DapE peptidase (succinil diaminopimelato dessuccinilase)	<i>dapE/msgB/C2999</i>
		M20 não-designada	Peptidases da subfamília M20A não-designada (proteína de YgeY)	<i>ygey</i>
	M20B	M20.003	Peptidase T	<i>pepT/Z1832/ECS1572</i>
	M20C	M20.007	X-His dipeptidase	<i>pepD/pepH/ECs0264</i>

Classe	Família	Código	Peptidase ou homólogo (subtipo)	Gene
	M20D	M20 não-designada	Peptidases da família M20D não-designada (proteína de YdaJ)	<i>ydaJ/ECs1922</i>
	M28A	M28 não-designada	Peptidases da subfamília M28A não-designada (proteína de YfbL)	<i>yfbL</i>
	M28C	M28.005	IAP aminopeptidase	<i>iap</i>
	M42	M42 não-designada	Peptidases da família M42 não-designada (proteína de YjhO)	<i>yjhO</i>
		M42 não-designada	Peptidases da família M42 não-designada (proteína de FrvX)	<i>frvX</i>
		M42 não-designada	Peptidases da família M42 não-designada (proteína de FrvX)	<i>frvX/b2384/ypdE</i>
	M38	M38.001	Beta-aspartil dipeptidase	<i>iadA</i>
	M22	M22.001	O-sialoglicoproteína endopeptidase	<i>ygiD</i>
		M22.002	Proteína de yeaZ	<i>ye-aZ/C2211/Z2850/ECS2516</i>
	M23B	M23.006	YibP Peptidase (proteína de YibP)	<i>yibP</i>
		M23 não-designada	Peptidases da subfamília M23B não-designada (proteína de YebA)	<i>yebA</i>
	M48B	M48.002	HtpX endopeptidase	<i>HrpX</i>
		M48 não-designada	Peptidases da subfamília M48B não-designada	<i>YGGG/C3521</i>
		M48 não-designada	Peptidases da subfamília M48B não-designada	<i>YFGC/C3011</i>
		M48 não-designada	Peptidases da subfamília M48B não-designada (proteína de YggG)	<i>YggG/Z4280/EC S3811</i>
		M48 não-designada	Peptidases da subfamília M48B não-designada (proteína de YcaL)	<i>ycaL/C1047/Z1255/ECS0992</i>
	M50A	M50.004	YaeL protease (proteína de	<i>ec-</i>

Classe	Família	Código	Peptidase ou homólogo (subtipo)	Gene
			Yael)	<i>fE/YAEL/B0176/Z</i> <i>0187/ECS0178/C</i> <i>0213</i>
	M52	M52.001	HybD endopeptidase (proteína de HybD)	<i>hybD/ECS3878</i>
		M52.002	HyaD endopeptidase (proteína de HyaD)	<i>hyaD</i>
		M52.003	Hycl endopeptidase (proteína de Hycl)	<i>hycl/C3277</i>
Serina peptidases	S1B	S01.260	B1598 endopeptidase	<i>B1598</i>
	S1C	S01.273	Protease Do	<i>htrA/degP</i>
		S01.274	DegQ	<i>hho-AI/degQ/ECS410</i> <i>7/Z4593</i>
		S01.275	DegS	<i>hhoB/degS</i>
	S6	S06.002	EspP g.p. ( <i>Escherichia coli</i> )	<i>espP/pssA</i>
		S06.003	Tsh peptidase ( <i>Escherichia coli</i> ) (proteína de Tsh)	<i>tsh/hbp</i>
		S06.003	Tsh peptidase ( <i>Escherichia coli</i> )	<i>C0393</i>
		S06.004	Pet endopeptidase	<i>Sat</i>
		S06.004	Pet endopeptidase	
		S06.005	Pic endopeptidase ( <i>Shigella flexneri</i> )	<i>she/pic</i>
		S6 não-designada	Peptidases da família S6 não-designada (proteína de eatA)	<i>eatA</i>
		S6 não-designada	Peptidases da família S6 não-designada (proteína de c0350)	<i>C0350</i>
		S6 não-designada	Peptidases da família S6 não-designada (proteína de EspC)	<i>espC</i>
		S6 não-designada	Peptidases da família S6 não-designada (proteína de epeA)	<i>epeA</i>
		S6 não-designada	Peptidases da família S6 não-designada	

Classe	Família	Código	Peptidase ou homólogo (subtipo)	Gene
	S8A	S8 não-designada	Peptidases da subfamília S8A não-designada	
	S9A	S09.010	Oligopeptidase B	<i>ptrB</i>
		S09.010	Oligopeptidase B	<i>ptrB/C2255</i>
	S9X	S9 não-designada	Peptidases da família S9 não-designada	<i>YF-HR/C3060/b2534/Z3802</i>
	S11	S11.002	Murein-DD-endopeptidase	<i>pbpG</i>
		S11.003	Proteína de ligação de penicilina 6	<i>dacC/Z1066/ECS0919</i>
		S11.003	Proteína de ligação de penicilina 6 (proteína de ligação de penicilina pbp-6B)	<i>DacD/phsE/ECs2812</i>
		S11.003	Proteína de ligação de penicilina 6	<i>dacA</i>
	S12	S12 não-designada	Peptidases da família S12 não-designada (proteína de c2452)	<i>C2452</i>
		S12 não-designada	Peptidases da família S12 não-designada (proteína de YaiH)	<i>yaiH/C0480</i>
	S13	S13.001	D-Ala-D-Ala peptidase C	<i>dacB/ECs4061</i>
	S14	S14.001	Endopeptidase Clp (tipo 1)	<i>clpP/lopP/ECS0491</i>
		S14 não-designada	Peptidases da família S14 não-designada (proteína de ECs0829)	<i>Z0967/ECS0829</i>
		S14 não-designada	Peptidases da família S14 não-designada (proteína de ECs2960)	<i>H0022/Z2112/EC2960/L34</i>
	S16	S16.001	Lon protease	<i>lon/deg/ECs0493</i>
		S16 não-designada	Peptidases da família S16 não-designada (proteína de ECS1039)	<i>lonB/Z1305/ECS1039</i>
		S16 não-designada	Peptidases da família S16 não-designada (proteína de c1091)	<i>C1091</i>
	S24	S24.001	Repressor LexA (proteína de LexA)	<i>lexA/exrA</i>

Classe	Família	Código	Peptidase ou homólogo (subtipo)	Gene
		S24.003	Proteína de UmuD	
		S24.003	Proteína de UmuD	<i>umuD/C1631</i>
	S26	S26A	S26.001	<i>Peptidase de sinal I</i>
		S26.014	Proteína de transferência de plasmídeo de traF (proteína de TraF)	<i>traF</i>
	S33	S33 não-designada	Peptidases da família S33 não-designada (proteína de BioH)	<i>bi-oH/C4189/Z4767/ECS4255</i>
	S41A	S41.001	Protease de processamento C-terminal 1	<i>prc/tsp/ECS2540/Z2877//C2239</i>
	S45	S45.001	Precursor de penicilina G acilase	<i>pac</i>
	S49	S49.001	Protease IV	<i>sp-pA/ECS2472//C2170</i>
		S49.002	sohB endopeptidase	<i>sohB/ECS1844/Z2538//C1737</i>
	S51	S51.001	Dipeptidase E	<i>pepE</i>
	S54	S54 não-designada	Peptidases da família S54 não-designada (proteína de c0741)	<i>C0741</i>
		S54 não-designada	Peptidases da família S54 não-designada (glicerofosfato desidrogenase)	<i>glpG/C4201/Z4784/ECS4267</i>
Treonina peptidases	T1B	T01.006	Componente de Hs1V de HsIUV peptidase	<i>hsIV</i>
	T2	T02.002	Asparaginase	<i>ybiK/Z1051m/C0913</i>
	T3	T03.001	Gama-glutamiltransferase 1 (bacteriana)	<i>ggt/C4246</i>
	S41A	S41.001	Protease de processamento C-terminal 1	<i>prc/tsp/ECS2540/Z2877//C2239</i>
Peptidases não-classificadas	U6	U06.001	Mureína endopeptidase	<i>me-pA/ECS3212//C2874</i>
		U32	U32 não-designada	<i>Peptidases da família U32</i>

Classe	Família	Código	Peptidase ou homólogo (subtipo)	Gene
				<i>não-designada (proteína de YdcP)</i>
		U32 não-designada	Peptidases da família U32 não-designada (proteína de YegQ)	<i>yegQ/C2611</i>
		U32 não-designada	Peptidases da família U32 não-designada (proteína de YhbU)	<i>YH-BU/C3911/Z4519 /ECS4039</i>
		U35	U35 não-designada	<i>Peptidases da família U35 não-designada</i>
		U35 não-designada	Peptidases da família U35 não-designada (proteína de ECs4973)	<i>ECs4973</i>
		U49	U49.001	<i>Lit protease (<i>Escherichia coli</i>)</i>
		U61	U61.001	<i>Muramoyil-tetrapeptídeo carboxipeptidase</i>
		U61 não-designada	Peptidases da família U61 não-designada (proteína de MccF)	<i>mccF</i>
		U62	U62.001	<i>Peptidase de processamento de microcina 1</i>
		U62.002	Peptidase de processamento de microcina 2	<i>tldD/ECs4117</i>
		M9G.035	Endopeptidase ECP 32 ( <i>Escherichia coli</i> )	

Certas proteases de origem de *S. Cerevisiae* são listadas na Tabela C.

TABELA C

Classe	Família	Código	Peptidase ou homólogo (subtipo)	Gene
Aspártico peptidases	A1	A01.015	Pepsina de barreira	<i>Bar1</i>
		A01.018	Sacaropepsina	<i>pep4/pho9</i>
		A01.030	lapsina 1	<i>iap3</i>
		A01.031	lapsina 2	<i>mkc7</i>
		A01.035	lapsina 3	<i>YPS3</i>
		A01.UPW	Peptidases da família A1 não-designada	<i>YPS7/D9476.8/YDR349C</i>
		A01.UPW	Peptidases da família A1 não-designada (proteína de YIR039C)	<i>YIR039C</i>
	A2D	A02.022	Transposon de Ty3 ( <i>Saccharomyces cerevisiae</i> ) endopeptidase (retrotransposon Ty3-1)	<i>POL3/TY3-2 orfB/TY3B</i>
	A11B	A11.003	Transposon de Ty1 ( <i>Saccharomyces cerevisiae</i> ) endopeptidase (proteína B de transposon Ty1-17)	<i>Ty1B</i>
		A11.003	Transposon de Ty1 ( <i>Saccharomyces cerevisiae</i> ) endopeptidase (proteína B de transposon Ty1)	<i>Ty1B</i>
		A11.003	Transposon de Ty1 ( <i>Saccharomyces cerevisiae</i> ) endopeptidase (proteína B de transposon Ty1)	<i>Ty1B</i>
	A11X	A11.UPW	Peptidases da família A11 não-designada (retrotransposon Ty4)	
	A22B	A22.008	Proteína de YKL100c ( <i>Saccharomyces cerevisiae</i> )	<i>YKL100c</i>
Cisteína peptidases	C1B	C01.085	Bleomicina hidrolase (levadura)	<i>GAL6/YCP1/LAP3</i>

Classe	Família	Código	Peptidase ou homólogo (subtipo)	Gene
	C2	C02.008	Calpaína-7	Y- <i>MR154C/Cpl1/Rim13</i>
	C12	C12.002	Ubiquitinil hidrolase YUH1	<i>yuh1</i>
	C13	C13.005	Glicosilfosfatidilinositol:proteína transamidase	<i>d9798.2</i>
	C19	C19.002	Ubp1 ubiquitina peptidase	<i>ubp1</i>
		C19.003	Ubp2 ubiquitina peptidase	<i>ubp2</i>
		C19.004	Ubp3 ubiquitina peptidase	<i>ubp3</i>
		C19.005	Doa4 ubiquitina peptidase	<i>DOA4</i>
		C19.006	Ubp5 ubiquitina peptidase	<i>ubp5</i>
		C19.079	UBP6 ( <i>Saccharomyces cerevisiae</i> ) (proteína de YFRO 10W)	<i>yfr010w</i>
		C19.UPW	Peptidases da família C19 não-designada (proteína de YNL186W)	<i>YNL186W</i>
		C19.UPW	Peptidases da família C19 não-designada (UBP9)	<i>ubp9</i>
		C19.UPW	Peptidases da família C19 não-designada (proteína de YBL067C)	<i>YBL067C</i>
		C19.UPW	Peptidases da família C19 não-designada (proteína de YBRO58C)	<i>UBP12/YBR058C</i>
		C19.UPW	Peptidases da família C19 não-designada (ubiquitina hidrolase 16 carbóxi-terminal)	<i>UBP16/YPL072W/LPF12W</i>
		C19.UPW	Peptidases da família C19 não-designada (proteína de YMR304W)	Y- <i>MR304W/ym995.2.06</i>
		C19.UPW	Peptidases da família C19 não-designada (proteína de YMR223W)	Y- <i>MR223W/ym995.9.05</i>
		C19.UPW	Peptidases da família C19 não-designada (UBP7)	<i>ubp7</i>
		C19.UPW	Peptidases da família C19 não-designada (UBP13)	<i>ubp13</i>

Classe	Família	Código	Peptidase ou homólogo (subtipo)	Gene
	C44	C44.971	Glucosami-na-frutose-6-fosfato amino-transferase	
		C44.971	Glicosami-na-frutose-6-fosfato amino-transferase (glicosami-na-frutose-6-fosfato amino-transferase)	<i>gfa1</i>
	C48	C48.001	Ulp1 endopeptidase	<i>YPL020c</i>
		C48.005	Ulp2 endopeptidase (proteína de Smt4p)	<i>SMT4</i>
	C50	C50.001	Separase	<i>ESP1/YGR098C</i>
	C54	C54.001	ATG4 peptidase ( <i>Saccharomyces cerevisiae</i> )	<i>Apg4/Aut2</i>
	C56	C56.004	YDR533C g.p. ( <i>Saccharomyces cerevisiae</i> )	<i>Y-DR533C/D9719.36</i>
		C56.UPW	Peptidases da família C56 não-designada (proteína de YPL280W)	<i>YPL280W</i>
		C56.UPW	Peptidases da família C56 não-designada (proteína de YOR391C)	<i>YOR391C</i>
	I34	I34.001	Inibidor de sacaropepsina	<i>PAI3/YMR174C/YM8010</i>
Metalopeptidases	M1	M01.006	Ape2 aminopeptidase	<i>lap1/ape2</i>
		M01.007	Aap1' aminopeptidase	<i>AAP1</i>
		M01.007	Aap1' aminopeptidase	
		M01.017	Yin7 g.p. ( <i>Saccharomyces cerevisiae</i> )	<i>Yil137C</i>
		M01.UPW	Peptidases da família M1 não-designada (proteína de ynl045w)	<i>inl045w</i>
	M3A	M03.003	Sacarolisina	<i>prd1</i>
		M03.006	Peptidase intermediária mitocondrial	<i>MIP1</i>
	M16A	M16.007	Axl1 peptidase	<i>axl1</i>
		M16.008	Ste23 peptidase	<i>ste23</i>

Classe	Família	Código	Peptidase ou homólogo (subtipo)	Gene
		M16.UPA	Peptidases da subfamília M16A não-designada (proteína de orf1)	<i>Orf1</i>
	M16B	M16.003	Beta-subunidade de peptidase de processamento mitocondrial (beta)	<i>mas1/mif1</i>
	M16C	M16.UPC	Peptidases da subfamília M16C não-designada (proteína de YDR430C)	<i>YDR430C</i>
		M16.UPC	Peptidases da subfamília M16C não-designada (proteína de YOL098C)	<i>YOL098C</i>
	M16X	M16.971	Subunidade alfa de não peptidase de peptidase de processamento mitocondrial (alfa)	<i>mas2/mif2</i>
		M16.974	UCR2_HUMANO (proteína de núcleo de ubiquinol citocromo c reductase 2)	<i>ucr2/cor2/qcr2</i>
	M18	M18.001	Aminopeptidase I	<i>ape1/lap4</i>
		M18.UPW	Peptidases da família M18 não-designada (proteína de YHR113W)	<i>YHR113W</i>
	M20A	M20.005	Dipeptidase não-específica citosólica	<i>YFR044C</i>
	M20E	M20.002	Gly-X Carboxipeptidase	<i>Cps1/cps</i>
		M20.002	Gly-X Carboxipeptidase (pseudogene; deduzido da seqüência de nucleotídeo por MEROPS)	<i>AOE110, AO-E264, AOE130</i>
	M22	M22.003	Mername-AA017 peptidase (proteína de YKR038C)	<i>YKR038C</i>
		M22.UPW	Peptidases da família M22 não-designada (proteína de QR17)	<i>QR17</i>
	M24A	M24.001	Metionil aminopeptidase 1	<i>map1</i>
		M24.002	Metionil aminopeptidase 2	<i>ybl091c</i>
	M24B	M24.009	Aminopeptidase P1	<i>YLL029w</i>
		M24.026	Homólogo de aminopepti-	<i>YER078C</i>

Classe	Família	Código	Peptidase ou homólogo (subtipo)	Gene
			dase P (proteína de YER078C)	
		M24.UPB	Peptidases da subfamília M24B não-designada (proteína de YFR006W)	<i>yfr006w</i>
	M28A	M28.001	Aminopeptidase Y	<i>ape3</i>
	M28E	M28.006	Mername-AA063 peptidase (proteína de YDR415c)	<i>YDR415c</i>
	M28X	M28.974	Glutaminil ciclase	<i>YFR018C</i>
		M28.UPW	Peptidases da família M28 não-designada (proteína de YBR074W)	<i>YBR074W</i>
	M41	M41.002	Afg3 g.p. ( <i>Saccharomyces cerevisiae</i> ) (proteína de AGF3)	<i>agf3/yta10</i>
		M41.003	m-AAA protease (proteína de RCA1)	<i>rca1/yta12</i>
		M41.004	i-AAA protease	<i>yme1/yta1/osd1</i>
	M48A	M48.001	Ste24 endopeptidase	<i>STE24</i>
	M48B	M48.018	Oma1 endopeptidase ( <i>Saccharomyces cerevisiae</i> ) (proteína de YKR087C)	<i>Y-KR087C/YKR407</i>
	M49	M49.001	Dipeptidil-peptidase III	<i>YOL057W</i>
		M49.UPW	Peptidases da família M49 não-designada	
	M67A	M67.001	Poh1 peptidase	<i>RPN11/MPR1/YFR004W</i>
		M67.002	Metaloenzima de domínio Jab1/MPN	<i>YDL216c/D0888</i>
		M67.973	Subunidade reguladora de não-ATPase de proteassoma 26S	<i>RPN8/YOR261C</i>
Serina peptidases	SIC	S01.434	Nma111 endopeptidase ( <i>Saccharomyces cerevisiae</i> ) (proteína YNL123W)	<i>ynl123w</i>
	S8A	S08.052	Cerevisina	<i>prb1</i>
		S08.UPA	Peptidases da subfamília S8A não-designada (proteína de YSP3)	<i>YSP3</i>

Classe	Família	Código	Peptidase ou homólogo (subtipo)	Gene
		S08.UPA	Peptidases da subfamília S8A não-designada (proteína de YCR54C)	<i>YCR54C</i>
	S8B	S08.070	Cexina	<i>kex2</i>
	S9B	S09.005	Dipeptidil aminopeptidase A	<i>ste13/yci1</i>
		S09.006	Dipeptidil aminopeptidase B (fungo)	<i>dap2</i>
	S9X	S09.UPW	Peptidases da família S9 não-designada (proteína de Ynl320w)	<i>YNL320W</i>
	S10	S10.001	Carboxipeptidase Y	<i>prc1</i>
		S10.007	Kex carboxipeptidase	<i>kex1</i>
		S10.UPW	Peptidases da família S10 não-designada (proteína de YBR139W)	<i>ybr139W</i>
	S16	S16.002	PIM1 endopeptidase	<i>lon/pim1</i>
	S26A	S26.002	Protease de membrana interna mitocondrial 1 (1)	<i>imp1</i>
		S26.012	Protease de membrana interna mitocondrial 2 (2)	<i>imp2</i>
	S26B	S26.010	Componente de 21 kDa de signalase (eucarioto)	<i>sec11</i>
		S33.UPW	Peptidases da família S33 não-designada	<i>ECM18/YDR125C</i>
		S33.UPW	Peptidases da família S33 não-designada	<i>ECM18/YDR125C</i>
	S54	S54.007	Proteína de Pcp1 ( <i>Saccharomyces cereviseae</i> ) (proteína de YGR101W)	<i>YGR101W</i>
	S59	S59.001	Nucleoporina 145	<i>Nup145</i>
Treonina peptidases	T1A	T01.010	Subunidade catalítica de proteassoma 1	<i>pre3</i>
		T01.011	Subunidade catalítica de proteassoma 2	<i>pup1</i>
		T01.012	Subunidade catalítica de proteassoma 3	<i>pre2/prg1</i>
		T01.983	Subunidade de proteassoma beta 3	<i>pup3</i>

Classe	Família	Código	Peptidase ou homólogo (subtipo)	Gene
		T01.984	Subunidade de proteassoma beta 2	<i>pre1</i>
		T01.986	Subunidade de proteassoma beta 1	<i>pre7/prs3</i>
		T01.987	Subunidade de proteassoma beta 4	<i>pre4</i>
	T1X	T01.971	Subunidade de proteassoma alfa 6	<i>prs2/prc2</i>
		T01.972	Subunidade de proteassoma alfa 2	<i>pre8/prs4</i>
		T01.973	Subunidade de proteassoma alfa 4	<i>pre9/prs5</i>
		T01.974	Subunidade de proteassoma alfa 7	<i>pre6</i>
		T01.975	Subunidade de proteassoma alfa 5	<i>pup2</i>
		T01.976	Subunidade de proteassoma alfa 1	<i>pre5</i>
		T01.977	Subunidade de proteassoma alfa 3	<i>pre10/prs1/prc1</i>
	T3	T03.012	Gama-glutamiltransferase (Saccharomyces) (proteína de YLR299w)	<i>L8003.4</i>
	T5	T05.002	Precursor de ornitina acetil-transferase	<i>arg7/emc40/YMR062C</i>
Peptidases não-classificadas	U48	U48.001	Prenil protease 2	<i>rce1</i>

### MODULADORES DE DOBRAMENTOS

Os genes ou produtos de gene sobre-regulados identificados podem ser um ou mais moduladores de dobramento. Moduladores de dobramento podem por exemplo ser proteínas de HSP70, proteínas de HSP110/SSE, proteínas de HSP40 (relacionadas a DNAJ), proteínas similares a GRPE, proteínas de HSP90, proteínas de CPN60 e de CPN10, chaperoninas citosólicas, proteínas de HSP100, HSPs Pequena, Calnexina e calreticulina, PDI e proteínas relacionadas a tiorredoxina, Peptidil-prolil isome-

5

rases, Ciclofilina PPIases, proteínas de ligação de FK-506, Parvulina PPIases, chaperoninas individuais, chaperonas proteína-específicas ou chaperonas intramoleculares. Moduladores de dobramentos são em geral descritos em "Guidebook to Molecular Chaperones and Protein-Folding Catalysts" (1997) 5 ed. M. Gething, Universidade de Melbourne, Austrália.

As chaperonas molecular mais bem caracterizadas no citoplasma de *E. coli* são sistemas dependentes de DnaK-DnaJ-GrpE ATP- e de GroEL-GroES. Com base em estudos *in vitro* e considerações de homologia, várias proteínas citoplásmicas adicionais foram propostas para funcionarem 10 como chaperonas moleculares em *E. coli*. Estas incluem ClpB, HtpG e IbpA/B que, como DnaK-DnaJ-GrpE e GroEL-GroES, são proteínas de choque térmico (Hsps) pertencendo ao regulon de tensão. A conformação *trans* de ligações Pro é energeticamente favorecida em cadeias de proteína nascentes; porém, ~5% de todas as ligações de prolil peptídeo são encontrados em uma 15 conformação *cis* nas proteínas nativas. A isomerização de *trans* para *cis* de ligações de X-Pro é limitativa em taxa no dobramento de muitos polipeptídeos e é catalisada *in vivo* através de peptidil prolil *cis/trans* isomerases (PPIases). Três PPIases citoplásmicas, SlyD, SlpA e fator desencadeador (TF), foram 20 identificadas até agora em *E. coli*. TF, uma proteína de 48 kDa associada às subunidades ribossômicas de 50 S que foram postuladas cooperar com as chaperonas em *E. coli* para garantir dobramento apropriado das proteínas recentemente sintetizadas. Pelo menos cinco proteínas (tiorredoxinas 1 e 2, e glutarredoxinas 1, 2 e 3, os produtos dos genes de *trxA*, *trxC*, *grxA*, *grxB* e *grxC*, respectivamente) estão envolvidos na redução de ligação em pontes de 25 dissulfeto que transientemente surgem nas enzimas citoplásmicas. Desse modo, genes identificados podem ser proteínas ou chaperonas de formação de ligação de dissulfeto que permitem formação de ligação de dissulfeto apropriada.

Certos moduladores de dobramento em *O. fluorescens* são listados na Tabela D. 30

Tabela D

RXF	Gene	Função	Família
<b>GroES/EL</b>			
rxf02095	<i>groES</i>	chaperona	Hsp10
rxf06767:: rxf02090	<i>groEL</i>	chaperona	Hsp60
RXF01748	<i>ibpA</i>	Proteína de choque térmico pequena (sHSP) IbpA PA3126; Age como um retentor para dobramento de GroESL	Hsp20
RXF03385	<i>hscB</i>	proteína de chaperona hscB	Hsp20
<b>Hsp70 (DnaK/J)</b>			
rxf05399	<i>dnaK</i>	chaperona	Hsp70
RXF06954	<i>dnaK</i>	chaperona	Hsp70
RXF03376	<i>hscA</i>	chaperona	Hsp70
RXF03987	<i>cbpA</i>	Proteína de ligação de DNA curvada, atividade similar a dnaJ	Hsp40
RXF05406	<i>dnaJ</i>	proteína de chaperona dnaJ	Hsp40
RXF03346	<i>dnaJ</i>	chaperonas moleculares (família de DnaJ)	Hsp40
<b>Hsp100 (Clp/Hsl)</b>			
RXF04587	<i>clpA</i>	subunidade de ligação de atp de clp protease atp-dependente ClpA	Hsp100
RXF08347	<i>clpB</i>	proteína de ClpB	Hsp100
RXF04654	<i>clpX</i>	subunidade de ligação de atp de clp protease atp-dependente clpX	Hsp100
RXF01957	<i>hslU</i>	subunidade de ligação de atp de hsl protease atp-dependente hslU	Hsp100
RXF01961	<i>hslV</i>	subunidade de ligação de atp de hsl protease atp-dependente hslV	Hsp100
<b>Hsp33</b>			
RXF04254	<i>yrfl</i>	chaperonina de 33 kDa (homólogo de proteína de choque térmico 33) (HSP33)	Hsp33
<b>Hsp90</b>			
RXF05455	<i>htpG</i>	proteína de chaperona htpG	Hsp90
<b>SecB</b>			

RXF	Gene	Função	Família
RXF022311	<i>secB</i>	chaperona específica de secreção SecB	SecB
<b>Isomerases de ligação de dissulfeto</b>			
RXF07017	<i>dsbA</i>	dissulfeto isomerase	DSBA oxidor-reductase
RXF08657	<i>ds-bA/dsbC/dsbG/fernA</i>	dissulfeto isomerase	DSBA oxidor-reductase
rxf01002	<i>ds-bA/dsbC</i>	dissulfeto isomerase	DSBA oxidor-reductase/ Tiorredoxina
rxf03307	<i>dsbC</i>	dissulfeto isomerase	glutarredoxina/ Tiorredoxina
Rxf04890	<i>dsbG</i>	dissulfeto isomerase	glutarredoxina/ Tiorredoxina
<b>Peptidiil-prolil cis-trans isomerases</b>			
RXF03768	<i>ppiA</i>	Peptidil-prolil cis-trans isomerase A (ec 5.2.1.8)	PPIase: tipo ciclofilina
RXF05345	<i>ppiB</i>	Peptidil-prolil cis-trans isomerase B	PPIase: tipo ciclofilina
RXF06034	<i>fklB</i>	Peptidil-prolil cis-trans isomerase FkIB	PPIase: tipo FKBP
RXF06591	<i>fklB/fkbP</i>	Peptidil-prolil cis-trans isomerase de proteína de ligação de fk506 (EC 5.2.1.8)	PPIase: tipo FKBP
RXF05753	<i>fklB/fkbP</i>	Peptidil-prolil cis-trans isomerase (ec 5.2.1.8)	PPIase: tipo FKBP
RXF01833	<i>slyD</i>	Peptidil-prolil cis-trans isomerase SlyD	PPIase: tipo FKBP
RXF04655	<i>tig</i>	fator desencadeador, ppiase (ec 5.2.1.8)	PPIase: tipo FKBP
RXF05385	<i>yaad</i>	peptidil-prolil cis-trans isomerase de 16 kDa do tipo FKBP provável (EC 5.2.1.8) (PPIase) (Rotamase)	PPIase: tipo FKBP
RXF00271	<i>yaad</i>	Peptidil-prolil cis-trans isomerase (ec 5.2.1.8)	PPIase: tipo FKBP

RXF	Gene	Função	Família
Chaperonas de montagem de pili (parecidas com papD)			
RXF06068	<i>cup</i>	Proteína de chaperona cup	papD de montagem de pili
RXF05719	<i>ecpD</i>	Proteína de chaperona ecpD	papD de montagem de pili
RXF03406	<i>ecpD;csuC</i>	Proteína de chaperona ecpD	papD de montagem de pili
RXF04296	<i>ecpD;cup</i>	Proteína de chaperona ecpD	papD de montagem de pili
RXF04553	<i>ecpD;cup</i>	Proteína de chaperona ecpD	papD de montagem de pili
RXF04554	<i>ecpD;cup</i>	Proteína de chaperona ecpD	papD de montagem de pili
RXF05310	<i>ecpD;cup</i>	Proteína de chaperona ecpD	papD de montagem de pili
RXF05304	<i>ecpD;cup</i>	Proteína de chaperona ecpD	papD de montagem de pili
RXF05073	<i>gltF</i>	função periplásica de chaperona de montagem de pili gram-negativa	papD de montagem de pili

Certos moduladores de dobramento em *E. coli* são listados na Tabela E.

Tabela E

Acesso de Uniprot	Uniprot ID	Anotação	Família
GroES/EL			
P05380	CH10_ECOLI	chaperonin de 10 kDa	Hsp10
P06139	CH60_ECOLI	chaperonin de 60 kDa	Hsp60
Hsp70(DnaK/J)			
P04475	DNAK_ECOLI	Proteína de chaperona dnaK	Hsp70
P77319	HSCC_ECOLI	Proteína de chaperona hscC	Hsp70
P36659	CBPA_ECOLI	Proteína de ligação de DNA curvada CbpA	Hsp40
P31680	DJLA_ECOLI	Proteína similar a DnaJ, rscG	Hsp40
P08622	DNAJ_ECOLI	Proteína de chaperona dnaJ	Hsp40

Acesso de Uniprot	Uniprot ID	Anotação	Família
P29131	FTSN_ECOLI	Proteína de divisão celular ftsN	Hsp40
P09372	GRPE_ECOLI	Proteína de GrpE	GrpE
P31658	HCHA_ECOLI	Proteína de chaperona hchA	Hsp31
<b>Hsp100 (Clp/Hsl)</b>			
P15716	CLPA_ECOLI	Subunidade de ligação de ATP de Clp proteases ATP-dependente	Hsp100
P03815	CLPB_ECOLI	Proteína de ClpB	Hsp100
P33138	CLPX_ECOLI	Subunidade de ligação de ATP de hsl proteases ATP-dependente clpX	Hsp100
P32168	HSLU_ECOLI	Subunidade de ligação de ATP de Clp protease ATP-dependente, clpY	Hsp100
<b>Proteína de choque térmico pequenas</b>			
P29209	IBPA_ECOLI	Proteína de choque térmico de 16 kDa A	Hsp16
P29210	IBPB_ECOLI	Proteína de choque térmico de 16 kDa B	Hsp16
<b>Proteína de choque térmico de 16 kDa B</b>			
P36662	TORD_ECOLI	Proteína de chaperona torD	TorD
P15040	SECB_ECOLI	Proteína de exportação de proteína secB	SecB
P45803	HSLO_ECOLI	Chaperonina de 33 kDa	Hsp33
P10413	HTPG_ECOLI	Proteína de chaperona htpG	Hsp90
HscAB			
P36541	HSCA_ECOLI	Proteína de chaperona hscA	Hsp66
P36540	HSCB_ECOLI	Proteína de Co-chaperona hscB	Hsp20
<b>Proteína veículo de lipoproteína</b>			
P61316	LOLA_ECOLI	Precursor de proteína veículo de lipoproteína de membrana externa	LolA
P61320	LOLB_ECOLI	Precursor de lolB de lipopro-	LolB

Acesso de Uniprot	Uniprot ID	Anotação	Família
		teína de membrana externa	
<b>Isomerases de ligação de dissulfeto</b>			
P24991	DSBA_ECOLI	Tiol: precursor de dsbA de proteína de interpermuta de dissulfeto	
P300018	DSBB_ECOLI	proteína de formação de ligação de dissulfeto	oxidoreductase de ligação de dissulfeto
P21892	DSBC_ECOLI	Tiol: precursor de dsbC de proteína de interpermuta de dissulfeto	
P36655	DSBD_ECOLI	Tiol: precursor de dsbD de proteína de interpermuta de dissulfeto (EC 1.8.1.8) (Proteína-dissulfeto reductase)	
P33926	DSBE_ECOLI	Tiol: precursor de dsbE de proteína de interpermuta de dissulfeto (proteína da biogênese de citocromo c ccmG)	
P77202	DSBG_ECOLI	Tiol: precursor de dsbG de proteína de interpermuta de dissulfeto	oxidoreductase de ligação de dissulfeto
<b>Peptidil-prolil cis-trans isomerases</b>			
P22257	TIG_ECOLI	fator desencadeador	PPIase: tipo FKBP
P45523	FKBA_ECOLI	precursor de fkpa de peptidil-prolil cis-trans isomerase do tipo FKBP	PPIase: tipo FKBP
P39311	FKBB_ECOLI	peptidil-prolil cis-trans isomerase de 22 kDa do tipo FKBP	PPIase: tipo FKBP
P22563	FKBX_ECOLI	peptidil-prolil cis-trans isomerase de 16 kDa do tipo FKBP	PPIase: tipo FKBP
P30856	SLYD_ECOLI	peptidil-prolil cis-trans isomerase do tipo FKBP slyD	PPIase: tipo FKBP
P20752	PPIA_ECOLI	precursor de peptidil-prolil cis-trans isomerase A	PPIase: tipo ciclofilina

Acesso de Uniprot	Uniprot ID	Anotação	Família
P23869	PPIB_ECOLI	peptidil-prolil cis-trans isomerase B -	PPIase: tipo ciclofilina
P39159	PPIC_ECOLI	peptidil-prolil cis-trans isomerase C	PPIase: tipo PPIC
P77241	PPID_ECOLI	peptidil-prolil cis-trans isomerase D	PPIase: tipo PPIC
P21202	SURA_ECOLI	precursor de surA de proteína de sobrevivência	PPIase: tipo parvulina
<b>Chaperonas de montagem de pili (similares a papD)</b>			
P53516	AFAB_ECOLI	precursor de afaB de proteína de chaperona	PapD de montagem de pili
P33128	ECPD_ECOLI	precursor de ecpD de proteína de chaperona	PapD de montagem de pili
P31697	FIMC_ECOLI	precursor de fimC de proteína de chaperona	PapD de montagem de pili
P77249	SFMC_ECOLI	precursor de sfmC de proteína de chaperona	PapD de montagem de pili
P75749	YBGP_ECOLI	precursor de ybgP de chaperona fimbrial hipotética	PapD de montagem de pili
P40876	YCBF_ECOLI	precursor de ycbF de chaperona fimbrial hipotética	PapD de montagem de pili
P75856	YCBR_ECOLI	precursor de ycbR de chaperona fimbrial hipotética	PapD de montagem de pili
P33342	YEHC_ECOLI	precursor de yehC de chaperona fimbrial hipotética	PapD de montagem de pili
P77599	YFCS_ECOLI	precursor de yfcS de chaperona fimbrial hipotética	PapD de montagem de pili
P28722	YHCA_ECOLI	precursor de yhcA de chaperona fimbrial hipotética	PapD de montagem de pili
P77616	YQIH_ECOLI	precursor de yqiH de chaperona fimbrial hipotética	PapD de montagem de pili
P42914	YRAI_ECOLI	precursor de yral de chaperona fimbrial hipotética	PapD de montagem de pili

Certos moduladores de dobramento de *S. cervisiae* são mostrados na tabela F.

Tabela F

Acesso Uniprot GroES/EL	Uniprot ID	Fonte GO	Anotação	Família
P19882	HS60_LEVE-DURA	GOA:interpro	proteína de choque térmico 60, precursor mitocondrial	Hsp60
P38228	TC62_LEVE-DURA	GOA:interpro	chaperona mitocondrial TCM62	Hsp60
P38910	CH10_LEVE-DURA	GOA:interpro	proteína de choque térmico de 10 kDa mitocondrial	Hsp10
<b>Hsp70 (DnaK/J)</b>				
P25491	MAS5_LEVE-DURA	GOA:interpro	proteína de importação de proteína mitocondrial MAS5, Ydj1	Hsp40
P10591	HS71_LEVE-DURA	PMID:9789005	proteína de choque térmico SSA1	Hsp70
P10592	HS72_LEVE-DURA	PMID:9448096	proteína de choque térmico SSA2	Hsp70
P11484	HS75_LEVE-DURA		proteína de choque térmico SSB1	Hsp70
P40150	HS76_LEVE-DURA		proteína de choque térmico SSB2	Hsp70
P09435	HS73_LEVE-DURA	PMID:7867784	proteína de choque térmico SSA3	Hsp70
P22202	HS74_LEVE-DURA		proteína de choque térmico SSA4	Hsp70
P25294	SIS1_LEVE-DURA	GOA:interpro	proteína de SIS1	Hsp40
P32527	ZUO1_LEVE-DURA	GO:0003754	Zuotina	Hsp40
P35191	MDJ1_LEVE-DURA	GOA:interpro	proteína de MDJ1, precursor mitocondrial	Hsp40
P12398	HS77_LEVE-DURA	PMID:8654364	proteína de choque térmico SSC1, precursor mitocondrial	Hsp70
P38523	GRPE_LEVEDURA	GOA:interpro	homólogo de proteína de GrpE, precursor mitocondrial, MGE1	GrpE
P14906	SC63_LEVE-	GOA:spkw	proteína de translocação	Hsp40

Acesso Uniprot GroES/EL	Uniprot ID	Fonte GO	Anotação	Família
	DURA		SEC63	
P16474	GR78_LEVE-DURA		GRP 78, BIP, Kar2	Hsp70
P25303	SCJ1_LEVE-DURA	GOA:interpro	proteína relacionada a DnaJ SCJ1	Hsp40
P39101	CAJ1_LEVE-DURA	GOA:interpro	proteína de CAJ1	Hsp40
P48353	HLJ1_LEVE-DURA	GOA:interpro	Proteína de HLJ1	Hsp40
P39102	XDJ1_LEVE-DURA	GOA:interpro	Proteína de XDJ1	Hsp40
P52868	YGM8_LEVEDURA	GOA:interpro	proteína de 41,0 kDa hipotética em região intergênica de CEG1-SOH1	Hsp40
P53940	YNH7_LEVE-DURA	GOA:interpro	proteína de 58,9 kDa hipotética em região intergênica de TPM1-MKS1	Hsp40
P38353	SSH1_LEVE-DURA		homólogo de proteína 61 de Séc	Hsp70
P36016	LHS1_LEVE-DURA	GOA:spkw	homólogo de proteína 70 de choque térmico LHS1, SSI1	Hsp70
P38788	YHMA_LEVEDURA	PMID:11054575	homólogo de proteína 70 de choque térmico YHR064C	Hsp70
<b>Hsp110/Sse</b>				
P32589	HS78_LEVE-DURA	PMID:10480867	homólogo de proteína de choque térmico SSE1	SSE
P32590	HS79_LEVE-DURA		homólogo de proteína de choque térmico SSE2	SSE
<b>Hsp100 (Clp/Hsl)</b>				
P31539	H104_LEVE-DURA	GOA:interpro	proteína de choque térmico 104	Hsp100
P33416	HSP7_LEVE-DURA		proteína de choque térmico 78, precursor mitocondrial	Hsp100
P38323	MCX1_LEVEDURA	GOA:interpro	chaperona similar a clpX mitocondrial MCX1	Hsp100
<b>Pequena proteína de choque térmico</b>				
P15992	HS26_LEVE-DURA	PMID:10581247	proteína de choque térmico 26	Hsp pequeño

Acesso Uniprot GroES/EL	Uniprot ID	Fonte GO	Anotação	Família
<b>Prefoldina</b>				
P48363	PFD3_LEVE-DURA	GOA:interpro	Subunidade de prefoldina provável 3	prefoldina
Q04493	PFD5_LEVE-DURA	GOA:interpro	subunidade de prefoldina 5	prefoldina
P43573	YFC3_LEVE-DURA	GOA:interpro	proteína de 91,4 kDa hipotética na região intergênica de STE2-FRS2	prefoldina
P46988	PFD1_LEVE-DURA	GOA:spkw	subunidade de prefoldina 1	KE2
P40005	PFD2_LEVE-DURA	GOA:spkw	subunidade de prefoldina 2	KE2
P53900	PFD4_LEVE-DURA	GOA:spkw	subunidade de prefoldina 4	KE2
P52553	PFD6_LEVE-DURA	GOA:spkw	subunidade de prefoldina 6	KE2
<b>Hsp 90</b>				
P02829	HS82_LEVE-DURA	GOA:interpro	proteína de choque térmico HSP82	Hsp90
P15108	HS83_LEVE-DURA	GOA:interpro	proteína de cognato de choque térmico HSC82	Hsp90
P06101	CC37_LEVE-DURA	GOA:spkw	co-chaperona de Hsp90 Cdc37	Cdc37
P33313	CNS1_LEVE-DURA	GOA:spkw	supressor 7 de ciclofinila 1	CNS1
P15705	STI1_LEVE-DURA	PMID:8972212	proteína de choque térmico STI1	
<b>Calnexina</b>				
P27825	CALX_LEVE-DURA	GOA:spkw	precursor de homólogo de calnexina	calnexina
<b>Complexo T de chaperoninas citosólicas</b>				
P12612	TCPA_LEVE-DURA	GOA:interpro	Proteína de complexo T 1, subunidade alfa	TCP-1, Hsp60
P39076	TCPB_LEVE-DURA	GOA:interpro	Proteína de complexo T 1, subunidade beta	TCP-1, Hsp60
P39078	TCPD_LEVE-DURA	GOA:interpro	Proteína de complexo T 1, subunidade delta	TCP-1, Hsp60
P40413	TCPE_LEVE-	GOA:interpro	Proteína de complexo T 1,	TCP-1,

Acesso Uniprot GroES/EL	Uniprot ID	Fonte GO	Anotação	Família
	DURA		subunidade epsilon	Hsp60
P39077	TCPG_LEVEDURA	GOA:interpro	Proteína de complexo T 1, subunidade gama	TCP-1, Hsp60
P42943	TCPH_LEVEDURA	GOA:interpro	Proteína de complexo T 1, subunidade eta	TCP-1, Hsp60
P47079	TCPQ_LEVEDURA	GOA:interpro	proteína de complexo T 1, subunidade teta	TCP-1, Hsp60
P39079	TCPZ_LEVEDURA	GOA:interpro	Proteína de complexo T 1, subunidade zeta	TCP-1, Hsp60
<b>Proteína específica</b>				
P48606	TBCA_LEVEDURA	GOA:spkw	Chaperona tubulina-específica A	Proteína específica
P53904	TBCB_LEVEDURA	GOA:spkw	Chaperona tubulina-específica B	Proteína específica
P46670	CIN2_LEVEDURA	GOA:spkw	co-fator de dobragem de tubulina C Cin2	Proteína específica
P40987	CIN1_LEVEDURA		co-fator de dobragem de tubulina D Cin1	Proteína específica
P39937	PAC2_LEVEDURA	GOA:spkw	co-fator de dobragem de tubulina E PAC2	Proteína específica
P21560	CBP3_LEVEDURA	GOA:spkw	proteína de CBP3, precursor mitocondrial	Proteína específica
Q12287	COXS_VEDURA	GOA:spkw	chaperona de cobre de oxidase de citocromo c	Proteína específica
P40202	LYS7_LEVEDURA	GOA:interpro	chaperona de cobre de superóxido dismutase 1	
Q02774	SHR3_LEVEDURA	PMID:1056425 5	Proteína de componente secretório SHR3	Proteína específica
P38293	UMP1_VEDURA	GOA:spkw	Fator de maturação de proteassoma UMP1	Proteína específica
P38784	VM22_LEVEDURA	PMID:7673216	proteína de montagem de ATPase vacuolar VMA22	Proteína específica
P38072	SCO2_VEDURA	GOA:spkw	Proteína de SCO2, precursor mitocondrial	proteína específica
P53266	SHY1_VEDURA	PMID:1138989 6	Proteína de SHY1	Proteína específica
P40046	VTC1_VEDURA	GOA:spkw	Chaperona de transportador vacuolar 1	Proteína específica

Acesso Uniprot GroES/EL	Uniprot ID	Fonte GO	Anotação	Família
P38958	PT00_LEVE-DURA	PMID:11498004	proteína de PET100, precursor mitocondrial	Proteína específica
<b>Isomerases de ligação de dissulfeto</b>				
P17967	PDI_LEVE-DURA	PMID:11157982	precursor de proteína dissulfeto isomerase	oxidoreductase de ligação de dissulfeto
P32474	EUG1_LEVE-DURA	PMID:11157982	precursor de EUG1 de proteína dissulfeto isomerase	oxidoreductase de ligação de dissulfeto
Q12404	MPD1_LEVEDURA	PMID:11157982	precursor de MPD1 de dissulfeto isomerase	oxidoreductase de ligação de dissulfeto
Q99316	MPD2_LEVEDURA	PMID:11157982	precursor de proteína dissulfeto isomerase MPD2 (EC 5.3.4.1)	oxidoreductase de ligação de dissulfeto
Q03103	ERO1_LEVE-DURA	PMID:9659913	precursor de oxidoreductina endoplásmica 1 (EC 1.8.4.-) (Proteína de oxidoreductase endoplásmica 1)	oxidoreductase de ligação de dissulfeto
P388866	FMO1_LEVEDURA	PMID:10077572	Monooxigenase tiol-específica (EC 1.14.13.-) (Monooxigenase flavina-dependente)	oxidoreductase de ligação de dissulfeto
<b>Peptidil-prolil cis-trans isomerases</b>				
P14832	CYPH_LEVEDURA	GOA:interpro	ciclofilina de peptidil-prolil cis-trans isomerase A/Cpr1/Cyp1/CPH1/Scc1	PPlase: tipo ciclofilina
P23285	CYPB_LEVE-DURA	GOA:interpro	Ciclofilina de peptidil-prolil cis-trans isomerase B/Cpr2/Cyp2	PPlase: tipo ciclofilina
P25719	CYPC_LEVEDURA	GOA:interpro	Peptidil-prolil cis-trans isomerase C/CYP3/CPR3, mitocondrial	PPlase: tipo ciclofilina
P25334	CYPR_LEVEDURA	GOA:interpro	Peptidil-prolil cis-trans isomerase CPR4/Scc3	
P35176	CYPD_LE-	GOA:interpro	Peptidil-prolil cis-trans iso-	PPlase:

Acesso Uniprot GroES/EL	Uniprot ID	Fonte GO	Anotação	Família
	VEDURA		merase D CypD/Cpr5	tipo ciclofilina
P53691	CYP6_LEVE-DURA	PMID:10942767	Peptidil-proline cis-trans isomerase CPR6	PPIase: tipo ciclofilina
P47103	CYP7_LEVE-DURA	PMID:10942767	Peptidil-proline cis-trans isomerase CYP7	PPIase: tipo ciclofilina
P53728	CYP8_LEVE-DURA	GOA:interpro	Peptidil-proline cis-trans isomerase CYP8	PPIase: tipo ciclofilina
Q02770	Q02770	GOA:interpro	Yp1064cp	PPIase: tipo ciclofilina
P20081	FKBP_LEVE-DURA	GOA:interpro	proteína de ligação de FK506 FKB1/RBP1	PPIase: tipo FKBP
P32472	FKB2_LEVE-DURA	GOA:interpro	proteína de ligação de FK506 FKBP-13/FKB-15/FKB2, FPR2	PPIase: tipo FKBP
P38911	FKB3_LEVE-DURA	GOA:interpro	proteína nuclear de ligação de FK506 FKB-70/Npi46/Fpr3	PPIase: tipo FKBP
Q06205	FKB4_LEVE-DURA	GOA:interpro	proteína de ligação de FK506 FPR4	PPIase: tipo FKBP
P22696	ES-SI_LEVEDURA	GOA:spkw	proteína de ESS1	PPIase: tipo parvulina
<b>Mista fracamente caracterizada</b>				
P27697	ABC1_LEVE-DURA	GOA:spkw	proteína de ABC1, precursor mitocondrial	ABC1
P53193	YGB8_LEVE-DURA	GOA:interpro	proteína de 21,8 kDa hipotética em região intergênica de CKB1-ATE1	Hsp20
P28707	YKL7_LEVE-DURA	PMID:9632755	proteína de 24,1 kDa em região intergênica de VMA12-APN1	P23/wos2
P38932	VP45_LEVE-DURA	PMID:11432826	proteína associada à classificação de proteína vacuolar 45	similar a SEC1
Q12019	MDN1_LEVEDURA	GOA:spkw	midasina	

## MANIPULAÇÃO GENÉTICA

Na etapa iii), o processo inclui alterar expressão do gene ou produto de gene compensatório identificado na célula recombinante através de modificação genética para fornecer uma célula recombinante modificada.

- 5 Após identificação de um ou mais genes sobre-regulados, proteínas ou processos metabólicos, o genoma do hospedeiro pode ser modificado. Certos genes ou produtos de gene, embora identificados como sobre-regulados, podem não estar disponíveis para modulação porque eles são essenciais à célula ou são conhecidos afetar outros processos que podem ser essenciais à 10 célula ou organismo.

O genoma pode ser modificado incluindo um gene exógeno ou elemento de promotor no genoma ou no hospedeiro com um vetor de expressão, intensificando a capacidade de um gene identificado para produzir mRNA ou proteína, ou deletando ou rompendo um gene ou elemento de

- 15 promotor, ou reduzindo a capacidade de um gene para produzir mRNA ou proteína. O código genético pode ser alterado, assim afetando a transcrição e/ou translação de um gene, por exemplo através de técnicas de substituição, deleção ("knock-out"), co-expressão ou de inserção ("knock-in"). Genes adicionais para uma proteína desejada ou seqüência reguladora que modula a 20 transcrição de uma seqüência existente podem também ser inseridos.

## RECOMBINAÇÃO

O genoma da célula hospedeira que expressa proteína ou peptídeo recombinante pode ser modificado por meio de um evento alvejante genético que pode ser por inserção ou recombinação por exemplo recombi-

- 25 nação homóloga. Recombinação homóloga refere-se ao processo de recombinação de DNA com base em homologia de seqüência. Recombinação homóloga permite modificações sítio-específicas em genes endógenos e desse modo novas alterações podem ser criadas em um genoma. Uma etapa em recombinação homóloga é permuta de filamento de DNA, que envolve um 30 emparelhamento de um dúplex de DNA de pelo menos um filamento de DNA que contém uma seqüência complementar para formar uma estrutura de recombinação intermediária contendo DNA de heterodúplex (ver, por exemplo

Radding, C. M. (1982) *Ann. Rev. Genet.* 16: 405; Pat. U.S. Nº 4.888.274). O DNA de heterodúplex pode ter várias formas, incluindo um filamento de três DNA contendo forma triplex em que um filamento complementar simples invade o dúplex de DNA (Hsieh, et al., *Genes and Development* 4: 1951 (1990); 5 Rao, et al., (1991) *PNAS* 88:2984)) e, quando dois filamentos de DNA complementares emparelham com um dúplex de DNA, uma juncção de recombinação de Holliday clássica ou estrutura qui (Holliday, R., *Genet. Res.* 5: 282 (1964)) pode formar, ou uma alça D dobrada ("Diagnostic Applications of Double-D Loop Formation" U.S. Nº 07/755.462, depositada em 4 de set, 1991).

10 Uma vez formada, uma estrutura de heterodúplex pode ser solucionada através de rompimento e permuta de filamento, de forma que toda ou uma porção de um filamento de DNA invasor é encaixada em um dúplex de DNA recipiente, adicionando ou substituindo um segmento do dúplex de DNA recipiente. Alternativamente, uma estrutura de heterodúplex pode resultar em 15 conversão do gene, em que uma seqüência de um filamento invasor é transferida para um dúplex de DNA recipiente mediante reparo das bases desemparelhadas usando o filamento invasor como um modelo (*Genes*, 3<sup>a</sup> Ed. (1987) Lewina, B., John Wiley, Nova Iorque, N. I.; Lopez, et al., *Nucleic Acids Res.* 15: 5643(1987)). Seja pelo mecanismo de rompimento e rejunção ou 20 pelo(s) mecanismo(s) de conversão de gene, formação de DNA de heterodúplex nas juntas homologamente emparelhadas pode servir para transferir informação de seqüência genética de uma molécula de DNA para outra.

Em recombinação homóloga, o DNA entrante interage e integra em um sítio no genoma que contém uma seqüência de DNA substancialmente 25 homóloga. Em integração não-homóloga ("aleatória" ou "ilícita"), o DNA entrante não integra com uma seqüência homóloga no genoma mas em outro lugar, em uma de um número grande de localizações potenciais. Vários documentos descrevem o uso de recombinação homóloga em células mamíferas.

30 Vários constructos podem ser preparados para recombinação homóloga em um lócus identificado. Usualmente, a construção pode incluir pelo menos 10 bp, 20 bp, 30 bp, 40 bp, 50 bp, 70 bp, 100 bp, 500 bp, 1kbp, 2

kbp, 4 kbp, 5 kbp, 10 kbp, 15 kbp, 20 kbp, ou 50 kbp de seqüência homólogo com o lócus identificado. Podem ser envolvidas várias considerações determinando a extensão de homologia de seqüências de DNA identificadas, como, por exemplo, o tamanho do lócus identificado, disponibilidade de seqüências, 5 eficiência relativa dos eventos de dupla troca no lócus identificado e a similaridade da seqüência identificada com outras seqüências.

O DNA alvo pode incluir uma seqüência em que o DNA substancialmente isogênico flanqueia as modificações de seqüência desejadas com uma seqüência identificada correspondente no genoma a ser modificado. A 10 seqüência substancialmente isogênica pode ser pelo menos cerca de 95%, 97-98%, 99,0-99,5%, 99,6-99,9% ou 100% idêntica à seqüência identificada correspondente (com exceção das modificações de seqüência desejadas). O DNA alvo e o DNA identificado podem compartilhar extensões de DNA de pelo menos cerca de 10, 20, 30, 50, 75, 150 ou 500 pares de base que são 100% 15 idênticos.

Os constructos de DNA podem ser projetados para modificar o produto de gene endógeno, identificado. A seqüência homóloga para identificar o constructo pode ter uma ou mais deleções, inserções, substituições ou combinações destas, projetadas para romper a função do produto de gene 20 resultante. Em uma modalidade, a alteração pode ser a inserção de um gene marcador selecionável fundido na estrutura de leitura com a seqüência a montante do gene identificado.

O genoma pode também ser modificado usando deleção insercional. Nesta modalidade, o genoma é modificado recombinando uma seqüência no gene que inibe a formação do produto de gene. Esta inserção ou 25 pode romper o gene inserindo um elemento separado, ou remover uma porção essencial do gene. Em uma modalidade, a deleção insercional inclui inserção de um gene que codifica para resistência a um estressor particular, como um antibiótico, ou para crescimento em um meio particular, por exemplo 30 para a produção de um aminoácido essencial.

O genoma pode também ser modificado pelo uso de transposons, que são elementos genéticos capazes de inserir em sítios em genomas

procariotes por mecanismos independentes de recombinação homóloga. Transposons podem incluir, por exemplo, Tn7 em *E. coli*, Tn554 em *S. aureus*, IS900 em *M. paratuberculosis*, IS492 de *Pseudomonas atlantica*, IS116 de *Streptomyces* e IS900 de *M. paratuberculosis*. Etapas acreditadas estar envolvidas em transposição incluem clivagem da terminação do transponson para render 3' OH; transferência do filamento, em que transposase reúne a terminação de 3'OH exposta do transponson e a seqüência identificada; e uma reação de transesterificação de etapa simples para render uma ligação covalente do transponson ao DNA identificado. A reação fundamental executada por transposase é em geral julgada ser corte ou permuta de filamento, o resto do processo é feito através de enzimas hospedeiro.

Em uma modalidade, um processo é fornecido para aumentar o nível de um gene identificado ou homólogo deste incorporando uma seqüência genética que codifica o gene ou homólogo no genoma através de recombinação. Em outra modalidade, um promotor é inserido no genoma para intensificar a expressão do gene identificado ou homólogo. Em uma modalidade separada, um processo é fornecido para diminuir a expressão de um gene identificado ou homólogo deste por recombinação com um gene inativo. Em outra modalidade, uma seqüência que codifica um gene diferente, que pode ter uma função separada na célula ou pode ser um gene repórter como um marcador de resistência ou um gene marcador do contrário detectável pode ser inserido em um genoma através de recombinação. Em ainda outra modalidade, uma cópia de pelo menos uma porção do gene identificado que foi mutado para uma ou mais localizações é inserida no genoma através de recombinação. A versão mutada do gene identificado não pode codificar uma proteína, ou a proteína codificada pelo gene mutado pode ser dada inativa, a atividade pode ser modulada (aumentada ou diminuída), ou a proteína mutante pode ter uma atividade diferente quando comparada à proteína nativa.

Há estratégias para knock-out genes em bactérias, que foram em geral exemplificadas em *E. coli*. Uma via é clonar um fragmento de DNA interno no gene em um vetor contendo um gene de resistência antibiótico (por exemplo ampicilina). Antes de as células serem transformadas por meio de

transferência conjugativa, transformação química ou eletroporação (Puehler, et al. (1984) *Advanced Molecular Genetics* Nova Iorque, Heidelberg, Berlim, Tóquio, Springer Verlag), uma origem de replicação, como a replicação de plasmídeo vegetativo (o lócus de *oriV*), é excisada e o fragmento de DNA restante é re-ligado e purificado (Sambrook, et al. (2000) *Molecular cloning: A laboratory manual*, terceira edição, Cold Spring Harbor, Nova Iorque, Cold Spring Harbor Laboratory Press). Alternativamente, plasmídeos resistentes a antibióticos tendo uma origem de replicação de DNA podem ser usados. Após transformação, as células foram banhadas por exemplo em placas de ágar LB contendo os antibióticos apropriados (por exemplo 200 µg/ml de ampicilina). As colônias que crescem nas placas contendo os antibióticos presumivelmente sofreram um evento de recombinação simples (Snyder, L., W. Champness, et al. (1997) *Molecular Genetics of Bacteria* Washington DC, ASM Press) que leva à integração do fragmento de DNA inteiro no genoma no lócus homólogo. Análise adicional das células resistentes a antibióticos para verificar que o knock-out de gene desejado ocorreu no lócus desejado é por exemplo por PCR diagnóstica (McPherson, M. J., P. Quirke, et al. (1991) *PCR: A Practical Approach* Nova Iorque, Oxford University Press). Aqui, pelo menos dois iniciadores de PCR são projetados: um que hibrida fora da região de DNA que foi usado para a construção do knock-out do gene; e um que hibridar dentro da cadeia principal de plasmídeo restante. Amplificação de PCR com sucesso do fragmento de DNA com o tamanho correto seguido por análise de seqüência de DNA verificará que o knock-out do gene ocorreu na localização correta no cromossomo bacteriano. O fenótipo da cepa mutante recentemente construída pode depois ser analisado por por exemplo eletroforese em gel de poliacrilamida de SDS (Simpson, R., J. (2003) *Proteins and Proteomics - A Laboratory Manual*, Cold Spring Harbor, Nova Iorque, Cold Spring Harbor Laboratory Press).

Uma via alternativa para gerar um knock-out de gene é pelo uso de um replicon sensível à temperatura, como o replicon de pSC101 para facilitar a substituição do gene (Hamilton, et al. (1989) New process for generation deletions and gene replacements in *Escherichia coli*. *Journal of Bac-*

*teriology* 171(9): 4617-22). O processo prossegue através de recombinação homóloga entre um gene em um cromossomo e seqüências homólogas realizadas em uma temperatura de plasmídeo sensível à replicação de DNA. Transformação do plasmídeo no hospedeiro apropriado foi possível selecionar para integração do plasmídeo no cromossomo a 44°C. Crescimento subsequente destes cointegrates a 30°C leva a um segundo evento de recombinação, resultando em sua resolução. Dependendo de onde o segundo evento de recombinação ocorre, o cromossomo terá sofrido uma substituição de gene ou terá retido a cópia original do gene.

Outras estratégias para inibir expressão de produtos de gene particulares foram desenvolvidas. Por exemplo, interferência de RNA (RNAi), particularmente usando interferência de RNA pequena (siRNA), foi desenvolvida para extensivamente reduzir ou mesmo eliminar a expressão de um produto de gene particular. siRNAs são moléculas de RNA curtas, bi-filamentares que podem alvejar mRNAs complementares para degradação. RNAi é o fenômeno em que a introdução de um RNA bifilamentar suprime a expressão do gene homólogo. Moléculas de dsRNA são reduzidas *in vivo* para siRNAs de 21-23 nt que são os mediadores do efeito de RNAi. Sob introdução, os RNAs bifilamentares são processados em siRNAs de 20-25 nucleotídeos por uma enzima similar à RNase III chamada Dicer (etapa de iniciação). Depois, os siRNAs rearranjam-se em complexos contendo endorribonuclease conhecidos como complexos de silenciamento induzido por RNA (RISCs), desenrolando no processo. Os filamentos de siRNA subsequentemente guiam os RISCs para as moléculas de RNA complementares onde eles clivam e destroem o RNA cognato (etapa efetadora). Clivagem de RNA cognato ocorre próxima do meio da região ligada pelo filamento de siRNA. RNAi foi de forma bem sucedida usado para reduzir a expressão de gene em uma variedade de organismos incluindo peixe-zebra, nematódeos (*C. elegans*), insetos (*Drosophila melanogaster*), planária, cnidária, tripanossomas, camundongos e células mamíferas.

### MUTAÇÃO

O genoma pode também ser modificado mediante mutação de um

ou mais nucleotídeos em uma estrutura de leitura aberta que codifica um gene identificado, particularmente uma protease identificada. Técnicas para mutação genética, por exemplo mutagênese loco-dirigida, são bem conhecidas na técnica. Alguns métodos focalizam na geração de mutações aleatórias em DNA cromossômico como aqueles induzidos por raios X e químicas. Mutagênese alvejada para uma região definida de DNA inclui muitas técnicas, mais popular que outras. Métodos *in vitro* para mutagênese dirigida podem ser em geral agrupados em três categorias: i) processos que reestruturam os fragmentos de DNA, como mutagênese de cassette; ii) mutagênese aleatória localizada; e iii) mutagênese oligonucleotídeo-direcionada.

Mutagênese oligonucleotídeo-direcionada é com base no conceito que um oligonucleotídeo que codifica uma(s) mutação(ões) desejada(s) é anelado em um filamento do DNA de interesse e serve como um iniciador para iniciação da síntese de DNA. Desta maneira, o oligonucleotídeo mutagênico é incorporado no filamento recentemente sintetizado. Oligonucleotídeos mutagênicos incorporam pelo menos uma alteração de base mas podem ser projetados para gerar múltiplas substituições, inserções ou deleções. Exemplos incluem processos com base em PCR e praticamente todos os processos não baseados em PCR em uso hoje. Estas técnicas incluem seleção antibiótica positiva (Lewis, M. K. e Thompson, D. V. (1990) *Nucl. Acids Res.* 18, 3439; Bohnsack, R. N. (1996) *Meth. Mol. Biol.* 57, 1; Vavra, S. e Brondyk, W. H. (1996) *Promega Notes* 58, 30; *Altered Sites® II in vitro Mutagenesis Systems Technical Manual #TM001*, Promega Corporation), seleção de sítio de restrição único (Deng, W. P. e Nickoloff, J. A. (1992) *Anal. Biochem.* 200, 81), incorporação de uracila (Kunkel, T. A. (1985) *Proc. Natl. Acad. Sci. USA* 82, 488; Kunkel, T. A., Roberts, J. D. e Zakour, R. A. (1987) *Meth. Enzymol.* 154, 367) e incorporação de fosforotioato (Taylor, J. W., Ott, J. e Ecksteina, F., (1985) *Nucl. Acids Res.* 13, 8764; Nakamaye, K. e Ecksteina, F., (1986) *Nucl. Acids Res.* 14, 9679). Oligonucleotídeos podem também codificar uma biblioteca de mutações randomizando a composição de base nos sítios durante a síntese química resultando em oligonucleotídeos degenerados ou "dopados". A habilidade para localizar e especificar mutações é

grandemente intensificada pelo uso de oligonucleotídeos sintéticos hibridados com vetor de plasmídeo contendo inserção de DNA.

- O formato geral para mutagênese dirigida é: desnaturação de plasmídeo DNA que contém o modelo de interesse (cDNA, promotor, etc.)
- 5 para produzir regiões unifilamentares; anelamento de um oligonucleotídeo mutante sintético no filamento identificado; síntese de um filamento complementar novo usando, por exemplo, T4 DNA polimerase; e vedação do corte resultante entre a terminação do filamento novo e do oligonucleotídeo, por exemplo usando T4 DNA Ligase. O heterodúplex resultante é propagado
- 10 através de transformação, por exemplo, em *E. coli*. Processos de seleção e enriquecimento foram incorporados nos processos de mutagênese para grandemente melhorar a eficiência de recuperação do filamento mutante e taxas que chegam a 80-90% são possíveis. Numerosos processos existem para gerar tipos diferentes de mutações e intensificar a seleção do mutante.
- 15 Exemplos de processos para intensificar a seleção do mutante incluem seleção antibiótica positiva do filamento mutante, uso de um filamento de DNA contendo uracila, que pode ser seletivamente degradado *in vivo* e incorporação de análogo de dNTP, que pode dar um filamento de DNA de heterodúplex impérvio para digestão. Alguns métodos podem ser combinados, como
- 20 mutagênese de cassette e o uso de oligonucleotídeos "dopados" para criar uma biblioteca de mutações aleatórias em uma região pequena, definida.

Uma extensão dos assim-chamados processos "padrão" de mutagênese dirigida inclui aqueles que contam com amplificação de DNA, especificamente a reação em cadeia de polimerase (PCR). O ponto comum principal em mutagênese dirigida é o uso de um oligonucleotídeo mutagênico. O oligonucleotídeo mutagênico deveria hibridar eficazmente com o modelo. Para hibridização eficiente, pode haver, por exemplo, 100% de emparelhamento de base em qualquer terminação da seqüência identificada sem formação de estrutura secundária, mas pode também ser menor que 100% de identidade, como 98%, 95%, 92%, 90%, 85%, 80%, 70% ou apenas uma porção da seqüência pode ser idêntica. Para substituições pequenas, 10-15 bases que hibridam com qualquer um dos lados da disparidade são usual-

mente suficientes. A composição da terminação 3' do iniciador é particularmente importante visto que as polimerases tipicamente não se estendem de uma terminação 3' desemparelhada ou fracamente hibridada.

A base para mutagênese dirigida através de seleção antibiótica positiva é que um oligonucleotídeo ou oligonucleotídeos de seleção são anelados simultaneamente, com o oligonucleotídeo mutagênico, para reparar um gene de resistência a antibióticos (10-13). Seleção pelo filamento mutante é permitida mediante resistência a antibióticos do DNA mutado e sensibilidade do filamento não-mutado. Este método oferece um meio muito eficiente para gerar um número indefinido das mutações desejadas com pouco tempo em mãos.

Mutagênese sítio-direcionada pelo uso de um sítio de restrição único é com base nos processos de Deng e Nickoloff (Deng, W. P. e Nickoloff, J. A. (1992) *Anal. Biochem.* 200, 81). Neste método, um oligonucleotídeo de seleção que contém uma seqüência mutada para um sítio de restrição único é anelado simultaneamente com o oligonucleotídeo mutagênico. O oligonucleotídeo de seleção transmite o sítio não-essencial imune para restrição pela enzima correspondente. Seleção pelo filamento mutante é intensificada digerindo o fundo geral resultante de plasmídeos com a enzima de restrição única. A digestão lineariza o plasmídeo parental assim eficazmente diminuindo sua habilidade para transformar bactérias.

Mutagênese sítio-direcionada através de incorporação de deoxiuridina contam com a habilidade de uma cepa hospedeira para degradar o DNA modelo contendo uracila (U) no lugar de timidina (T). Um número pequeno de dUTPs está incorporado no filamento modelo no lugar de dTTP em um hospedeiro que carece de atividades de dUTPase (*dut-*) e de uracil N-deglicosidase (*ung-*). (Uracila por si não é mutagênica e seus pares de base com adenina.) Normalmente, dUTPase degrada deoxiuridina e uracil N-deglicosidase remove qualquer uracila incorporada. Replicação de pós-mutação em uma cepa de *dut+* *ung+* é usada depois para degradar o filamento de DNA não-identificado. Este método requer que o DNA unifilamentado seja usado de forma que apenas um filamento contenha os Us que

são suscetíveis à degradação.

O método de incorporação de fosforotioato para mutagênese dirigida conta com a habilidade de um análogo de dNTP que contém um grupo tiol de dar DNA de heterodúplex resistente à digestão da enzima de restrição.

- 5 O filamento de mutante é estendido do oligonucleotídeo mutagênico e sintetizado na presença de dCTPalphaS. DNA modelo não-usado é removido através de digestão com uma exonuclease. Teoricamente, apenas DNA circular, heterodúplex permanece. O heterodúplex é depois entalhado, mas não cortado, no(s) sítio(s) de restrição. Exonuclease III é usada para digerir o
- 10 filamento entalhado e o fragmento restante depois age como um iniciador para repolimerização, criando um homodúplex mutante.

No método com base em reação em cadeia de polimerase (PCR) para gerar uma mutação em DNA, um modelo é amplificado usando um conjunto de iniciadores de oligonucleotídeo gene-específicos exceto que um

- 15 oligonucleotídeo, ou mais nos protocolos que usam amplificações múltiplas, contenha a mutação desejada. Variações incluem alterar o sítio de hibridização dos oligonucleotídeos para produzir múltiplos fragmentos de PCR de sobreposição com a mutação na sobreposição e o método de "mega-iniciador", que usa três oligonucleotídeos e duas rodadas de amplificação
- 20 em que um filamento do produto da primeira amplificação serve como um iniciador na segunda amplificação.

No método de extensão de sobreposição, iniciadores de oligodeoxirribonucleotídeo complementares (oligo) são usados e a reação em cadeia de polimerase para gerar dois fragmentos de DNA tendo terminações de sobreposição. Estes fragmentos são combinados em uma reação de 'fusão' subsequente em que as terminações de sobreposição anelam-se, permitindo a sobreposição de 3' de cada filamento servir como um iniciador para a extensão de 3' do filamento complementar. O produto de fusão resultante é amplificado também por PCR. Alterações específicas na seqüência de nucleotídeo (nt) podem ser introduzidas incorporando alterações de nucleotídeo nos iniciadores de oligo de sobreposição.

## CONSTRUCTOS DE VETOR

Em uma modalidade separada, a célula hospedeira é modificada incluindo um ou mais vetores que codificam um gene identificado, tipicamente um modulador de dobramento ou um co-fator de um modulador de dobramento. Em outra modalidade, a célula hospedeira é modificada intensificando um promotor para um modulador de dobramento ou um co-fator para um modulador de dobramento, incluindo acrescentar um promotor exógeno ao genoma de célula hospedeira.

Em outra modalidade, a célula hospedeira é modificada incluindo um ou mais vetores que codificam um inibidor de um gene compensatório identificado, como um inibidor de protease. Um tal inibidor pode ser uma molécula anti-sentido que limita a expressão do gene compensatório identificado, um co-fator do gene identificado ou um homólogo do gene identificado. Anti-sentido é em geral usado para referir-se a uma molécula de ácido nucléico com uma seqüência complementar para pelo menos uma porção do gene identificado. Além disso, o inibidor pode ser um RNA de interferência ou um gene que codifica um RNA de interferência. Em organismos eucarióticos, um tal RNA de interferência pode ser um RNA de interferência pequena ou uma ribozima, como descrito, por exemplo, em Burn, A. et al. ( 1998) *Nature* 391:806-11, Elbashir et al. (2001) *Genes & Development* 15(2):188-200, Elbashir et al. (2001) *Nature* 411(6836):494-8, Patentes U.S. Nºs 6.506.559 de Carnegie Institute, 6.573.099 de Benitec, Pedidos de Patente U.S. Nºs 2003/0108923 de Whitehead Inst. e 2003/0114409, Publicações do PCT Nºs WO03/006477, WO03/012052, WO03/023015, WO 03/056022, WO 03/064621 e WO 03/070966. O inibidor pode também ser outra proteína ou peptídeo. O inibidor pode, por exemplo, ser um peptídeo com uma seqüência de consenso para a protease ou proteína de protease. O inibidor pode também ser uma proteína ou peptídeo que pode produzir uma molécula inibidora direta ou indireta para a protease ou proteína de protease no hospedeiro. Inibidores de protease podem incluir Amastatina, E-64, Antipaína, Elastatinal, APMSF, Leupeptina, Bestatina, Pepstatina, Benzamidina, 1,10-Fenantrolina, Quimiostatina, Fosforamidon, 3,4-dicloroisocumarina, TLCK, DFP, TPCK.

Mais de 100 inibidores de protease de proteína de ocorrência natural foram identificados até agora. Eles foram isolados em uma variedade de organismos de bactérias para animais e plantas. Eles se comportam como inibidores de ligação firme reversíveis ou pseudo-irreversíveis de proteases impedindo

5 acesso do substrato ao sítio ativo através de obstáculo estérico. Seu tamanho é também extremamente variável de 50 resíduos (por exemplo BPTI: Inibidor de Tripsina Pancreático Bovino) até 400 resíduos (por exemplo alfa-1PI: Inibidor de Proteinase alfa-1). Eles são estritamente classe-específicos exceto as proteínas da família de alfa-macroglobulina (por exemplo alfa-2 macro-

10 globulina) que liga e inibe a maioria das proteases através de um mecanismo de captura molecular.

Um constructo vetor ou DNA exógeno pode ser transfeccionado ou transformado na célula hospedeira. Técnicas para transfeccionar e transformar células eucarióticas e procarióticas respectivamente com ácidos nu-  
15cléicos exógenos são bem conhecidas na técnica. Estas podem incluir absorção mediada por vesícula de lipídio, transfecção mediada por fosfato de cálcio (co-precipitação de fosfato de cálcio/DNA), infecção viral, particularmente usando vírus modificados como, por exemplo, adenovírus modificados, microinjeção e eletroporação. Para transformação procariótica, as técnicas  
20 podem incluir absorção mediada por choque térmico, fusão de protoplasto bacteriano com células intactas, microinjeção e eletroporação. Técnicas para transformação de planta incluem transferência mediada por *Agrobacterium*, como por *A. tumefaciens*, microprojéteis de tungstênio ou ouro rapidamente impelidos, eletroporação, microinjeção e absorção mediada por polietileno  
25 glicol. O DNA pode ser DNA uni ou bifilamentar, linear ou circular, relaxado ou super-enrolado. Para várias técnicas para transfeccionar células mamíferas, ver, por exemplo, Keown et al. (1990) *Processes in Enzymology* Vol. 185, págs. 527-537.

Para eventos de recombinação, os constructos podem incluir uma  
30 ou mais seqüências de inserção que podem inserir ou transpor uma ou mais seqüência de ácido nucléico em uma seqüência diferente. Porém, a construção pode ser projetada para expressão exógena de um gene compensa-

tório identificado ou homólogo deste sem incorporação no DNA/genoma celular existente.

Os constructos podem conter um, ou mais que um, sítio de entrada de ribossoma interno (IRES). O constructo pode também conter um promotor operavelmente ligado à seqüência de ácido nucléico que codifica pelo menos uma porção do gene identificado, ou um co-fator do gene identificado, uma versão mutante de pelo menos uma porção do gene compensatório identificado, ou no caso de proteases, um inibidor do gene identificado. Alternativamente, o constructo pode ser sem promotor. Em casos em que o constructo não é projetado para incorporar no DNA/genoma celular, o vetor tipicamente contém pelo menos um elemento promotor. Além das seqüências de ácido nucléico, o vetor de expressão pode conter seqüências de marcador selecionáveis. Os constructos de expressão pode também conter sítios para iniciação, terminação de transcrição e/ou sítios de ligação de ribossoma. Os constructos identificados podem ser inseridos e expressados em qualquer célula procariótica ou eucariótica, incluindo, mas não limitada às células bacterianas, como *P. fluorescens* ou *E. coli*, células de levedura, células mamíferas, como células de CHO, ou células de planta.

Vetores de clonagem podem incluir por exemplo plasmídeo p-BR322 (Bolivar, Rodriguez et al. 1977), a série de pUC dos plasmídeos (Vieira e Messing 1982), pBluescript (Short, Fernandez et al. 1988), pACYC177 e pACYC184 (Chang e Cohen 1978). Promotores exógenos para o uso em tais constructos, incluem, mas não são limitados a, o promotor de PL de fago lambda, promotores lac de *E. coli*, trp de *E. coli*, phoA de *E. coli*, tac de *E. coli*, SV40 tardia, LTRs de SV40 tardia, retroviral, genes de PGK1, GAL1, de GAL10, CYC1, PH05, TRP1, ADH1, ADH2, fórmico fosfato desidrogenase, hexocinase, piruvato descarboxilase, fosfofructocinase, triose fosfato isomerase, fosfoglucone isomerase, feromônio de fator de emparelhamento de glucocinase alfa, promotor de PRB1, GUT2, de GPDI, promotor de metalotironína e/ou promotores virais mamíferos, como aqueles derivados de adenovírus e vírus de vacina. Outros promotores serão conhecidos a alguém versado na técnica.

Promotores para vetores exógenos, ou promotores exógenos projetados para ser inseridos no genoma pode ser com base em elementos de resposta específicos em uma célula. Por exemplo, os promotores podem ser responsivos aos compostos químicos, por exemplo a antranilato ou benzoato,

5 como descritos na Publicação do PCT Nº WO 2004/005221. Os constructos podem incluir um ou mais promotores. Estes podem ser independentes, ou podem estar em tandem. Por exemplo os promotores podem ser projetados de forma que um gene compensatório identificado seja sobre ou sub-regulado em um prazo particular com a proteína ou peptídeo recombinante. Por exemplo, em um caso em que o gene identificado é um modulador de dobramento, o modulador ou co-fator de dobramento pode ser induzido logo antes da indução da proteína ou peptídeo recombinante. Promotores podem incluir, mas não são limitados aos seguintes:

10

Promotor	Fonte	Regulação	Indução
lac	<i>E. coli</i>	lacI, lacI <sup>Q</sup>	IPTG
lacUV5	<i>E. coli</i>	lacI, lacI <sup>Q</sup>	IPTG
tac (híbrido)	<i>E. coli</i>	lacI, lacI <sup>Q</sup>	IPTG
trc (híbrido)	<i>E. coli</i>	lacI, lacI <sup>Q</sup>	IPTG
P <sub>syn</sub> (sintético)	<i>E. coli</i>	lacI, lacI <sup>Q</sup>	IPTG
trp	<i>E. coli</i>		inanição de triptofano
araBAD	<i>E. coli</i>	araC	1-arabinose
lpp <sup>a</sup>	<i>E. coli</i>	IPTG	lactose
1pp-lac (híbrido)	<i>E. coli</i>	lacI	IPTG
phoA	<i>E. coli</i>	phoB (positivo) phoR (negativo)	inanição de fosfato
recA	<i>E. coli</i>	lexA	ácido nalidíxico
proU	<i>E. coli</i>		osmolaridade
cst-1	<i>E. coli</i>		inanição de glicose
tetA	<i>E. coli</i>		tetraciclina
cadA	<i>E. coli</i>	cadR	pH
nar	<i>E. coli</i>	fnr	condições anaeróbicas
P <sub>L</sub>	λ	λ clts857	térmico (alteração para

Promotor	Fonte	Regulação	Indução
			42°C)
cspA	<i>E. coli</i>		térmico (alteração para abaixo de 20°C)
T7	T7	$\lambda$ clts857	térmico
Operador de T7-lac	T7	lacI <sup>q</sup>	IPTG
Operador de T3-lac	T3	lacI <sup>q</sup>	IPTG
Operador de T5-lac	T5	lacI, lacI <sup>q</sup>	IPTG
Gene de T4 32	T4		Infecção de T4
Operador de n-prM-lac	<i>Bacillus</i>	lacI <sup>q</sup>	IPTG
VHb	<i>Vitreoscilla</i>		oxigênio
Proteína A	<i>S. aureus</i>		

Constructos podem incluir marcadores de seleção para identificar células modificadas. Genes marcadores selecionáveis adequados incluem, mas não são limitados a: genes que conferem a habilidade para crescer em certos substratos de meios, como o gene tk (timidina cinase) ou o gene hprt (hipoxantina fosforibosiltransferase) que confere a habilidade para crescer em meio HAT (hipoxantina, aminopterina e timidina); o gene gpt bacteriano (guanine/xantina fosforibosiltransferase) que permite crescimento em meio MAX (ácido micofenólico, adenina e xantina). Ver, por exemplo, Song, K-Y., et al. (1987) *Proc. Nat'l Acad. Sci. U. S. A.* 84:6820-6824; Sambrook, J. et al. (1989) *Molecular Cloning - A Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y., Capítulo 16. Outros exemplos de marcadores selecionáveis incluem: genes que conferem resistência a compostos como antibióticos, genes que conferem a habilidade para crescer em substratos selecionados, genes que codificam proteínas que produzem sinais detectáveis como luminescência, como proteína fluorescente verde, proteína fluorescente verde intensificada (eGFP). Uma ampla variedade de tais marcadores é conhecida e disponível, incluindo, por exemplo, genes de resistência

a antibióticos como o gene de resistência à neomicina (neo) (Southern, P. e P. Berg, (1982) *J. Mol. Appl. Genet.* 1:327-341); e o gene de resistência à higromicina (hyg) ((1983) *Nucleic Acids Research* 11:6895-6911, e Te Riele, H., et al. (1990) *Nature* 348:649-651). Outros genes marcadores selecionáveis incluem: ácido acetoidróxi sintase (AHAS), fosfatase alcalina (AP), galactosidase beta (LacZ), glucoronidase beta (GUS), cloranfenicol acetiltransferase (CAT), proteína fluorescente verde (GFP), proteína fluorescente vermelha (RFP), proteína fluorescente amarela (YFP), proteína fluorescente ciana (CFP), peroxidase de raiz-forte (HRP), luciferase (Luc), nopalina sintase (NOS), octopina sintase (OCS) e derivados destes. Marcadores selecionáveis múltiplos estão disponíveis que confere resistência a ampicilina, bleomicina, cloranfenicol, gentamicina, higromicina, canamicina, lincomicina, metotrexato, fosfinotricina, puromicina e tetraciclina. Genes marcadores selecionáveis adicionais úteis nesta invenção, por exemplo, são descritos nas patentes U. S. Nos: 6.319.669; 6.316.181; 6.303.373; 6.291.177; 6.284.519; 6.284.496; 6.280.934; 6.274.354; 6.270.958; 6.268.201; 6.265.548; 6.261.760; 6.255.558; 6.255.071; 6.251.677; 6.251.602; 6.251.582; 6.251.384; 6.248.558; 6.248.550; 6.248.543; 6.232.107; 6.228.639; 6.225.082; 6.221.612; 6.218.185; 6.214.567; 6.214.563; 6.210.922; 6.210.910; 6.203.986; 6.197.928; 6.180.343; 6.172.188; 6.153.409; 6.150.176; 6.146.826; 6.140.132; 6.136.539; 6.136.538; 6.133.429; 6.130.313; 6.124.128; 6.110.711; 6.096.865; 6.096.717; 6.093.808; 6.090.919; 6.083.690; 6.077.707; 6.066.476; 6.060.247; 6.054.321; 6.037.133; 6.027.881; 6.025.192; 6.020.192; 6.013.447; 6.001.557; 5.994.077; 5.994.071; 5.993.778; 5.989.808; 5.985.577; 5.968.773; 5.968.738; 5.958.713; 5.952.236; 5.948.889; 5.948.681; 5.942.387; 5.932.435; 5.922.576; 5.919.445; e 5.914.233.

Deleções podem ser pelo menos cerca de 5 bp, 10 bp, 20 bp, 30 bp, 40 bp ou 50 bp, comumente pelo menos cerca de 100 bp, e em geral não mais que cerca de 20 kbp, onde a deleção normalmente pode incluir pelo menos uma porção da região de codificação incluindo uma porção de ou um ou mais éxons, uma porção de ou um ou mais íntrons, e pode ou não pode incluir uma porção das regiões de não-codificação de flanqueamento, particularmente a região de não-codificação de 5' (região reguladora transcri-

cional). Desse modo, a região homóloga pode estender-se além da região de codificação na região de não-codificação de 5' ou alternativamente na região de não-codificação de 3'. Inserções podem em geral não exceder a 10 kbp, usualmente não exceder a 5 kbp, em geral sendo pelo menos 50 bp, mais 5 usualmente pelo menos 200 bp.

A(s) região(ões) de homologia pode(m) incluir mutações onde as mutações podem também inativar o gene identificado, fornecendo um deslocamento de estrutura, ou alterando um aminoácido fundamental, ou a mutação pode corrigir um alelo disfuncional, etc. Usualmente, a mutação pode 10 ser uma alteração sutil, não excedendo cerca de 5% das seqüências de flanqueamento homólogas.

O constructo pode ser preparado de acordo com os processos conhecidos na técnica, vários fragmentos podem ser reunidos, introduzidos em vetores apropriados, clonados, analisados e depois manipulados também 15 até que o constructo desejado tenha sido alcançado (ver, por exemplo Figuras 5-11). Várias modificações podem ser feitas à seqüência, para permitir análise de restrição, excisão, identificação de sondas, etc. Mutações silenciosas podem ser introduzidas, como desejado. Em vários estágios, análise de restrição, seqüenciação, amplificação com a reação em cadeia de polimerase, 20 reparo de iniciador, mutagênese *in vitro*, etc. podem ser empregados. Processos para a incorporação de genes de resistência a antibióticos e fatores de seleção negativos são familiarizados àqueles de habilidade usual na técnica (ver, por exemplo, WO 99/15650; patente U.S. Nº 6.080.576; patente U.S. Nº 25 6.136.566; Niwa, et al., *J. Biochem.* 113:343-349 (1993); e Yoshida, et al., *Transgenic Research*, 4:277-287 (1995)).

O constructo pode ser preparado usando um vetor bacteriano, incluindo um sistema de replicação procariótico, por exemplo uma origem reconhecível por uma célula procariótica como *P. fluorescens* ou *E. coli*. Um marcador, igual ou diferente do marcador a ser usado para inserção, pode ser 30 empregado que pode ser removido antes da introdução na célula identificada. Uma vez o vetor contendo o constructo foi completado, ele pode ser também manipulado, como por deleção de certas seqüências, linearização, ou intro-

duzindo mutações, deleções ou outras seqüências na seqüência homóloga. Após manipulação final, o constructo pode ser introduzido na célula.

O processo pode ser iterativo. Em uma modalidade, após modificação do hospedeiro e da expressão da proteína recombinante no hospedeiro modificado, um perfil genético da célula hospedeira modificada é analisado para identificar um ou mais outros genes identificados na expressão que é alterada na célula hospedeira modificada. Em particular, genes compensatórios podem ser aqueles que mostram expressão aumentada no hospedeiro modificado expressando proteína recombinante quando comparada a uma célula hospedeira modificada que não expressa a proteína ou peptídeo recombinante, ou quando comparada a uma célula hospedeira integralmente alterada. O processo também inclui alterar a expressão do gene ou outros genes identificados e expressar a proteína ou peptídeo na célula duplamente modificada. Estas etapas podem ser iteradas para melhorar a expressão de proteína e podem ser repetidas uma, duas, três, quatro, cinco, seis, sete, oito, nove, ou pelo menos dez vezes.

#### PRODUÇÃO DE PROTEÍNA

O processo da invenção otimamente leva à produção aumentada de proteína ou peptídeo recombinante em uma célula hospedeira. A produção aumentada pode incluir uma quantidade aumentada de proteína por grama de proteína do hospedeiro em uma quantidade dada de tempo, ou pode incluir um aumento no comprimento de tempo que a célula hospedeira está produzindo proteína ou peptídeo recombinante. A produção aumentada pode também incluir uma melhoria nos requerimentos para crescimento da célula hospedeira recombinante. A produção aumentada pode ser uma produção aumentada de proteína ou peptídeo de comprimento total. Se a melhoria for em níveis aumentados de proteína, a proteína ou peptídeo pode ser produzido em um ou mais corpos de inclusão em uma célula hospedeira.

A produção aumentada pode ser alternativamente um nível aumentado de proteína ou peptídeo ativo por grama de proteína produzida, ou por grama de proteína do hospedeiro. A produção aumentada pode também ser um nível aumentado de proteína ou peptídeo recuperável, como proteína

solúvel, produzida por grama de recombinante ou por grama de proteína de célula hospedeira. A produção aumentada pode também ser qualquer combinação de nível total aumentado e nível de ativo ou solúvel aumentado de proteína.

5                  Produção aumentada é tipicamente medida comparando o nível de produção após um certo período de indução em uma célula modificada para a mesma indução na célula inalterada.

#### SOLÚVEL/INSOLÚVEL

A expressão melhorada de proteína recombinante pode ser um  
10 aumento na solubilidade da proteína. A proteína ou peptídeo recombinante pode ser produzido e restabelecido do citoplasma, periplasma ou meio extracelular da célula hospedeira. A proteína ou peptídeo pode ser insolúvel ou solúvel. A proteína ou peptídeo pode incluir uma ou mais seqüências alvos ou seqüências para ajudar a purificação.

15                  Em certas modalidades, a invenção fornece um processo para melhorar a solubilidade de uma proteína ou peptídeo recombinante em uma célula hospedeira. O termo "solúvel" como aqui usado significa que a proteína não é precipitada através de centrifugação entre cerca de 5.000 e 20.000 x gravidade quando girado durante 10-30 minutos em um tampão sob condições fisiológicas. Proteínas solúveis, ativas são capazes de exibir a função, e  
20 não fazem parte de um corpo de inclusão ou outra massa precipitada.

A invenção pode também melhorar a recuperação das proteínas ou peptídeos recombinantes ativos. Por exemplo, a interação entre um polipeptídeo identificado e um de origem, polipeptídeo variante, polipeptídeo  
25 segmento-substituído e/ou polipeptídeo resíduo-substituído pode ser medida por qualquer ensaio conveniente *in vitro* ou *in vivo*. Desse modo, ensaios *in vitro* podem ser usados para determinar qualquer interação detectável entre um polipeptídeo identificado e, por exemplo entre enzima e substrato, entre hormônio e receptor de hormônio, entre anticorpo e antígeno, etc. Tal detecção pode incluir a medição das alterações colorimétricas, alterações na radioatividade, alterações na solubilidade, alterações no peso molecular quando medido por eletroforese em gel e/ou processos de exclusão em gel,  
30

etc. Ensaios *in vivo* incluem, mas não são limitados a, ensaios para detectar efeitos fisiológicos, por exemplo ganho de peso, alteração no equilíbrio de eletrólito, alteração no tempo de coágulo sanguíneo, alterações na dissolução de coágulo e a indução de resposta antigênica. Em geral, qualquer ensaio *in vivo* pode ser usado desde que um parâmetro variável exista para detectar uma alteração na interação entre o polipeptídeo identificado e o de interesse. Ver, por exemplo, patente US Nº 5.834.250.

#### CITOPLÁSMICO/PERIPLÁSMICO/SECRETADO

Em certas modalidades, a proteína pode também ser segregada no periplasma se fundida em uma seqüência de secreção de sinal apropriada. Em uma modalidade, a seqüência sinal pode ser uma proteína de ligação de fosfato, um peptídeo sinal de secreção de proteína de ligação de Lys-Arg-Orn (LAObp ou KRObp), um peptídeo de sinal de secreção de Porina de Membrana Externa E (OprE), um peptídeo de sinal de secreção de azurina, um peptídeo de sinal de secreção de proteína de ligação de ferro (III) [Fe(III)bp] ou um peptídeo de sinal de secreção de lipoproteína B (LprB).

Em uma modalidade, nenhuma condição ou agente de promoção de ligação de dissulfeto adicional é requerido para restabelecer o polipeptídeo contendo ligação de dissulfeto identificado na forma ativa, solúvel da célula hospedeira modificada ou célula duplamente ou múltipla modificada. Em uma modalidade, o peptídeo, polipeptídeo, proteína transgênica(o) ou fragmento deste tem uma conformação intramolecular dobrada em seu estado ativo. Foi observado que as proteínas mamíferas complexas solúveis no citoplasma podem configurar-se apropriadamente ao posicionamento apropriado dos grupos tiol para formação de ligação de dissulfeto posterior no periplasma. Em uma modalidade, o peptídeo, polipeptídeo, proteína transgênica(o) ou fragmento contém pelo menos uma ligação de dissulfeto intramolecular em seu estado ativo; e talvez até 2, 4, 6, 8, 10, 12, 14, 16, 18 ou 20 ou mais ligações de dissulfeto.

Em uma modalidade, mais que 50% do peptídeo, polipeptídeo, proteína transgênica(o) ou fragmento deste expresso produzido será produzido como peptídeos, polipeptídeos, proteínas simples funcionais, ou frag-

mentos destes em forma solúvel, ativa ou forma insolúvel facilmente renaturada no citoplasma ou periplasma. Em outra modalidade cerca de 60%, 70%, 75%, 80%, 85%, 90%, 95% da proteína expressa são obtidos ou facilmente renaturados em forma ativa.

5

### EXEMPLOS

As cepas bacterianas usadas no estudo atual estão listadas na TABELA 1. Cepas de *P. fluorescens* foram crescidas em frascos de agitação a 30°C. OD<sub>575</sub> foi registrada para cada cepa em vários pontos de tempo.

**TABELA 1. VISÃO GERAL DAS CEPAS BACTERIANAS**

Cepa	Genótipo de Cepa Relevante	Plasmídeo	Proteína recombinante
MB214	cepa hospedeira de <i>P. Fluorescens</i>		
MB206	<i>pyrF</i>		
DC240	<i>pyrF</i>	pDOW2415	Nitrilase
DC271	<i>pyrF</i>	pDOW1323	pbp:hGH
DC280	<i>pyrF</i>	pDOW1339	plasmídeo apenas de vetor
DC369	<i>pyrF</i>	pDOW1426	hGH
DC462	<i>pyrF</i>	pDOW3501	GrpE, DnaKJ
DC463	<i>pyrF</i>	pDOW3501, pDOW1426	GrpE, DnaKJ, hGH
HJ104	<i>pyrF</i>	pDOW1349	hGH-COP
DC370	<i>pyrF, hslU</i>		
DC372	<i>pyrF, hslU</i>	pDOW1426	hGH
DC373	<i>pyrF, hslU</i>	pDOW1323	pbp:hGH
HJ105	<i>pyrF, hslU</i>	pDOW1349	hGH-COP
DC417	<i>pyrF, hslUV</i>		
HJ115	<i>pyrF, hslUV</i>	pDOW1426	hGH
HJ117	<i>pyrF, hslUV</i>	pDOW1349	hGH-COP

10

Os plasmídeos usados nos experimentos a seguir estão listados na Tabela 2.

**TABELA 2: VISÃO GERAL DOS PLASMÍDEOS**

Plasmídeos	Relevância
pDOW2236	vetor de clonagem
pDOW2240	Ptac <i>grpE-dnaKJ, pyrF<sup>+</sup></i>
pDOW2247	Pmtl nenhum gene recombinante; vetor vazio
pDOW3501	Pmti <i>grpE-dnaKJ, pyrF<sup>+</sup></i>
pDOW1349	<i>pyrF<sup>+</sup>, hGH::COP</i>
pDOW1426	<i>pyrF<sup>+</sup>, hGH</i>
pDOW1261-2	vetor suicida, <i>pyrF<sup>+</sup></i>
pDOW2050	usado para construção das cepas de deleção de <i>hsf/UV</i>

**COLETÂNEA DE AMOSTRA E ISOLAMENTO DE RNA**

Todas as amostras foram colhidas de uns experimentos de frascos de agitação de 250 ml padrão. As amostras foram tiradas em pontos de tempo diferentes como indicados nas figuras. A cada ponto de tempo, 10 ml de cultura de célula dos frascos de agitação foram colhidos e misturados com 10 ml de reagente *RNAlater* (Ambion, Austina, TX) para estabilizar o RNA.

**HIBRIDIZAÇÃO DE MICROARRANJO E ANÁLISE DE DADOS**

Para cada amostra de RNA, os nucleotídeos fluorescentes Cy3-dUTP ou Cy5-dUTP (Amersham Pharmacia, Piscataway, NJ) foram incorporados no cDNA em uma reação de transcrição reversa (RT) usando iniciador de hexâmero aleatório (Amersham). Os dois fundos gerais de cDNA marcados foram combinados e aplicados a uma lâmina de microarranjo. As lâminas de microarranjo contêm oligodeoxirribonucleotídeos amino-modificados de 50 mer (oligos) representando cada um por ORF de *P. fluorescens*. Cada oligo foi impresso duas vezes para manchas em duplicata em localização diferente usando o robô SDDC-2 (Virtek, Toronto, Canadá - agora distribuído através de Bio-Rad Laboratories, Hercules, CA) e pinos de SMP3 (TeleChem Internacional Inc., Sunnyvale, CA). As lâminas de microscópio usadas foram revestidas com uma resina epóxi positivamente carregada para ligação de DNA eficiente (MWG Inc, Alameda, CA). Após imprimir, as lâminas foram pós-processadas de acordo com as especificações de MWG. Um pa-

cote de software de BioDiscovery Inc. (El Segundo, CA) foi usado para facilitar a análise de dados. Este pacote consiste em módulos CloneTracker®, ImaGene®, GeneSight® e a base de dados de GeneDirector®. Cada lâmina hibridada foi varrida usando ScanArray 5000 (Packard BioScience, Billerica, MA)

5 para medir a fluorescência do cDNA Cy3- e Cy5-marcado ligado ao microarranjo. As imagens adquiridas foram quantificadas em ImaGene® e os dados brutos foram processados em GeneSight®. Durante a preparação dos dados, a intensidade das manchas para cada gene foi corrigida com a base; o sinal para o canal de Cy5 foi normalizado ao canal de Cy3 usando a intensidade de

10 sinal total para o arranjo inteiro; a razão normalizada de Cy5 para Cy3 para cada gene foi transformada de log2, e as réplicas foram combinadas.

#### ANÁLISE DA EXPRESSÃO DE PROTEÍNA ATRAVÉS DE SDS-PAGE

Alíquotas de cultura foram colhidas em vários pontos de tempo após indução de IPTG, normalizada para OD<sub>600</sub> de 10. Os lisados de célula

15 foram separados em frações solúveis e insolúveis através de centrifugação a 11000 g durante 5 minutos. As alíquotas de 2,5 ul foram combinadas com 5 ul de tampão de amostra de 2X NuPAGE LDS (Invitrogen, San Diego, CA), 50 uM de DTT e H<sub>2</sub>O para 10 µl, depois aquecidas para 95°C durante 5 minutos.

20 As proteínas foram separadas e visualizadas em 12% de géis de Nupage tingidos com Coomassie Blue usando Simply Blue Safestain (Invitrogen, San Diego, CA).

#### MEDIÇÃO DA ATIVIDADE DE FLUORESCÊNCIA

Rendimento da proteína foi também medido por atividade de fluorescência da fusão de proteína de fluorescência verde (COP) e hormônio de crescimento humano (hGH). O constructo de fusão de hgh::COP foi transformado nas cepas do tipo selvagem ou mutantes de *hsfU* e selecionadas na placa de ágar de glicose M9 sem uracila. A cultura de célula induzida por IPTG foi normalizada para OD<sub>600</sub> de cinco. Atividade de fluorescência relativa (RF) foi medida usando o espectrofluorímetro de microplaca de Spectramax Ge-

25 mini (Molecular Devices, Sunnyvale, CA) sob o ajuste apropriado (Ex485, filtro de passagem de banda Em538530).

30

EXEMPLO 1: ANÁLISE DE EXPRESSÃO DE GENE DE CEPAS QUE PRODUZEM PROTEÍNAS CITOPLÁSMICAS E PERIPLÁSMICAS - COMPARAÇÃO DE DIFERENTES PNOTOS DE TEMPO

Para estudo dos FMs e expressão de gene de protease durante a produção de proteína heteróloga, cepas de *P. fluorescens* DC206, 280, 240 e 271 foram usadas nos experimentos de microarranjo iniciais. DC206 é a cepa hospedeira e foi usada como um controle para crescimento de célula; DC280 tem um plasmídeo apenas de vetor e foi usado como um controle para o experimento de microarranjo; DC240 é DC206 com uma enzima citoplásica de nitrilase de codificação de plasmídeo que é solúvel; DC271 é DC206 com um plasmídeo que codifica o hormônio de crescimento humano periplásico (pbp::hGH) que é parcialmente insolúvel. Cepas foram crescidas em 200 ml de meio de frasco de agitação e crescimento das células foi monitorado medindo a OD<sub>575</sub>. Indução de IPTG foi executada 24 h após inoculação. Todas as cepas cresceram similarmente e as amostras de cultura foram tiradas logo antes (0 h) e 4 h após a indução para isolamento de RNA e perfilação transcripcional (TxP) usando microarranjo de DNA (Figura 1).

Os perfis genéticos, isto é, perfis transcripcionais foram com base na comparação da amostra de ponto de tempo de indução de 4 h com aquele da amostra de 0 h, as duas amostras foram rotuladas com corantes fluorescentes, Cy3-dUTP ou Cy5-dUTP, e co-hibridadas na mesma lâmina para cada cepa. Cada hibridização foi duplicada com experimentos de troca de corante (isto é, amostras foram marcadas com Cy3-dUTP ou Cy5-dUTP) (Tabela 3, lâminas 1 a 6). As lâminas hibridadas foram varridas usando um escâner a laser confocal. Intensidade de sinal para cada gene foi determinada e processada usando o pacote de software de microarranjo de Biodiscovery (El Segundo, CA). A razão de expressão dos dois pontos de tempo para cada gene foi calculada e as razões para todos os genes ao longo das cepas foram agrupadas com base no valor da razão e tendência entre as três cepas (DC280, DC240 e DC271) (Figura 2).

**TABELA 3.** Sumário dos experimentos de microarranjo executados nos Exemplos 1-3

Experi- mento	Lâmina	Cy3	Cy5	Experi- mento	Lâmina	Cy3	Cy5
DC280	1	amostra de 4 h	amostra de 0 h	0 h	7	DC240	DC271
	2	amostra de 0 h	amostra de 4 h		8	DC271	DC240
DC240	3	amostra de 4 h	amostra de 0 h	4 h	9	DC240	DC271
	4	amostra de 0 h	amostra de 4 h		10	DC271	DC240
DC271	5	amostra de 4 h	amostra de 0 h	DC369	11	amostra de 4 h	amostra de 0 h
	6	amostra de 0 h	amostra de 4 h		12	amostra de 0 h	amostra de 4 h

Para focalizar em FM e expressão de gene de protease em *P. fluorescens* sob a tensão imposta pela produção de proteína recombinante de nível alto, uma lista de FM e genes de protease foi comparada à análise de agrupamento. Após análise de agrupamento hierárquica de todos os genes de DC280, DC240 e DC271, FMs e proteases foram identificadas em dois agrupamentos (linhas nos agrupamentos 6 e 7; Figura 2).

Quatro genes no agrupamento 7 mostram expressão mais alta significativa em DC271 que expressa principalmente hormônio de crescimento humano periplásrmico insolúvel quando comparado a DC240 que produz nitrilase citoplásrmica solúvel ou DC280 que não superproduz nenhuma proteína. Os quatro genes são HslV de codificação de rxf01961, HslU de codificação de rxf01957, CbpA de codificação de rxf03987 e HtpG de codificação de rxf05455. O HslV de *E. coli* (ClpQ) e HslU (ClpY) juntos formam uma protease citoplásrmica. A subunidade pequena, HslV, é uma peptidase relacionada às α-subunidades proteassomais de eucariotes. A subunidade grande, HslU, é um ATPase com homologia para outra família de ATPases de Clp como ClpA e ClpX. CbpA de *E. coli* é um análogo da co-chaperona bem caracterizada DnaJ como julgado de não só sua estrutura mas também sua

função. O fenótipo de lesões em DnaJ, como sensibilidade de temperatura para crescimento, é restabelecido sob introdução do gene de *cbpA* em um plasmídeo de multicópias. HtpG de *E. coli* funciona como uma chaperona molecular ATP-independente *in vitro*. Ela reconhece e transientemente liga 5 intermediários de dobramentos estranhos, reduzindo sua concentração livre na solução e desse modo impedindo a agregação não específica.

Os genes foram agrupados no agrupamento 6 da Figura 2 novamente usando agrupamento hierárquico para identificar efeitos menos pronunciados. Figura 3 mostra que FMs e proteases foram identificados em 10 dois agrupamentos principais (linhas no agrupamento 6 e 8). Os dois FMs no agrupamento 8 são DnaK e DnaJ, duas chaperonas principais que são bem conhecidas trabalhar juntas para dobrar numerosas proteínas. Outra análise dos valores de expressão de genes do agrupamento 6 identificou um FM adicional, ClpX que é mais alta expressa em DC271 produzindo pbp::hGH 15 quando comparada a DC240 produzindo nitrilase ou DC280 que não superproduz nenhuma proteína. A proteína de choque térmico de ClpX de *E. coli* é homóloga aos membros da família de ATPases de HSP100/Clp procarióticos e eucarióticos. ClpX de *E. coli* foi isolado como um componente específico das proteases de Clp ATP-dependentes, que mantêm certos polipeptídeos em 20 uma forma competente para proteólise pela subunidade de protease de ClpP. ClpX pode agir como uma chaperona molecular, na ausência de ClpP, ativando as proteínas de iniciação envolvidas na replicação de DNA. FMs e proteases identificados importantes para produção de hGH periplásмico estão listadas na TABELA 4.

25 **TABELA 4.** Lista de FM e genes de protease cujos níveis de razão de mRNA de estado constante são mais altos em DC271 quando comparado a DC240 e DC280. Os valores listados são a razão de 4 h após indução de IPTG por 0 h.

Gene ID	DC280 (4 h vs. 0 h)	DC240 (4 h vs. 0 h)	DC271 (4 h vs. 0 h)	Gene	Função
RXF05455_1	0,8	0,6	5,3	<i>htpG</i>	Proteína de chaperona HtpG
RXF03987_1	1,0	0,5	5,2	<i>cbpA</i>	Proteína de ligação de DNA curvada

Gene ID	DC280 (4 h vs. 0 h)	DC240 (4 h vs. 0 h)	DC271 (4 h vs. 0 h)	Gene	Função
RXF01961_1	0,9	0,4	5,0	<i>hslV</i>	HslV Protease ATP-dependente (ec 3.4.25.-)
RXF01957_1	1,0		4,8	<i>hslU</i>	Hsl Protease ATP-dependente, subunidade de ligação de ATP HslU
RXF05399	1,0	0,6	3,3	<i>dnaK*</i>	Proteína de chaperona DnaK
RXF05399_1	1,3	0,6	3,0	<i>dnaK*</i>	Proteína de chaperona DnaK
RXF05406_1	1,2	0,7	3,0	<i>dnaJ</i>	Proteína de chaperona DnaJ
RXF04654_1	1,1	0,9	2,0	<i>clpX</i>	Clp protease ATP-dependente, subunidade de ligação de ATP ClpX

\* Para *dnaK*, duas sondas estão presentes na fatia de microarranjo e desse modo são fornecidos dois valores de expressão de gene.

#### EXEMPLO 2 - ANÁLISE DE EXPRESSÃO DE GENE DE CEPAS QUE PRODUZEM PROTEÍNAS CITOPLÁSMICAS E PERIPLÁSMICAS - COM-

##### 5 PARAÇÃO DIRETA DE CEPAS DIFERENTES

Para confirmar os resultados obtidos acima, experimentos de microarranjo adicionais foram executados por comparação direta das duas cepas DC271 e DC240 (lâminas 7 a 10 na Tabela 3). A comparação das duas cepas em 4 h após ponto de tempo de indução confirmou que um conjunto

10 quase idêntico de FM e genes de protease foi sobre-regulado em células que expressam pbp::hGH parcialmente solúveis (Tabela 5). Todos os genes listados na Tabela 5 são significativamente (isto é > 2 vezes) mais alto expressos nas cepas que produzem hGH parcialmente insolúvel quando comparado às células que produzem nitrilase completamente solúvel. Na comparação direta de DC271 a DC240, algumas proteínas adicionais foram identificadas quando comparadas à comparação de ponto de tempo (ver Ta-

bela 4) que mostrou valores de expressão de gene significativamente mais altos durante a produção de hGH parcialmente insolúvel. Aqueles genes incluíram ClpB de codificação de rxf08347, ClpA de codificação de rxf04587 e FkbP de codificação de rxf05753. O homólogo de ClpB de *E. coli* está envol-

5 vido na reativação dos corpos de inclusão juntos com DnaKJ-GrpE. ClpA de *E. coli* tem uma função de chaperona ou, quando junto com ClpP, degrada as proteínas. Em *E. coli*, FkbP funciona como uma peptidil-prolil isomerase.

10 **TABELA 5.** Lista de FM e genes de protease cujos níveis de mRNA de estado constante são mais altos em DC271 quando comparados a DC240. Os valores listados são a razão de DC271 para DC240 em 4 h após indução de IPTG.

Gene ID	DC271 vs. DC240 a 0 h	DC271 vs. DC240 a 4 h	Gene	Função
RXF03987_1	0,8	10,8	<i>cbpA</i>	Proteína de ligação de DNA curvada
RXF01957_1	0,9	10,0	<i>hslU</i>	Hsl protease ATP-dependente, subunidade de ligação de ATP HslU
RXF01961_1	0,7	10,0	<i>hsIV</i>	Protease ATP-dependente HsIV (ec 3.4.25.-)
RXF05455_1	0,7	7,8	<i>htpG</i>	Proteína de chaperona HtpG
RXF05406_1	1,0	4,7	<i>dnaJ</i>	Proteína de chaperona DnaJ
RXF08347_1	0,6	3,8	<i>clpB</i>	Proteína de ClpB
RXF05399_1	1,0	3,7	<i>dnaK*</i>	Proteína de chaperona DnaK
RXF05399	0,9	2,9	<i>dnaK*</i>	Proteína de chaperona DnaK
RXF04587_1	0,9	2,8	<i>clpA</i>	Clp protease ATP-dependente, subunidade de ligação de ATP ClpA

Gene ID	DC271 vs. DC240 a 0 h	DC271 vs. DC240 a 4 h	Gene	Função
RXF05753_1	1,1	2,1	<i>fkbP</i>	Peptidil-prolil cis-trans isomerase (ec 5.2.1.8)
RXF04654_1	1,2	2,0	<i>cpx</i>	Clp protease ATP-dependente, subunidade de ligação de ATP ClpX

\*Para *dnaK*, duas sondas estão presentes na fatia de microarranjo e desse modo dois valores de expressão de gene são fornecidos.

**EXEMPLO 3: ANÁLISE DE EXPRESSÃO DE GENE DE UMA CEPA QUE PRODUZ UMA PROTEÍNA CITOPLÁSMICA INSOLÚVEL**

5           Uma vez que DC271 parcialmente expressa hormônio de crescimento humano periplásrmico (pbp::hGH), ele foi investigado se FMs e genes de protease similares ou diferentes fossem sobre-regulados em uma cepa que principalmente expressa hGH citoplásrmico insolúvel. DC369 foi usado neste experimento. As 4 h após indução, a amostra foi comparada à amostra de ponto de tempo 0 h, e experimentos de microarranjo foram executados como mostrado na Tabela 3 (lâminas 11 e 12). Novamente, FM e genes de protease similares foram observados ser sobre-regulados indicando que os genes identificados estão envolvidos em dobramento citoplásrmico ao invés periplásrmico e degradação de proteína (Tabela 6). Um resumo de quais genes que foram identificados em qual experimento junto com a sobre-regulação de dobramento é mostrado no diagrama de Venn da Figura 4.

10           

15           

**TABELA 6.** Lista de FM e genes de protease cujos níveis de mRNA de estado constante foram mais altos em DC369 em 4 h após indução quando comparados a zero no tempo. Os valores listados são a razão de 4 h após indução

20 de IPTG para 0 h (logo antes da indução).

Gene ID	DC369 (4 h vs. 0 h)	Gene	Função
RXF01961_1	4,8	<i>hsIV</i>	Protease ATP-dependente HsIV (ec 3.4.25.-)
RXF01957_1	4,3	<i>hsIU</i>	HsI protease ATP-dependente, subunidade de ligação de ATP HsIU

Gene ID	DC369 (4 h vs. 0 h)	Gene	Função
RXF03987_1	4,1	<i>cbpA</i>	Proteína de ligação de DNA curvada
RXF05455_1	3,3	<i>htpG</i>	Proteína de chaperona HtpG
RXF05406_1	2,3	<i>dnaJ</i>	Proteína de chaperona DnaJ
RXF08347_1	2,2	<i>clpB</i>	Proteína de ClpB
RXF05399_1	2,1	<i>dnaK*</i>	Proteína de chaperona DnaK
RXF02095_1	2,0	<i>groES</i>	Chaperonina de 10 kDa GroES
RXF06767_1	2,0	<i>groEL</i>	Chaperonina de 10 kDa GroEL
RXF05399	1,8	<i>dnaK*</i>	Proteína de chaperona DnaK
RXF04587_1	1,7	<i>clpA</i>	Clp protease ATP-dependente, subunidade de ligação de ATP ClpA

\* Para *dnaK*, duas sondas estão presentes na fatia de microarranjo e desse modo são fornecidos dois valores de expressão de gene.

**EXEMPLO 4: GERAÇÃO DE UMA CEPA MUTANTE DE *hslU* EM DC206 DE *P. fluorescens***

5 O dois genes *hslU* foram encontrados estar entre os genes identificados mais altamente sobre-regulados. HslU é uma ATPase citoplasmica. A proteína homóloga em *E. coli* pode agir em combinação com uma segunda proteína para promover degradação de proteína energia-dependente em *E. coli*. HslU interage com HslV, uma proteína com homologia para as α-subunidades de proteasoma. Os homólogos de HslVU de *E. coli* foram relatados estar envolvidos em proteólise geral das proteínas desdobradas em Missiakas, D., et al. (1996) Identification and characterization of HslV HslU (ClpQ ClpY) proteins involved in overall proteolysis of misfolded proteins in *Escherichia coli*. *Embo J*15:6899-909. análise de seqüência de DNA sugeriu 10 que os genes de *hslVU* de *P. fluorescens* são prováveis de fazer parte de um operon bicistrônico (Figura 5).

15

Para verificar se HslVU estão de fato envolvidos na degradação de hGH, um cepa knock-out de *hslU* foi construída. Uma tal cepa foi gerada por inativação insercional de *hslU* (Figura 6). Um fragmento de DNA interno 20 de cerca de 550 bp para *hslU* foi clonado no vetor de pCR2.1-TOPO resistente

à canamicina. Uma vez que este vetor tem uma origem de replicação (ColE1) que é funcional em *E. coli* mas não em *P. fluorescens*, os plasmídeos construídos integrarão no cromossomo de DC206 através de recombinação homóloga para conferir resistência à canamicina. O sítio de inserção correto para as colônias resistentes à canamicina foi confirmado através de PCR de diagnóstico de colônia usando iniciadores que hibridam fora da região originalmente amplificada e dentro da cadeia principal de plasmídeo (Tabela 3). A cepa mutante de *hslU* construída foi designada DC370.

Os iniciadores foram projetados que amplificariam uma região interna de ~550 bp do gene de *hslU* (Tabela 7). O fragmento interno foi amplificado usando Taq Polimerase (Promega), purificado e clonado em vetor de pCR2.1-TOPO (Invitrogen, San Diego, CA). Os plasmídeos foram transformados em DC206 de *P. fluorescens* competente e selecionados nas placas de ágar de glicose de M9 suplementada com 250 µg/ml de uracila e 50 µg/ml de canamicina.

**TABELA 7. INICIADORES**

Iniciadores	Seqüência	Propósito
hsI_U_sens	accgaagtccgtatgtggg (SEQ ID NO: 1)	usado na amplificação de PCR do fragmento de <i>hslU</i> interno
Hsl_U_antis	aatcgcgctgcacgccttcg (SEQ ID NO: 2)	
hsI_F2	tcatcaaggcgtcgaaagcg (SEQ ID NO: 3)	
hsI_R2	tcaatcttgcaccatgcc (SEQ ID NO: 4)	usado em PCR de diagnóstico
M13_R	caggaaacagctatgac (SEQ ID NO: 5)	
M13_F	taaaaacgacggccag (SEQ ID NO: 6)	
hsI_UP	gtggcagccaccaaggctgc (SEQ ID NO: 7)	
hsI_middleUp	ccccacattggatggaggctacaaggggagagtctccacg (SEQ ID NO: 8)	usado em PCR de SOE o fragmento de DNA acima e abaixo de <i>hslUV</i>
hsI_middleDown	cgtggagactctcccttgtaaggcctactcaatgtggg (SEQ ID NO: 9)	
hsI_down	ggccaatggttggccacgcg (SEQ ID NO: 10)	
hsI_UpUp	tgccgacgccacagggtgc (SEQ ID NO: 11)	

Iniciadores	Seqüência	Propósito
hsl_DownDown	gcctggtactgcgactcg (SEQ ID NO: 12)	usado em PCR de diagnóstico
RC199	atatactagtaggaggtaacttatggctgacgaaca-gacgca (SEQ ID NO: 13)	usado na clonagem de <i>gr-pE-DnaKJ</i>
RC200	atattctagattacaggtcgccgaagaagc (SEQ ID NO: 14)	

#### COMPARAÇÃO DE EXPRESSÃO DA PROTEÍNA POR ANÁLISE DE SDS-PAGE

Para estudar o efeito do knockout do gene de *hslU*, duas expressões de proteína exógena foram comparadas entre a cepa de origem DC206 e a cepa mutante recentemente construída DC370. Os plasmídeos abrigando o gene codificando pbp::hGH (pDOW1323) e hGH (pDOW1426) foram cada um transformados em células de DC370 competentes e resultaram em cepas DC373 e DC372, respectivamente. Os experimentos de crescimento padrão em frasco de agitação foram executados com as quatro cepas.

Figura 7 mostra que as cepas do tipo selvagem e mutantes têm taxas de crescimento similares. As amostras foram operadas em géis de SDS-PAGE (Figura 8 e 9). Os resultados sugerem que o mutante produziu quantidades mais altas de proteínas devido à deleção da subunidade de protease HslU.

#### COMPARAÇÃO DE EXPRESSÃO DA PROTEÍNA POR ATIVIDADE DE FLUORESCÊNCIA

Uma vez que o efeito observado da falta de HslU no rendimento de hGH é difícil para quantificar usando análises de SDS-PAGE, o perfil temporal de produção de proteína foi monitorado pela fluorescência de uma proteína de fusão entre proteína fluorescente verde COP e hGH. Um plasmídeo contendo uma fusão de hGH::COP foi construído e transformado na cepa de origem DC206 e na cepa de deleção de gene de *hslU* DC370 resultando em cepas HJ104 e HJ105 (Tabela 1). Os experimentos de frasco de agitação padrão foram executados e as amostras tiradas em vários pontos de tempo para medições da fluorescência (Figura 10). As leituras do fluorímetro claramente mostraram que a cepa mutante de protease de *hslU* teve níveis de

expressão de proteína significativamente mais altos comparados aos da cepa parental (Figura 11). Esta descoberta confirma os resultados obtidos por análise de SDS-PAGE. Comparando à cepa de tipo selvagem, o mutante de *hslU* aumentou 33,05% da fluorescência relativa em 24 h após indução (ver 5 inserção na Figura 11).

#### EXEMPLO 5: CONSTRUÇÃO DE UMA CEPA KNOCKOUT LIMPA DE *hslUV*

A protease de Hsl consiste em duas subunidades: uma subunidade de ligação de ATP codificada por *hslU* e uma subunidade de protease codificada por *hslV*. A cepa knock-out de protease de Hsl previamente construída é uma inativação insercional do gene de *hslU*. Para remover a preocupação que *HslV* poderia ainda funcionar como uma protease podendo acoplar a uma subunidade de ligação de ATP de outra protease, uma cepa de deleção foi construída que teve os genes de *hslU* e *hslV* removidos do cromossomo.

15 Como mostrado na Figura 13, plasmídeo pDOW2050 foi construído por amplificação de PCR de dois fragmentos de DNA flanqueando a região de *hslUV*, os dois fragmentos foram subsequentemente fundidos usando o método de PCR de Encaixe através de Extensão de Sobreposição (SOE) (ver Ho, S. N. (1991) Method for gene splicing by overlap extension 20 using the polymerase chain reaction. Pedido: Patente US 89-3920955023171). Os fragmentos de DNA fundidos foram depois ligados no sítio de *SrfI* de vetor pDOW1261-2. O plasmídeo de deleção foi nomeado pDOW2050 após a inserção ter sido confirmada através de seqüenciamento de DNA.

Plasmídeo pDOW2050 foi eletroporado em DC206 e banhado 25 sobre placas de ágar de M9 suplementadas com 1% de glicose e 15 pg/ml de tetraciclina. Resistência à tetraciclina é devido a um evento de integração que recombina o plasmídeo inteiro no cromossomo com uma das duas regiões homólogas dentro do genoma (Figura 13). Para selecionar as células que têm uma deleção dos genes de *hslUV*, que resulta de uma segunda recombinação 30 homóloga entre o plasmídeo integrado e a região de DNA homóloga no cromossomo, as colônias resistentes à tetraciclina foram crescidas para fase estacionária em meio de LB suplementado com 250 µg/ml de uracila. As cé-

Iulas são depois banhadas sobre placas de ágar de LB suplementadas com 500 pg/ml de ácido 5-fluoroorótico (5-FOA). As células que perderam o plasmídeo integrado por um segundo evento de recombinação também perderam o gene de *pyrF* e desse modo são resistentes a 5-FOA, resultando na 5 cepa de deleção de *hsl/UV* cromossômico desejada, chamada DC417.

#### ANÁLISE FENOTÍPICA DA CEPA DE DELEÇÃO DE *hsl/UV*

Análise de SDS-PAGE da cepa de deleção de *hsl/UV* expressando a proteína de hGH (cepa HJ115) mostrou rendimento de proteína muito mais alto que a cepa do tipo selvagem DC369, similar ao que foi observado mais 10 cedo usando a cepa mutante insercional de *hsl/U* DC372 (dados não mostrados).

Rendimento da proteína foi também medido por atividade de fluorescência da fusão de hGH::COP usando o mesmo método descrito mais cedo. Plasmídeo pDOW1349 contendo a fusão de hGH::COP foi transformado nas cepas do tipo selvagem e mutante que resultam em cepas HJ104 e 15 HJ117, respectivamente. Experimentos de frasco de agitação padrão foram executados e as amostras foram tiradas em vários pontos de tempo para medições de fluorescência relativa (Figura 14). As leituras do fluorímetro indicaram que a cepa de deleção de protease de *hsl/UV* teve níveis de expressão de proteínas significativamente mais altos (cerca de 50% de aumento de rendimento) quando comparada aos da cepa do tipo selvagem. Este resultado é similar ao que foi previamente observado com a cepa knock-out de *hsl/U*. 20

#### EXEMPLO 6: IDENTIFICAÇÃO DE ALVO ITERATIVO USANDO TECNO- 25 LOGIA DE MICROARRANJO DE DNA

Para investigar se um conjunto novo de proteases é sobre-regulado na cepa de deleção de protease de *hsl/UV*, experimentos de microarranjo de DNA foram conduzidos. Experimentos de frasco de agitação padrão foram executados usando a cepa do tipo selvagem (DC369) e mutante 30 (HJ115) para expressar hGH. Para cada cepa, as 4 h após as amostras de indução term sido comparadas à amostra de ponto de tempo 0 h (logo antes da indução de proteína heteróloga) os experimentos de microarranjo de DNA

foram executados. Comparando a razão de dois pontos de tempo entre as cepas do tipo selvagem e mutante, uma lista nova de genes de protease que são sobre-regulados na cepa de deleção de protesase de *hs/UV* foi identificada (Tabela 8). Estes genes recentemente identificados codificando proteases podem agora ser os alvos para uma segunda rodada de eventos de deleção de gene para também melhorar o rendimento de produção de proteína heteróloga.

**TABELA 8:** Genes de protease cujos níveis de mRNA de estado constante são mais altos na cepa de deleção de protease de *hsfUV* (HJ115) quando comparados aos da cepa do tipo selvagem (DC369), com base na razão de 4 h após indução de IPTG para 0 h (logo antes da indução).

ORFID	Função Curada	Seqüênci a
rxf00133	D-alanil-me- so-diaminop- imelato en- dopeptidase (ec 3.4.-.)	atgcacttggaaaatggttcacaccagcacccctgtggcggcttgagttttgtctggggcgg ctgcgcgcaggcttcacaaacctccaccccgcaaccctggataagctgttgagcgacccgg cgctgcacaggcgcaccgtctcgatggcgtatggccgcacaggcaccacgtgtat cagcacaacccacgcacccggctgtggccgcgttcaacctaagctgttgaccacggcgg cagccatggatgtatggggccgcagttaccgcgtccacgcacactgtctggatggctac gccaggcgcaccggctgactggcaacctgtaccgcgtggcttggccgaccggatgttac tttgcgcactatcaggcgcgtccgcgcgaattggccaggcaggcgtgcgcaggcagg gtgaccctgggtgtcgacgacacttgggtcgatgcgcaggcggctggactgttccatg atgatgaaccacactactacggcgcgcagatttcagcgcgtgaccgtggcccaataccact tttgcgtggcagcgtgtgtccacccgcacccggccatgcgcgtggcgcggcgg gggagatctacccgcacccgcacacttgcacactgaataaccgcgcgtcagcggccgg gtaacagctatggatcaaccgtcgccatggccaccaacctgtcgacgtcagcggccgg gcccgtggccggcagggcagcactgttgcgtggagccgacgcacactgttggcca acctgttgcgtggccacggcatcaaggcgtctggccgtgggtatggcc ggggcaagtctgtgggtgacgggtctggccgagcacaatggccgcgtcaggagg tgatcgtgcgcgtgtcaagctcgacaacaacgcctatgcgcgtgcgtcaaggccatgg gcgcgcacggccacgcggcgcacggccggccgtggccgtggccgcacit ctcaagcgcacgggtgtggacaccgcgtgtgagccaaatggacggcgtccgcgtgc ggcgttaacctgggtgtcgcaacccgcacccgcgtgtggccgcacaaacacc tgggtcgacgcgtgttgcatacgcgcgtccgggtgcggccatgcgcacgtatgaccgg cagccgtgggttaccgcgtcgccgcacggcgtgcggaaaataacctgcgtcaagcgg ccatggccggcgtgtcgatgtggatgcacccgcgtcgtccacggccacgcacgg cgcgtgttgcaccaactatgtggcgtggccgcaggtaaaagccgtggaaaaccgc tcgcgtggccgtccacagcgaagactga

ORFID	Função Curada	Seqüência
rf01918	Zinco pro-tease (ec 3.4.99.-)	atgagtatcgcaaaaacagccgcctgatctggccggcctgatcgccgtcacccgtatggcgcc ccagcgccgtttaacttcttgcgcggcaggcgatcgccgcaggccctggacaaggc tcaaaccggccgcacccctgcaattccctggggaaactggatggcaaggcaccgaccaaccgc aagtgcacgtacaacacctggaccacccggcgaaggcggcaagggtgttgtcgtaagcc tgaggcgcatgttcgacatgcgcctgtgttcgcggccggcaggcagccaggatggcgacgt gccaggcctggcgctgtatgaccaacgcctatgtcaacgaaggcgtggccggcaaggacgt agccagatgcgcaggatggcttcgaaggcctggggccgactttggcaacggcgcctaccgc acatggcgctggigaccctggcgccgtggccggcggccggcggccggccggcc actgttcaaccagggtatcgccgcacttcccccggcagactcactggcacgcaccaagaa ccagatctggccgggttgcgatgtaccagaaggcagaaccccccggcaactggcgagcatcg tgttcaagcgccgtacggcgaccacccctacgcacacccggagaaggcaccggag cgtggccagaattaccctggcgcatgtgcaggcggttccacgcggccatgtcagcgggt acggcggtatgcgatgtgtggcgaccgtggccggcggccggcggcc gggttcccgccgtcgccaaaggccggctatggccaagatgcggccggccggcc aaaggccggctgagccgtatcgatgttccgtcaagcaacccactgttgttgcgcagtg ggcatcgaccgtggcgaccggattacgcggccgttccgtggtaaccagatctcgcc cggtggcttcggccaccgcgttgcgatggcgaaatgcggccatgtcggcc gtgttaatccgggttctccatggcgccggccgttgcgtatgcacccgc ccgaaatggcggtggccaccctggccgtgtggaggacgtactggctactaccctcaagacc ggccgcggccacaaaggactggatgcggccaaatgcggccggccggcc ccaccgcggccacgcggatatgcggccgttggccatgggttctacaacccgc ctggatctggaaatgttcatggaaacaaatcccgccgttgcgatggccggcc caatgttcaacacttgagccgacaaatggcgatgttgcggccggcc caaaaggccactaccggccccactgataaaacccatgcggccggcc attaa
rf02689	Dipeptidase microsso- mal (ec 3.4.13.19)	tttgtcggttttgcgcgttccctgtcggttttgcaccggctccgtcggtgcctgg gcataatgcgtggcccaaaaggcgccgcaggcgattcgccgcgtatgcataatgcgg gcctgtatgcggccatgcggatgtgcacgcgcacatgttgcacggctgtatgtccaa gtggaaacccgcgcactgttgcgaaacatgcgcggatgcggccatgtgcacgg tgtcggtgtggaaagggttcccgccatgcataataacatgttgcgcggcc gcggaaaacgcgcacccgttgcgtcccgatggccatgcgttgcacccatgc atgcggccaggactggcatatctcggttcccgatgcgcggccatgcgttgc ctatgtcgatctcaagcgcgttgcgtggcgatgttgcgatgttgc cctgggtggccaccgggttgcgtacgcgcgtatggccgttgcgggttgc tcggcgatgttgcgtcccgatcatgtgcgcacccatgtggccgttgc gaagaggatccctcgatggaaacaaaggccgggtgtactccactgttgc aaggccatccctcgatggaaatgcgttgcgttgcgttgc aaggccatccctcgatggaaatgcgttgcgttgcgttgc tcaccaggccatggccaggatccctcgatggaaatgcgttgcgttgc gcctgaccagatcccgatgcgttgcgttgcgttgcgttgc acccatccctcgatggaaatgcgttgcgttgcgttgc cgagaactgggtcaacgtgtcaaggacgttgcgttgcgttgc attaa



EXEMPLO 7: CO-SOBRE-EXPRESSÃO DE MODULADORES DE DOBRAMENTO AUMENTA A SOLUBILIDADE DA PROTEÍNA ALVO hGH

Com base nos dados de perfilção transcripcional mostrados na Figura 4, expressão de moduladores de dobramentos (FMs) DnaK e DnaJ foi aumentada em cepas que produzem proteína recombinante comparadas às cepas de controle (ver Tabelas 4 e 5). Uma cepa que co-superproduziu GrpE, DnaK e DnaJ junto com hGH foi produzida e testada para identificar se isto resultou na acumulação de níveis aumentados de hGH solúvel.

CONSTRUCTO DE PLASMÍDEO CONTENDO grpE-dnaKJ PARA CO-SOBRE-EXPRESSÃO COM hGH

Os genes de *grpE-dnaKJ* de *P. fluorescens* foram amplificados usando o DNA cromossômico isolado de MB214 (DNeasy; Qiagen, Valença, CA) como um modelo, RC199 (5'-ATATACTAGTAGGAGGTAACTTATGGCTGACGAACAGACGCA-3') e RC200 (5'-ATATTCTAGATTACAGGTCGCCGA-  
15 AGAACG-3') como iniciadores, *PfuTurbo* (Stratagene, La Jolla, CA) foi usado seguindo as recomendações do fabricante. O produto de PCR resultante (4 kb) foi digerido com *Spel* e *XbaI* (sítios de restrição sublinhados acima nos iniciadores) e ligado a pDOW2236 para criar pDOW2240 contendo os genes de *grpE-dnaKJ* sob o controle do promotor de tac. Plasmídeo pDOW2240 foi digerido com *Spel* e *HindIII* e o fragmento de 4,0 kb contendo *grpE-dnaKJ* resultante foi purificado em gel usando Qiaquick (Qiagen) e ligado a pDOW2247 também digerido com *Spel* e *HindIII*. O plasmídeo resultante, pDOW3501, contendo *grpE-dnaKJ* sob o controle do promotor de manitol, foi transformado em DC388 mediante seleção em placas de glicose de M9 suplementadas com 250 µg/ml de uracila. Por fim, pDOW1426 foi eletroporado na cepa acima (DC462) e selecionado em placas de glicose de M9, resultando em cepa DC463 com dois plasmídeos induzíveis: 1) pDOW1426 carregando  $P_{tac}$  hGH e 2) pDOW3501 carregando  $P_{mts}$  *grpE-dnaKJ*.

FERMENTAÇÃO EM FRASCO DE AGITAÇÃO, COLETA DE AMOSTRA E ANÁLISE

Culturas em duplicata de DC463 foram crescidas em frascos de agitação. Indução de proteína foi realizada pela adição de 0,1 mM de IPTG

para hGH e 0,5% de manitol para GrpE-DnaKJ em 24 hs após inoculação. As amostras foram colhidas a 0, 4, 8, 24 e 48 horas após indução. Em cada ponto de tempo, 20 células de OD<sub>600</sub> normalizada em 1 ml foram colhidas, lisadas usando Easylse® (Epicentro, Madison, WI) e separadas em frações solúveis e insolúveis mediante centrifugação a 14000 rpm durante 30 minutos. Volumes iguais das amostras foram combinados com BioRad (Hercules, CA) 2x tampão de Laemmli, aquecidos para 95°C durante 5 minutos com 30 µL carregados em um gel de Criterion a 15% de Tris HCl de BioRad usando tampão de operação de SDS de 1x Tris Glicina corrente (BioRad). As proteínas foram visualizadas com Simply Blue Safestain (Invitrogen, Carlsbad, CA) como mostrado na Figura 15. Os géis tingidos com Coomassie resultantes foram varridos usando um Densitômetro Pessoal de Molecular Devices (Molecular Devices, Sunnyvale, CA) com análises executadas usando Image-Quant e Excel. Como mostrado na Figura 15, a co-sobre-expressão de GrpE, DnaKJ significativamente aumentou a solubilidade de hGH, convertendo quase 100% da proteína alvo na fração solúvel, embora a um rendimento de proteína total inferior. Experimentos adicionais repetindo crescimento e indução de DC463 usando a adição simultânea de IPTG e manitol de forma bem próxima imitaram os resultados mostrados aqui, embora com um grau variado de solubilidade de hGH (entre 50-100%; dados não mostrados), quando GrpE DnaKJ foram co-superproduzidos. Estas descobertas também demonstram que engenharia de cepa alvejada com base em perfilamento transcripcional pode levar a um projeto de cepa racional para aumentar a solubilidade e/ou rendimento de uma proteína recombinante.

A invenção foi descrita com referência a certas modalidades e exemplos não-limitativos. Estará claro a alguém de habilidade na técnica que outras modalidades da invenção são também possíveis.

## REIVINDICAÇÕES

1. Processo para melhorar a expressão de uma proteína ou peptídeo recombinante em uma célula ou organismo hospedeiro, compreendendo:
  - 5            i) expressar a proteína ou peptídeo recombinante na célula ou organismo recombinante hospedeiro;
  - ii) analisar um perfil genético da célula recombinante para identificar um ou mais genes ou produtos de gene compensatórios que são expressos a um nível mais alto na célula recombinante que em ou uma célula hospedeira que não foi modificada para expressar a proteína recombinante ou uma célula recombinante que não está expressando a proteína recombinante; e
  - 10           iii) alterar a expressão do gene ou produto de gene compensatório identificado na célula recombinante através de modificação genética para fornecer uma célula recombinante modificada que alcança um aumento em expressão, atividade ou solubilidade da proteína recombinante.
2. Processo de acordo com a reivindicação 1, ainda compreendendo expressar a proteína ou peptídeo na célula recombinante modificada.
3. Processo de acordo com a reivindicação 1, ainda compreendendo:
  - 20           (a) expressar a proteína ou peptídeo recombinante na célula recombinante modificada;
  - (b) analisar um perfil genético da célula recombinante modificada para identificar pelo menos um/uns segundo(s) gene(s) ou produto(s) de gene que é/são diferencialmente expresso(s) na célula recombinante modificada;
  - 25           (c) alterar a expressão do segundo produto de gene identificado na célula recombinante modificada para fornecer uma célula duplamente modificada; e
  - (d) expressar a proteína ou peptídeo na célula recombinante duplamente modificada.
4. Processo de acordo com a reivindicação 3, ainda compreendendo repetir as etapas a) a d).

5. Processo de acordo com a reivindicação 4, compreendendo repetir as etapas a) a d) até a viabilidade da célula ser afetada alterando a expressão do(s) gene(s) ou produto(s) de gene modificado(s).
6. Processo de acordo com a reivindicação 4, compreendendo 5 repetir as etapas a) a d) até a expressão da proteína ou peptídeo recombinante alcançar um ponto terminal alvejado.
7. Processo de acordo com a reivindicação 1, em que um perfil genético é analisado comparando um perfil genético da célula recombinante a um segundo perfil genético da célula hospedeira.
- 10 8. Processo de acordo com a reivindicação 1, em que o perfil genético é um perfil de transcriptoma.
9. Processo de acordo com a reivindicação 8, em que o perfil de transcriptoma é determinado por um microarranjo.
10. Processo de acordo com a reivindicação 1, em que o perfil 15 genético é um perfil de proteoma.
11. Processo de acordo com a reivindicação 10, em que o perfil de proteoma é determinado por eletroforese em gel bidimensional, ICAT ou LC/MS.
12. Processo de acordo com a reivindicação 10, em que o perfil de 20 proteoma é determinado por um arranjo de peptídeo.
13. Processo de acordo com a reivindicação 12, em que o arranjo de peptídeo é um arranjo de anticorpo.
14. Processo de acordo com a reivindicação 1, em que o produto de gene identificado é uma protease, uma subunidade de uma protease, um 25 co-fator de uma protease, um modulador celular ou um genético que afeta a expressão de uma protease.
15. Processo de acordo com a reivindicação 14, em que o produto de gene identificado é uma protease.
16. Processo de acordo com a reivindicação 14, em que o produto 30 de gene identificado é uma subunidade de uma protease.
17. Processo de acordo com a reivindicação 14, em que o produto de gene identificado é um co-fator de uma protease.

18. Processo de acordo com a reivindicação 14, em que o produto de gene identificado é um modulador celular ou genético que afeta a expressão de uma protease.

19. Processo de acordo com a reivindicação 14, em que o produto de gene identificado é selecionado do grupo que consiste em D-alanil-meso-diaminopimelato endopeptidase, zinco protease, dipeptidase microssomal, precursor de metaloprotease extracelular, proteína de divisão de célula ftsH e produtos de gene derivados dos genes *hsIV*, *hsIU*, *clpX*, *clpA* e *clpB*.

20. Processo de acordo com a reivindicação 14, em que o nível de rRNA do produto de gene identificado é sobre-regulado quando a proteína ou peptídeo recombinante for expressado na célula hospedeira.

21. Processo de acordo com a reivindicação 14, em que o produto de gene identificado é removido de um genoma de célula hospedeira.

22. Processo de acordo com a reivindicação 21, em que o produto de gene identificado é removido através de recombinação homóloga.

23. Processo de acordo com a reivindicação 1, em que o produto de gene identificado é um modulador de dobra, uma subunidade de um modulador de dobra, um co-fator de um modulador de dobra ou um modulador celular ou genético que afeta a expressão de um modulador de dobra.

24. Processo de acordo com a reivindicação 23, em que o produto de gene identificado é um modulador de dobra.

25. Processo de acordo com a reivindicação 23, em que o produto de gene identificado é uma subunidade de um modulador de dobra.

26. Processo de acordo com a reivindicação 23, em que o produto de gene identificado é um co-fator de um modulador de dobra.

27. Processo de acordo com a reivindicação 23, em que o produto de gene identificado é um modulador celular ou genético que afeta a expressão de um modulador de dobra.

28. Processo de acordo com a reivindicação 23, em que o modulador de dobra é uma proteína de chaperona.

29. Processo de acordo com a reivindicação 23, em que o modulador de dobra é selecionado do grupo que consiste em produtos de gene

dos genes *cbpA*, *htpG*, *dnak*, *dnaJ*, *fkbP2*, *groES* e *groEL*.

30. Processo de acordo com a reivindicação 23, em que a expressão do produto de gene identificado é alterada aumentando a expressão do gene identificado, um co-fator de um gene identificado ou um modulador 5 celular ou genético do gene identificado.

31. Processo de acordo com a reivindicação 30, em que a expressão aumentada é por inclusão de um DNA que codifica o produto de gene identificado.

10 32. Processo de acordo com a reivindicação 30, em que a expressão aumentada é por inserção de um promotor em um genoma da célula hospedeira.

33. Processo de acordo com a reivindicação 30, em que a expressão aumentada é por inclusão de um vetor exógeno na célula hospedeira.

15 34. Processo de acordo com a reivindicação 1, em que a célula hospedeira é uma célula microbiana.

35. Processo de acordo com a reivindicação 1, em que a célula hospedeira é uma Pseudomonada.

36. Processo de acordo com a reivindicação 1, em que a célula hospedeira é uma célula de *P. fluorescens*.

20 37. Processo de acordo com a reivindicação 1, em que a célula hospedeira é uma célula de *E. coli*.

38. Processo de acordo com a reivindicação 1, em que a célula hospedeira é selecionada do grupo que consiste em uma célula de inseto, uma célula mamífera, uma célula de levedura, uma célula fúngica e uma cé- 25 lula de planta.

39. Processo de acordo com a reivindicação 9, em que o micro-arranjo compreende amostras de pares de ligação para pelo menos 50% de um genoma da célula hospedeira.

30 40. Processo de acordo com a reivindicação 9, em que a técnica de microarranjo compreende amostras de pares de ligação para pelo menos 80% de um genoma da célula hospedeira.

41. Processo de acordo com a reivindicação 9, em que o micro-

arranjo compreende amostras de pares de ligação para pelo menos 90% de um genoma da célula hospedeira.

42. Processo de acordo com a reivindicação 9, em que o micro-arranjo compreende amostras de pares de ligação para pelo menos 95% de 5 um genoma da célula hospedeira.

43. Processo de acordo com a reivindicação 1, em que a expressão melhorada é um aumento na quantidade de proteína ou peptídeo recombinante.

44. Processo de acordo com a reivindicação 1, em que a expressão melhorada é uma solubilidade aumentada da proteína ou peptídeo recombinante. 10

45. Processo de acordo com a reivindicação 1, em que a expressão melhorada é uma atividade aumentada da proteína ou peptídeo recombinante.

46. Processo de acordo com a reivindicação 1, em que o perfil genético é um perfil de genes em uma família de gene. 15

47. Processo de acordo com a reivindicação 1, em que o perfil compreende proteases e moduladores de dobra.

48. Processo de acordo com a reivindicação 46, em que o perfil 20 consiste essencialmente em proteases.

49. Célula ou organismo hospedeiro, em que expressa uma proteína recombinante que foi geneticamente modificada para reduzir a expressão de pelo menos duas proteases.

50. Célula ou organismo hospedeiro, em que expressa uma proteína recombinante que foi geneticamente modificada para reduzir a expressão de pelo menos uma protease selecionada do grupo que consiste em D-alanil-meso-diaminopimelato endopeptidase, zinco protease, dipeptidase microssomal, precursor de metaloprotease extracelular, proteína de divisão de célula ftsH e produtos de gene derivados dos genes *hsIV*, *hsIU*, *clpX*, *clpA* 25 e *clpB*. 30

51. Célula ou organismo hospedeiro, em que expressa uma proteína derivada mamífera recombinante que foi geneticamente modificada para

reduzir a expressão de pelo menos uma protease.

52. Célula ou organismo hospedeiro de acordo com a reivindicação 52, em que a proteína recombinante é hormônio de crescimento humano.

53. Célula ou organismo hospedeiro, em que expressa uma proteína recombinante que foi geneticamente modificada para aumentar a expressão de pelo menos dois moduladores de dobra que não sejam subunidades de modulador de dobra.

FIG. 1

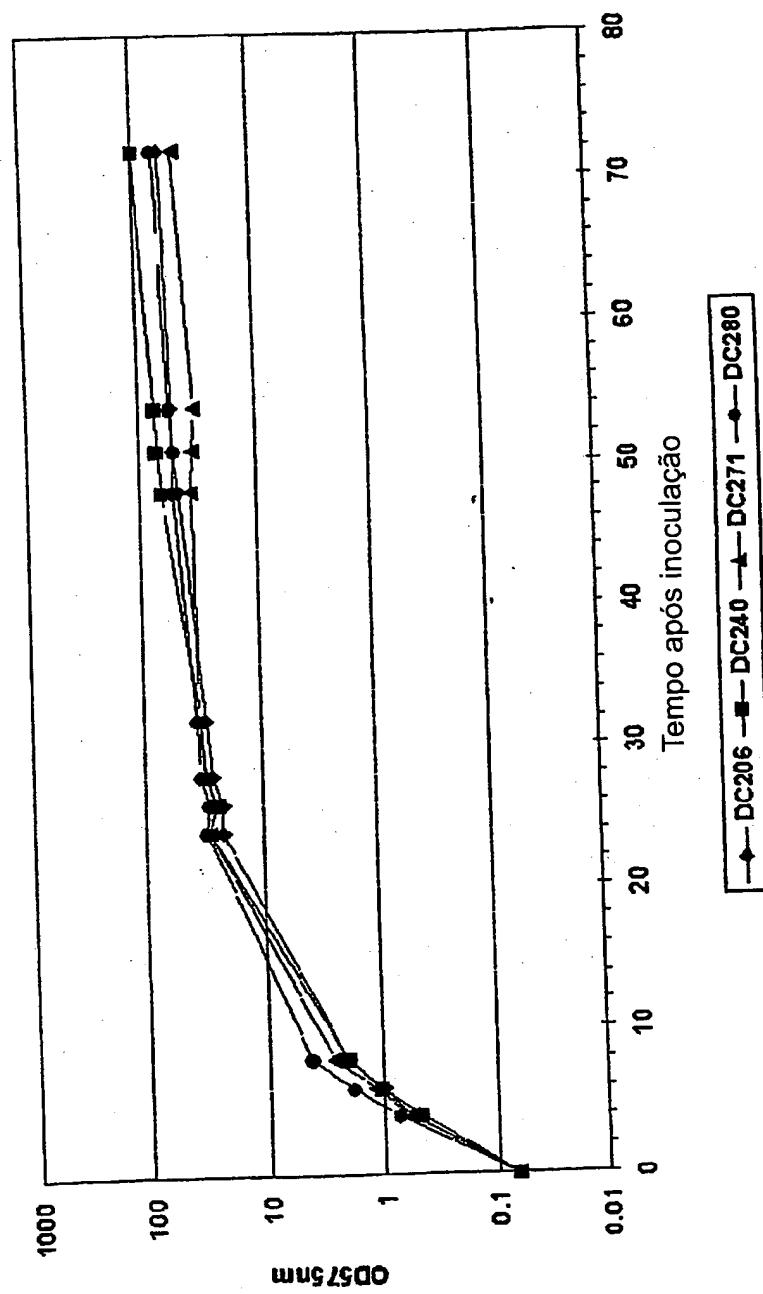


FIG. 2

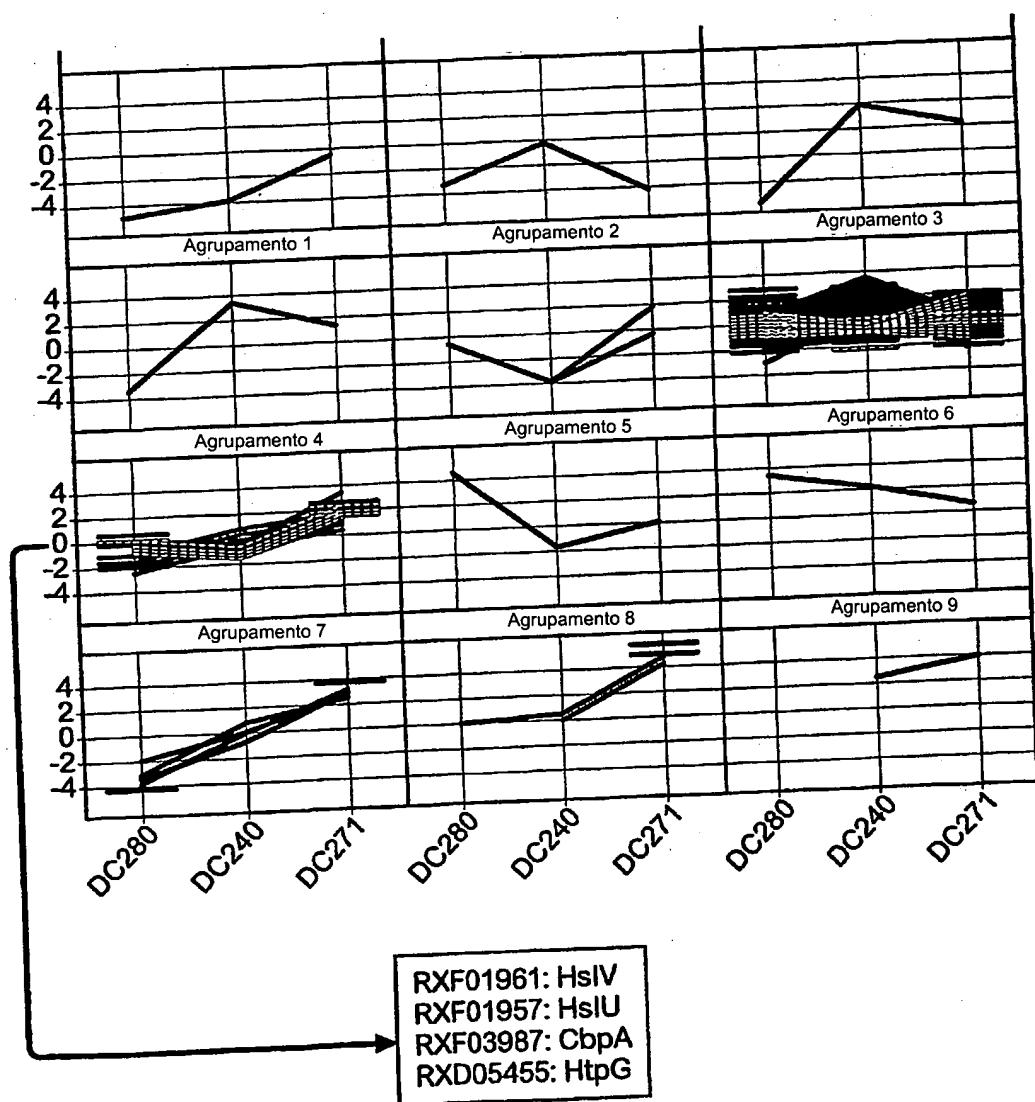


FIG. 3

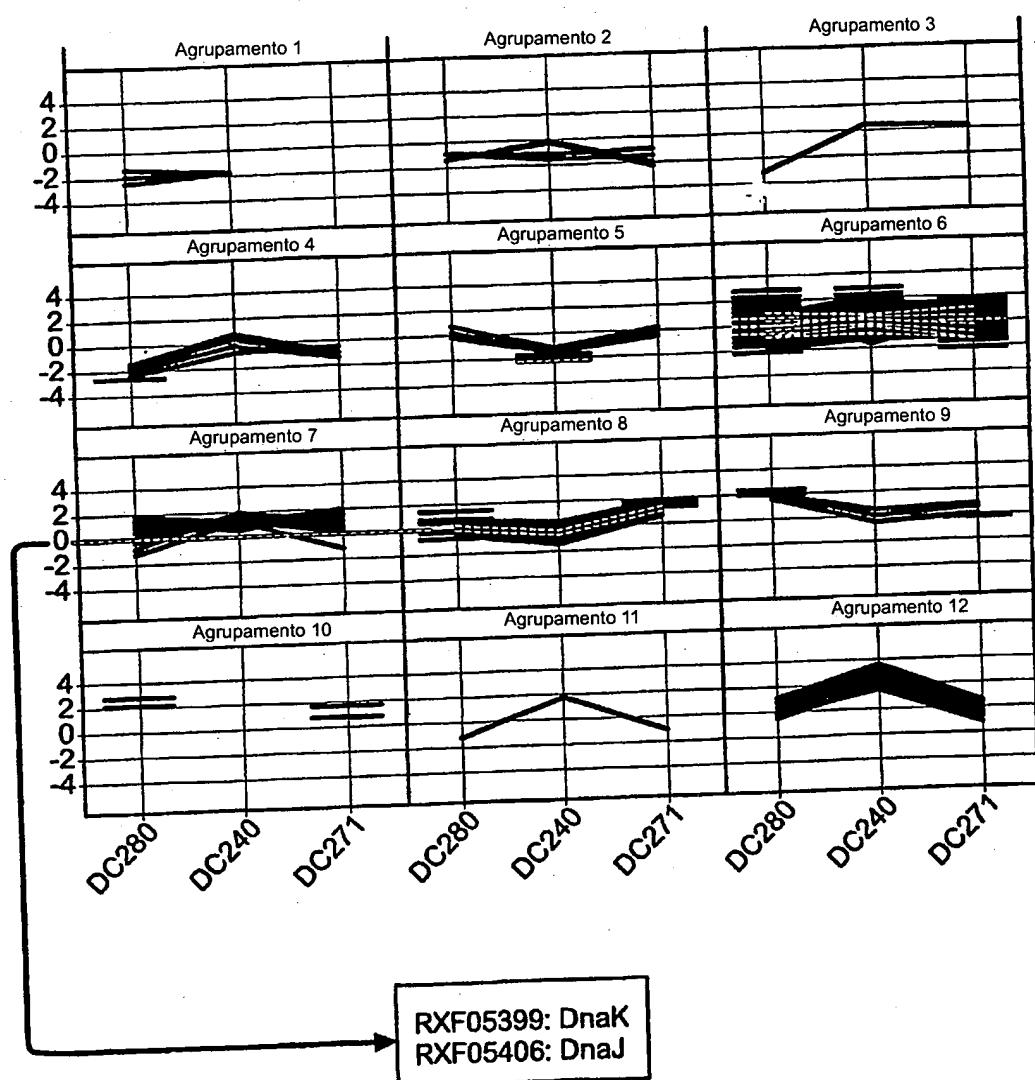
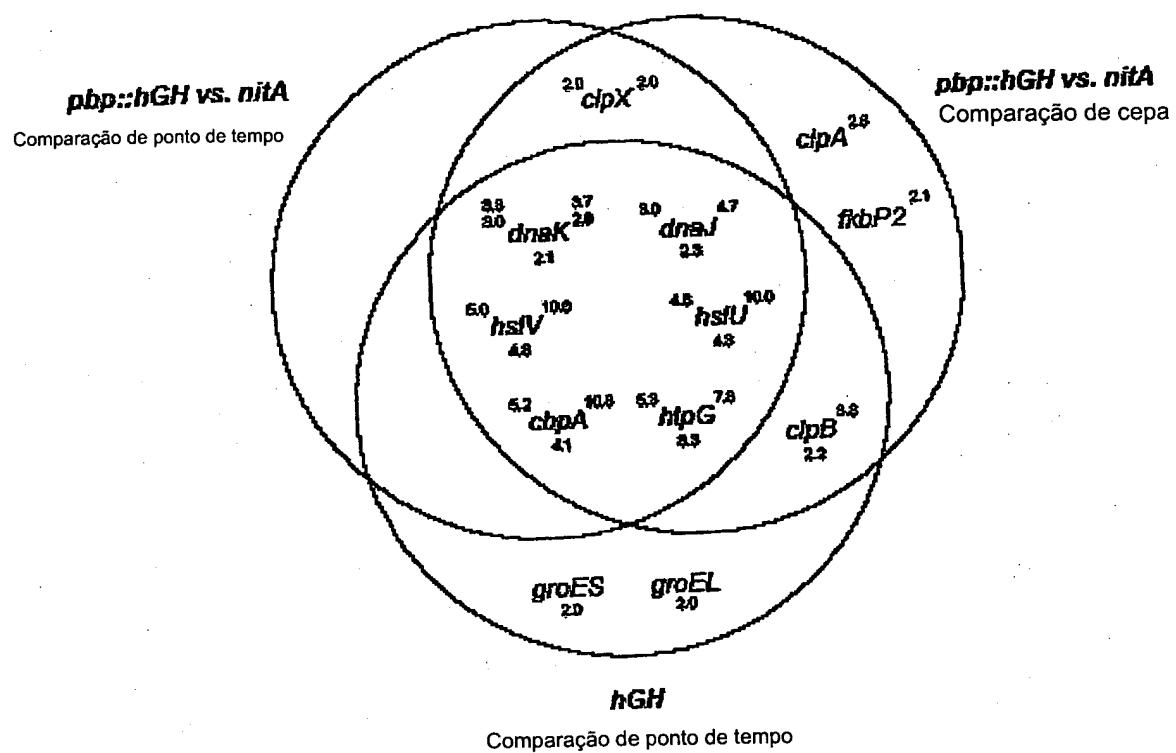


FIG. 4



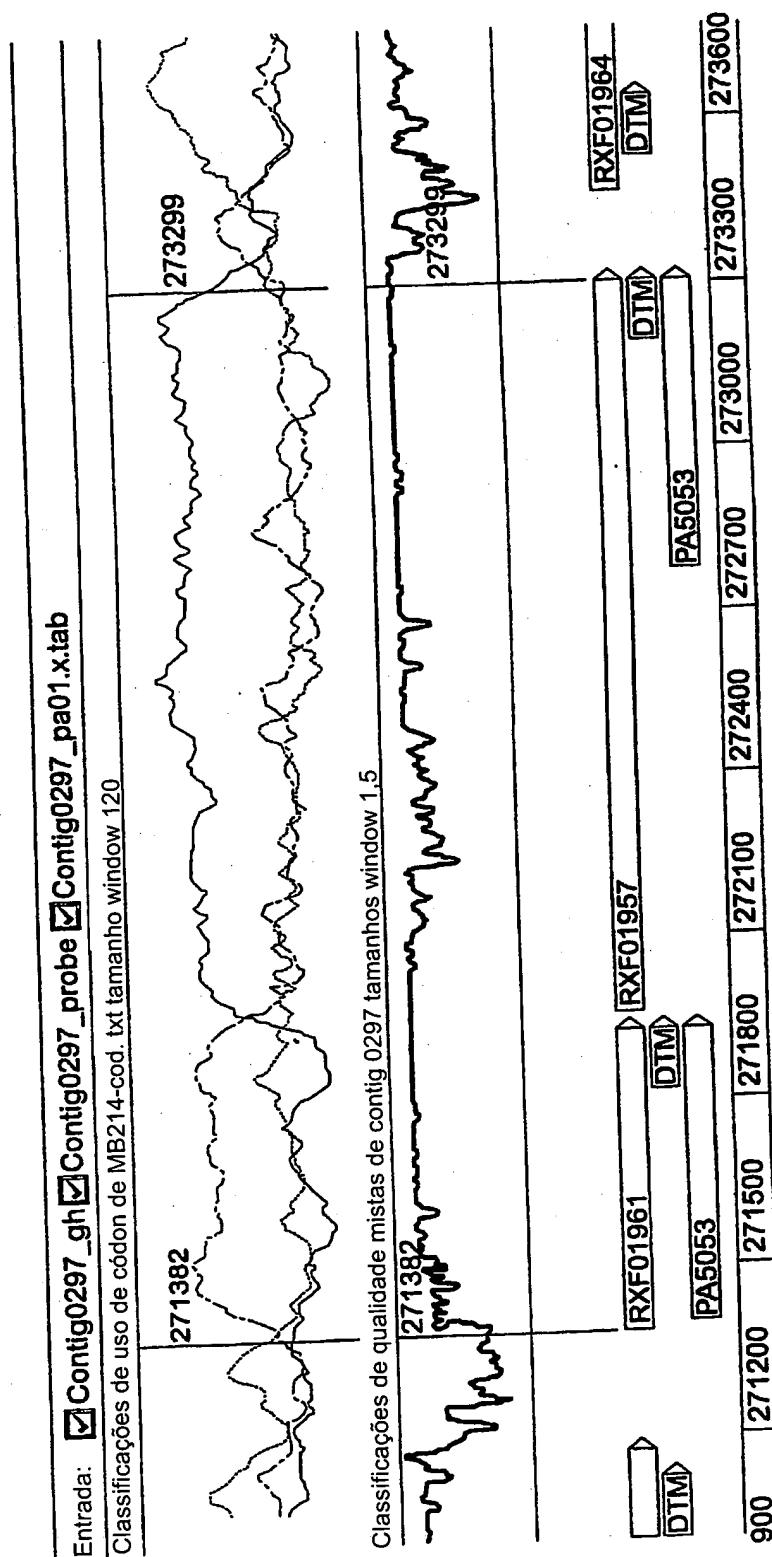


FIG. 5

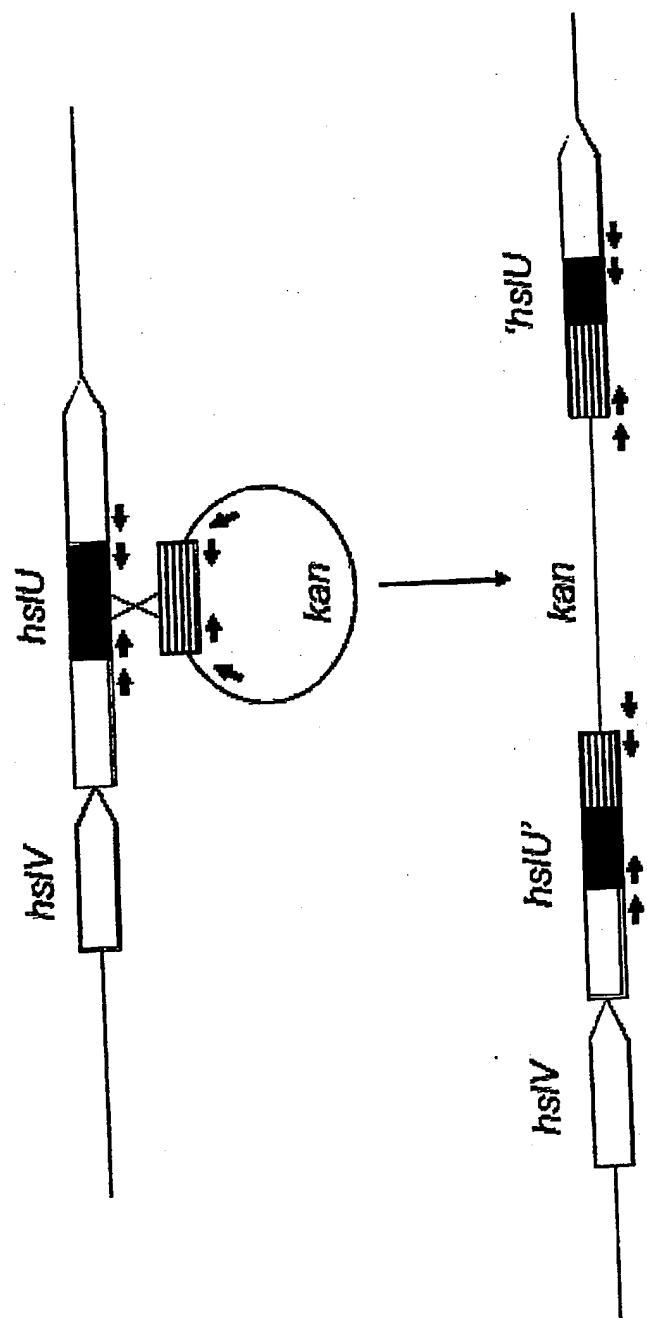
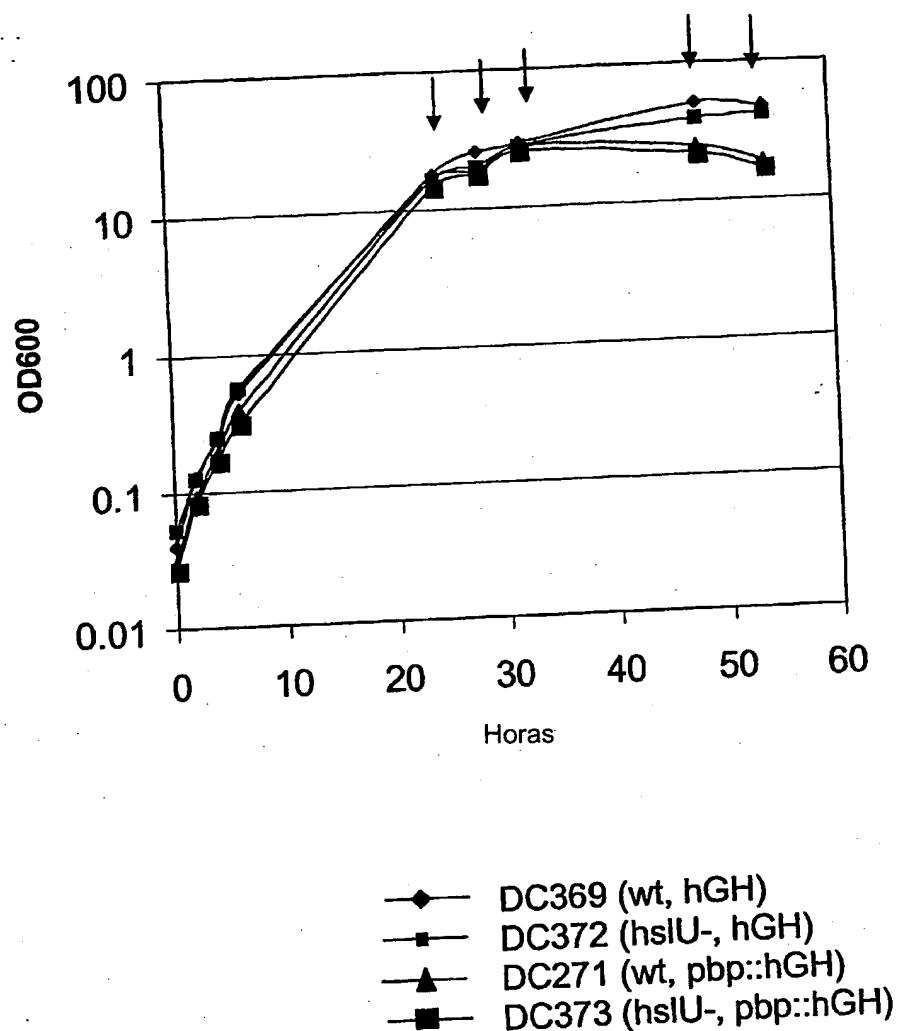


FIG. 6

7/15

FIG. 7



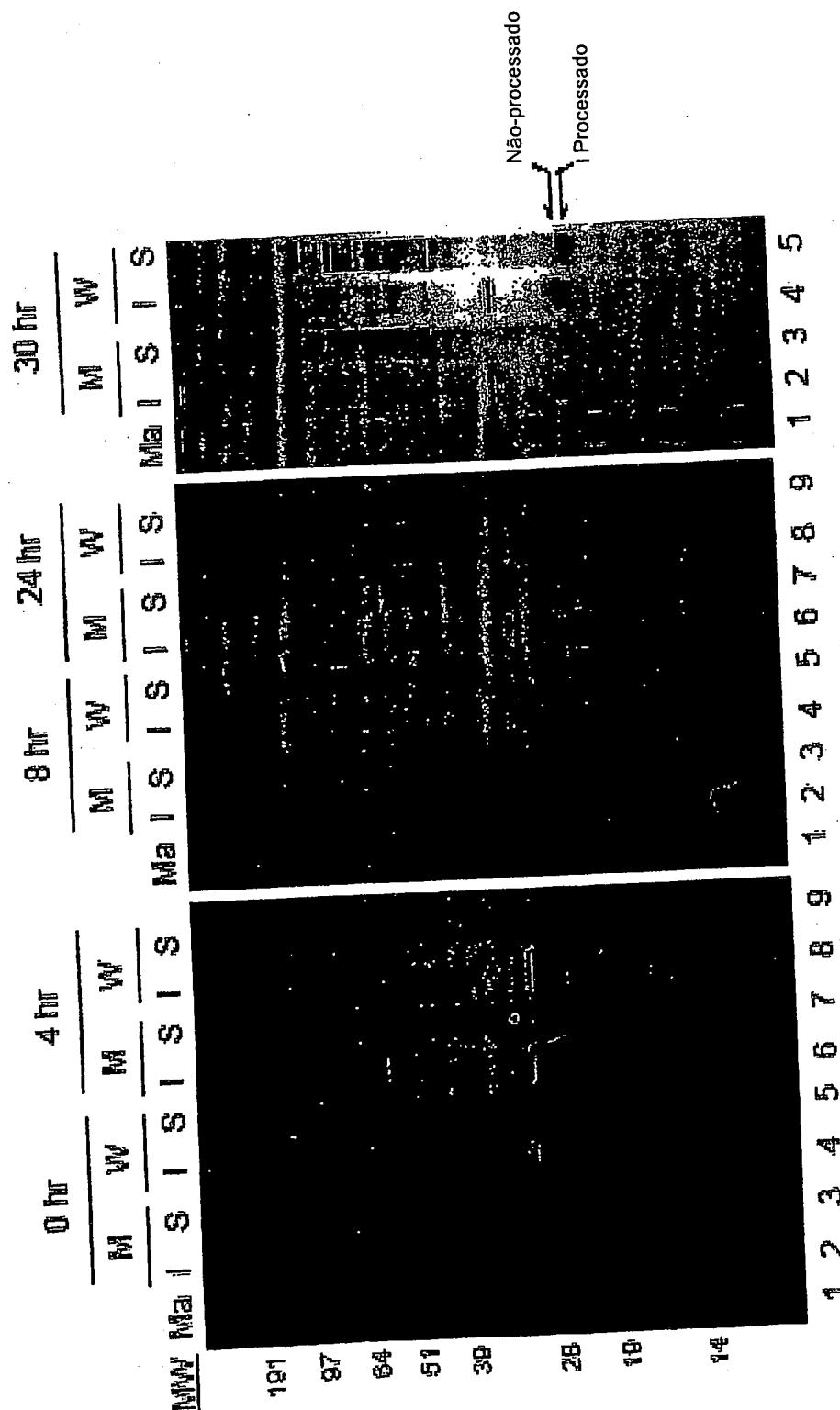


FIG. 8

9/15

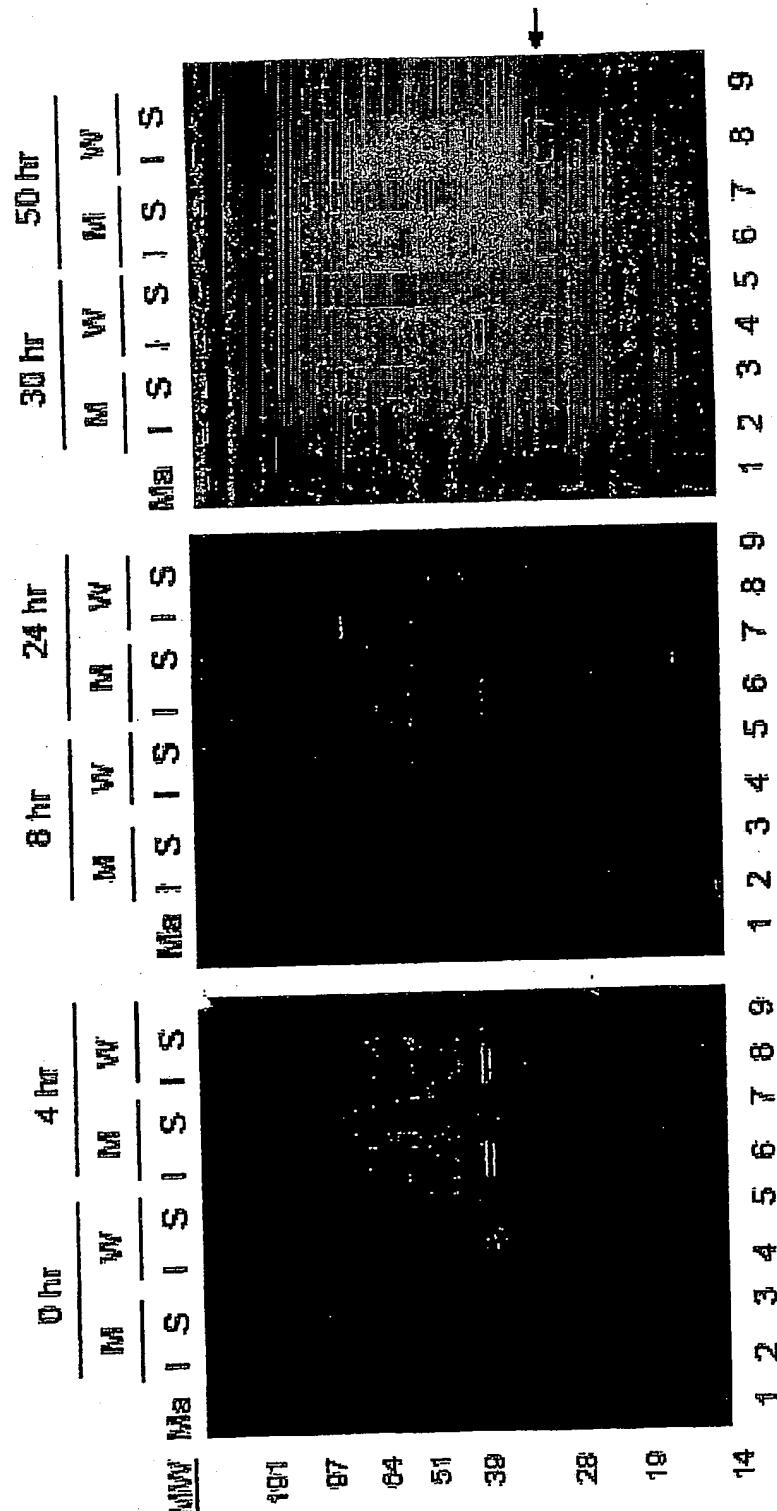


FIG. 9

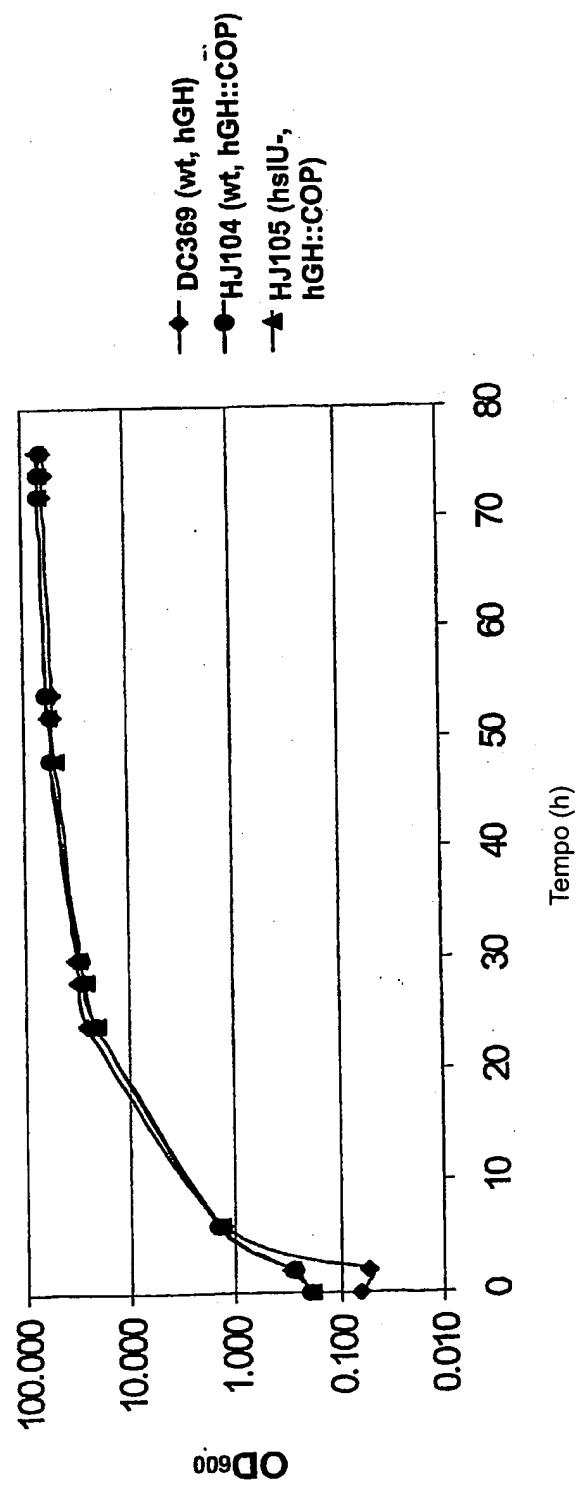


FIG. 10

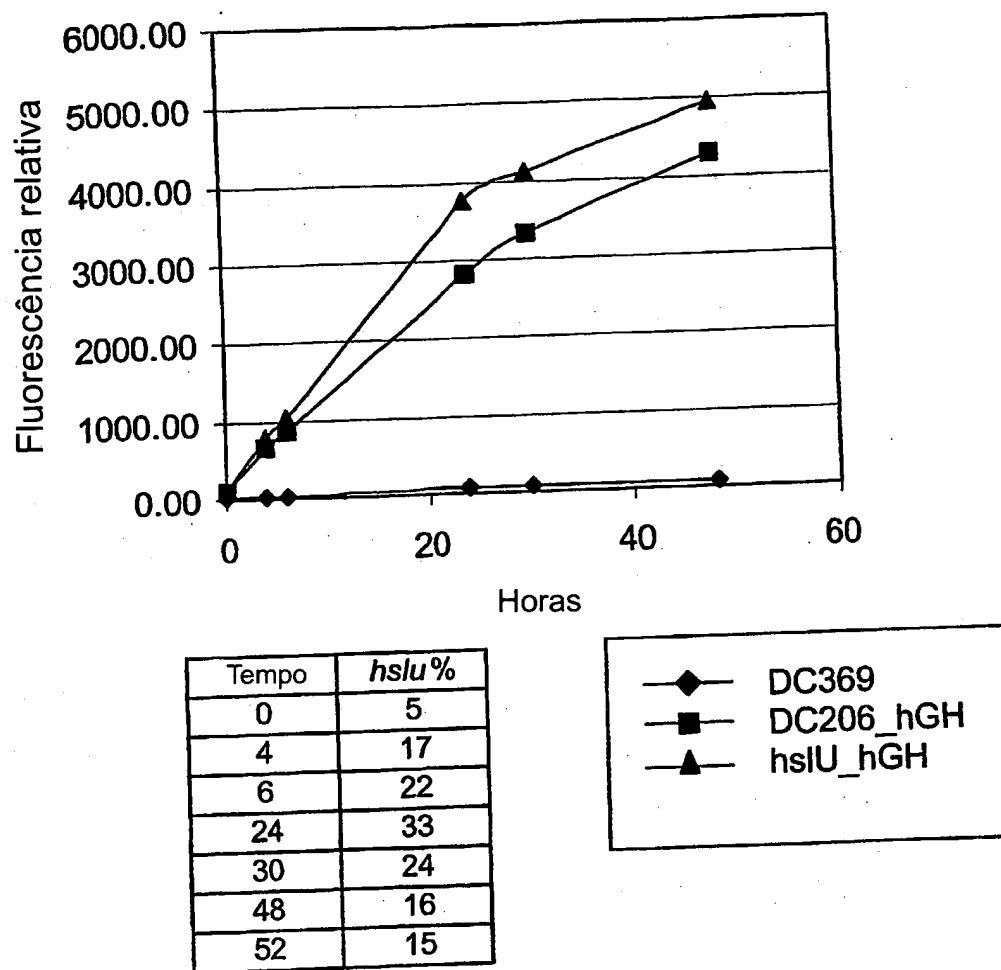


FIG. 11

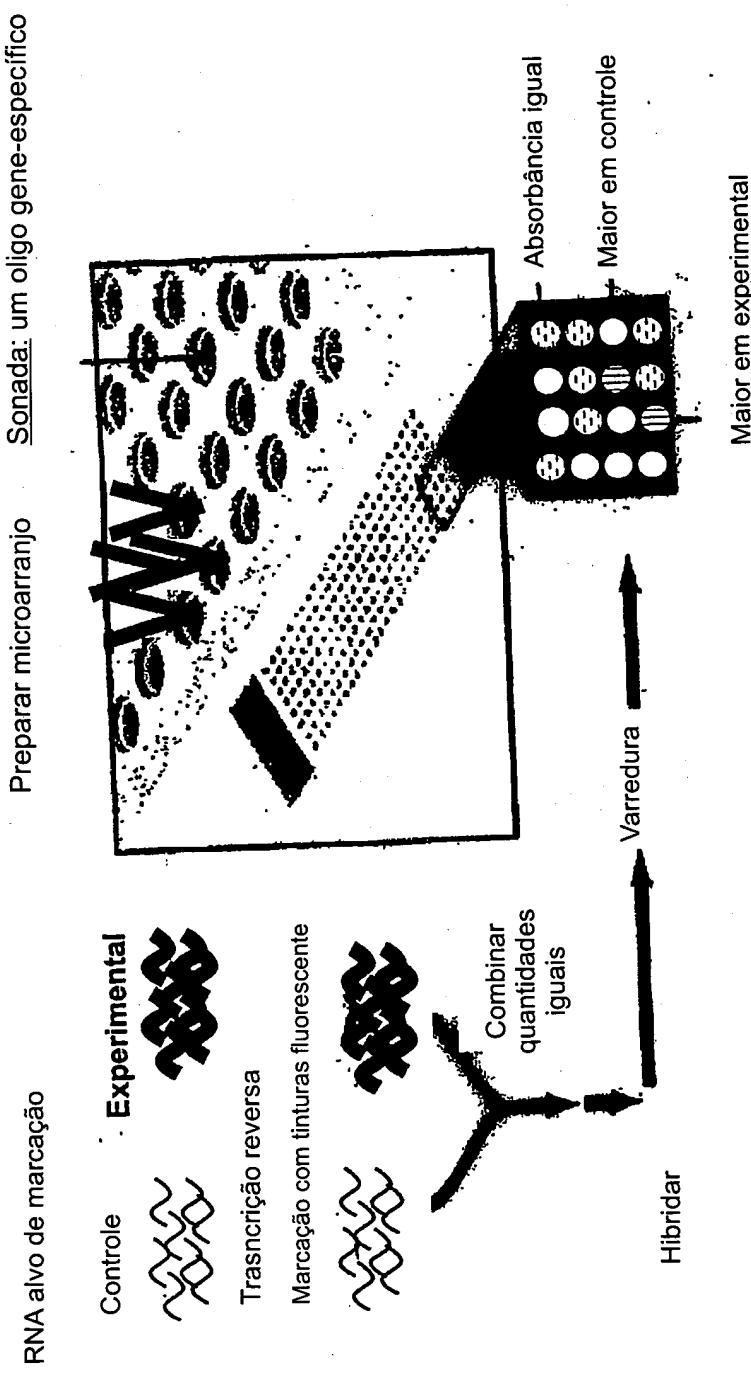


FIG. 12

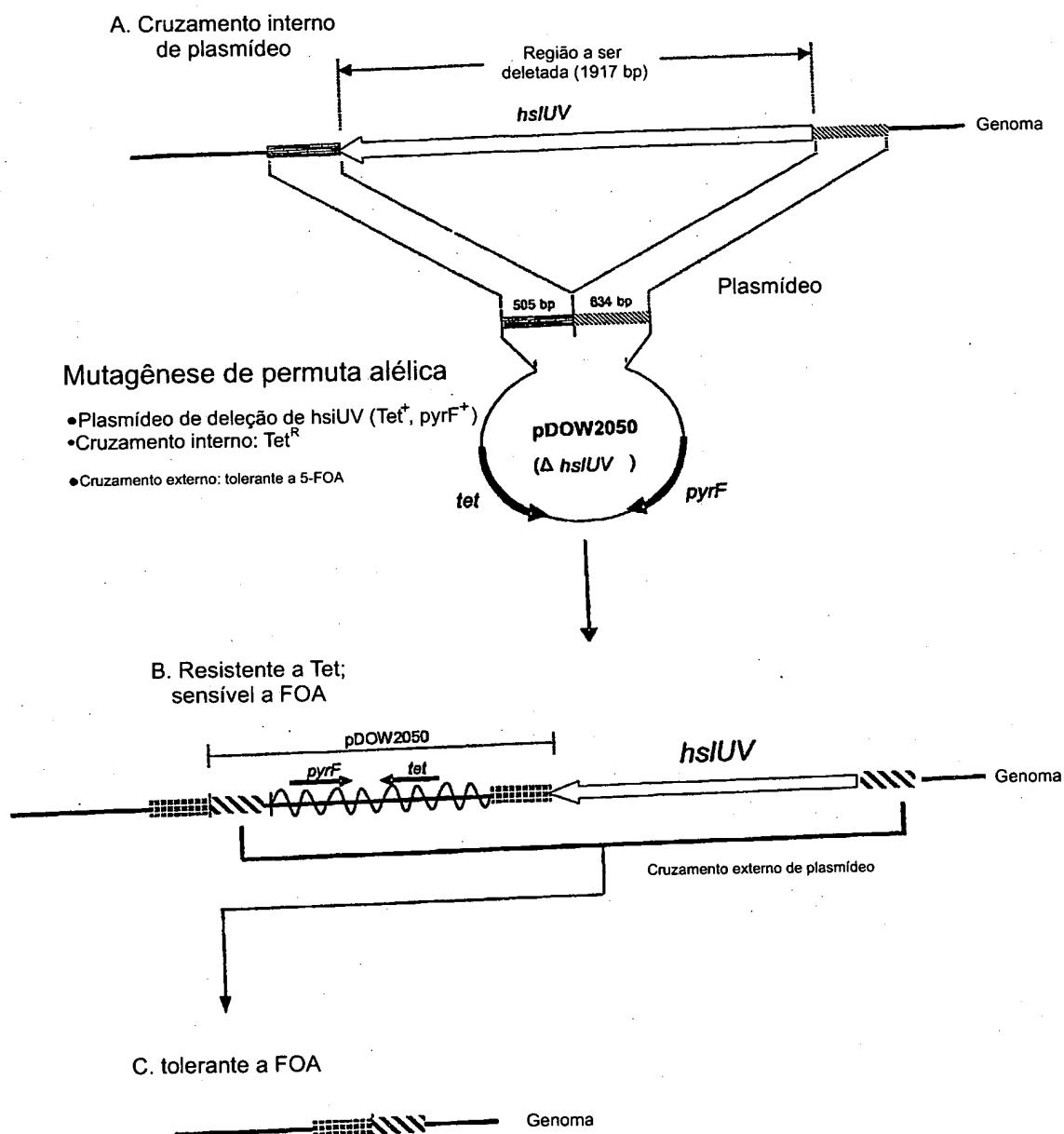


FIG. 13

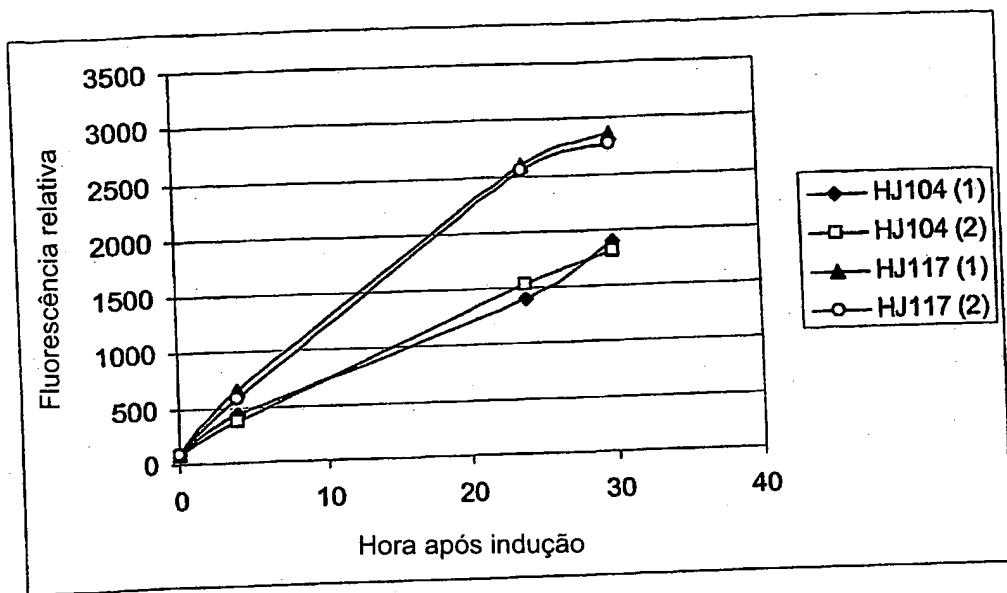


FIG. 14

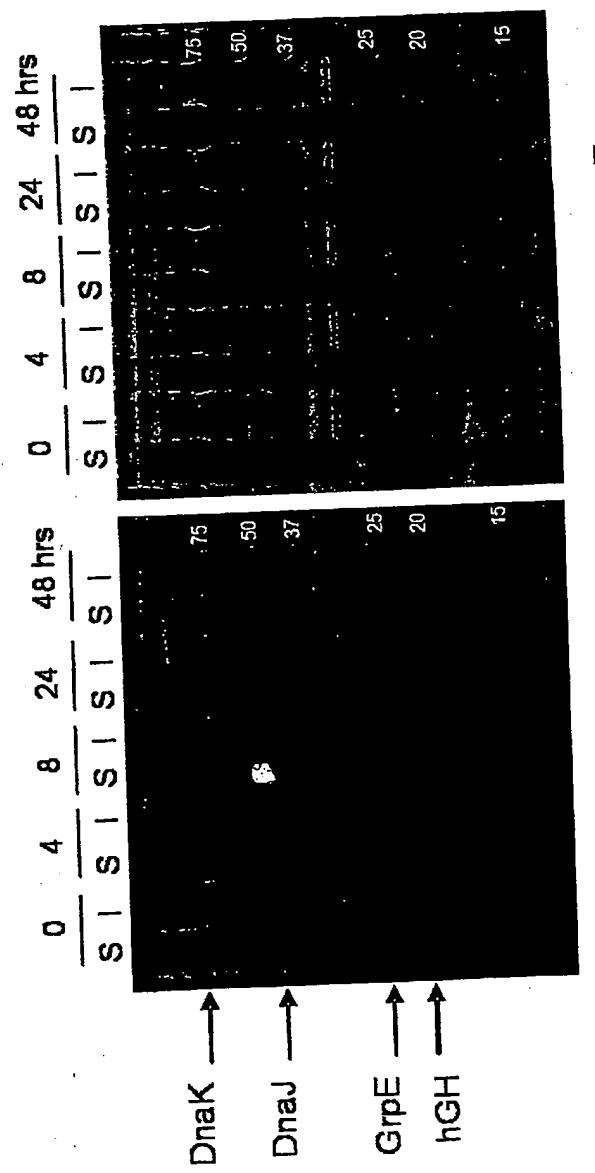


FIG. 15

Sem moduladores de dobramento

DnaKJ-GrpE

## RESUMO

### Patente de Invenção: "PROCESSO PARA EXPRESSÃO DE PROTEÍNA MELHORADA ATRAVÉS DE ENGENHARIA DE CEPA".

A presente invenção refere-se a um processo para melhorar os níveis de produção de proteínas ou peptídeos recombinantes ou melhorar o nível de proteínas ou peptídeos recombinantes ativos expressos em células hospedeiras. A invenção é um processo de comparar dois perfis genéticos de uma célula que expressa uma proteína recombinante e modificar a célula para alterar a expressão de um produto de gene que é sobre-regulado em resposta à expressão da proteína recombinante. O processo pode melhorar a produção de proteína ou pode melhorar a qualidade da proteína, por exemplo, aumentando a solubilidade de uma proteína recombinante.