

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

(21) PI 0919031-7 A2



* B R P I 0 9 1 9 0 3 1 A 2 *

(22) Data de Depósito: 28/09/2009
(43) Data da Publicação: 23/09/2014
(RPI 2281)

(51) Int.Cl.:
C12N 1/36
C12P 21/00

(54) Título: MICRO-ORGANISMOS E VACINAS
DEPENDENTES DE REPLICAÇÃO ATRAVÉS DO USO
DE AMINOÁCIDOS NÃO NATURAIS

(57) Resumo:

(30) Prioridade Unionista: 26/09/2008 US 61100688

(73) Titular(es): Ambrx, Inc

(72) Inventor(es): Brad Hehl, Feng Tian

(74) Procurador(es): NELLIE ANNE DAIEL-SHORES

(86) Pedido Internacional: PCT US2009058668 de
28/09/2009

(87) Publicação Internacional: WO 2010/037062de
01/04/2010

“MICRO-ORGANISMOS E VACINAS DEPENDENTES DE REPLICAÇÃO ATRAVÉS DO USO DE AMINOÁCIDOS NÃO NATURAIS”

CAMPO DA PRESENTE INVENÇÃO

A invenção refere-se a vacinas. Em algumas modalidades, a presente invenção refere-se a composições e métodos de produzir vacinas, incluindo vacinas de organismo total com capacidades de replicação limitadas ou ausentes, através do uso de aminoácidos não naturais ou não naturalmente codificados.

FUNDAMENTOS DA PRESENTE INVENÇÃO

Até agora, houve progresso técnico relativamente lento e, mesmo após 100 anos da morte de Louis Pasteur, seu protocolo “III” (Isolamento, Inativação, Injeção) continua a ser aplicável. Com o advento da biologia molecular e tecnologia de proteína recombinante, entretanto, pesquisadores começaram a desenvolver vacinas de subunidades e vacinas de organismos totais geneticamente construídas. Mais recentemente, avanços na imunologia tornaram várias classes de imunopotenciadores disponíveis, incluindo ligantes do Receptor do tipo Toll (TLR). Geralmente, gerações mais novas de vacinas e seus adjuvantes estão se tornando cada vez mais química e geneticamente definidas.

O desenvolvimento de produtos terapêuticos e, em particular, vacinas dirigidas contra patógenos, tais como vírus, bactérias, protozoários e fungos, é contínuo. Tal pesquisa provou ser inestimável na prevenção da propagação de doenças em animais, incluindo seres humanos. De fato, na medicina moderna, a imunoterapia, incluindo a vacinação, erradicou a varíola e, na prática, erradicou doenças, tais como poliomielite, tétano, tuberculose, catapora e sarampo.

Geralmente, vacinas ideais têm um prazo de validade longo, são capazes de induzir imunidade duradoura contra um patógeno pré-selecionado e todas as variantes fenotípicas, são incapazes de causar a doença contra a qual são dirigidas, são terapêutica e profilaticamente eficazes, são facilmente preparadas usando metodologias padrão econômicas e podem ser facilmente administradas no campo.

Existem quatro classes principais de vacinas comercialmente disponíveis. Elas incluem vacinas não vivas de organismo total, vacinas vivas atenuadas, vacinas de vetores e vacinas de subunidades. A vacinação com materiais não vivos, tais como proteínas, geralmente leva a uma resposta do anticorpo ou resposta de células T auxiliares CD4+, enquanto que a vacinação com materiais vivos (por exemplo, vírus infecciosos) geralmente leva a uma resposta de linfócitos T citotóxicos CD8+ (CTL). Uma resposta de CTL é crucial para a proteção contra patógenos do tipo vírus e bactérias infecciosos. Isto impõe um problema prático e a única maneira correta para obter uma resposta de CTL é usar agentes vivos que são por si só patogênicos. O problema é geralmente contornado através do uso de cepas virais e bacterianas atenuadas ou da morte de células totais que podem ser usadas para a vacina-

ção. Estas estratégias foram bem sucedidas, porém o uso de cepas atenuadas sempre acarreta o risco de que o agente atenuado possa se recombinar geneticamente no hospedeiro e se transformar em uma cepa virulenta. Desse modo, há uma necessidade quanto a produtos terapêuticos e métodos que possam produzir uma resposta de CTL CD8+ através de vacinação com materiais não vivos, tais como proteínas, em uma maneira específica.

Vacinas de subunidades proporcionaram meios para lidar com alguns destes problemas. Tais vacinas geralmente compreendem um componente subcelular derivado de um patógeno de interesse. Um componente de subunidade pode ser produzido a partir de uma fração subcelular definida do patógeno ou ser uma proteína, ácido nucleico ou polissacarídeo purificado. Todos estes elementos têm um determinante antigênico capaz de estimular uma resposta imune contra o patógeno de interesse. Geralmente, o componente subcelular da vacina de subunidade é obtido purificando-se uma preparação de patógeno rompido ou sintetizado usando procedimentos bem conhecidos.

Entretanto, existem várias limitações associadas às vacinas de subunidades. Primeiro, uma exigência quanto à produção de tal vacina é que o(s) determinante(s) antigênico(s) deve(m) ser caracterizado(s) e identificado(s). Isto impõe limitações em seu uso, particularmente contra determinantes antigênicos altamente variáveis. Segundo, vacinas de subunidades são geralmente ineficazes em estimular respostas de células T citotóxicas. Terceiro, a imunidade conferida por vacinas de subunidades frequentemente apresenta vida curta e, portanto, exige injeções de reforço contínuas. Poucas vacinas de subunidades recombinantes expressadas foram mostradas induzir imunidade forte e duradoura em animais vacinados (incluindo seres humanos). Uma exceção notável é a vacina contra a Hepatite B de antígeno de superfície recombinante usada em seres humanos. Um dos problemas associados ao uso de tais vacinas parece estar em apresentar corretamente os antígenos ao sistema imune, tal que a imunidade humoral forte e a imunidade mediada por células forte sejam induzidas. Em particular, vacinas recombinantes existentes (subunidades) não parecem resultar em respostas de “memória” fortes, tal que os animais vacinados reagem muito rapidamente quando são expostos a infecções naturais causadas por um patógeno.

Por via de exemplo apenas, deficiências nas vacinas de subunidades correntes preparadas a partir do vírus da diarreia viral bovina do tipo pestivírus (BVDV) foram extensivamente relatadas. Estes estudos mostraram que, ainda que grandes quantidades de proteína recombinante tenham sido usadas nas vacinas, taxas de proteção deficientes observadas foram apresentadas, mostrando que as vacinas falhariam na proteção a partir do desafio com isolados de BVDV vivos (tanto proteção homóloga quanto heteróloga).

Existem muitas doenças infecciosas para as quais uma vacina eficaz ainda não foi desenvolvida e muitas das vacinas correntemente disponíveis fornecem apenas proteção parcial contra a doença. Além disso, existem lacunas no campo das vacinas. Vacinas vivas

produzem imunidade mais forte, mais ampla e mais duradoura em relação a outros tipos de vacinas. Há uma necessidade quanto a um veículo de vacina viva mais seguro, que será incapaz de causar a doença, mesmo em indivíduos imunossuprimidos. Também há uma necessidade quanto a vacinas que induzem imunidade mediada por célula e não apenas imunidade com base em anticorpos. Além disso, há uma necessidade quanto à indução de respostas imunes protetoras diretamente nas superfícies mucosas do corpo, onde ocorre a entrada da maioria dos patógenos. Desse modo, há uma necessidade quanto a vacinas aperfeiçoadas. A presente invenção visa fornecer uma vacina terapêutica aperfeiçoada que melhora pelo menos algumas das desvantagens sobre a técnica anterior existente.

SUMÁRIO DA PRESENTE INVENÇÃO

A presente invenção fornece organismos totais dependentes de replicação quimicamente modulada. Em algumas modalidades, a presente invenção fornece um organismo total quimicamente modulado com uma modificação através da qual sua capacidade de replicar é dependente da presença de um aminoácido não natural. Em algumas modalidades, a presente invenção fornece um organismo total com uma ou mais modificações através das quais a capacidade de o organismo replicar depende se ele está sendo cultivado na presença de um aminoácido não natural. Em algumas modalidades, a presente invenção fornece um organismo total com uma ou mais modificações através das quais a capacidade de o organismo replicar depende se ele está sendo cultivado na presença de um ou mais aminoácidos não naturais. Em algumas modalidades, a presente invenção fornece um organismo total com duas ou mais modificações através das quais a capacidade de o organismo replicar depende se ele está sendo cultivado na presença de um ou mais aminoácidos não naturais. Em algumas modalidades da presente invenção, o organismo total é uma vacina atenuada - uma vacina modulada por um ou mais aminoácidos não naturais sítio-especificamente incorporados, em que a replicação ou expressão do agente imunizador é modulada pela presença ou ausência de um aminoácido não natural.

Em algumas modalidades da presente invenção, a replicação ou expressão de um organismo total é modulada pela presença ou ausência de um aminoácido não natural através de uma modificação genética única. Em algumas modalidades da presente invenção, a replicação ou expressão de um organismo total é modulada pela presença ou ausência de um aminoácido não natural através de uma ou mais modificações genéticas. Em algumas modalidades da presente invenção, a replicação ou expressão de um organismo total é modulada pela presença ou ausência de um aminoácido não natural através de duas modificações genéticas. Em algumas modalidades da presente invenção, a replicação ou expressão de um organismo total é modulada pela presença ou ausência de um aminoácido não natural através de três modificações genéticas. Em algumas modalidades da presente invenção, a replicação ou expressão de um organismo total é modulada pela presença ou ausência de

um aminoácido não natural através de quatro modificações genéticas. Em algumas modalidades da presente invenção, a replicação ou expressão de um organismo total é modulada pela presença ou ausência de um aminoácido não natural através de cinco modificações genéticas. Em algumas modalidades da presente invenção, a replicação ou expressão de um organismo total é modulada pela presença ou ausência de um aminoácido não natural através de quatro ou mais modificações genéticas.

A capacidade de induzir seletivamente uma resposta imune contra organismos, antígenos, proteínas ou proteínas próprias ou aumentar a imunogenicidade de epítomos específicos de antígenos estranhos é significativa na produção de vacinas para vários estados de doença (incluindo, mas não limitados ao câncer e doenças relacionadas à dobra de proteína) e doenças infecciosas (por exemplo, infecções bacterianas ou virais). A presente invenção utiliza a incorporação de aminoácidos não naturais em organismos para produzir imunógenos dependentes de replicação e/ou deficientes de replicação a serem usados em vacinações de organismo total ou para produzir anticorpos a serem usados na imunização passiva. A presente invenção também utiliza a incorporação de aminoácidos não naturais em proteínas, antígenos e/ou polipeptídeos para produzir imunógenos a serem usados em vacinações ou para produzir anticorpos a serem usados na imunização passiva. Na invenção, os imunógenos aos quais os aminoácidos não naturais são adicionados correspondem às porções alvo (por exemplo, porções relacionadas à doença) no paciente a ser vacinado/imunizado ou correspondem às porções alvo (por exemplo, porções relacionadas à doença) que são capazes de estar no paciente. Em modalidades onde o imunógeno com o aminoácido não natural é administrado a um paciente, a presença do aminoácido não natural evoca uma resposta imunológica contra o imunógeno que é reativo de forma cruzada contra a porção alvo (por exemplo, relacionada à doença).

Em um primeiro aspecto, a invenção fornece métodos de produzir ou acentuar uma resposta imunológica, por exemplo, uma resposta mediada por células B e/ou uma resposta mediada por células T, em um paciente, contra uma porção alvo, por exemplo, um polipeptídeo, um carboidrato ou uma combinação dos mesmos, que está no paciente ou que é capaz de estar no paciente. Os métodos incluem fornecer um organismo dependente de aminoácido não natural geneticamente modificado e administrar o organismo ao paciente. Os métodos também incluem fornecer um imunógeno não natural geneticamente modificado que compreende um ou mais aminoácidos não naturais e administrar o imunógeno não natural ao paciente. O paciente (por exemplo, ser humano, macaco, camundongo, rato, animal de criação, porco, vaca, galinha, pássaro de gaiola, pássaro de aviário, animal doméstico, cão, gato, réptil e/ou anfíbio) produz um ou mais anticorpos contra o imunógeno não natural e tais anticorpos reagem de forma cruzada contra a porção alvo (desse modo, produzindo ou acentuando a resposta imunogênica contra o alvo).

Em certas modalidades, o organismo total geneticamente modificado pode ser qualquer organismo total contra o qual é desejável imunizar o paciente, por exemplo, bactéria, vírus, fungo, *Mycoplasma*, protozoário, helminto ou príon. Uma vacina de organismo total pode opcionalmente incluir um ou mais entre: um antígeno bacteriano, antígeno viral, antígeno fúngico, antígeno de micoplasma, antígeno de protozoário, antígeno de helminto, antígeno de príon, antígeno de HIV, HIVgpl20, HIV gp41, HIV gag, HIV pol, HIV env, HTV tat, HIV nef, HTV rev, antígeno do capsídeo de calicivírus, antígeno do núcleo da hepatite B, antígeno de superfície da hepatite B, agente delta da hepatite, glicoproteína do vírus do herpes simples, glicoproteína do vírus varicela zoster, hemaglutinina do vírus influenza, neuraminidase do vírus influenza, nucleoproteína do vírus influenza, proteína do capsídeo do HPV, hemaglutinina/neuraminidase do vírus parainfluenza, polipeptídeo do capsídeo de poliovírus, antígeno da Hep A, polipeptídeo do vírus da vacínia, glicoproteína G do vírus da raiva, OspA de *B. burgdorferi*, proteína da membrana externa de *H. influenzae* tipo b, lipoarabinomannana de micobactéria, mAPG de micobactéria, proteína M de *S. pyogenes*, polisacarídeo capsular de *S. pneumoniae*, FI de *Y. pestis*, antígeno V de *Y. pestis*, circunsporozoito de *P. falciparum* (PfCSP), proteína 2 da superfície de esporozoito de *P. falciparum* (PfSSP2), terminal carboxila de *P. falciparum* do antígeno 1 no estado vivo (c-term de PfL-SAI), proteína 1 exportada de *P. falciparum* (PfExp-1), Pfs 48/45, Pfs 28, Pfs 25 ou Pfs 230.

O organismo dependente de replicação de aminoácido não natural geneticamente modificado pode ser um ou mais entre: uma bactéria, vírus, fungo, *Mycoplasma*, protozoário, helminto, príon, *Actinomyces*, *Bacillus*, *Bacteroides*, *Bordetella*, *Bartonella*, *Borrelia*, *Brucella*, *Campylobacter*, *Capnocytophaga*, *Chlamydia*, *Clostridium*, *Corynebacterium*, *Coxiella*, *Dermatophilus*, *Enterococcus*, *Ehrlichia*, *Escherichia*, *Francisella*, *Fusobacterium*, *Haemobartonella*, *Haemophilus*, *Helicobacter*, *Klebsiella*, uma bactéria na forma de L, *Leptospira*, *Listeria*, *Mycobacterium*, *Mycoplasma*, *Neisseria*, *Neorickettsia*, *Nocardia*, *Pasteurella*, *Peptococcus*, *Peptostreptococcus*, *Pneumococcus*, *Proteus*, *Pseudomonas*, *Rickettsia*, *Rochalimaea*, *Salmonella*, *Shigella*, *Staphylococcus*, *Streptococcus*, *Treponema*, *Yersinia*, adenovírus, alfavírus, calicivírus, coronavírus, CMV, vírus da cinomose, vírus Ebola, enterovírus, EBV, flavivírus, vírus da Hep C, hepadnavírus, vírus da Hep B, agente delta da hepatite, vírus da Hep E ou F, GBV-C, herpesvírus, vírus do herpes simples, vírus varicela-zoster, vírus da imunodeficiência, HIV, vírus da peritonite infecciosa, vírus influenza, vírus influenza A, vírus da leucemia, vírus Marburg, ortomixovírus, papilomavírus, HPV, vírus parainfluenza, paramixovírus, RSV, parvovírus, pestivírus, picornavírus, poliovírus, poxvírus, vírus da vacínia, vírus da raiva, reovírus, retrovírus, rotavírus, *Absidia*, *Acremonium*, *Alternaria*, *Aspergillus*, *Basidiobolus*, *Bipolaris*, *Blastomyces*, *Candida*, *Coccidioides*, *Conidiobolus*, *Cryptococcus*, *Curvalaria*, *Epidermophyton*, *Exophiala*, *Geotrichum*, *Histoplasma*, *Madurella*, *Malassezia*, *Microsporum*, *Moniliella*, *Mortierella*, *Mucor*, *Paecilomyces*, *Penicillium*, *Phialemonium*,

Phialophora, Prototheca, Pseudallescheria, Pseudomicrodochium, Pythium, Rhizopus, Scolecobasidium, Sporothrix, Stemphylium, Trichophyton, Trichosporon, Xylohypha, Babesia, Balantidium, Besnoitia, Cryptosporidium, Eimeria, Encephalitozoon, Entamoeba, Giardia, Hammondia, Hepatozoon, Isospora, Leishmania, Microsporidia, Neospora, Nosema, Pentatrichomonas, Plasmodium, P. falciparum, Pneumocystis, Sarcocystis, Schistosoma, Theileria, Toxoplasma, Trypanosoma, Acanthocheilonema, Aelurostrongylus, Ancylostoma, Angiostrongylus, Ascaris, Brugia, Bunostomum, Capillaria, Chabertia, Cooperia, Crenosoma, Dictyocaulus, Dioctophyme, Dipetalonema, Diphyllbothrium, Diplydium, Dirofilaria, Dracunculus, Enterobius, Filaroides, Haemonchus, Lagochilascaris, Loa polipeptídeo, Mansonella, Muellerius, Nanophyetus, Necator, Nematodirus, Oesophagostomum, Onchocerca, Opisthorchis, Ostertagia, Parafilaria, Paragonimus, Parascaris, Physaloptera, Protoststrongylus, Setaria, Spirocerca, Spirometra, Stephanofilaria, Strongyloides, Strongylus, Thelazia, Toxascaris, Toxocara, Trichinella, Trichostrongylus, Trichuris, Uncinaria ou Wuchereria.

Em um outro aspecto, a invenção fornece métodos de tratar profilática ou terapêuticamente um estado de doença em um paciente, por exemplo, através da produção de uma resposta mediada por células B e/ou uma resposta mediada por células T no paciente. Em várias modalidades, o estado de doença pode ser, mas não é limitado a um ou mais entre: uma infecção bacteriana, infecção viral, infecção fúngica, infecção por *Mycoplasma*, infecção por prion, infecção por protozoário ou infecção por helminto. Um conjunto de métodos do aspecto inclui administrar a um paciente, por exemplo, ser humano, macaco, camundongo, rato, porco, vaca, galinha, pássaro de gaiola, pássaro de aviário, réptil ou anfíbio, um organismo total geneticamente modificado. O organismo total geneticamente modificado estimula a produção de anticorpos no paciente. Em um segundo conjunto de métodos, deste aspecto, a invenção compreende tratar profilática ou terapêuticamente um estado de doença em um paciente através da produção de um anticorpo contra uma ou mais doenças, condições ou organismos que envolve produzir uma resposta do anticorpo e isolar o anticorpo ou anticorpos que depois são administrados a um paciente.

Com a última ferramenta de engenharia de proteínas, a tecnologia da Ambrx, a incorporação sítio-específica de aminoácidos não naturais apresenta uma grande oportunidade no desenvolvimento de vacinas. Esta pode criar nanoestruturas quimicamente bem definidas através da combinação de antígenos com imunopotenciadores. Estas nanoestruturas, que podem ser usadas para o desenvolvimento de vacinas de subunidades, possuem informação do antígeno e funcionalidade de imunopotenciação.

Além disso, a tecnologia da Ambrx também pode ser usada para criar novas vacinas de organismo total atenuadas, em que os aminoácidos não naturais servirão como modificações para a expressão de genes essenciais em organismos patogênicos, ao invés de

veículos para introduzir grupos funcionais não naturais na proteína. Os organismos resultantes emularão o organismo do tipo selvagem, mas serão apenas capazes de replicação na presença de aminoácidos não naturais da Ambrx, que não existem na natureza.

Vacina de subunidade:

5 Devido à baixa eficácia de vacinas de subunidades, adjuvantes são exigidos para estimular uma resposta imune. Tradicionalmente, os adjuvantes são formulados como misturas com antígenos. Sob administração, os antígenos serão absorvidos pela célula dendrítica (DC) e apresentados às células T. Os adjuvantes estimularão as células dendríticas para liberação de citocina, apresentação de antígeno acentuada e maturação de células dendríticas. Uma resposta imune específica do antígeno eficaz será obtida quando estes dois processos independentes, apresentação de antígeno e imunoativação, convergem na mesma DC. Uma dosagem relativamente alta é exigida. Entretanto, a dosagem alta de adjuvante evocará a ativação imune independente de antígeno, um efeito colateral indesejado. Portanto, intrinsecamente, este método não é ideal e confere uma baixa eficácia e alta toxicidade à vacina.

15 Ao contrário de adjuvantes tradicionais, tais como alume, diversos adjuvantes recém descobertos são molecularmente bem definidos. Com sua incorporação de aminoácidos não naturais sítio-específica única e suas capacidades de conjugação sítio-específicas, a tecnologia da Ambrx nos fornece as ferramentas necessárias para criar a próxima geração de vacinas com alta potência e baixa toxicidade. A característica essencial das vacinas de subunidades da Ambrx será a combinação de antígeno e imunopotenciador em uma nano-estrutura quimicamente definida. Quando é absorvida por uma DC, os dois processos, apresentação de antígeno e ativação imune, convergem na mesma DC.

20 A incorporação sítio-específica de aminoácidos não naturais é uma tecnologia que nos permite incorporar aminoácidos não naturais com porções altamente imunogênicas, tais como mononitrofenila e dinitrofenila, diretamente nos antígenos de proteína. Recentemente, imunopotenciadores de molécula pequena, tais como ligante imidazoquinolina do Receptor do tipo Toll 7 (TLR7), tornaram-se disponíveis. É possível projetar aminoácidos não naturais com estes imunopotenciadores de molécula pequena como suas cadeias laterais e incorporá-los em um antígeno de proteína.

30 A tecnologia de conjugação sítio-específica da Ambrx fornece ainda mais flexibilidade. Ela nos permite conjugar não apenas as moléculas pequenas anteriormente mencionadas à superfície de antígenos de proteína, mas também a uma variedade de outros imunopotenciadores (por exemplo, lipídeos, lipopeptídeos, polissacarídeos, DNA, RNA e nanopartículas).

35 A conjugação sítio-específica do DNA de fita única em superfícies de proteínas cria uma outra dimensão de facilidade para o projeto de uma vacina. DNAs com CpG não meti-

lado são ligantes de TLR-9, que, de forma interessante, não estão presentes na superfície celular, mas dentro das células. O DNA conjugado ao antígeno pode servir não apenas como imunopotenciador, mas também como um bloco de construção para criar uma a três dimensões e nanoestruturas valentes de multiantígeno através dos processos de hibridização específicos da sequência. Este esquema geral também pode ser usado para combinar antígenos com outros elementos, tais como reagentes alvejantes de APC e outros ligantes de TLR. Foi mostrado que alvejantes de APC (anticorpo ou peptídeo) podem acentuar a eficácia da vacina e promovem apresentação cruzada. A combinação de dois ligantes de TLR diferentes terá maior efeito sinérgico.

A conjugação sítio-específica da Ambrx possibilita o projeto preciso e otimização de vacinas. Sem a tecnologia da Ambrx, um processo de conjugação não específica apresenta controle limitado ou nenhum controle sobre os sítios de modificação no antígeno de proteína. A conjugação não específica pode alterar os epítomos T e B. Ela também pode modificar a conformação 3D do antígeno que é crítica para o reconhecimento do anticorpo. Além disso, algumas modificações podem impedir o processamento do antígeno. Todos estes fatores tornam a vacina não especificamente conjugada menos eficaz.

Vacina de organismo total:

A vacinologia teve início com vacinas de organismo total e, ainda hoje, estas são as vacinas mais bem sucedidas, classificadas como vivas atenuadas ou mortas. A essência das vacinas de organismo total é que a vacina deve ser o mais parecida possível ao organismo patogênico por si só, de modo a provocar uma resposta imune protetora, porém, com capacidades de replicação muito limitadas ou ausentes no hospedeiro.

Portanto, se a replicação limitada ou ausente é desejável, é possível usar moléculas pequenas para controlar o ciclo de vida de micro-organismos para criar vacinas melhores e mais seguras? A resposta, usando a tecnologia da Ambrx, é que agora é possível. Se o aminoácido não natural é incorporado em um gene em uma bactéria, vírus ou ainda um parasita que é essencial à replicação dos micro-organismo, a vida destes organismos será dependente da presença do aminoácido não natural. Pelo fato de o aminoácido não natural incorporado não existir na natureza, o resultado será organismos, tais como vírus e bactérias, tendo componentes e estruturas virais e celulares quase exatos, sem a capacidade de replicar no hospedeiro. Além disso, o vírus ou bactéria modificado pela tecnologia da Ambrx pode servir não apenas como a vacina por si só, mas também como vetores para o gene e liberação de antígeno.

A plataforma da vacina de subunidade da Ambrx tem uma ampla variedade de aplicações em doenças infecciosas e áreas de vacina terapêutica contra o câncer. Vacinas correntes contra o câncer são geralmente ineficazes. Abordagens da ativação sistêmica de células T e inibição da via regulatória negativa foram propostas e algumas delas estão em ex-

perimentos clínicos. O desacoplamento da ativação de células T e reconhecimento do antígeno pode causar sérios efeitos colaterais. A ineficácia de vacinas contra o câncer é devido à ausência de tecnologia para induzir respostas imunes fortes específicas do tumor sem causar efeitos colaterais sérios. A Ambrx fornece a tecnologia para criar nanoestruturas que
 5 podem invocar respostas imunes fortes específicas do câncer, incluindo respostas de células T e células B, desse modo, proporcionando possibilidades vantajosas, tanto terapêuticas quanto preventivas.

A plataforma da vacina de organismo total pode encontrar incontáveis aplicações na área de doenças infecciosas. Vírus, bactérias ou outros parasitas que podem ser cultivados e têm um genoma capaz de mutação são bons candidatos. Vacinas da presente invenção
 10 aperfeiçoarão a administração de risco no desenvolvimento da vacina e as vacinas serão mais definidas, tanto química quanto geneticamente, e fornecerão eficácias inéditas.

Para expandir o código genético, a invenção fornece composições e métodos para produzir tRNAs ortogonais. Aminoacil-tRNA sintetases aminoacilam tRNAs da presente invenção com um aminoácido não naturalmente codificado. Estes componentes de tradução
 15 podem ser usados para incorporar um aminoácido selecionado em uma posição específica em uma cadeia polipeptídica crescente (durante a tradução de ácido nucleico) em resposta a um códon seletor que é reconhecido pelo tRNA.

Métodos de produzir uma proteína em uma célula com um aminoácido selecionado em uma posição específica também são uma característica da presente invenção. Por exemplo, um método inclui cultivar, em um meio apropriado, uma célula, em que a célula compreende um ácido nucleico que compreende pelo menos um códon seletor e codifica uma proteína, e fornecer o aminoácido selecionado. A célula compreende ainda: um tRNA ortogonal (O-tRNA) que funciona na célula e reconhece o códon seletor e uma aminoacil-
 20 tRNA sintetase ortogonal (O-RS) que preferencialmente aminoacila o O-tRNA com o aminoácido selecionado. Tipicamente, o O-tRNA compreende atividade de supressão na presença de uma sintetase cognata. Uma proteína produzida por este método também é uma característica da presente invenção.

BREVE DESCRIÇÃO DOS DESENHOS

30 Figura 1 - A estrutura de folha de trevo de J17 tRNA com sítios de mutação em tronco TΨC é mostrada.

Figura 2 - A supressão de uma mutação âmbar no hormônio de crescimento humano é mostrada usando J17 ou mutantes de J17 (F12, F13, F14). O lisato celular total para cada amostra foi analisado por SDS PAGE.

35 Figura 3 - A supressão de uma mutação âmbar no hormônio de crescimento humano é mostrada em linhagens celulares diferentes usando F13.

Figura 4 - Mostra um diagrama do vetor pVK10-camR contendo o Sistema de Su-

pressão da Ambrx.

Figura 5 - Mostra um diagrama do vetor pKD46 contendo componentes Lambda Red (bet, gam, exonuclease).

Figura 6 - Mostra um esquema de *primers* sobrepostos e os métodos usados no

Exemplo 5. Os *primers* são indicados por setas e identificador numérico, e, por exemplo, *primers* 1 para os clones Y31 e N51 seriam a SEQ ID NO: 18 e 26, respectivamente; *primers* 2 para os clones Y31 e N51 seriam a SEQ ID NO: 19 e 27, respectivamente; *primers* 3 para os clones Y31 e N51 seriam a SEQ ID NO: 20 e 28, respectivamente; *primers* 4 para os clones Y31 e N51 seriam a SEQ ID NO: 21 e 29, respectivamente; *primers* 5 para os clones Y31 e N51 seriam a SEQ ID NO: 22 e 30, respectivamente; *primers* 6 para os clones Y31 e N51 seriam a SEQ ID NO: 23 e 31, respectivamente; *primers* 7 para os clones Y31 e N51 seriam a SEQ ID NO: 24 e 32, respectivamente; *primers* 8 para os clones Y31 e N51 seriam a SEQ ID NO: 25 e 33, respectivamente. Os produtos da PCR são identificados por uma letra. O sítio âmbar contendo metiona amino peptidase foi gerado sobrepondo-se PCR sobre a mutação âmbar (A e B), tornando-a condicionalmente letal a *E. coli*, como um organismo total. O mutante metiona amino peptidase depois foi fundido em um gene completo (C) e combinado com um marcador selecionável (D) e uma parte parcial do mutante metiona amino peptidase (E) para a seleção a jusante.

Figura 7 - Mostra um esquema de triagem para dependência de aminoácido artificial. As três etapas mostradas são 1) escolha dos transformantes, tais como transformantes positivos KanR, AmpR, tratando-se as placas; 2) cultivo dos sobreviventes em KanR, AmpR, aminoácido não natural +; e 3) crescimento de colônias individuais em placas com e sem o aminoácido não natural e seleção das colônias que crescem na presença do aminoácido não natural e morrem quando cultivadas sem o aminoácido não natural.

Figura 8 - Mostra duas placas de cultura celular de *E. coli* que estão sendo triadas para uma mutação âmbar no gene metiona amino peptidase em Y31, cada uma revestida com LB ágar e com meio contendo Ampicilina 50 µg/mL e Canamicina 50 µg/mL, a placa da esquerda adicionalmente contendo para-acetilfenilalanina 2 mM. A12, C5 e E3 são exemplos de *E. coli* dependente de replicação quimicamente modulada.

Figura 9 - Mostra duas placas de cultura celular de *E. coli* que estão sendo triadas para uma mutação âmbar no gene metiona amino peptidase em N51, cada uma revestida com LB ágar e com meio contendo Ampicilina 50 µg/mL e Canamicina 50 µg/mL, a placa da esquerda adicionalmente contendo para-acetilfenilalanina 2 mM. A12, C5 e E3 são exemplos de *E. coli* dependente de replicação quimicamente modulada.

Figura 10A e 10B - Mostra o crescimento em meio permissivo e não permissivo. Durante a noite, as culturas foram inoculadas em meio de glicose (losangos) ou arabinose (círculos) nas razões de diluição ideais mostradas no canto direito superior de cada painel. A

OD600 foi monitorada em uma placa de 96 poços em um leitor de placa Spectramax com agitação a 37 °C com remoção periódica de amostras para microscopia e laminação em placas de arabinose. A mudança na viabilidade, fornecida no eixo direito (triângulos), foi calculada usando-se a concentração de CFU normalizada dividindo-se pela OD600. Para cada ponto no tempo, este valor foi dividido pelo valor em T = 0 e o log desse valor é apresentado neste relatório. Os números de CFU por mililitro por unidade de OD em T = 0 medidos a partir do inóculo saturado e corrigidos para diluição foram os seguintes: WT, $4,6 \times 10^9$; mutante *frr*, $3,2 \times 10^9$; mutante *gcpE*, $3,4 \times 10^9$; mutante *lpxC*, $2,2 \times 10^9$; mutante *map*, $2,9 \times 10^9$; mutante *murA*, $8,7 \times 10^8$; mutante *ppa*, $1,5 \times 10^9$; mutante *rpsA*, $1,2 \times 10^9$. A OD é mostrada em uma escala linear ao invés de uma escala logarítmica para mostrar, da melhor forma, a dinâmica das culturas hesitantes, e a escala do eixo direito difere para *lpxC* e *map* para evitar a sobreposição das linhas de tendência. Para os últimos dois pontos no tempo da medição de *murA*, não houve crescimento de colônias. Do Herring e Blattner JOURNAL OF BACTERIOLOGY, 2004, páginas 2673-2681, incorporado como referência por meio deste relatório.

Figura 11 - Mostra um esquema da caracterização de taxas de reversão em que células dependentes de replicação através de aminoácido não natural geneticamente modificadas são cultivadas na densidade óptica de 1,0 em meio de aminoácido não natural +, depois essas células são plaqueadas em diluições em série em meio com aminoácido não natural adicionado para calcular unidades formadoras de colônia e as células remanescentes são cultivadas sem a presença do aminoácido não natural para detectar revertentes.

Figura 12 - Mostra um esquema de uma estratégia de mutagênese. Um fragmento de DNA linear é produzido a partir DNA genômico de molde WT por sobreposição da extensão da PCR. As posições dos *primers* são indicadas por setas unilaterais. O produto de fusão da PCR é eletroporado em células e as integrações resultantes de um evento de recombinação dupla são selecionadas. Depois da identificação de um clone portando o códon de parada âmbar *tagalong*, o gene I-SceI é induzido, resultando na remoção do gene que codifica Camr. A recombinação dentro de uma região duplicada curta leva à geração de um mutante âmbar que é, de outro modo, sem marcas.

Figura 13 - Mostra plasmídeos usados na mutagênese *tagalong*.

Figura 14 - Mostra um esquema das etapas de criação de um mutante letal condicional em *E. coli*, partindo do princípio, 1) criar um molde de mutante condicional; 2) transformar em cepa competente de supressão âmbar lambda RED recombinante; e 3) induzir a recombinação lambda RED.

Figura 15 - Mostra um esquema das etapas na caracterização de um mutante letal condicional em *E. coli*, partindo do princípio, 1) caracterizar o crescimento na presença e ausência do aminoácido não natural sob o qual um organismo transformado com êxito seria

dependente; 2) caracterizar a mutação como bactericida ou bacteriostática, e 3) caracterizar a frequência do revertente.

Figura 16 - Mostra uma vista estrutural das seleções de resíduo de MAP.

Figura 17 - Mostra uma vista estrutural das seleções de resíduo de MAP.

5

DEFINIÇÕES

Antes de descrever a invenção em detalhes, deve ser entendido que esta invenção não é limitada a sistemas biológicos particulares, que podem, certamente, variar. Também deve ser entendido que a terminologia usada neste relatório é para o propósito de descrever apenas modalidades particulares, e não é intencionada a limitar o escopo da presente invenção, que será limitado apenas pelas reivindicações anexas. Conforme usado neste relatório e nas reivindicações anexas, as formas no singular “um”, “uma” e “o”, “a” incluem as referências no plural, a menos que o conteúdo claramente indique de outro modo. Desta forma, por exemplo, referência a “uma célula” inclui uma combinação de duas ou mais células e inclui equivalentes da mesma conhecidos àqueles de habilidade comum na técnica, e assim por diante. Referência a “bactérias” inclui misturas de bactérias e semelhantes.

10

15

A menos que definido neste relatório e abaixo no restante do relatório descritivo, todos os termos técnicos e científicos usados neste relatório têm o mesmo significado, conforme comumente entendido por uma pessoa de habilidade comum na técnica a qual a invenção pertence.

20

Todas as publicações e patentes mencionadas neste relatório são incorporadas no mesmo como referência para o propósito de descrever e divulgar, por exemplo, os construtos e metodologias que são descritos nas publicações, que podem ser usados em relação à invenção presentemente descrita. As publicações debatidas neste relatório são fornecidas somente para sua divulgação antes da data de depósito do presente pedido. Nada neste relatório deve ser interpretado como uma admissão de que os inventores estejam habilitados a prever tal divulgação em virtude da invenção anterior ou por qualquer outra razão.

25

Proteínas e/ou sequências de proteínas são “homólogas” quando elas são derivadas, natural ou artificialmente, de uma proteína ou sequência de proteína ancestral comum. Similarmente, ácidos nucleicos e/ou sequências de ácidos nucleicos são homólogos quando eles são derivados, natural ou artificialmente, de um ácido nucleico ou sequência de ácido nucleico ancestral comum. Por exemplo, qualquer ácido nucleico que ocorre naturalmente pode ser modificado por qualquer método de mutagênese disponível para incluir um ou mais códons seletores. Quando expressado, este ácido nucleico mutagenizado codifica um polipeptídeo compreendendo um ou mais aminoácidos selecionados, por exemplo, aminoácidos não naturais. O processo de mutação pode, adicionalmente, alterar um ou mais códons padrão, desse modo, modificando um ou mais aminoácidos padrão na proteína mutante resultante. Um ou mais aminoácidos padrão podem ser modificados para um aminoácido não

30

35

natural ou um aminoácido natural. A homologia é geralmente inferida a partir da similaridade de sequência entre dois ou mais ácidos nucleicos ou proteínas (ou sequências dos mesmos). A porcentagem precisa de similaridade entre sequências que é útil no estabelecimento da homologia varia de acordo com o ácido nucleico e proteína em questão, porém, 25 % de similaridade de sequência são rotineiramente usados para estabelecer a homologia. Níveis mais altos de similaridade de sequência, por exemplo, 30 %, 35 %, 40 %, 45 %, 50 %, 55 %, 60 %, 65 %, 70 %, 75 %, 80 %, 85 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 % ou mais, também podem ser usados para estabelecer a homologia. Métodos para determinar as porcentagens de similaridade de sequência (por exemplo, BLASTP e BLASTN usando parâmetros pré-determinados) são descritos neste relatório e são geralmente disponíveis.

Conforme usado neste relatório, o termo “ortogonal” refere-se a uma molécula (por exemplo, um tRNA ortogonal (O-tRNA) e/ou uma aminoacil tRNA sintetase ortogonal (O-RS)) que é usada com eficácia reduzida por um sistema de interesse (por exemplo, um sistema de tradução, por exemplo, uma célula). Ortogonal refere-se à incapacidade ou eficácia reduzida, por exemplo, menos do que 20 % eficaz, menos do que 10 % eficaz, menos do que 5 % eficaz ou, por exemplo, menos do que 1 % eficaz, de um tRNA ortogonal e/ou RS ortogonal funcionar no sistema de tradução de interesse. Por exemplo, um tRNA ortogonal em um sistema de tradução de interesse é aminoacilado por qualquer RS endógena de um sistema de tradução de interesse com eficácia reduzida ou nula em comparação à aminoacilação de um tRNA endógeno por uma RS endógena. Em um outro exemplo, uma RS ortogonal aminoacila qualquer tRNA endógeno no sistema de tradução de interesse com eficácia reduzida ou nula em comparação à aminoacilação do tRNA endógeno por uma RS endógena. Uma segunda molécula ortogonal pode ser introduzida na célula, a qual funciona com a primeira molécula ortogonal. Por exemplo, um par de tRNA/RS ortogonal inclui componentes complementares introduzidos que funcionam juntos na célula com uma eficácia, por exemplo, cerca de 50 % de eficácia, cerca de 60 % de eficácia, cerca de 70 % de eficácia, cerca de 75 % de eficácia, cerca de 80 % de eficácia, cerca de 85 % de eficácia, cerca de 90 % de eficácia, cerca de 95 % de eficácia, cerca de 99 % ou mais de eficácia, em relação àquela de um par de tRNA/RS endógeno correspondente. O “aperfeiçoamento na ortogonalidade” refere-se à ortogonalidade acentuada comparada a um material de partida ou um tRNA ou RS que ocorre naturalmente.

O termo “cognato” refere-se aos componentes que funcionam juntos, por exemplo, um tRNA e uma aminoacil-tRNA sintetase. Os componentes também podem ser referidos como complementares.

O termo “preferencialmente aminoacila” refere-se a uma eficácia, por exemplo, cerca de 70 % eficaz, cerca de 75 % eficaz, cerca de 80 % eficaz, cerca de 85 % eficaz, cerca

de 90 % eficaz, cerca de 95 % eficaz, cerca de 99 % ou mais eficaz, em que uma ORS aminoacila um O-tRNA com um aminoácido selecionado, por exemplo, um aminoácido não natural, em comparação à O-RS que aminoacila um tRNA que ocorre naturalmente ou um material de partida usado para gerar o O-tRNA. O aminoácido não natural depois é incorporado em uma cadeia polipeptídica crescente com alta fidelidade, por exemplo, mais do que cerca de 70 % de eficácia para um códon seletor fornecido, mais do que cerca de 75 % de eficácia para um códon seletor fornecido, mais do que cerca de 80 % de eficácia para um códon seletor fornecido, mais do que cerca de 85 % de eficácia para um códon seletor fornecido, mais do que cerca de 90 % de eficácia para um códon seletor fornecido, mais do que cerca de 95 % de eficácia para um códon seletor fornecido ou mais do que cerca de 99 % de eficácia para um códon seletor fornecido.

Um “tratamento profilático” é um tratamento administrado a um paciente que não apresenta sinais ou sintomas de uma doença, patologia ou distúrbio médico ou apresenta apenas sinais ou sintomas precoces de uma doença, patologia ou distúrbio, tal que o tratamento é administrado para o propósito de diminuir ou prevenir o risco de desenvolvimento da doença, patologia ou distúrbio médico. Um tratamento profilático funciona como um tratamento preventivo contra uma doença ou distúrbio. Uma “atividade profilática” é uma atividade de um agente, tal como um imunógeno e/ou anticorpo não natural, ou composição dos mesmos, que, quando administrados a um paciente que não apresenta sinais ou sintomas de uma patologia, doença ou distúrbio (ou que apresenta apenas sinais ou sintomas precoces de tal patologia, doença ou distúrbio) diminui ou previne o risco de o paciente desenvolver a patologia, doença ou distúrbio. Um agente ou composto “profilaticamente útil” (por exemplo, um imunógeno e/ou anticorpo não natural da invenção) refere-se a um agente ou composto que é útil para diminuir, prevenir ou tratar o desenvolvimento de uma patologia, doença ou distúrbio.

Tratamentos, vacinas e/ou tratamentos profiláticos podem ser administrados a um paciente em necessidade dos mesmos. Tratamentos, vacinas e/ou tratamentos profiláticos também podem ser administrados a uma variedade de animais, incluindo, mas não limitados a animais de criação, tais como vacas, porcos, cabras, ovelhas, galinhas e/ou outros animais de criação comuns e animais domésticos comuns, por exemplo, gatos, cães, papagaios, periquitos, etc.

O termo “códon seletor” refere-se a códons reconhecidos pelo OtRNA no processo de tradução e não reconhecidos por um tRNA endógeno. A alça anticódon de O-tRNA reconhece o códon seletor no mRNA e incorpora seu aminoácido, por exemplo, um aminoácido selecionado, tal como um aminoácido não natural, neste sítio no polipeptídeo. Códons seletores podem incluir, mas não são limitados, por exemplo, a códons sem sentido, tais como, códons de parada, incluindo, mas não limitados a códons âmbar, ocre e opala; códons de

quatro ou mais bases; códons raros; códons derivados de pares de base naturais ou não naturais e/ou semelhantes. Para um sistema fornecido, um códon seletor também pode incluir um dos códons de três bases naturais, em que o sistema endógeno não usa (ou raramente usa) o dito códon de três bases natural. Por exemplo, isto inclui um sistema que é

5 desprovido de um tRNA que reconhece o códon de três bases natural e/ou um sistema em que o códon de três bases natural é um códon raro.

Um tRNA supressor é um tRNA que altera a leitura de um RNA mensageiro (mRNA) em um sistema de tradução fornecido, por exemplo, fornecendo-se um mecanismo para incorporar um aminoácido em uma cadeia polipeptídica em resposta a um códon seletor.

10 Por exemplo, um tRNA supressor pode ler através de um códon, incluindo, mas não limitado a um códon de parada, um códon de quatro bases ou um códon raro.

O termo “atividade de supressão” refere-se à capacidade de um tRNA, por exemplo, um tRNA supressor, ler através de um códon seletor. A atividade pode ser expressada como uma porcentagem de atividade observada em comparação a um controle (por exemplo, falta de uma sintetase cognata).

15

O termo “sistema de tradução” refere-se aos componentes necessários para incorporar um aminoácido que ocorre naturalmente em uma cadeia polipeptídica crescente (proteína). Os componentes de um sistema de tradução podem incluir, por exemplo, ribossomos, tRNAs, sintetases, mRNA e semelhantes. Os componentes da presente invenção podem ser adicionados a um sistema de tradução *in vitro* ou *in vivo*. Exemplos dos sistemas de

20 tradução incluem, mas não são limitados a uma célula não eucariótica, por exemplo, uma bactéria (tal como *E. coli*), uma célula eucariótica, por exemplo, uma célula de levedura, célula de mamífero, célula de planta, célula de alga, célula de fungo, célula de inseto, um sistema de tradução livre de célula, por exemplo, um lisato celular e/ou semelhantes.

Sistemas de tradução podem ser celulares ou livres de célula e podem ser procarióticos ou eucarióticos. Sistemas de tradução celulares incluem, mas não são limitados a preparações de células totais, tais como células ou culturas celulares permeabilizadas em que uma sequência de ácido nucleico desejada pode ser transcrita em mRNA e o mRNA traduzido. Sistemas de tradução livres de célula são comercialmente disponíveis e tipos e sistemas diferentes são bem conhecidos. Exemplos de sistemas livres de célula incluem, mas

30 não são limitados a lisatos procarióticos, tais como lisatos de *Escherichia coli*, e lisatos eucarióticos, tais como extratos de germe de trigo, lisatos de células de inseto, lisatos de reticulócitos de coelho, lisatos de oócitos de coelho e lisatos de células humanas. Extratos ou lisatos eucarióticos podem ser preferidos quando a proteína resultante é glicosilada, fosforilada ou, de outro modo, modificada, pois tais modificações são apenas possíveis em sistemas eucarióticos. Alguns destes extratos e lisatos são comercialmente disponíveis (Promega; Madison, Wis.; Stratagene; La Jolla, Calif.; Amersham; Arlington Heights, Ill.;

35

GIBCO/BRL; Grand Island, N.Y.). Extratos membranosos, tais como os extratos pancreáticos caninos contendo membranas microssômicas, também são disponíveis, os quais são úteis para traduzir proteínas secretórias.

Sistemas de tradução reconstituídos também podem ser usados. Misturas de fatores de tradução purificados também foram usadas com êxito para traduzir mRNA em proteína, assim como combinações de lisatos ou lisatos suplementados com fatores de tradução purificados, tais como fator 1 de iniciação (IF-1), IF-2, IF-3 (α ou β), fator de alongação T (EF-Tu) ou fatores de terminação. Sistemas livres de célula também podem ser sistemas de transcrição/tradução acoplados em que o DNA é introduzido ao sistema, transcrito em mRNA e o mRNA traduzido, conforme descrito em *Current Protocols in Molecular Biology* (F. M. Ausubel *et al.* editores, Wiley Interscience, 1993), que é especificamente incorporado como referência por meio deste relatório. O RNA transcrito no sistema de transcrição eucariótico pode estar na forma de RNA heteronuclear (hnRNA) ou *caps* de extremidade 5' (7-metil guanosina) e mRNA maduro com cauda poli A de extremidade 3', que pode ser uma vantagem em certos sistemas de tradução. Por exemplo, mRNAs capeados são traduzidos com alta eficácia no sistema de lisato de reticulócitos.

O termo “aminoácido selecionado” refere-se a qualquer aminoácido que ocorre naturalmente ou aminoácido não natural desejado. Conforme usado neste relatório, o termo “aminoácido não natural” ou “aminoácido não naturalmente codificado” refere-se a qualquer aminoácido, aminoácido modificado e/ou análogo de aminoácido que não é um dos 20 aminoácidos que ocorrem naturalmente comuns ou selenocisteína ou pirrolisina. Outros termos que podem ser sinonimicamente usados com o termo “aminoácido não naturalmente codificado” e “aminoácido não natural” são “aminoácido não natural,” “aminoácido que não ocorre naturalmente,” e versões variadamente hifenizadas e não hifenizadas dos mesmos. O termo “aminoácido não naturalmente codificado” também inclui, mas não é limitado aos aminoácidos que ocorrem por modificação (por exemplo, modificações pós-tradução) de um aminoácido naturalmente codificado (incluindo, mas não limitado aos 20 aminoácidos comuns ou pirrolisina e selenocisteína), porém não são por si só naturalmente incorporados em uma cadeia polipeptídica crescente pelo complexo de tradução. Exemplos de tais aminoácidos que não ocorrem naturalmente incluem, mas não são limitados a *N*-acetilglucosaminil-L-serina, *N*-acetilglucosaminil-L-treonina e O-fosfotirosina.

Aminoácido não natural: Conforme usado neste relatório, um aminoácido não natural refere-se a qualquer aminoácido, aminoácido modificado ou análogo de aminoácido, exceto selenocisteína e/ou pirrolisina, e os vinte alfa-aminoácidos geneticamente codificados canônicos seguintes: alanina, arginina, asparagina, ácido aspártico, cisteína, glutamina, ácido glutâmico, glicina, histidina, isoleucina, leucina, lisina, metionina, fenilalanina, prolina, serina, treonina, triptofano, tirosina, valina. Em várias modalidades da invenção, um ou mais

aminoácidos não naturais que são incorporados no imunógeno não natural podem ser qualquer aminoácido não natural. Desse modo, será avaliado que a citação de aminoácidos não naturais específicos neste relatório não deve ser necessariamente limitante na invenção. Uma ampla variedade de aminoácidos não naturais foi incorporada em proteínas codifican-

5 do-se os mesmos *in vivo*, por exemplo, usando sistemas de tradução que compreendem elementos ortogonais. Veja, por exemplo, Liu, *et al.* (2007) "Genetic incorporation of unnatural amino acids into proteins in mammalian cells" *Nat Methods* 4:239-244; Wang, *et al.* (2006) "Expanding the genetic code" *Annu Rev Biophys Biomol Struct* 35:225-249; Xie & Schultz (2006) "A chemical toolkit for proteins-an expanded genetic code" *Nat Rev Mol Cell*

10 *Biol* 7:775-782; Wang e Schultz "Expanding the Genetic Code," *Angewandte Chemie Int. Ed*, 44(I):34-66 (2005) e Chin, *et al.* (2003) "An expanded eukariotic genetic code" *Science* 301:964-967 para uma revisão.

Em algumas modalidades da presente invenção, é desejável usar aminoácidos não naturais que não são um dos 20 aminoácidos que ocorrem naturalmente comuns ou os amino-

15 ácido raros que ocorrem naturalmente, por exemplo, selenocisteína ou pirrolisina. Por exemplo, os aminoácidos não naturais, p-nitrofenilalanina, p-sulfotirosina e p-carboxifenilalanina encontram uso em várias modalidades neste relatório. Em algumas modalidades, o aminoácido não natural pode incluir, mas não é limitado a: p-nitrofenilalanina; o-nitrofenilalanina; m-nitrofenilalanina; p-boronilfenilalanina; o-boronilfenilalanina; m-

20 boronilfenilalanina; p-aminofenilalanina; o-aminofenilalanina; m-aminofenilalanina; p-acilfenilalanina; o-acilfenilalanina; m-acilfenilalanina; p-OMe fenilalanina; o-OMe fenilalanina; m-OMe fenilalanina; p-sulfofenilalanina; o-sulfofenilalanina; m-sulfofenilalanina; 5-nitro His; 3-nitro Tyr; 2-nitro Tyr; Leu nitro substituída; His nitro substituída; De nitro substituída; Trp nitro substituída; 2-nitro Trp; 4-nitro Trp; 5-nitro Trp; 6-nitro Trp; 7-nitro Trp; 3-amino tirosina,

25 2-aminotirosina, O-sulfotirosina, 2-sulfo-oxifenilalanina, 3-sulfo-oxioxifenilalanina ou p-carboxifenilalanina, o-carboxifenilalanina e m-carboxifenilalanina. Novamente, será avaliado que a invenção não é limitada a aminoácidos não naturais particulares.

Além disso, em várias modalidades da presente invenção, aminoácidos não naturais podem ser incorporados em imunógenos *in vitro*, por exemplo, usando métodos biossintéticos em que um tRNA supressor é quimicamente acilado com um aminoácido não natural

30 desejado e é adicionado a um extrato *in vitro* capaz de suportar a biossíntese do imunógeno. Para uma descrição de tais métodos sintéticos *in vitro*, veja, por exemplo, V. W. Cornish, D. Mendel e P. G. Schultz, *Angew. Chem. Int. Ed. Engl.* 34:621 (1995); C.J. Noren, S.J. Anthony-Cahill, M.C. Griffith, P. G. Schultz, "A general method for site-specific incorporation of unnatural amino acids into proteins", *Science* 244 182-188 (1989); e, J.D. Bain, C.G. Glabe, T.A. Dix, A.R. Chamberlin, E.S. Diala, "Biosynthetic site-specific incorporation of a

35 unnatural amino acid into a polypeptide," *J. Am. Chem. Soc.* 111 8013-8014 (1989). Os ami-

noácidos não naturais também podem ser adicionados a proteínas natural ou sinteticamente produzidas através de químicas de peptídeo sintético disponíveis (ou aminoácidos naturais podem ser convertidos em aminoácidos não naturais através de tais métodos), ou através de processamento pós-tradução. Novamente, entretanto, será avaliado que tais modificações pós-tradução e químicas são tipicamente feitas em combinação com, ou além da incorporação de um ou mais aminoácidos não naturais durante a síntese de uma molécula (por exemplo, incorporação direta, tal como tradução ortogonal, síntese em fase sólida, etc.). Desse modo, modificações por adição pós-tradução ou química de aminoácidos são tipicamente feitas apenas em moléculas já tendo aminoácidos não naturais que foram adicionados durante a síntese da molécula. Informação adicional sobre a incorporação não ortogonal de aminoácidos não naturais em imunógenos é fornecida abaixo.

Conforme usado neste relatório, o termo “derivado de” refere-se a um componente que é isolado ou fabricado usando informação a partir de uma molécula ou organismo específico.

Uma “célula hospedeira recombinante” ou “célula hospedeira” refere-se a uma célula que inclui um polinucleotídeo exógeno, independentemente do método usado para inserção, por exemplo, absorção direta, transdução, *f-mating* ou outros métodos conhecidos na técnica para criar células hospedeiras recombinantes. O polinucleotídeo exógeno pode ser mantido como um vetor não integrado, por exemplo, um plasmídeo, ou, alternativamente, pode ser integrado no genoma hospedeiro.

Conforme usado neste relatório, o termo “meio” ou “meios” inclui qualquer meio de cultura, solução, suporte sólido, semissólido ou rígido que pode suportar ou conter qualquer célula hospedeira, incluindo células hospedeiras bacterianas, células hospedeiras de levedura, células hospedeiras de inseto, células hospedeiras de planta, células hospedeiras eucarióticas, células hospedeiras de mamífero, células de CHO, células hospedeiras procarióticas, células hospedeiras de *E. coli* ou *Pseudomonas* e teores celulares. Desse modo, o termo pode abranger o meio em que a célula hospedeira foi cultivada, por exemplo, o meio em que uma cultura, organismo total ou célula está crescendo, e o meio pode ter um aminoácido não natural incluído para suportar o crescimento de organismos ou células deficientes de replicação, incluindo o meio antes ou depois de uma etapa de proliferação. O termo também pode abranger tampões ou reagentes que contêm lisatos de células hospedeiras, tais como no caso onde os antígenos são intracelularmente produzidos e as células hospedeiras são lisadas ou rompidas para liberar os antígenos ou antígenos recombinantes.

“Agente redutor,” conforme usado neste relatório com respeito à redobra de proteína, é definido como qualquer composto ou material que mantém grupos sulfidríla no estado reduzido e reduz ligações dissulfeto intra ou intermoleculares. Agentes redutores adequados incluem, mas não são limitados a ditioneitol (DTT), 2-mercaptoetanol, ditioeritritol, cisteína,

cisteamina (2-aminoetanotiol) e glutathione reduzida. É evidente àqueles de habilidade comum na técnica que uma ampla variedade de agentes redutores é adequada para o uso nos métodos e composições da presente invenção.

5 “Agente oxidante,” conforme usado neste relatório com respeito à redobra de proteína, é definido como qualquer composto ou material que é capaz de remover um elétron a partir de um composto sendo oxidado. Agentes oxidantes adequados incluem, mas não são limitados a glutathione oxidada, cistina, cisteína, ditioneitol oxidado, eritritol oxidado e oxigênio. É evidente àqueles de habilidade comum na técnica que uma ampla variedade de agentes oxidantes é adequada para o uso nos métodos da presente invenção.

10 “Redobra,” conforme usado neste relatório, descreve qualquer processo, reação ou método que transforma polipeptídeos contendo ligação dissulfeto a partir de um estado imprópriamente dobrado ou não dobrado em uma conformação nativa ou propriamente dobrada com respeito às ligações dissulfeto.

15 “Codobra,” conforme usado neste relatório, refere-se especificamente aos processos, reações ou métodos de redobra que utilizam pelo menos dois polipeptídeos que interagem entre si e resultam na transformação de polipeptídeos não dobrados ou imprópriamente dobrados em polipeptídeos nativos e propriamente dobrados.

20 Um “grupo de modificação no terminal amino” refere-se a qualquer molécula que pode ser ligada ao terminal amino de um polipeptídeo. Similarmente, um “grupo de modificação no terminal carboxi” refere-se a qualquer molécula que pode ser ligada ao terminal carboxi de um polipeptídeo. Grupos de modificação no terminal incluem, mas não são limitados a vários polímeros solúveis em água, peptídeos ou proteínas, tais como albumina sérica, ou outras porções que aumentam a meia-vida no soro de peptídeos.

25 Os termos “grupo funcional”, “porção ativa”, “grupo de ativação”, “grupo de partida”, “sítio reativo”, “grupo quimicamente reativo” e “porção quimicamente reativa” são usados na técnica e neste relatório e referem-se a porções ou unidades distintas e definíveis de uma molécula. Os termos são, de certa forma, sinônimos na técnica da química e são usados neste relatório para indicar as porções de moléculas que realizam alguma função ou atividade e são reativas com outras moléculas.

30 O termo “ligação” ou “ligador” é usado neste relatório para se referir a grupos ou ligações que normalmente são formados como o resultado de uma reação química e tipicamente são ligações covalentes. Ligações hidroliticamente estáveis significam que as ligações são substancialmente estáveis em água e não reagem com água em valores de pH úteis sob condições fisiológicas durante um período de tempo prolongado, talvez até por
35 tempo indeterminado. Ligações hidroliticamente instáveis ou degradáveis significam que as ligações são degradáveis em água ou em soluções aquosas, incluindo, por exemplo, sangue. Ligações enzimaticamente instáveis ou degradáveis significam que a ligação pode ser

degradada por uma ou mais enzimas. Conforme entendido na técnica, PEG e polímeros relacionados podem incluir ligações degradáveis na cadeia principal do polímero ou no grupo ligador entre a cadeia principal do polímero e um ou mais entre os grupos funcionais terminais da molécula de polímero. Por exemplo, ligações éster formadas pela reação de ácidos carboxílicos de PEG ou ácidos carboxílicos de PEG ativados com grupos álcool em um agente biologicamente ativo geralmente hidrolisam sob condições fisiológicas para liberar o agente. Outras ligações hidroliticamente degradáveis incluem, mas não são limitadas às ligações carbonato; ligações imina resultantes da reação de uma amina e um aldeído; ligações fosfato éster formadas através da reação de um álcool com um grupo fosfato; ligações hidrazona que são produto de reação de uma hidrazida e um aldeído; ligações acetal que são o produto de reação de um aldeído e um álcool; ligações ortoéster que são o produto de reação de um formiato e um álcool; ligações peptídicas formadas por um grupo amina em uma terminação de um polímero, tal como PEG, e um grupo carboxila de um peptídeo; e ligações oligonucleotídicas formadas por um grupo fosforamidita na terminação de um polímero e um grupo 5' hidroxila de um oligonucleotídeo.

O termo “molécula biologicamente ativa”, “porção biologicamente ativa” ou “agente biologicamente ativo”, quando usado neste relatório, significa qualquer substância que pode afetar quaisquer propriedades físicas ou bioquímicas de um sistema biológico, via, molécula ou interação referindo-se a um organismo, incluindo, mas não limitado a vírus, bactérias, bacteriófago, transposon, prion, insetos, fungos, plantas, animais e seres humanos. Em particular, conforme usado neste relatório, moléculas biologicamente ativas incluem, mas não são limitadas a qualquer substância intencionada para diagnose, cura, mitigação, tratamento ou prevenção de doença em seres humanos ou outros animais, ou para, de outro modo, acentuar o bem-estar físico ou mental de seres humanos ou animais. Exemplos de moléculas biologicamente ativas incluem, mas não são limitados a peptídeos, proteínas, enzimas, fármacos de molécula pequena, vacinas, imunógenos, fármacos duros, fármacos moles, carboidratos, átomos ou moléculas inorgânicos, pigmentos, lipídeos, nucleosídeos, radionuclídeos, oligonucleotídeos, toxoides, toxinas, células procarióticas e eucarióticas, vírus, polissacarídeos, ácidos nucleicos e porções dos mesmos obtidas ou derivadas de vírus, bactérias, insetos, animais ou qualquer outra célula ou tipo celular, lipossomas, micropartículas e micelas. Classes de agentes biologicamente ativos que são adequadas para o uso com a invenção incluem, mas não são limitadas a fármacos, pró-fármacos, radionuclídeos, agentes de imagem, polímeros, antibióticos, fungicidas, agentes antivirais, agentes anti-inflamatórios, agentes antitumor, agentes cardiovasculares, agentes anti-ansiedade, hormônios, fatores de crescimento, agentes esteroidais, toxinas derivadas de micro-organismos e semelhantes.

O termo “arila” significa, a menos que de outro modo estabelecido, um substituinte de hidrocarboneto poli-insaturado e aromático, que pode ser um anel único ou anéis múlti-

plos (de 1 a 3 anéis) que são fundidos ou covalentemente ligados. O termo “heteroarila” refere-se a grupos arila (ou anéis) que contêm de um a quatro heteroátomos selecionados de N, O e S, em que os átomos de nitrogênio e enxofre são opcionalmente oxidados e os átomos de nitrogênio são opcionalmente quaternizados. Um grupo heteroarila pode ser ligado ao restante da molécula através de um heteroátomo. Exemplos não limitantes de grupos arila e heteroarila incluem fenila, 1-naftila, 2-naftila, 4-bifenila, 1-pirrolila, 2-pirrolila, 3-pirrolila, 3-pirazolila, 2-imidazolila, 4-imidazolila, pirazinila, 2-oxazolila, 4-oxazolila, 2-fenil-4-oxazolila, 5-oxazolila, 3-isoxazolila, 4-isoxazolila, 5-isoxazolila, 2-tiazolila, 4-tiazolila, 5-tiazolila, 2-furila, 3-furila, 2-tienila, 3-tienila, 2-piridila, 3-piridila, 4-piridila, 2-pirimidila, 4-pirimidila, 5-benzotiazolila, purinila, 2-benzimidazolila, 5-indolila, 1-isoquinolila, 5-isoquinolila, 2-quinoxalinila, 5-quinoxalinila, 3-quinolila e 6-quinolila. Substituintes para os sistemas de anel arila e heteroarila observados acima são selecionados do grupo de substituintes aceitáveis descritos abaixo.

Para brevidade, o termo “arila”, quando usado em combinação com outros termos (incluindo, mas não limitados a arilóxi, ariltióxi, arilalquila), inclui anéis arila e heteroarila, conforme definido acima. Desse modo, o termo “arilalquila” inclui aqueles radicais em que um grupo arila é ligado a um grupo alquila (incluindo, mas não limitado a benzila, fenetila, piridilmetila e semelhantes), incluindo aqueles grupos alquila em que um átomo de carbono (incluindo, mas não limitado a um grupo metileno) foi substituído, por exemplo, por um átomo de oxigênio (incluindo, mas não limitado a fenoximetila, 2-piridiloximetila, 3-(1-naftilóxi)propila e semelhantes).

Cada um dos termos acima (incluindo, mas não limitados a “alquila,” “heteroalquila,” “arila” e “heteroarila”) inclui formas substituídas e não substituídas do radical indicado. Substituintes exemplares para cada tipo de radical são fornecidos abaixo.

Substituintes para os radicais alquila e heteroalquila (incluindo aqueles grupos frequentemente referidos como alquilenos, alquenila, heteroalquilenos, heteroalquenila, alquinila, cicloalquila, heterocicloalquila, cicloalquenila e heterocicloalquenila) podem ser um ou mais entre uma variedade de grupos selecionados de, mas não limitados a: $-OR'$, $=O$, $=NR'$, $=N-OR'$, $-NR'R''$, $-SR'$, $-halogênio$, $-SiR'R''R'''$, $OC(O)R'$, $-C(O)R'$, $-CO_2R'$, $-CONR'R''$, $-OC(O)NR'R''$, $-NR''C(O)R'$, $NR'C(O)NR''R'''$, $-NR''C(O)_2R'$, $-NR-C(NR'R''R''')=NR''''$, $NR-C(NR'R'')=NR''''$, $-S(O)R'$, $-S(O)_2R'$, $-S(O)_2NR'R''$, $NRSO_2R'$, $-CN$ e $-NO_2$ em um número variando de zero a $(2m' + 1)$, onde m' é o número total de átomos de carbono em tal radical. R' , R'' , R''' e R'''' independentemente referem-se a hidrogênio, heteroalquila substituído ou não substituído, arila substituído ou não substituído, incluindo mas não limitado a arila substituído com 1 a 3 halogênios, alquila substituído ou não substituído, grupos alcóxi ou tioalcóxi ou grupos arilalquila. Quando um composto da invenção inclui mais do que um grupo R , por exemplo, cada um dos grupos R é independentemente selecionado como são grupos R' .

R'', R''' e R'''' quando mais do que um destes grupos está presente. Quando R' e R'' são ligados ao mesmo átomo de nitrogênio, eles podem ser combinados com o átomo de nitrogênio para formar um anel de 5, 6 ou 7 membros. Por exemplo, -NR'R'' inclui, mas não é limitado a 1-pirrolidinila e 4-morfolinila. A partir do debate acima de substituintes, uma pessoa habilitada na técnica entenderá que o termo "alquila" inclui grupos, incluindo átomos de carbono ligados a grupos, exceto grupos hidrogênio, tais como haloalquila (incluindo, mas não limitado a -CF₃ e -CH₂CF₃) e acila (incluindo, mas não limitado a -C(O)CH₃, -C(O)CF₃, -C(O)CH₂OCH₃ e semelhantes).

Similar aos substituintes descritos para o radical alquila, substituintes para os grupos arila e heteroarila são variados e são selecionados de, mas não são limitados a: halogênio, OR', =O, =NR', =N-OR', -NR'R'', -SR', -halogênio, -SiR'R''R''', OC(O)R', -C(O)R', CO₂R', -CONR'R'', -OC(O)NR'R'', -NR''C(O)R', NR' C(O)R''R''', -NR''C(O)2R', NR-C(NR'R''R''')=NR''', NR C(NR'R'')=NR''', -S(O)R', -S(O)2R', -S(O)2NR'R'', NRSO₂R', -CN e -NO₂, -R', -N₃, -CH(Ph)₂, fluoro(C1-C4)alcóxi, e fluoro(C1-C4)alquila, em um número variando de zero ao número total de valências abertas no sistema de anel aromático; e onde R', R'', R''' e R'''' são independentemente selecionados de hidrogênio, alquila, heteroalquila, arila e heteroarila. Quando um composto da invenção inclui mais do que um grupo R, por exemplo, cada um dos grupos R é independentemente selecionado como são os grupos R', R'', R''' e R'''' quando mais do que um destes grupos está presente.

O termo "ácido nucleico" refere-se a desoxirribonucleotídeos, desoxirribonucleosídeos, ribonucleosídeos ou ribonucleotídeos e polímeros dos mesmos na forma de fita única ou dupla. A menos que especificamente limitado, o termo abrange ácidos nucleicos contendo análogos conhecidos de nucleotídeos naturais que têm propriedades de ligação similares em relação ao ácido nucleico de referência e são metabolizados em uma maneira similar aos nucleotídeos que ocorrem naturalmente. A menos que especificamente limitado de outro modo, o termo também refere-se a análogos de oligonucleotídeos, incluindo PNA (ácido peptidonucleico), análogos de DNA usados na tecnologia antissentido (fosforotioatos, fosforoamidatos e semelhantes). A menos que de outro modo indicado, uma sequência de ácido nucleico particular também abrange implicitamente variantes conservativamente modificadas da mesma (incluindo, mas não limitada a substituições de códon degenerado) e sequências complementares, assim como a sequência explicitamente indicada. Especificamente, substituições de códon degenerado podem ser obtidas gerando-se sequências em que a terceira posição de um ou mais códons selecionados (ou todos) é substituída com base mista e/ou resíduos de desóxi-inosina (Batzer *et al.*, Nucleic Acid Res. 19:5081 (1991); Ohtsuka *et al.*, J. Biol. Chem. 260:2605-2608 (1985); Rossolini *et al.*, Mol. Cell. Probes 8:91-98 (1994)).

Onde grupos substituintes são especificados por suas fórmulas químicas convencionais, escritas da esquerda para a direita, eles abrangem igualmente os substituintes qui-

micamente idênticos que resultariam da escrita da estrutura da direita para a esquerda, por exemplo, a estrutura CH₂O é equivalente à estrutura -OCH₂.

O termo “substituintes” inclui, mas não é limitado a “substituintes não interferentes”. “Substituintes não interferentes” são aqueles grupos que produzem compostos estáveis.

- 5 Substituintes não interferentes ou radicais adequados incluem, mas não são limitados a halo, alquila C1-C10, alquenila C2-C10, alquinila C2-C10, alcóxi C1-C10, aralquila C1-C12, alcarila C1-C12, cicloalquila C3-C12, cicloalquenila C3-C12, fenila, fenila substituído, toluoila, xilenila, bifenila, alcoxialquila C2-C12, alcoxiarila C2-C12, ariloxialquila C7-C12, oxiarila C7-C12, alquilsulfonila C1-C6, alquilsulfonila C1-C10, -(CH₂)_m -O-(alquila C1-C10), em que
- 10 m é de 1 a 8, arila, arila substituído, alcóxi substituído, fluoroalquila, radical heterocíclico, radical heterocíclico substituído, nitroalquila, -NO₂, -CN, -NRC(O)-(alquila C1-C10), -C(O)- (alquila C1-C10), alquiltioalquila C2-C10, -C(O)O-(alquila C1-C10), -OH, -SO₂, =S, -COOH, -NR₂, carbonila, -C(O)-(alquila C1-C10)-CF₃, -C(O)-CF₃, -C(O)NR₂, -(arila C1-C10)-S-(arila C6-C10), -C(O)-(arila C1-C10), -(CH₂)_m -O-(-(CH₂)_m-O-(alquila C1-C10)), em que cada m é
- 15 de 1 a 8, -C(O)NR₂, -C(S)NR₂, -SO₂NR₂, -NRC(O) NR₂, -NRC(S) NR₂, sais dos mesmos e semelhantes. Cada R, conforme usado neste relatório, é H, alquila ou alquila substituído, arila ou arila substituído, aralquila ou alcarila.

O termo “halogênio” inclui flúor, cloro, iodo e bromo.

- O termo “alquila” por si só ou como parte de um outro substituinte, significa, a me-
- 20 nos que de outro modo estabelecido, uma cadeia reta ou ramificada, ou radical hidrocarboneto cíclico, ou combinação dos mesmos, que pode ser completamente saturado, mono ou poli-insaturado e pode incluir radicais di e multivalentes, tendo o número de átomos de carbono designado (isto é C1-C10 significa um a dez carbonos). Exemplos de radicais hidrocarboneto saturados incluem, mas não são limitados a grupos, tais como metila, etila, n-
- 25 propila, isopropila, n-butila, t-butila, isobutila, sec-butila, ciclo-hexila, (ciclo-hexil)metila, ciclo-propilmetila, homólogos e isômeros de, por exemplo, n-pentila, n-hexila, n-heptila, n-octila e semelhantes. Um grupo alquila insaturado é aquele tendo uma ou mais ligações duplas ou ligações triplas. Exemplos de grupos alquila insaturados incluem, mas não são limitados a
- 30 vinila, 2-propenila, crotila, 2-isopentenila, 2-(butadienila), 2,4-pentadienila, 3-(1,4-pentadienila), etinila, 1- e 3-propinila, 3-butenila e os homólogos e isômeros superiores. O termo “alquila,” a menos que de outro modo observado, também inclui aqueles derivados de alquila definidos em mais detalhes abaixo, tais como “heteroalquila,” Grupos alquila que são limitados a grupos hidrocarboneto são denominados “homoalquila”.

- O termo “alquilenos” por si só ou como parte de um outro substituinte significa um
- 35 radical divalente derivado de um alcano, conforme exemplificado, mas não limitado pelas estruturas -CH₂CH₂- e -CH₂CH₂CH₂CH₂-, e ainda inclui aqueles grupos descritos abaixo como “heteroalquilenos.” Tipicamente, um grupo alquila (ou alquilenos) terá de 1 a 24 átomos

de carbono, com esses grupos tendo 10 ou menos átomos de carbono sendo uma modalidade particular dos métodos e composições descritos neste relatório. Um “alquila inferior” ou “alquilen inferior” é um grupo alquila ou alquilen de cadeia mais curta, geralmente tendo oito ou menos átomos de carbono.

5 Os termos “alcóxi,” “alquilamino” e “alquiltio” (ou tioalcóxi) são usados em seu sentido convencional e referem-se àqueles grupos alquila ligados ao restante da molécula por intermédio de um átomo de oxigênio, um grupo amino ou um átomo de enxofre, respectivamente.

O termo “heteroalquila,” por si só ou em combinação com um outro termo, significa, 10 a menos que de outro modo estabelecido, uma cadeia reta ou ramificada estável, ou radical hidrocarboneto cíclico, ou combinações dos mesmos, consistindo do número estabelecido de átomos de carbono e pelo menos um heteroátomo selecionado do grupo que consiste de O, N, Si e S e, em que os átomos de nitrogênio e enxofre podem opcionalmente ser oxidados e o heteroátomo de nitrogênio pode opcionalmente ser quaternizado. Os heteroátomos 15 O, N e S e Si podem ser colocados em qualquer posição interna do grupo heteroalquila ou na posição em que o grupo alquila é ligado ao restante da molécula. Exemplos incluem, mas não são limitados a, $-\text{CH}_2-\text{CH}_2-\text{O}-\text{CH}_3$, $-\text{CH}_2-\text{CH}_2-\text{NH}-\text{CH}_3$, $-\text{CH}_2-\text{CH}_2-\text{N}(\text{CH}_3)-\text{CH}_3$, $-\text{CH}_2-\text{S}-\text{CH}_2-\text{CH}_3$, $-\text{CH}_2-\text{CH}_2-$, $-\text{S}(\text{O})-\text{CH}_3$, $-\text{CH}_2-\text{CH}_2-\text{S}(\text{O})_2-\text{CH}_3$, $-\text{CH}=\text{CH}-\text{O}-\text{CH}_3$, $-\text{Si}(\text{CH}_3)_3$, $-\text{CH}_2-\text{CH}=\text{N}-\text{OCH}_3$ e $-\text{CH}=\text{CH}-\text{N}(\text{CH}_3)-\text{CH}_3$. Até dois heteroátomos podem ser consecutivos, 20 tais como, por exemplo, $-\text{CH}_2-\text{NH}-\text{OCH}_3$ e $-\text{CH}_2-\text{O}-\text{Si}(\text{CH}_3)_3$. Similarmente, o termo “heteroalquilen” por si só ou como parte de um outro substituinte significa um radical divalente derivado de heteroalquila, conforme exemplificado, mas não limitado por $-\text{CH}_2-\text{CH}_2-\text{S}-\text{CH}_2-\text{CH}_2-$ e $-\text{CH}_2-\text{S}-\text{CH}_2-\text{CH}_2-\text{NH}-\text{CH}_2-$. Para grupos heteroalquilen, os mesmos heteroátomos ou heteroátomos diferentes também podem ocupar um ou ambos terminais de cadeia (incluindo, mas não limitados a alquilen-óxi, alquilenodióxi, alquilenamino, alquilenodiamino, 25 amino-oxialquilen e semelhantes). Além disso, para grupos de ligação alquilen e heteroalquilen, nenhuma orientação do grupo de ligação é indicada pela direção em que a fórmula do grupo de ligação é escrita. Por exemplo, a fórmula $-\text{C}(\text{O})_2\text{R}'$ representa tanto $-\text{C}(\text{O})_2\text{R}'$ quanto $-\text{R}'\text{C}(\text{O})_2$.

30 Os termos “cicloalquila” e “heterocicloalquila”, por si só ou em combinação com outros termos, representam, a menos que de outro modo estabelecido, versões cíclicas de “alquila” e “heteroalquila”, respectivamente. Desse modo, um cicloalquila ou heterocicloalquila inclui ligações de anel saturadas, parcialmente insaturadas e completamente insaturadas. Adicionalmente, para o heterocicloalquila, um heteroátomo pode ocupar a posição em que o 35 heterociclo é ligado ao restante da molécula. Exemplos de cicloalquila incluem, mas não são limitados a ciclopentila, ciclo-hexila, 1-ciclo-hexenila, 3-ciclo-hexenila, ciclo-heptila e semelhantes. Exemplos de heterocicloalquila incluem, mas não são limitados a 1-(1,2,5,6-tetra-

hidropiridila), 1-piperidinila, 2-piperidinila, 3-piperidinila, 4-morfolinila, 3-morfolinila, tetraidro-
furan-2-ila, tetraidrofuran-3-ila, tetraidrotien-2-ila, tetraidrotien-3-ila, 1-piperazinila, 2-
piperazinila e semelhantes. Adicionalmente, o termo abrange estruturas de anel bicíclicas e
tricíclicas. Similarmente, o termo “heterocicloalquileno” por si só ou como parte de um outro
5 substituinte significa um radical divalente derivado de heterocicloalquila, e o termo “cicloal-
quileno” por si só ou como parte de um outro substituinte significa um radical divalente deri-
vado de cicloalquila.

O termo “aminoácido” refere-se a aminoácidos que ocorrem naturalmente e que
não ocorrem naturalmente, assim como análogos de aminoácidos e miméticos de aminoáci-
10 dos que funcionam em uma maneira similar aos aminoácidos que ocorrem naturalmente.
Aminoácidos naturalmente codificados são os 20 aminoácidos comuns (alanina, arginina,
asparagina, ácido aspártico, cisteína, glutamina, ácido glutâmico, glicina, histidina, isoleuci-
na, leucina, lisina, metionina, fenilalanina, prolina, serina, treonina, triptofano, tirosina e vali-
na) e pirrolisina e selenocisteína. Análogos de aminoácidos referem-se a compostos que
15 têm a mesma estrutura química básica em relação a um aminoácido que ocorre naturalmen-
te, isto é, um α carbono que é ligado a um hidrogênio, um grupo carboxila, um grupo amino
e um grupo R, tal como homoserina, norleucina, sulfóxido de metionina, metionina metil
sulfônio. Tais análogos têm grupos R modificados (tais como, norleucina) ou cadeias princi-
pais de peptídeo modificadas, porém retêm a mesma estrutura química básica em relação a
20 um aminoácido que ocorre naturalmente. Referência a um aminoácido inclui, por exemplo,
L-aminoácidos proteogênicos que ocorrem naturalmente; D-aminoácidos, aminoácidos qui-
micamente modificados, tais como variantes e derivados de aminoácidos; aminoácidos não
proteogênicos que ocorrem naturalmente, tais como α -alanina, ornitina, etc.; e compostos
quimicamente sintetizados tendo propriedades conhecidas na técnica como características
25 de aminoácidos. Exemplos de aminoácidos que não ocorrem naturalmente incluem, mas
não são limitados a α -metil aminoácidos (por exemplo, α -metil alanina), D-aminoácidos, a-
minoácidos do tipo histidina (por exemplo, 2-amino-histidina, α -hidróxi-histidina, homo-
histidina, α -fluorometil-histidina e α -metil-histidina), aminoácidos tendo um metileno extra na
cadeia lateral (“homo” aminoácidos), e aminoácidos em que um grupo funcional de ácido
30 carboxílico na cadeia lateral é substituído com um grupo ácido sulfônico (por exemplo, ácido
cisteico).

Aminoácidos podem ser referidos neste relatório por seus símbolos de três letras
comumente conhecidos ou pelos símbolos de uma letra recomendados pela IUPAC-IUB
Biochemical Nomenclature Commission. Nucleotídeos, do mesmo modo, podem ser referi-
35 dos por seus códigos de letra única comumente aceitos.

“Variantes conservativamente modificadas” aplicam-se a sequências de aminoáci-
dos e ácidos nucleicos. Com respeito a sequências de ácidos nucleicos particulares, “varian-

tes conservativamente modificadas” referem-se àqueles ácidos nucleicos que codificam sequências de aminoácidos idênticas ou essencialmente idênticas, ou onde o ácido nucleico não codifica uma sequência de aminoácido, em sequências essencialmente idênticas. Por causa da degeneração do código genético, um grande número de ácidos nucleicos funcionalmente idênticos codifica qualquer proteína fornecida. Por exemplo, os códons GCA, GCC, GCG e GCU codificam o aminoácido alanina. Desse modo, em cada posição onde uma alanina é especificada por um códon, o códon pode ser alterado para qualquer um dos códons correspondentes descritos sem alterar o polipeptídeo codificado. Tais variações de ácido nucleico são “variações silenciosas,” que são uma espécie de variação conservativamente modificada. Cada sequência de ácido nucleico neste relatório que codifica um polipeptídeo também descreve cada variação silenciosa possível do ácido nucleico. Uma pessoa de habilidade comum na técnica reconhecerá que cada códon em um ácido nucleico (exceto AUG, que é comumente o único códon para metionina, e TGG, que é comumente o único códon para triptofano) pode ser modificado para produzir uma molécula funcionalmente idêntica. Consequentemente, cada variação silenciosa de um ácido nucleico que codifica um polipeptídeo está implícita em cada sequência descrita.

Quanto às sequências de aminoácidos, uma pessoa de habilidade comum na técnica reconhecerá que substituições, deleções ou adições individuais a uma sequência de ácido nucleico, peptídeo, polipeptídeo ou proteína que altera, adiciona ou deleta um aminoácido único ou uma pequena porcentagem de aminoácidos na sequência codificada é uma “variante conservativamente modificada” onde a alteração resulta na deleção de um aminoácido, adição de um aminoácido ou substituição de um aminoácido com um aminoácido quimicamente similar. Tabelas de substituição conservativa fornecendo aminoácidos funcionalmente similares são conhecidas àqueles de habilidade comum na técnica. Tais variantes conservativamente modificadas são além de e não excluem variantes polimórficas, homólogos interespécies e alelos da invenção.

Tabelas de substituição conservativa fornecendo aminoácidos funcionalmente similares são conhecidas àqueles de habilidade comum na técnica. Os oito grupos seguintes contêm aminoácidos que são substituições conservativas entre si:

- 1) Alanina (A), Glicina (G);
- 2) Ácido aspártico (D), Ácido glutâmico (E);
- 3) Asparagina (N), Glutamina (Q);
- 4) Arginina (R), Lisina (K);
- 5) Isoleucina (I), Leucina (L), Metionina (M), Valina (V);
- 6) Fenilalanina (F), Tirosina (Y), Triptofano (W);
- 7) Serina (S), Treonina (T); e
- 8) Cisteína (C), Metionina (M)

(veja, por exemplo, Creighton, *Proteins: Structures and Molecular Properties* (W H Freeman & Co.; 2ª edição (Dezembro 1993)

Os termos “idêntica” ou “identidade de porcentagem”, no contexto de duas ou mais sequências de ácidos nucleicos ou polipeptídeos, referem-se a duas ou mais sequências ou subsequências que são as mesmas. Sequências são “substancialmente idênticas” se elas têm uma porcentagem de resíduos de aminoácidos ou nucleotídeos que é a mesma (isto é, cerca de 60 % de identidade, cerca de 65 %, cerca de 70 %, cerca de 75 %, cerca de 80 %, cerca de 85 %, cerca de 90 % ou cerca de 95 % de identidade sobre uma região específica), quando comparada e alinhada para máxima correspondência sobre uma janela de comparação, ou região designada, conforme medido, usando um dos algoritmos de comparação de sequência seguintes (ou outros algoritmos disponíveis a pessoas de habilidade comum na técnica) ou por alinhamento manual e inspeção visual. Esta definição também refere-se ao complemento de uma sequência de teste. A identidade pode existir sobre uma região que apresenta pelo menos cerca de 50 aminoácidos ou nucleotídeos no comprimento, ou sobre uma região que apresenta 75 a 100 aminoácidos ou nucleotídeos no comprimento, ou, onde não especificado, através da sequência total de um polinucleotídeo ou polipeptídeo. Um polinucleotídeo que codifica um polipeptídeo da presente invenção, incluindo homólogos da espécie, exceto humana, pode ser obtido por um processo compreendendo as etapas de triar um biblioteca sob condições de hibridização severas com uma sonda marcada tendo uma sequência de polinucleotídeos da invenção ou um fragmento da mesma, e isolar cDNA de comprimento total e clones genômicos contendo a dita sequência de polinucleotídeos. Tais técnicas de hibridização são bem conhecidas ao técnico habilitado.

Para comparação de sequência, tipicamente, uma sequência atua como uma sequência de referência, a qual as sequências de teste são comparadas. Quando do uso de um algoritmo de comparação de sequência, sequências de teste e de referência são introduzidas em um computador, coordenadas de subsequência são designadas, se necessário, e parâmetros de programa de algoritmo de sequência são designados. Parâmetros de programa pré-determinados podem ser usados, ou parâmetros alternativos podem ser designados. O algoritmo de comparação de sequência depois calcula a porcentagem de identidades de sequência para as sequências de teste em relação à sequência de referência, com base nos parâmetros do programa.

Uma “janela de comparação”, conforme usado neste relatório, inclui referência a um segmento de qualquer um dos números de posições contíguas selecionadas do grupo que consiste de 20 a 600, usualmente cerca de 50 a cerca de 200, mais usualmente cerca de 100 a cerca de 150, em que uma sequência pode ser comparada a uma sequência de referência do mesmo número de posições contíguas depois que as duas sequências são idealmente alinhadas. Métodos de alinhamento de sequências para comparação são conhecidos

àqueles de habilidade comum na técnica. O alinhamento ideal de sequências para comparação pode ser conduzido pelo algoritmo de homologia local de Smith e Waterman (1970) *Adv. Appl. Math.* 2:482c, pelo algoritmo de alinhamento de homologia de Needleman e Wunsch (1970) *J. Mol. Biol.* 48:443, pela pesquisa para o método de similaridade de Pearson e Lipman (1988) *Proc. Nat'l. Acad. Sci. USA* 85:2444, por implementações computadorizadas destes algoritmos (GAP, BESTFIT, FASTA e TFASTA no Wisconsin Genetics Software Package, Genetics Computer Group, 575 Science Dr., Madison, WI), ou por alinhamento manual e inspeção visual (veja, por exemplo, Ausubel *et al.*, *Current Protocols in Molecular Biology* (1995 suplemento)).

Um exemplo de um algoritmo que é adequado para determinar a porcentagem de identidade de sequência e similaridade de sequência são os algoritmos BLAST e BLAST 2.0, que são descritos em Altschul *et al.* (1997) *Nuc. Acids Res.* 25:3389-3402, e Altschul *et al.* (1990) *J. Mol. Biol.* 215:403-410, respectivamente. O software para realizar análises BLAST está publicamente disponível através do National Center for Biotechnology Information disponível na World Wide Web em ncbi.nlm.nih.gov. Os parâmetros do algoritmo BLAST W, T e X determinam a sensibilidade e velocidade do alinhamento. O programa BLASTN (para sequências de nucleotídeos) usa como padrões um comprimento de palavra (W) de 11, uma expectativa (E) de 10, M = S, N = -4 e uma comparação de ambas as fitas. Para sequências de aminoácidos, o programa BLASTP usa como padrões um comprimento de palavra de 3, e expectativa (E) de 10, e os alinhamentos (B) de matriz de pontuação BLOSUM62 (veja Henikoff e Henikoff (1992) *Proc. Natl. Acad. Sci. USA* 89: 10915) de 50, expectativa (E) de 10, M = 5, N = -4, e uma comparação de ambas as fitas. O algoritmo BLAST é tipicamente realizado com o filtro de “baixa complexidade” desativado.

O algoritmo BLAST também realiza uma análise estatística da similaridade entre duas sequências (veja, por exemplo, Karlin e Altschul (1993) *Proc. Natl. Acad. Sci. USA* 90:5873-5787). Uma medição de similaridade fornecida pelo algoritmo BLAST é a menor probabilidade de soma (P(N)), que fornece uma indicação da probabilidade pela qual uma comparação entre duas sequências de nucleotídeos ou aminoácidos ocorreria por acaso. Por exemplo, um ácido nucleico é considerado similar a uma sequência de referência se a menor probabilidade de soma em uma comparação do ácido nucleico de teste ao ácido nucleico de referência é menor do que cerca de 0,2, ou menor do que cerca de 0,01, ou menor do que cerca de 0,001.

A frase “seletivamente (ou especificamente) hibridiza” refere-se à ligação, duplexação ou hibridização de uma molécula apenas em uma sequência de nucleotídeos particular sob condições de hibridização severas, quando tal sequência está presente em uma mistura complexa (incluindo, mas não limitada ao DNA ou RNA celular total ou biblioteca).

A frase “condições de hibridização severas” refere-se à hibridização de sequências

de DNA, RNA, PNA, ou outras imitações de ácido nucleico, ou combinações das mesmas, sob condições de baixa força iônica e alta temperatura, conforme é conhecido na técnica. Tipicamente, sob condições severas, uma sonda hibridizará sua subsequência alvo em uma mistura complexa de ácido nucleico (incluindo, mas não limitada a DNA ou RNA celular total ou biblioteca), porém não hibridiza outras sequências na mistura complexa. Condições severas são dependentes da sequência e serão diferentes em circunstâncias diferentes. Sequências mais longas hibridizam especificamente em temperaturas mais altas. Um guia extensivo para a hibridização de ácidos nucleicos é encontrado em Tijssen, Laboratory Techniques in Biochemistry and Molecular Biology - Hybridization with Nucleic Probes, "Overview of principles of hybridization and the strategy of nucleic acid assays" (1993). Geralmente, condições severas são selecionadas em aproximadamente 5 a 10 °C inferiores ao ponto de fusão térmico (T_m) para a sequência específica em um pH de força iônica definida. O T_m é a temperatura (sob força iônica, pH e concentração nucleica definidos) na qual 50 % das sondas complementares ao alvo hibridizam à sequência alvo em equilíbrio (como as sequências alvo estão presentes em excesso, no T_m, 50 % das sondas são ocupadas em equilíbrio). Condições severas podem ser aquelas em que a concentração de sal é menor do que cerca de 1,0 M de íon sódio, tipicamente cerca de 0,01 a 1,0 M de concentração de íon sódio (ou outros sais) ao pH 7,0 a 8,3 e a temperatura é pelo menos cerca de 30 °C para sondas curtas (incluindo, mas não limitadas entre 10 a 50 nucleotídeos) e pelo menos cerca de 60 °C para sondas longas (incluindo, mas não limitadas a mais do que 50 nucleotídeos). Condições severas também podem ser obtidas com a adição de agentes desestabilizantes, tais como formamida. Para a hibridização seletiva ou específica, um sinal positivo pode ser pelo menos duas vezes o ruído, opcionalmente 10 vezes a hibridização de ruído. Condições de hibridização severas exemplares podem ser as seguintes: 50 % de formamida, SSC 5X e 1 % de SDS, incubação a 42 °C ou SSC 5X, 1 % de SDS, incubação a 65 °C, com lavagem em SSC 0,2X e 0,1 % de SDS a 65 °C. Tais lavagens podem ser realizadas durante 5, 15, 30, 60, 120 minutos ou mais.

Conforme usado neste relatório, o termo "eucarionte" refere-se a organismos pertencentes ao domínio filogenético *Eucarya*, tais como animais (incluindo, mas não limitados a mamíferos, insetos, répteis, pássaros, etc.), ciliados, plantas (incluindo, mas não limitadas a monocotiledôneas, dicotiledôneas, algas, etc.), fungos, leveduras, flagelados, microsporídios, protistas, etc.

Conforme usado neste relatório, o termo "não eucarionte" refere-se a organismos não eucarióticos. Por exemplo, um organismo não eucariótico pode pertencer ao domínio filogenético *Eubacteria* (incluindo, mas não limitado a *Escherichia coli*, *Thermus thermophilus*, *Bacillus stearothermophilus*, *Pseudomonas fluorescens*, *Pseudomonas aeruginosa*, *Pseudomonas putida*, etc.), ou ao domínio filogenético *Archaea* (incluindo, mas não limitado

a *Methanococcus jannaschii*, *Methanobacterium thermoautotrophicum*, *Halobacterium*, tal como *Haloferax volcanii* e *Halobacterium* espécie NRC-1, *Archaeoglobus fulgidus*, *Pyrococcus furiosus*, *Pyrococcus horikoshii*, *Aeuropyrum pernix*, etc.).

5 O termo “paciente”, conforme usado neste relatório, refere-se a um animal, em algumas modalidades, um mamífero, e, em outras modalidades, um ser humano, que é o objeto do tratamento, observação ou experimento. Um animal pode ser um animal de estimação (por exemplo, cães, gatos e semelhantes), animal de criação (por exemplo, vacas, ovelhas, porcos, cavalos e semelhantes) ou um animal de laboratório (por exemplo, ratos, camundongos, porquinhos da índia e semelhantes).

10 O termo “quantidade eficaz”, conforme usado neste relatório, refere-se àquela quantidade do polipeptídeo de aminoácido não natural modificado sendo administrada que, de certa forma, atenuará um ou mais dos sintomas da doença, condição ou distúrbio sendo tratado. Composições contendo o polipeptídeo de aminoácido não natural modificado descrito neste relatório podem ser administradas para tratamentos profiláticos, acentuadores e/ou
15 terapêuticos.

Os termos “acentuar” ou “acentuando” significam aumentar ou prolongar a potência ou duração de um efeito desejado. Desse modo, com respeito a acentuar o efeito de agentes terapêuticos, o termo “acentuando” refere-se à capacidade de aumentar ou prolongar a potência ou duração do efeito de outros agentes terapêuticos em um sistema. Uma “quantidade eficaz acentuadora”, conforme usado neste relatório, refere-se a uma quantidade adequada para acentuar o efeito de um outro agente terapêutico em um sistema desejado. Quando usadas em um paciente, quantidades eficazes para este uso dependerão da severidade e curso da doença, distúrbio ou condição, terapia prévia, estado de saúde do paciente e resposta aos fármacos e julgamento do médico.

25 Conforme usado neste relatório, o termo “seleção positiva” ou “marcador de triagem” refere-se a um marcador que, quando presente, por exemplo, expressado, ativado ou semelhantes, resulta na identificação de uma célula com o marcador de seleção positiva a partir daquela sem o marcador de seleção positiva.

Conforme usado neste relatório, o termo “seleção negativa” ou “marcador de triagem” refere-se a um marcador que, quando presente, por exemplo, expressado, ativado ou semelhantes, permite a identificação de uma célula que não possui a propriedade desejada (por exemplo, em comparação a uma célula que possui a propriedade desejada).

Conforme usado neste relatório, o termo “repórter” refere-se a um componente que pode ser usado para selecionar componentes alvo de um sistema de interesse. Por exemplo, um repórter pode incluir uma proteína, por exemplo, uma enzima, que confere resistência ou sensibilidade ao antibiótico (incluindo, mas não limitada a β -lactamase, cloranfenicol acetiltransferase (CAT) e semelhantes), um marcador de triagem fluorescente (incluindo,
35

mas não limitado à proteína verde fluorescente (por exemplo, GFP), YFP, EGFP, RFP, um marcador luminescente (incluindo, mas não limitado a uma proteína luciferase de vaga-lume), um marcador de triagem com base na afinidade, ou genes marcadores selecionáveis positivos ou negativos, tais como lacZ, β -gal/lacZ (β -galactosidase), ADH (álcool desidrogenase), his3, ura3, leu2, lys2 ou semelhantes.

Conforme usado neste relatório, o termo “eucarionte” refere-se a organismos pertencentes ao domínio filogenético *Eucarya*, tais como animais (incluindo, mas não limitados a mamíferos, insetos, répteis, pássaros, etc.), ciliados, plantas (incluindo, mas não limitadas a monocotiledôneas, dicotiledôneas, algas, etc.), fungos, leveduras, flagelados, microsporídios, protistas, etc.

Conforme usado neste relatório, o termo “não eucarionte” refere-se a organismos não eucarióticos. Por exemplo, um organismo não eucariótico pode pertencer ao domínio filogenético *Eubacteria* (incluindo, mas não limitado a *Escherichia coli*, *Thermus thermophilus*, *Bacillus stearothermophilus*, *Pseudomonas fluorescens*, *Pseudomonas aeruginosa*, *Pseudomonas putida*, etc.), ou ao domínio filogenético *Archaea* (por exemplo, *Methanococcus jannaschii*, *Methanobacterium thermoautotrophicum*, *Halobacterium*, tal como *Haloferax volcanii* e *Halobacterium* espécie NRC-1, *Archaeoglobus fulgidus*, *Pyrococcus furiosus*, *Pyrococcus horikoshii*, *Aeuropyrum pernix*, etc.).

Também incluídos são os exemplos não limitantes seguintes de proteína envelope viral; proteínas envelope a partir de filovírus (tais como vírus Ebola), ortomixovírus (tais como vírus influenza), VSV-G, alfa vírus (tais como o vírus Semliki forest e vírus Sindbis), arenavírus (tais como vírus da coriomeningite linfocítica), flavivírus (tais como vírus da encefalite transmitida por carrapato e vírus da Dengue), rabdovírus (tais como vírus da estomatite vesicular e vírus da raiva), vírus da leucemia de Moloney, HSV, VZV, vírus da Caxumba, Rinovírus, Sarampo, Rubéola, Arbovírus, Enterovírus (tais como Poliomielite, Coxsackie, Ecovírus), vírus da Poliomielite, Coxsackie B, A e Ecovírus, Rinovírus, vírus da Hepatite, vírus Norwalk, Astrovírus, Togavírus, Alfavírus, Pestivírus, Coronavírus, Parainfluenza, vírus da Caxumba, vírus do Sarampo, Vírus Respiratório Sincicial (RSV), Bunyaviridae, Reoviridae, Reovírus, Rotavírus, HTLV, Poliomavírus, Papilomavírus, Adenovírus, Parvovírus, EBV, CMV, vírus Varicella Zoster, herpesvírus e Poxvírus.

Variante conservativa: O termo “variante conservativa” refere-se a um componente de tradução, por exemplo, um tRNA de variante conservativa ou uma ORS de variante conservativa, que funcionalmente atua como o componente a partir do qual a variante conservativa é baseada, por exemplo, um tRNA ou ORS, porém com variações na sequência. Por exemplo, uma O-RS aminoacilará um tRNA complementar ou um tRNA de variante conservativa com um aminoácido selecionado, por exemplo, um aminoácido não natural, embora o tRNA e o tRNA de variante conservativa não tenham a mesma sequência. Si-

milarmemente, um tRNA será aminoacilado com um aminoácido selecionado, por exemplo, um aminoácido não natural, por uma O-RS complementar ou uma O-RS de variante conservativa, embora a O-RS e a O-RS de variante conservativa não tenham a mesma sequência. A variante conservativa pode ter, por exemplo, uma variação, duas variações, três variações, quatro variações, cinco ou mais variações na sequência, contanto que a variante conservativa seja complementar ao O-tRNA ou O-RS correspondente.

Agente de seleção ou triagem: Conforme usado neste relatório, o termo “agente de seleção ou triagem” refere-se a um agente que, quando presente, possibilita uma seleção/triagem de certos componentes a partir de uma população. Por exemplo, um agente de seleção ou triagem inclui, mas não é limitado a, por exemplo, um nutriente, um antibiótico, um comprimento de onda de luz, um anticorpo, um polinucleotídeo expressado ou semelhantes. O agente de seleção pode ser variado, por exemplo, por concentração, intensidade, etc.

O termo “não eficazmente reconhecido” refere-se a uma eficácia, por exemplo, menos do que cerca de 10 %, menos do que cerca de 5 % ou menos do que cerca de 1 %, em que uma RS a partir de um organismo aminoacila O-tRNA.

DESCRIÇÃO DETALHADA

Até agora, as vacinas foram limitadas àquelas produzidas com organismos mortos ou atenuados. A morte de micro-organismos pelo tratamento de calor, UV, formaldeído resulta em epítomos nativos reduzidos, e vírus atenuados são tipicamente produzidos através da deleção e truncamento de gene que leva à replicação extremamente baixa, porém não nula, um risco ao paciente e, particularmente, aos pacientes mais jovens, pacientes mais velhos e aqueles com sistemas imunes comprometidos.

Em uma modalidade, a presente invenção fornece vacinas de organismo total modificadas. Em uma outra modalidade, a presente invenção fornece vacinas de organismos totais geneticamente modificadas dependentes de replicação de um ou mais aminoácidos não naturais. A presente invenção fornece vacinas que incorporam micro-organismos nativos, que podem ser usados com ou sem adjuvantes. Estes podem fornecer adicionalmente uma alta densidade de epítomos nativos que são altamente imunogênicos e acionarão respostas humorais e mediadas por células T, desse modo, fornecendo uma vacina mais eficaz. A presente invenção fornece vacinas que incorporam um ou mais aminoácidos não naturais ou não naturalmente codificados em um ou mais sítios no vírus ou bactéria. Em algumas modalidades da presente invenção, o aminoácido não naturalmente codificado é incluído em uma parte do micro-organismo exigido para replicação, desse modo, eliminando os riscos associados aos vírus atenuados e eliminando a necessidade quanto à morte do vírus/bactéria. A vacina com aminoácidos não naturalmente codificados incorporados fornecerá um micro-organismo que, na estrutura, é extremamente próximo ao micro-organismo na-

tivo, entretanto, incapaz de replicação, ou apenas capaz de replicação limitada em um meio natural. Os controles absolutos da replicação do micro-organismo são realizados usando incorporação sítio-específica de aminoácidos, detalhada abaixo, para controlar a expressão do gene essencial (função) usando supressão do códon de parada. Para vírus, uma linha-
 5 gem celular de produção de hospedeiro sítio-especificamente incorpora um aminoácido não natural exigido para desenvolvimento. Para bactérias, o genoma é construído para incluir um ou mais aminoácidos não naturais sítio-especificamente incorporados.

O controle da função do gene essencial pode ser realizada em níveis funcionais genéticos e estruturais. Por exemplo, Tabela 2, no nível genético, o gene essencial de comprimento total é expressado na presença do aminoácido não naturalmente incorporado. No
 10 nível funcional estrutural, o gene essencial é construído para ser funcional apenas na presença do aminoácido não natural em um sítio específico. A incorporação de qualquer aminoácido natural neste sítio resultará na eliminação de sua função nativa. A liberdade de escolha de aminoácidos não naturais permite que uma pessoa de habilidade na técnica module a imunogenicidade da vacina desejada e a liberdade de replicação excessiva ou super-
 15 controlada, onde tradicionalmente foi necessário o equilíbrio entre atenuação de replicação e produção, a incorporação sítio-específica de um aminoácido não natural fornece as ferramentas de desenvolvimento para uma vacina com capacidades de replicação nulas na ausência do aminoácido não natural.

Bactérias e vírus através dos quais vacinas podem ser desenvolvidas para usar esta tecnologia incluem vírus conhecidos. Para exemplos não limitantes, isto inclui, vírus que afetam o trato respiratório superior, tais como rinovírus humano (HRV), adenovírus, coxsackievírus, influenza, parainfluenza, vírus respiratório sincicial (RSV), Vírus Epstein-Barr (EBV) e citomegalovírus (CMV); vírus que afetam o trato gastrointestinal (GI), tais como rotavírus, agente de Norwalk, hepatite A (HAV), vírus da poliomielite e outros picornavírus; vírus sexualmente transmitidos, incluindo, mas não limitados ao vírus da imunodeficiência humana (HIV), papilomavírus humano (HPV), vírus do herpes simples (HSV) 1, HSV-2, VZV, CMV, EBV, HHV-6, HHV-7 e HHV-8; CMV, vírus da hepatite B (HBV), vírus da hepatite C (HCV). Exemplos não limitantes de bactérias incluem bactérias gram positivas e gram negativas, incluindo *Staphylococcus*, *Streptococcus*, *Mycobacterium* (por exemplo, *Mycobacterium avium* e *Mycobacterium avium* subespécie *paratuberculosis*), *Enterococcus*, *Corynebacterium*, *Borrelia*, *Bacillus*, *Chlamydia*, *Mycoplasma* e semelhantes.
 20
 25
 30

Sistemas de tradução que são adequados para fabricar proteínas que incluem um ou mais aminoácidos selecionados, por exemplo, um aminoácido não natural, são descritos nos Pedidos de Patente U.S. 10/126.931, intitulada "METHODS AND COMPOSITION FOR THE PRODUCTION OF ORTHOGONAL tRNA-AMINOACYL tRNA SYNTHETASE PAIRS" e 10/126.927, intitulada "IN VIVO INCORPORATION OF UNNATURAL AMINO ACIDS". Além
 35

disso, veja USSN 10/825.867 intitulado “EXPANDING THE EUKARIOTIC GENETIC CODE.” Cada um destes pedidos é integralmente incorporado neste relatório como referência. Tais sistemas de tradução geralmente compreendem células que incluem um tRNA ortogonal (O-tRNA), uma aminoacil tRNA sintetase ortogonal (O-RS) e um aminoácido selecionado, por exemplo, um aminoácido não natural, onde a O-RS aminoacila o O-tRNA com o aminoácido selecionado. Um par ortogonal da presente invenção é composto de um O-tRNA, por exemplo, um tRNA supressor, um tRNA de mudança de fase de leitura ou semelhantes, e uma O-RS. O O-tRNA reconhece um primeiro códon seletor e tem atividade de supressão na presença de uma sintetase cognata em resposta a um códon seletor. A célula usa os componentes para incorporar o aminoácido selecionado em uma cadeia polipeptídica crescente. Por exemplo, um ácido nucleico que compreende um polinucleotídeo que codifica um polipeptídeo de interesse também pode estar presente, onde o polinucleotídeo compreende um códon seletor que é reconhecido pelo O-tRNA. O sistema de tradução também pode ser um sistema *in vitro*. Moléculas de tRNA da presente invenção são úteis em qualquer sistema de tradução, incluindo sistemas que utilizam ribossomos na tradução.

O sistema de tradução também pode ser um sistema de tradução livre de célula (*in vitro*). Nestes sistemas, que podem incluir tanto mRNA como um molde (tradução *in vitro*) quanto DNA como um molde (transcrição e tradução *in vitro* combinadas), a síntese *in vitro* é dirigida pelos ribossomos. Esforço considerável foi aplicado para o desenvolvimento de sistemas de expressão de proteína livres de células. Veja, por exemplo, Kim, D.M. e J.R. Swartz, *Biotechnology and Bioengineering*, 74:309-316 (2001); Kim, D.M. e J.R. Swartz, *Biotechnology Letters*, 22, 1537-1542, (2000); Kim, D.M., e J.R. Swartz, *Biotechnology Progress*, 16, 385-390, (2000); Kim, D.M., e J.R. Swartz, *Biotechnology and Bioengineering*, 66, 180-188, (1999); e Patnaik, R. e J.R. Swartz, *Biotechniques* 24, 862-868, (1998); Patente U.S. Nº 6.337.191; Publicação de Patente U.S. Nº 2002/0081660; WO 00/55353; WO 90/05785, os quais são incorporados como referência neste relatório. Um outro método que pode ser aplicado inclui a técnica de fusão mRNA-peptídeo. Veja, por exemplo, R. Roberts e J. Szostak, *Proc. Natl Acad. Sci. (USA)* 94: 12297-12302 (1997); A. Frankel, *et al.*, *Chemistry & Biology* 10:1043-1050 (2003). Neste método, um molde de mRNA ligado à puromicina é traduzido em peptídeo no ribossomo. Se uma ou mais moléculas de tRNA forem modificadas, os aminoácidos não naturais também podem ser incorporados no peptídeo. Depois que o último códon do mRNA for lido, a puromicina captura o C-terminal do peptídeo. Se for descoberto que o conjugado mRNA-peptídeo resultante tem propriedades interessantes em um ensaio *in vitro*, sua identidade pode ser facilmente revelada a partir da sequência de mRNA. Deste modo, uma pessoa habilitada pode triar bibliotecas de polipeptídeos compreendendo um ou mais aminoácidos não naturalmente codificados para identificar polipeptídeos tendo propriedades desejadas. Mais recentemente, traduções ribossomais *in vitro* com componen-

tes purificados foram relatadas, as quais permitem a síntese de peptídeos substituídos com aminoácidos não naturalmente codificados. Veja, por exemplo, A. Forster *et al.*, *Proc. Natl Acad. Sci. (USA)* 100:6353 (2003).

Em certas modalidades, a presente invenção fornece uma célula em que um aminoácido não natural foi incorporado em um produto de gene essencial. Em outras modalidades da presente invenção, uma célula foi geneticamente modificada para incluir um aminoácido não natural em um produto de gene que é exigido para replicação. Em certas modalidades, a presente invenção fornece uma bactéria em que um aminoácido não natural foi incorporado em um produto de gene essencial. Em outras modalidades da presente invenção, uma bactéria foi geneticamente modificada para incluir um aminoácido não natural em um produto de gene que é exigido para replicação. Em certas modalidades, a presente invenção fornece um vírus em que um aminoácido não natural foi incorporado em um produto de gene essencial. Em outras modalidades da presente invenção, um vírus foi geneticamente modificado para incluir um aminoácido não natural em um produto de gene que é exigido para replicação. Em certas modalidades, a presente invenção fornece uma célula de *Mycobacterium avium* subespécie *paratuberculosis* em que um aminoácido não natural foi incorporado em um produto de gene essencial. Em outras modalidades da presente invenção, uma célula de *Mycobacterium avium* subespécie *paratuberculosis* foi geneticamente modificada para incluir um aminoácido não natural em um produto de gene que é exigido para replicação. Em certas modalidades, a presente invenção fornece uma célula da meningite em que um aminoácido não natural foi incorporado em um produto de gene essencial. Em outras modalidades da presente invenção, uma célula da meningite foi geneticamente modificada para incluir um aminoácido não natural em um produto de gene que é exigido para replicação. Em certas modalidades, a presente invenção fornece uma célula da raiva em que um aminoácido não natural foi incorporado em um produto de gene essencial. Em outras modalidades da presente invenção, uma célula da raiva foi geneticamente modificada para incluir um aminoácido não natural em um produto de gene que é exigido para replicação. Em certas modalidades, a presente invenção fornece uma célula de alga em que um aminoácido não natural foi incorporado em um produto de gene essencial. Em outras modalidades da presente invenção, uma célula de alga foi geneticamente modificada para incluir um aminoácido não natural em um produto de gene que é exigido para replicação. Em certas modalidades, a presente invenção fornece uma célula viral em que um aminoácido não natural foi incorporado em um produto de gene essencial. Em outras modalidades da presente invenção, uma célula viral foi geneticamente modificada para incluir um aminoácido não natural em um produto de gene que é exigido para replicação. Em certas modalidades, a presente invenção fornece uma célula bacteriana em que um aminoácido não natural foi incorporado em um produto de gene essencial. Em outras modalidades da presente inven-

ção, uma célula bacteriana foi geneticamente modificada para incluir um aminoácido não natural em um produto de gene que é exigido para replicação. Em certas modalidades, a presente invenção fornece uma célula fúngica em que um aminoácido não natural foi incorporado em um produto de gene essencial. Em outras modalidades da presente invenção, uma célula fúngica foi geneticamente modificada para incluir um aminoácido não natural em um produto de gene que é exigido para replicação.

Em certas modalidades, a presente invenção fornece uma célula da poliomielite em que um aminoácido não natural foi incorporado em um produto de gene essencial. Em outras modalidades da presente invenção, uma célula da poliomielite foi geneticamente modificada para incluir um aminoácido não natural em um produto de gene que é exigido para replicação. Em certas modalidades, a presente invenção fornece uma célula de *E. coli* em que um aminoácido não natural foi incorporado em um produto de gene essencial. Em outras modalidades da presente invenção, uma célula de *E. coli* foi geneticamente modificada para incluir um aminoácido não natural em um produto de gene que é exigido para replicação. Em certas modalidades, a presente invenção fornece uma célula de micobactéria em que um aminoácido não natural foi incorporado em um produto de gene essencial. Em outras modalidades da presente invenção, uma célula de micobactéria foi geneticamente modificada para incluir um aminoácido não natural em um produto de gene que é exigido para replicação.

Em certas modalidades, células dependentes de aminoácido não natural geneticamente modificadas da presente invenção podem ser usadas para produzir uma vacina. Em certas modalidades, células dependentes de aminoácido não natural geneticamente modificadas da presente invenção podem ser usadas para produzir anticorpos que podem ser administrados como uma vacina. Em certas modalidades, células dependentes de aminoácido não natural geneticamente modificadas da presente invenção podem ser usadas em uma inoculação.

Em certas modalidades, uma célula de *E. coli* compreendendo o tRNA da presente invenção inclui tal sistema de tradução. Por exemplo, a célula de *E. coli* da presente invenção inclui um tRNA ortogonal (O-tRNA), onde o O-tRNA compreende a atividade de supressão na presença de uma sintetase cognata em resposta a um códon seletor; uma aminoacil-tRNA sintetase ortogonal (O-RS); um aminoácido selecionado; e, um ácido nucleico que compreende um polinucleotídeo que codifica um polipeptídeo de interesse, onde o polinucleotídeo compreende um códon seletor que é reconhecido pelo O-tRNA.

A invenção também caracteriza múltiplos pares de O-tRNA/O-RS em uma célula, que permite a incorporação de mais do que um aminoácido selecionado. Em certas modalidades, a célula pode ainda incluir um par de O-tRNA/O-RS diferente adicional e um segundo aminoácido selecionado, onde o O-tRNA reconhece um segundo códon seletor e a O-RS

preferencialmente aminoacila o O-tRNA com o segundo aminoácido selecionado. Por exemplo, uma célula pode ainda compreender, por exemplo, um par de tRNA supressor âmbar-aminoacil tRNA sintetase derivado da tirosil-tRNA sintetase de *Methanococcus jannaschii*.

O O-tRNA e/ou a O-RS podem ocorrer naturalmente ou podem ser derivados por mutação de um tRNA e/ou RS que ocorrem naturalmente, por exemplo, que geram bibliotecas de tRNAs e/ou bibliotecas de RSs, a partir de uma variedade de organismos. Por exemplo, uma estratégia de produzir um par de tRNA/aminoacil-tRNA sintetase ortogonal envolve importar um par de tRNA/sintetase heterólogo, por exemplo, a partir de uma fonte, exceto a célula hospedeira, ou múltiplas fontes, na célula hospedeira. As propriedades da sintetase heteróloga candidata incluem, por exemplo, aquelas que não carregam qualquer tRNA de célula hospedeira, e as propriedades do tRNA heterólogo candidato incluem, por exemplo, aquelas que não são aminoaciladas por qualquer sintetase de célula hospedeira. Além disso, o tRNA heterólogo é ortogonal a todas as sintetases de células hospedeiras.

Uma segunda estratégia para gerar um par ortogonal envolve gerar bibliotecas mutantes para triagem e/ou seleção de um O-tRNA ou O-RS. Estas estratégias também podem ser combinadas.

Em várias modalidades, o O-tRNA e a O-RS são derivados de pelo menos um organismo. Em uma outra modalidade, o O-tRNA é derivado de um tRNA que ocorre naturalmente ou tRNA que ocorre naturalmente modificado a partir de um primeiro organismo e a O-RS é derivada da RS que ocorre naturalmente ou RS que ocorre naturalmente modificada a partir de um segundo organismo. Em uma modalidade, o primeiro e segundo organismos são diferentes. Por exemplo, um par ortogonal pode incluir uma tRNA sintetase derivada de *Methanobacterium thermoautotrophicum*, e um tRNA derivado de um tRNA de archae (por exemplo, a partir de *Halobacterium sp. NRC-1*). Alternativamente, o primeiro e segundo organismos são os mesmos. Veja a seção intitulada “Fontes e Organismos Hospedeiros” neste relatório para informação adicional.

Em certas modalidades da presente invenção, um O-tRNA da presente invenção compreende ou é codificado por uma sequência de polinucleotídeos apresentada na SEQ ID NO.: 1, 2 ou 3, ou uma sequência de polinucleotídeos complementar da mesma, ou uma variação conservativa da mesma. Veja também a seção intitulada “Sequência de Ácidos Nucleicos e Polipeptídeos e Variantes” neste relatório.

tRNA ortogonal (O-tRNA)

Um tRNA ortogonal (O-tRNA) medeia a incorporação de um aminoácido selecionado em uma proteína que é codificada por um polinucleotídeo que compreende um códon seletor que é reconhecido pelo O-tRNA, por exemplo, *in vivo* ou *in vitro*. Um O-tRNA da presente invenção pode ser aminoacilado com um aminoácido desejado por qualquer método ou técnica, incluindo, mas não limitado à aminoacilação química ou enzimática. O O-tRNA

aminoacilado da presente invenção pode ser diretamente adicionado a um sistema de tradução. Um O-tRNA da presente invenção pode ser aminoacilado por uma RS com um aminoácido selecionado *in vitro* ou *in vivo*. Além disso, a RS pode ser uma O-RS. Um O-tRNA da presente invenção pode ser diretamente fornecido ao sistema de tradução (por exemplo, componentes de tradução *in vitro*, ou uma célula), ou fornecendo-se um polinucleotídeo que codifica um O-tRNA ou uma porção do mesmo. Por exemplo, um O-tRNA, ou uma porção do mesmo, é codificado por uma sequência de polinucleotídeos apresentada na SEQ ID NO.: 1, 2, 3, ou uma sequência de polinucleotídeos complementar da mesma, ou uma variação conservativa da mesma.

Um O-tRNA da presente invenção compreende atividade de supressão na presença de um sintetase cognata em resposta a um códon seletor. A atividade de supressão pode ser determinada por qualquer um dos vários ensaios conhecidos na técnica. Por exemplo, um ensaio de β -galactosidase repórter pode ser usado. Um derivado de um plasmídeo que expressa o gene *lacZ* sob o controle do promotor é usado, por exemplo, onde a Leu-25 do peptídeo VVLQRRDWEN de *lacZ* é substituída por um códon seletor, por exemplo, códons TAG, TGA, AGGA, etc., ou códons sentido (como um controle) para tirosina, serina, leucina, etc. O plasmídeo de *lacZ* derivatizado é introduzido em células a partir de um organismo apropriado (por exemplo, um organismo onde os componentes ortogonais podem ser usados) junto com o plasmídeo compreendendo um O-tRNA da presente invenção. Uma sintetase cognata também pode ser introduzida (como um polipeptídeo ou um polinucleotídeo que codifica a sintetase cognata quando expressada). As células são cultivadas no meio em uma densidade desejada, por exemplo, em uma OD₆₀₀ de cerca de 0,5, e os ensaios de β -galactosidase são realizados, por exemplo, usando o BetaFluor™ β -Galactosidase Assay Kit (Novagen). A porcentagem de supressão é calculada como a porcentagem da atividade para uma amostra em relação a um controle comparável, por exemplo, o valor observado a partir do constructo de *lacZ* derivatizado, onde o constructo tem um códon sentido correspondente na posição desejada ao invés de um códon seletor.

Na molécula de tRNA, Timina (T) é substituída por Uracila (U). Além disso, modificações adicionais nas bases podem estar presentes. A invenção também inclui variações conservativas de O-tRNA. Por exemplo, variações conservativas de O-tRNA incluem aquelas moléculas que funcionam como o O-tRNA e mantêm a estrutura na forma de L do tRNA, porém não têm a mesma sequência (e são outras que não moléculas de tRNA do tipo selvagem). Veja também a seção neste relatório intitulada “Sequência de Ácidos Nucleicos e Polipeptídeos e Variantes”.

A composição compreendendo um O-tRNA pode ainda incluir uma aminoacil-tRNA sintetase ortogonal (O-RS), onde a O-RS preferencialmente aminoacila o O-tRNA com um aminoácido selecionado (por exemplo, um aminoácido não natural). Em certas modalidades,

uma composição incluindo um O-tRNA pode ainda incluir um sistema de tradução (por exemplo, um sistema de tradução *in vitro* ou *in vivo*). Um ácido nucleico compreendendo um polinucleotídeo que codifica um polipeptídeo de interesse, em que o polinucleotídeo compreende um ou mais códons seletores reconhecidos pelo O-tRNA, ou uma combinação de um ou mais entre estes, também pode estar presente na célula ou outro sistema de tradução. Veja também a seção neste relatório intitulada “Aminoacil-tRNA Sintetases Ortogonais (O-RS)”.

Métodos de produzir um tRNA ortogonal (O-tRNA), por exemplo, um O-tRNA, também são uma característica da presente invenção. Um tRNA, por exemplo, um O-tRNA, produzido pelo método também é uma característica da presente invenção.

Métodos de produzir um tRNA ortogonal incluem modificar a alça anticódon de cada um dos reservatórios de tRNAs para permitir o reconhecimento de um códon seletor (por exemplo, um códon âmbar, um códon opala, um códon de quatro bases, etc.), desse modo, fornecendo uma pluralidade de O-tRNAs potenciais; e analisar a estrutura secundária de um membro da pluralidade de O-tRNAs potenciais para identificar pares de base não canônicos na estrutura secundária e, opcionalmente, modificar os pares de base não canônicos (por exemplo, os pares de base não canônicos são modificados para pares de base canônicos). Os pares de base não canônicos podem estar localizados na região de tronco da estrutura secundária. Um O-tRNA pode possuir um aperfeiçoamento de uma ou mais características ou atividades, tal como um aperfeiçoamento na ortogonalidade para um organismo desejado em comparação ao material de partida, por exemplo, a pluralidade de sequências de tRNA, enquanto preserva sua afinidade em relação a uma RS desejada.

Alternativamente, OtRNAs podem ser desenvolvidos modificando-se um tRNA conhecido para modular sua interação ou afinidade de ligação a uma ou mais moléculas que influenciam na tradução ou são componentes do mecanismo de tradução. Tais componentes incluem, mas não são limitados aos fatores de alongação. O fator de alongação bacteriano EF-Tu desempenha um papel essencial na etapa de alongação na síntese de proteínas. Após a aminoacilação do tRNA em tRNA sintetase, EF-Tu se liga ao tRNA aminoacilado e leva o mesmo ao sítio A do ribossomo. A ligação éster entre o aminoácido carregado e o tRNA é protegida da hidrólise espontânea devido à ligação entre EF-Tu e tRNA aminoacilado. Stortchevoi *et al.* investigaram mutantes do par de base oscilante U50:G64 de tRNA^{Met} de iniciação de *E. coli* no tronco TΨC, visto que seu par de base foi descoberto ser um determinante negativo secundário bloqueando a atividade de tRNAs na alongação, presumivelmente devido a uma interação enfraquecida entre EF-Tu.GTP e tRNA aminoacilado (JBC 2003 278(20): 17672-17679). Além disso, LaRiviere *et al.*, descreveram em Science 2001 Oct 5;294(5540): 165-8 as contribuições termodinâmicas do aminoácido e corpo de tRNA para a afinidade de ligação global para EF-Tu. Eles indicaram que as contribuições do corpo

de tRNA e do aminoácido são independentes umas das outras e que elas se compensam quando os tRNAs são corretamente acilados. Alterações para a interação entre EF-Tu.GTP e o tRNA aminoacilado com o aminoácido não natural podem afetar a eficácia do carregamento do tRNA no sítio A do ribossomo. Sítios de mutação potenciais também podem ser encontrados analisando-se as estruturas cristalinas de complexos entre tRNA e outros componentes do mecanismo de tradução, tais como EF-Tu. Por exemplo, Nissen *et al.* indicaram que EF-Tu.GTP se liga diretamente à cadeia principal do fosfato do tronco TΨC de fenilalanila-RNA de transferência (Phe-tRNA) de levedura (Science 1995 270(5241):1464-1472).

Os métodos incluem opcionalmente analisar a homologia de sequências de tRNAs e/ou aminoacil-tRNA sintetases para determinar os candidatos potenciais para um O- tRNA, O-RS e/ou pares dos mesmos, que parecem ser ortogonais para um organismo específico. Programas de computador conhecidos na técnica e descritos neste relatório podem ser usados para a análise. Em um exemplo, para escolher componentes de tradução ortogonais potenciais para o uso em um organismo procariótico, uma sintetase e/ou um tRNA são escolhidos, os quais não apresentam homologia incomum aos organismos procarióticos.

Um reservatório de tRNAs também pode ser produzido por uma estratégia consenso. Por exemplo, o reservatório de tRNAs é produzido alinhando-se uma pluralidade de sequências de tRNA; determinando-se uma sequência consenso; e gerando-se uma biblioteca de tRNAs usando pelo menos uma porção, a maioria ou a sequência consenso total. Por exemplo, uma sequência consenso pode ser compilada com um programa de computador, por exemplo, o programa GCG *pileup*. Opcionalmente, posições degeneradas determinadas pelo programa são modificadas na base mais frequente em tais posições. Uma biblioteca é sintetizada por técnicas conhecidas no ramo usando a sequência consenso. Por exemplo, a sobreposição da extensão de oligonucleotídeos em que cada sítio do gene de tRNA pode ser sintetizado como uma mistura dopada de 90 % da sequência consenso e 10 % de uma mistura das outras 3 bases pode ser usada para fornecer a biblioteca com base na sequência consenso. Outras misturas também podem ser usadas, por exemplo, 75 % da sequência consenso e 25 % de uma mistura das outras 3 bases, 80 % da sequência consenso e 20 % de uma mistura das outras 3 bases, 95 % da sequência consenso e 5 % de uma mistura das outras 3 bases, etc.

Bibliotecas de tRNAs mutantes podem ser geradas usando várias técnicas de mutagênese conhecidas no ramo. Por exemplo, os tRNAs mutantes podem ser gerados por mutações sítio-específicas, mutações de ponto aleatórias, recombinação homóloga, embaralhamento do DNA ou outros métodos de mutagênese recursivos, construção quimérica ou qualquer combinação dos mesmos.

Mutações adicionais podem ser introduzidas em uma posição específica, por exemplo, em uma posição não conservativa, ou em uma posição conservativa, em uma posi-

ção randomizada, ou uma combinação das mesmas, em uma alça ou região desejada de um tRNA, por exemplo, uma alça anticódon, o tronco aceitante, ramo ou alça D, alça variável, ramo ou alça TΨC, outras regiões da molécula de tRNA ou uma combinação dos mesmos. As mutações podem incluir pares de base combinados na região de tronco.

5 Tipicamente, um O-tRNA é obtido submetendo-se uma população de células de uma primeira espécie à seleção negativa, onde as células compreendem um membro da pluralidade de O-tRNAs potenciais. A seleção negativa elimina células que compreendem um membro da pluralidade de O-tRNAs potenciais que é aminoacilada por uma aminoacil-tRNA sintetase (RS) que é endógena às células. Isto fornece um reservatório de tRNAs que
10 são ortogonais à célula da primeira espécie.

 Em certas modalidades da seleção negativa, um códon seletor é introduzido no polinucleotídeo que codifica um marcador de seleção negativa, por exemplo, uma enzima que confere resistência ao antibiótico, por exemplo, β -lactamase, uma enzima que confere um produto detectável, por exemplo, β -galactosidase, cloranfenicol acetiltransferase (CAT), por
15 exemplo, um produto tóxico, tal como barnase, em uma posição não essencial, etc. A triagem/seleção pode ser feita através do crescimento da população de células na presença de um agente de seleção (por exemplo, um antibiótico, tal como ampicilina). Em uma modalidade, a concentração do agente de seleção é variada.

 Por exemplo, para medir a atividade de tRNAs supressores, um sistema de seleção
20 é usado, o qual é fundamentado na supressão *in vivo* do códon seletor, por exemplo, mutações sem sentido ou de mudança de fase de leitura introduzidas em um polinucleotídeo que codifica um marcador de seleção negativa, por exemplo, um gene para β -lactamase (*bla*). Por exemplo, variantes de polinucleotídeos, por exemplo, variantes de *bla*, por exemplo, com TAG, AGGA e TGA em uma determinada posição, são construídas. Células, por exemplo, bactérias, são transformadas com estes polinucleotídeos. No caso de um tRNA ortogonal, que não pode ser eficazmente carregado por sintetases endógenas de *E. coli*, a resistência ao antibiótico, por exemplo, resistência à ampicilina, deve ser cerca de ou menor do que aquela para uma bactéria transformada sem plasmídeo. Se o tRNA não é ortogonal, ou se uma sintetase heteróloga capaz de carregar o tRNA é coexpressada no sistema, um nível
25 mais alto de resistência ao antibiótico, por exemplo, ampicilina, deve ser observado. Células, por exemplo, bactérias, são escolhidas, as quais são incapazes de crescer em placas de LB ágar com concentrações de antibiótico aproximadamente iguais às células transformadas sem plasmídeos.

 No caso de um produto tóxico (por exemplo, ribonuclease barnase), quando um
35 membro da pluralidade de tRNAs potenciais é aminoacilado pelo hospedeiro endógeno, por exemplo, sintetases de *Escherichia coli* (isto é, ele não é ortogonal ao hospedeiro, por exemplo, sintetases de *Escherichia coli*), o códon seletor é suprimido e o produto de polinu-

cleotídeo tóxico produzido leva à morte celular. Células que hospedam tRNA ortogonal ou tRNAs não funcionais sobrevivem.

Em uma modalidade, o reservatório de tRNAs que são ortogonais a um organismo desejado depois é submetido a uma seleção positiva em que um códon seletor é colocado em um marcador de seleção positiva, por exemplo, codificado por um gene de resistência ao fármaco, tal como um gene de β -lactamase. A seleção positiva é realizada na célula compreendendo um polinucleotídeo que codifica ou compreende um membro do reservatório de tRNAs, um polinucleotídeo que codifica um marcador de seleção positiva e um polinucleotídeo que codifica RS cognata. Estes polinucleotídeos são expressados na célula e a célula é cultivada na presença de um agente de seleção, por exemplo, ampicilina. O tRNA depois é selecionado quanto a sua capacidade de ser aminoacilado pela sintetase cognata coexpressada e para inserir um aminoácido em resposta a este códon seletor. Tipicamente, estas células mostram uma acentuação na eficácia de supressão comparadas às células que hospedam tRNAs não funcionais, ou tRNAs que não podem ser eficazmente reconhecidos pela sintetase de interesse. A célula que hospeda os tRNAs não funcionais ou tRNAs que não são eficazmente reconhecidos pela sintetase de interesse é sensível ao antibiótico. Portanto, tRNAs que: (i) não são substratos para o hospedeiro endógeno, por exemplo, sintetases de *Escherichia coli*; (ii) podem ser aminoacilados pela sintetase de interesse; e (iii) são funcionais em tradução e sobrevivem em ambas as seleções.

A estringência da seleção, por exemplo, a seleção positiva, a seleção negativa ou seleção positiva e negativa, nos métodos acima descritos, pode ser opcionalmente variada. Por exemplo, pelo fato de a barnase ser uma proteína extremamente tóxica, a estringência da seleção negativa pode ser controlada introduzindo-se números diferentes de códons seletores no gene da barnase e/ou usando-se um promotor induzível. Em um outro exemplo, a concentração do agente de seleção ou triagem é variada (por exemplo, ampicilina). Em um aspecto, a estringência é variada, pois a atividade desejada pode ser baixa durante as rodadas iniciais. Desse modo, critérios de seleção menos severos são aplicados em rodadas iniciais e critérios mais severos são aplicado em rodadas finais de seleção. Em certas modalidades, a seleção negativa, a seleção positiva ou a seleção negativa e positiva podem ser repetidas diversas vezes. Múltiplos marcadores de seleção negativa, marcadores de seleção positiva ou marcadores de seleção negativa e positiva diferentes podem ser usados. Em certas modalidades, o marcador de seleção positiva e negativa pode ser o mesmo.

Outros tipos de seleções/triagem podem ser usados na invenção para produzir componentes de tradução ortogonais, por exemplo, um O-tRNA, uma O-RS e um par de O-tRNA/O-RS. Por exemplo, o marcador de seleção negativa, o marcador de seleção positiva ou marcadores de seleção positiva e negativa podem incluir um marcador que fluoresce ou catalisa uma reação luminescente na presença de um reagente adequado. Em uma outra

modalidade, um produto do marcador é detectado por separador celular ativado por fluorescência (FACS) ou por luminescência. Opcionalmente, o marcador inclui um marcador de triagem com base na afinidade. Veja, Francisco, J. A., *et al.*, (1993) *Production and fluorescence-activated cell sorting of Escherichia coli expressing a functional antibody fragment on the external surface*. Proc Natl Acad Sci U.S.A. 90: 10444-8.

Métodos adicionais para produzir um tRNA ortogonal recombinante podem ser encontrados, por exemplo, nos Pedidos de Patente U.S. 10/126.931, intitulada "Methods and Compositions for the Production of Orthogonal tRNA-Aminoacyl tRNA Synthetase Pairs" e 10/126.127, intitulada "In vivo Incorporation of Unnatural Amino Acids," e USSN 10/825.867 intitulado "EXPANDING THE EUKARYOTIC GENETIC CODE". Veja também, Forster *et al.*, (2003) *Programming peptidomimetic synthetases by translating genetic codes designed de novo*. PNAS 100(II):6353-6357; e, Feng *et al.*, (2003), *Expanding tRNA recognition of a tRNA synthetase by a single amino acid change*, PNAS 100(10): 5676-5681.

Um tRNA da presente invenção pode ser aminoacilado com um aminoácido desejado por qualquer método ou técnica, incluindo, mas não limitado à aminoacilação química ou enzimática.

A aminoacilação pode ser realizada por aminoacil tRNA sintetases ou por outras moléculas enzimáticas, incluindo, mas não limitadas às ribozimas. O termo "ribozima" é permutável com "RNA catalítico". Cech e colaboradores (Cech, 1987, Science, 236:1532-1539; McCorkle *et al.*, 1987, Concepts Biochem. 64:221-226) demonstraram a presença de RNAs que ocorrem naturalmente que podem atuar como catalisadores (ribozimas). Entretanto, embora estes catalisadores de RNA naturais tenham sido mostrados atuando em substratos de ácido ribonucleico para clivagem e *splicing*, o desenvolvimento recente de evolução artificial de ribozimas expandiu o repertório de catálise para várias reações químicas. Estudos identificaram moléculas de RNA que podem catalisar ligações de aminoacil-RNA em seu próprio (2') 3'-terminal (Illangakekare *et al.*, 1995 Science 267:643-647), e uma molécula de RNA que pode transferir um aminoácido a partir de uma molécula de RNA para uma outra (Lohse *et al.*, 1996, Nature 381:442-444).

A Publicação do Pedido de Patente U.S. 2003/0228593, que é incorporado como referência neste relatório, descreve métodos para construir ribozimas e seu uso na aminoacilação de tRNAs com aminoácidos naturalmente codificados e não naturalmente codificados. Formas imobilizadas em substrato de moléculas enzimáticas que podem aminoacilar tRNAs, incluindo, mas não limitadas às ribozimas, podem permitir purificação de afinidade eficaz dos produtos aminoacilados. Exemplos de substratos adequados incluem agarose, sepharose e pérolas magnéticas. A produção e uso de uma forma imobilizada em substrato de ribozima para aminoacilação são descritos em Chemistry and Biology 2003, 10: 1077-1084 e Publicação de Pedido de Patente U.S. 2003/0228593, os quais são incorporados

como referência neste relatório.

Métodos de aminoacilação química incluem, mas não são limitados àqueles introduzidos por Hecht e colaboradores (Hecht, S. M. *Ace. Chem. Res.* 1992, 25, 545; Heckler, T. G.; Roesser, J. R.; Xu, C; Chang, P.; Hecht, S. M. *Biochemistry* 1988, 27, 7254; Hecht, S. M.; Alford, B. L.; Kuroda, Y.; Kitano, S. J. *Biol. Chem.* 1978, 253, 4517) e por Schultz, Chamberlin, Dougherty e outros (Cornish, V. W.; Mendel, D.; Schultz, P. G. *Angew. Chem. Int. Ed. Engl.* 1995, 34, 621; Robertson, S. A.; Ellman, J. A.; Schultz, P. G. *J. Am. Chem. Soc.* 1991, 113, 2722; Noren, C. J.; Anthony-Cahill, S. J.; Griffith, M. C; Schultz, P. G. *Science* 1989, 244, 182; Bain, J. D.; Glabe, C. G.; Dix, T. A.; Chamberlin, A. R. *J. Am. Chem. Soc.* 1989, 111, 8013; Bain, J. D. *et al. Nature* 1992, 356, 537; Gallivan, J. P.; Lester, H. A.; Dougherty, D. A. *Chem. Biol.* 1997, 4, 740; Turcatti, *et al. J. Biol. Chem.* 1996, 271, 19991; Nowak, M. W. *et al. Science*, 1995, 268, 439; Saks, M. E. *et al. J. Biol. Chem.* 1996, 271, 23169; Hohsaka, T. *et al. J. Am. Chem. Soc.* 1999, 121, 34), para evitar o uso de sínteses na aminoacilação. Tais métodos ou outros métodos de aminoacilação química podem ser usados para aminoacilar moléculas de tRNA da invenção.

Métodos biossintéticos que utilizam aminoacil-tRNAs quimicamente modificados foram usados para incorporar várias sondas biofísicas em proteínas sintetizadas *in vitro*. Veja as publicações e referências seguintes citadas em: Brunner, J. *New Photolabeling and crosslinking methods*, Annu. Rev Biochem, 62:483-514 (1993); e, Krieg, U.C., Walter, P., Hohnson, A.E. *Photocrosslinking of the signal sequence of nascent preprolactin of the 54-kilodalton polypeptide of the signal recognition particle*, Proc. Natl. Acad. Sci, 83(22):8604-8608 (1986).

Previamente, foi mostrado que os aminoácidos não naturais podem ser sítio-especificamente incorporados em proteínas *in vitro* pela adição de tRNAs supressores quimicamente aminoacilados para reações de síntese de proteína programadas com um gene contendo uma mutação sem sentido âmbra desejada. Usando estas abordagens, uma pessoa habilitada pode substituir os vinte aminoácidos comuns com homólogos estruturais próximos, por exemplo, fluorofenilalanina para fenilalanina, usando cepas auxotróficas para um aminoácido particular. Veja, por exemplo, Noren, C.J., Anthony-Cahill, Griffith, M. C, Schultz, P. G. *A general method for site-specific incorporation of unnatural amino acids into proteins*, *Science*, 244: 182-188 (1989); M. W, Nowak, *et al.*, *Science* 268:439-42 (1995); Bain, J.D., Glabe, C.G., Dix, T.A., Chamberlin, A.R., Diala, E.S. *Biosynthetic site-specific Incorporation of a non-natural amino acid into a polypeptide*, *J. Am Chem Soc*, 111:8013-8014 (1989); N. Budisa *et al.*, *FASEB J.* 13:41-51 (1999); Ellman, J.A., Mendel, D., Anthony-Cahill, S., Noren, C.J., Schultz, P.G. *Biosynthetic method for introducing unnatural amino acids site-specifically into proteins*. Methods in Enz., vol. 202, 301-336 (1992); e, Mendel, D., Cornish, V.W. & Schultz, P.G. *Site-Directed Mutagenesis with an Expanded Genetic Code*, Annu Rev Bio-

phys. Biomol Struct. 24, 435-62 (1995).

Por exemplo, um tRNA supressor foi preparado, o qual reconheceu o códon de parada UAG e foi quimicamente aminoacilado com um aminoácido não natural. Mutagênese sítio-dirigida convencional foi usada para introduzir o códon de parada TAG no sítio de interesse no gene da proteína. Veja, por exemplo, Sayers, J.R., Schmidt, W. Eckstein, F. 5' -3' *Exonucleases in phosphorothioate-based oligonucleotide-directed mutagenesis*, Nucleic Acids Res. 16(3):791-802 (1988), Quando o tRNA supressor acilado e o gene mutante foram combinados em um sistema de transcrição/tradução *in vitro*, o aminoácido não natural foi incorporado em resposta ao códon UAG que forneceu uma proteína contendo tal aminoácido na posição específica. Experimentos usando [³H]-Phe e experimentos com ácidos α-hidróxi demonstraram que apenas o aminoácido desejado é incorporado na posição especificada pelo códon UAG e que este aminoácido não é incorporado em qualquer outro sítio na proteína. Veja, por exemplo, Noren, *et al.*, *supra*; Kobayashi *et al.*, (2003) *Nature Structural Biology* 10(6):425-432; e, Ellman, J.A., Mendel, D., Schultz, P.G. *Site-specific incorporation of novel backbone structures into proteins*, *Science*, 255(5041):197-200 (1992).

Métodos para gerar RNA catalítico podem envolver gerar reservatórios separados de sequências de ribozima randomizadas, realizar evolução dirigida sobre os reservatórios, triar os reservatórios para atividade de aminoacilação desejável e selecionar sequências de tais ribozimas que exibem atividade de aminoacilação desejada.

Ribozimas podem compreender motivos e/ou regiões que facilitam a atividade de acilação, tal como um motivo GGU e uma região rica em U. Por exemplo, foi relatado que regiões ricas em U podem facilitar o reconhecimento de um substrato de aminoácido, e um motivo GGU pode formar pares de base com os terminais 3' de um tRNA. Em combinação, o motivo GGU e a região rica em U facilitam o reconhecimento simultâneo do aminoácido e tRNA e, desse modo, facilitam a aminoacilação do terminal 3' do tRNA.

Ribozimas podem ser geradas por seleção *in vitro* usando um r24mini parcialmente randomizado conjugado com tRNA^{Asn}_{CCCG}, seguido por engenharia sistemática de uma sequência consenso encontrada nos clones ativos. Uma ribozima exemplar obtida por este método é denominada "ribozima Fx3" e é descrita em U.S. Pub. App. Nº 2003/0228593, os conteúdos do qual são incorporados como referência neste relatório, atua como um catalisador versátil para a síntese de vários aminoacil-tRNAs carregados com aminoácidos não naturais cognatos.

A imobilização em um substrato pode ser usada para permitir a purificação de afinidade eficaz dos tRNAs aminoacilados. Exemplos de substratos adequados incluem, mas não são limitados à agarose, sepharose e pérolas magnéticas. Ribozimas podem ser imobilizadas em resinas tomando-se vantagem da estrutura química do RNA, tal como o 3'-cis-diol na ribose do RNA pode ser oxidado com periodato para produzir o dialdeído correspon-

dente para facilitar a imobilização do RNA na resina. Vários tipos de resinas podem ser usados, incluindo resinas de hidrazida baratas em que a aminação redutiva torna a interação entre a resina e a ribozima uma ligação irreversível. A síntese de aminoacil-tRNAs pode ser significativamente facilitada por sua técnica de aminoacilação em coluna. Kourouklis *et al.* Methods 2005; 36:239-4 descrevem um sistema de aminoacilação com base em coluna.

O isolamento dos tRNAs aminoacilados pode ser realizado em uma variedade de formas. Um método adequado é eluir os tRNAs aminoacilados a partir de uma coluna com um tampão, tal como uma solução de acetato de sódio com EDTA 10 mM, um tampão contendo N-(2-hidroxietil)piperazino-N'-(ácido 3-propanossulfônico) 50 mM, KCl 12,5 mM, pH 7,0, EDTA 10 mM ou simplesmente água tamponada com EDTA (pH 7,0).

Os tRNAs aminoacilados da presente invenção podem ser adicionados às reações de tradução, de modo a incorporar o aminoácido com o qual o tRNA foi aminoacilado em uma posição de escolha em um polipeptídeo preparado pela reação de tradução. Exemplos dos sistemas de tradução em que os tRNAs aminoacilados da presente invenção podem ser usados incluem, mas não são limitados aos lisatos celulares. Lisatos celulares fornecem componentes de reação necessários para tradução *in vitro* de um polipeptídeo a partir de um mRNA de entrada. Exemplos de tais componentes de reação incluem, mas não são limitados às proteínas ribossomais, rRNA, aminoácidos, tRNAs, GTP, ATP, iniciação de tradução e fatores de alongação e fatores adicionais associados à tradução. Adicionalmente, os sistemas de tradução podem ser traduções em lote ou tradução compartimentalizada. Sistemas de tradução em lote combinam componentes de reação em um compartimento único, enquanto os sistemas de tradução compartimentalizados separam os componentes da reação de tradução dos produtos de reação que podem inibir a eficácia da tradução. Tais sistemas de tradução são comercialmente disponíveis.

Além disso, um sistema de transcrição/tradução acoplado pode ser usado. Sistemas de transcrição/tradução acoplados incluem a transcrição de um DNA de entrada em um mRNA correspondente, que é, por sua vez, traduzido pelos componentes de reação. Um exemplo de um sistema de transcrição/tradução acoplado comercialmente disponível é o Rapid Translation System (RTS, Roche Inc.). O sistema inclui uma mistura contendo lisato de *E. coli* para fornecer componentes de tradução, tais como ribossomos e fatores de tradução. Adicionalmente, uma RNA polimerase é incluída para a transcrição do DNA de entrada em um molde de mRNA para o uso na tradução. O RTS pode usar compartimentalização dos componentes de reação por via de uma membrana interposta entre os compartimentos de reação, incluindo um compartimento de fornecimento/gasto e um compartimento de transcrição/tradução.

A aminoacilação de tRNA pode ser realizada por outros agentes, incluindo, mas não limitados às transferases, polimerases, anticorpos catalíticos, proteínas multifuncionais

e semelhantes.

Aminoacil-tRNA Sintetases Ortogonais (O-RS)

Uma O-RS preferencialmente aminoacila um O-tRNA da presente invenção com um aminoácido selecionado *in vitro* ou *in vivo*. Uma O-RS da presente invenção pode ser fornecida ao sistema de tradução (por exemplo, componentes de tradução *in vitro* ou uma célula) por um polipeptídeo que inclui uma O-RS e/ou por um polinucleotídeo que codifica uma O-RS ou uma porção da mesma. Uma O-RS, ou uma porção da mesma, é codificada por uma sequência de polinucleotídeos ou uma sequência de polinucleotídeos complementar da mesma, ou uma variação conservativa da mesma. Um O-tRNA da presente invenção pode ser aminoacilado por várias moléculas de O-RS diferentes, incluindo, mas não limitadas àquelas divulgadas neste relatório.

Métodos para identificar uma aminoacil-tRNA sintetase ortogonal (O-RS), por exemplo, uma O-RS, para o uso com um O-tRNA, por exemplo, um O-tRNA, também são uma característica da presente invenção. Por exemplo um método inclui submeter uma população de células de uma primeira espécie à seleção positiva, onde as células compreendem: 1) um membro de uma pluralidade de aminoacil-tRNA sintetases (RSs), onde a pluralidade de RSs compreende RSs mutantes, RSs derivadas de uma espécie, exceto a primeira espécie, ou as RSs mutantes e RSs derivadas de uma espécie, exceto a primeira espécie; 2) o tRNA ortogonal (O-tRNA) a partir de uma segunda espécie; e 3) um polinucleotídeo que codifica um marcador de seleção positiva e compreende pelo menos um códon seletor. As células são selecionadas ou triadas quanto àquelas que mostram uma acentuação na eficácia de supressão em comparação às células desprovidas ou com uma quantidade reduzida do membro da pluralidade de RSs. As células que apresentam uma acentuação na eficácia de supressão compreendem uma RS ativa que aminoacila o O-tRNA. Um nível de aminoacilação (*in vitro* ou *in vivo*) pela RS ativa de um primeiro conjunto de tRNAs a partir da primeira espécie é comparado ao nível de aminoacilação (*in vitro* ou *in vivo*) pela RS ativa de um segundo conjunto de tRNAs a partir da segunda espécie. O nível de aminoacilação pode ser determinado por uma substância detectável (por exemplo, um aminoácido marcado ou aminoácido não natural). A RS ativa que mais eficazmente aminoacila o segundo conjunto de tRNAs em comparação ao primeiro conjunto de tRNAs é selecionada, desse modo, fornecendo a aminoacil-tRNA sintetase ortogonal para o uso com o O-tRNA. Uma O-RS, por exemplo, uma O-RS identificada pelo método, também é uma característica da presente invenção.

Qualquer um dos vários ensaios pode ser usado para determinar a aminoacilação. Estes ensaios podem ser realizados *in vitro* ou *in vivo*. Por exemplo, ensaios de aminoacilação *in vitro* são descritos, por exemplo, em Hoben, P., e Solo, D. (1985) Methods Enzymol. 113:55-59 e na Publicação de Pedido de Patente U.S. Nº 2003/0228593. A aminoacilação

também pode ser determinada usando-se um repórter com os componentes de tradução ortogonais e detectando-se o repórter em uma célula que expressa um polinucleotídeo compreendendo pelo menos um códon seletor que codifica uma proteína. Veja também, Pedido de Patente U.S. 10/126.927, intitulado "IN VIVO INCORPORATION OF UNNATURAL AMINO ACIDS", e USSN 10/825.867 intitulado "EXPANDING THE EUKARIOTIC GENETIC CODE".

Uma O-RS identificada pode ser ainda manipulada para alterar a especificidade do substrato da sintetase, de modo que apenas um aminoácido desejado não natural, porém não qualquer um dos 20 aminoácidos comuns, é carregado em O-tRNA. Métodos para gerar uma aminoacil tRNA sintetase ortogonal com uma especificidade de substrato para um aminoácido não natural incluem modificar a sintetase, por exemplo, no sítio ativo na sintetase, no sítio do mecanismo de edição na sintetase, em sítios diferentes combinando-se domínios diferentes de sintetases ou semelhantes, e aplicando-se um processo de seleção. Uma estratégia é usada, a qual é fundamentada na combinação de uma seleção positiva, seguido por uma seleção negativa. Na seleção positiva, a supressão do códon seletor introduzido em uma posição não essencial de um marcador positivo permite que as células sobrevivam sob pressão de seleção positiva. Na presença de aminoácidos naturais e não naturais, os sobreviventes codificam sintetases ativas carregando o tRNA supressor ortogonal com um aminoácido natural ou não natural. Na seleção negativa, a supressão de um códon seletor introduzido em uma posição não essencial de um marcador negativo remove as sintetases com as especificidades do aminoácido natural. Os sobreviventes da seleção negativa e positiva co-codificam as sintetases que aminoacilam (carregam) o tRNA supressor ortogonal apenas com aminoácidos não naturais. Estas sintetases depois podem ser submetidas à mutagênese adicional, por exemplo, embaralhamento do DNA ou outros métodos de mutagênese recursivos.

A biblioteca de O-RSs mutantes pode ser gerada usando várias técnicas de mutagênese conhecidas no ramo. Por exemplo, as RSs mutantes podem ser geradas por mutações sítio-específicas, mutações de ponto aleatório, recombinação homóloga, embaralhamento do DNA ou outros métodos de mutagênese recursivos, construção quimérica ou qualquer combinação dos mesmos. Por exemplo, uma biblioteca de RSs mutantes pode ser produzida a partir de duas ou mais outras, por exemplo, "sub-bibliotecas" menores e menos diversas. Bibliotecas quiméricas de RSs também são incluídas na invenção. Deve ser observado que as bibliotecas de tRNA sintetases a partir de vários organismos (por exemplo, micro-organismos, tais como eubactérias ou archaeobactérias), tais como bibliotecas que compreendem diversidade natural (veja, por exemplo, Patente U.S. Nº 6.238.884 de Short *et al.*; Patente U.S. Nº 5.756.316 de Schallenger *et al.*; Patente U.S. Nº 5.783.431 de Petersen *et al.*; Patente U.S. Nº 5.824.485 de Thompson *et al.*; Patente U.S. Nº 5.958.672 de

Short *et al.*), são opcionalmente construídas e triadas para pares ortogonais.

Uma vez que as sintetases são submetidas à estratégia de seleção/triagem positiva e negativa, estas sintetases depois podem ser submetidas à mutagênese adicional. Por exemplo, um ácido nucleico que codifica a O-RS pode ser isolado; um conjunto de polinucleotídeos que codificam as O-RSs modificadas (por exemplo, por mutagênese aleatória, mutagênese sítio-específica, recombinação ou qualquer combinação das mesmas) pode ser gerado a partir do ácido nucleico; e, estas etapas individuais ou uma combinação destas etapas podem ser repetidas até que uma O-RS modificada seja obtida, a qual preferencialmente aminoacila o O-tRNA com o aminoácido não natural. Em um aspecto da presente invenção, as etapas são realizadas diversas vezes, por exemplo, pelo menos duas vezes.

Níveis adicionais de estringência de seleção/triagem também podem ser usados nos métodos da presente invenção para produzir O-tRNA, O-RS ou pares dos mesmos. A estringência de seleção ou triagem pode ser variada em uma ou ambas as etapas do método para produzir uma O-RS. Isto pode incluir, por exemplo, variar a quantidade do agente de seleção/triagem que é usado, etc. Rodadas adicionais de seleções positivas e/ou negativas também podem ser realizadas. A seleção ou triagem também pode compreender uma ou mais seleções ou triagens positivas ou negativas que incluem, por exemplo, uma mudança na permeabilidade do aminoácido, uma mudança na eficácia da tradução, uma mudança na fidelidade de tradução, etc. Tipicamente, uma ou mais mudanças se baseiam em uma mutação em um ou mais genes em um organismo, em que um par de tRNA-tRNA sintetase ortogonal é usado para produzir a proteína.

Outros tipos de seleções podem ser usados na presente invenção, por exemplo, para O-RS, O-tRNA e par de O-tRNA/O-RS. O marcador de seleção positiva pode ser qualquer um entre uma variedade de moléculas, incluindo, mas não limitado a um produto que fornece um suplemento nutricional para o crescimento, e a seleção é realizada em um meio que é desprovido do suplemento nutricional. Exemplos de polinucleotídeos que codificam marcadores de seleção positiva incluem, mas não são limitados a, por exemplo, um gene repórter com base na complementação da auxotrofia de aminoácido de uma célula, um gene *his3* (por exemplo, onde o gene *his3* codifica uma imidazol glicerol fosfato desidratase, detectada fornecendo-se 3-aminotriazol (3-AT)), gene *ura3*, gene *leu2*, gene *lys2*, gene *lacZ*, gene *adh*, etc. Veja, por exemplo, G. M. Kishore, & D. M. Shah, (1988), Amino acid biosynthesis inhibitors as herbicides, *Annual Review of Biochemistry* 57:627-663. Em uma modalidade, a produção de *lacZ* é detectada pela hidrólise de orto-nitrofenil- β -D-galactopiranosídeo (ONPG). Veja, por exemplo, I. G. Serebriiskii, & E. A. Golemis, (2000). Uses of *lacZ* to study gene function: evaluation of beta-galactosidase assays employed in the yeast two-hybrid system, *Analytical Biochemistry* 285:1-15. Marcadores de seleção positiva adicionais incluem, por exemplo, luciferase, proteína verde fluorescente (GFP), YFP,

EGFP, RFP, o produto de um gene resistente a antibiótico (por exemplo, cloranfenicol acetiltransferase (CAT)), uma proteína moduladora transcricional (por exemplo, GAL4), etc. Opcionalmente, um polinucleotídeo que codifica um marcador de seleção positiva compreende um códon seletor.

5 Um polinucleotídeo que codifica o marcador de seleção positiva pode ser, de maneira operável, ligado a um elemento de resposta. Um polinucleotídeo adicional que codifica uma proteína moduladora transcricional que modula a transcrição a partir do elemento de resposta, e compreende pelo menos um códon seletor, também pode estar presente. A incorporação do aminoácido não natural na proteína moduladora transcricional pelo O-tRNA
10 aminoacilado com o aminoácido não natural resulta na transcrição do polinucleotídeo (por exemplo, gene repórter) que codifica o marcador de seleção positiva. Opcionalmente, o códon seletor está localizado ou está substancialmente próximo a uma porção do polinucleotídeo que codifica um domínio de ligação de DNA da proteína moduladora transcricional.

Um polinucleotídeo que codifica o marcador de seleção negativa também pode ser,
15 de maneira operável, ligado a um elemento de resposta a partir do qual a transcrição é mediada pela proteína moduladora transcricional. Veja, por exemplo, A. J. DeMaggio, *et al.*, (2000), The yeast split-hybrid system, *Method Enzymol.* 328: 128-137; H. M. Shih, *et al.*, (1996), A positive genetic selection for disrupting protein-protein interactions: identification of CREB mutations that prevent association with the coactivator CBP, *Proc. Natl. Acad. Sci. U.S.A.* 93:13896-13901; M. Vidal, *et al.*, (1996), Genetic characterization of a mammalian protein-protein interaction domain by using a yeast reverse two-hybrid system, *Proc. Natl. Acad. Sci. U.S.A.* 93:10321-10326; e, M. Vidal, *et al.*, (1996), Reverse two-hybrid and one-hybrid systems to detect dissociation of protein-protein and DNA-protein interactions (*Proc. Natl. Acad. Sci. U.S.A.* 93: 10315-10320). A incorporação de um aminoácido natural na pro-
25 teína moduladora transcricional pelo O-tRNA aminoacilado com um aminoácido natural resulta na transcrição do marcador de seleção negativa. Opcionalmente, o marcador de seleção negativa compreende um códon seletor. O marcador de seleção positiva e/ou marcador de seleção negativa da invenção podem compreender pelo menos dois códons seletores, os quais podem compreender pelo menos dois códons seletores diferentes ou pelo menos dois
30 dos mesmos códons seletores.

A proteína moduladora transcricional é uma molécula que se liga (direta ou indiretamente) a uma sequência de ácido nucleico (por exemplo, uma elemento de resposta) e modula a transcrição de uma sequência que é, de maneira operável, ligada ao elemento de resposta. Uma proteína moduladora transcricional pode ser uma proteína ativadora transcri-
35 cional (por exemplo, GAL4, receptores nucleares de hormônio, AP1, CREB, membros da família LEF/tcf, SMADs, VP 16, SPI, etc.), uma proteína transcricional repressora (por exemplo, receptores nucleares de hormônio, família Groucho/tle, família Engrailed, etc), ou

uma proteína que pode ter ambas as atividades dependendo do meio (por exemplo, LEF/tcf, proteínas homobox, etc.). Um elemento de resposta é tipicamente uma sequência de ácido nucleico que é reconhecida pela proteína moduladora transcricional ou um agente adicional que atua em combinação com a proteína moduladora transcricional.

5 Um outro exemplo de uma proteína moduladora transcricional é a proteína ativadora transcricional, GAL4. Veja, por exemplo, A. Laughon, *et al.*, (1984), Identification of two proteins encoded by the *Saccharomyces cerevisiae* GAL4 gene, *Molecular & Cellular Biology* 4:268-275; A. Laughon, & R. F. Gesteland, (1984), Primary structure of the *Saccharomyces cerevisiae* GAL4 gene, *Molecular & Cellular Biology* 4:260-267; L. Keegan, *et al.*,
10 (1986), Separation of DNA binding from the transcription-activating function of a eukariotic regulatory protein, *Science* 231:699-704; e, M. Ptashne, (1988), How eukariotic transcriptional activators work, *Nature* 335:683-689. Os 147 aminoácidos N-terminais desta proteína de 881 aminoácidos formam um domínio de ligação de DNA (DBD) que se liga especificamente à sequência de DNA. Veja, por exemplo, M. Carey, *et al.*, (1989), An amino-terminal
15 fragment of GAL4 binds DNA as a dimer, *J. Mol. Biol.* 209:423-432; e, E. Giniger, *et al.*, (1985), Specific DNA binding of GAL4, a positive regulatory protein of yeast, *Cell* 40:767-774. O DBD é ligado, por uma sequência de proteína de intervenção, a um domínio de ativação (AD) de 113 aminoácidos C-terminais que pode ativar a transcrição quando ligado ao DNA. Veja, por exemplo, J. Ma, & M. Ptashne, (1987), Deletion analysis of GAL4 defines two
20 transcriptional activating segments, *Cell* 48:847-853; e, J. Ma, & M. Ptashne, (1987), The carboxy-terminal 30 amino acids of GAL4 are recognized by GAL80, *Cell* 50:137-142. Colocando-se códons âmbar, por exemplo, em direção ao DBD N-terminal de um único polipeptídeo que contém tanto o DBD N-terminal de GAL4 e seu AD C-terminal, a supressão âmbar pelo par O-tRNA/O-RS pode ser ligada à ativação transcricional por GAL4. Genes repórteres
25 ativados por GAL4 podem ser usados para realizar seleções positivas e negativas com o gene.

O meio usado para seleção negativa pode compreender um agente de seleção ou triagem que é convertido em uma substância detectável pelo marcador de seleção negativa. Em um aspecto da invenção, a substância detectável é uma substância tóxica. Um polinucleotídeo que codifica um marcador de seleção negativa pode ser, por exemplo, um gene
30 *ura3*. Por exemplo, o repórter *URA3* pode ser colocado sob controle de um promotor que contém sítios de ligação GAL4 DNA. Quando o marcador de seleção negativa é produzido, por exemplo, por tradução de um polinucleotídeo que codifica a GAL4 com códons seletores, GAL4 ativa a transcrição de *URA3*. A seleção negativa é realizada em um meio que
35 compreende ácido 5-fluoroorótico (5-FOA), que é convertido em uma substância detectável (por exemplo, uma substância tóxica que mata a célula) pelo produto de gene do gene *ura3*. Veja, por exemplo, J. D. Boeke, *et al.*, (1984), A positive selection for mutants lacking

orotidine-5'-phosphate decarboxylase activity in yeast: 5-fluoroorotic acid resistance, *Molecular & General Genetics* 197:345-346); M. Vidal, *et al.*, (1996), Genetic characterization of a mammalian protein-protein interaction domain by using a yeast reverse two-hybrid system, *Proc. Natl. Acad. Sci. U.S.A.* 93:10321-10326; e, M. Vidal, *et al.*, (1996), Reverse two-hybrid and one-hybrid systems to detect dissociation of protein-protein and DNA-protein interactions, *Proc. Natl. Acad. Sci. U.S.A.* 93:10315-10320.

Tal como o marcador de seleção positiva, o marcador de seleção negativa também pode ser qualquer um entre uma variedade de moléculas. O marcador de seleção positiva e/ou o marcador de seleção negativa podem ser um polipeptídeo que fluoresce ou catalisa uma reação luminescente na presença de um reagente adequado. Por exemplo, marcadores de seleção negativa incluem, mas não são limitados a, por exemplo, luciferase, proteína verde fluorescente (GFP), YFP, EGFP, RFP, o produto de um gene resistente ao antibiótico (por exemplo, cloranfenicol acetiltransferase (CAT)), o produto de um gene lacZ, proteína moduladora transcricional, etc. O marcador de seleção positiva e/ou o marcador de seleção negativa podem ser detectados por separador celular ativado por fluorescência (FACS) ou por luminescência. O marcador de seleção positiva e/ou marcador de seleção negativa podem compreender um marcador de triagem com base na afinidade. O mesmo polinucleotídeo pode codificar tanto o marcador de seleção positiva quanto o marcador de seleção negativa. Por exemplo, a etapa de seleção positiva, a etapa de seleção negativa ou tanto a etapa de seleção positiva quanto negativa podem incluir usar um repórter, em que o repórter é detectado pelo separador celular ativado por fluorescência (FACS). Por exemplo, uma seleção positiva pode ser feita primeiro com um marcador de seleção positiva, por exemplo, gene cloranfenicol acetiltransferase (CAT), onde o gene CAT compreende um códon seletor, por exemplo, um códon de parada âmbar, no gene CAT, que, seguido por uma triagem de seleção negativa, que é baseada na incapacidade de suprimir um códon seletor, por exemplo, dois ou mais, em posições dentro de um marcador negativo, por exemplo, gene T7 RNA polimerase. O marcador de seleção positiva e o marcador de seleção negativa podem ser encontrados no mesmo vetor, por exemplo, plasmídeo. A expressão do marcador negativo direciona a expressão do repórter, por exemplo, proteína verde fluorescente (GFP). A estratêgia da seleção e triagem pode ser variada, por exemplo, a intensidade da luz necessária para fluorescência do repórter pode ser variada. Uma seleção positiva pode ser feita com um repórter como um marcador de seleção positiva, que é triado por FACS, seguido por uma triagem de seleção negativa, que é baseada na incapacidade de suprimir um códon seletor, por exemplo, dois ou mais, em posições dentro de um marcador negativo, por exemplo, gene barnase.

Opcionalmente, o repórter é exibido em uma superfície celular, por exemplo, em uma exibição de fago ou semelhantes. A exibição de superfície celular, por exemplo, o sis-

tema de exibição de superfície celular com base em OmpA, depende da expressão de um epítipo particular, por exemplo, um peptídeo C3 de poliovírus fundido a uma porina OmpA de membrana externa, na superfície da célula de *Escherichia coli*. O epítipo é exibido na superfície celular apenas quando um códon seletor na mensagem da proteína é suprimido durante a tradução. O peptídeo exibido então contém o aminoácido reconhecido por uma das aminoacil-tRNA sintetases mutantes na biblioteca, e a célula contendo o gene sintetase correspondente pode ser isolada com anticorpos recrutados contra peptídeos contendo aminoácidos não naturais específicos. O sistema de exibição de superfície celular com base em OmpA foi desenvolvido e otimizado por Georgiou *et al.* como uma alternativa para a exibição do fago. Veja, Francisco, J. A., Campbell, R., Iverson, B. L., & Georgiou, G. Production and fluorescence-activated cell sorting of *Escherichia coli* expressing a functional antibody fragment on the external surface. *Proc. Natl. Acad. Sci. U.S.A.* 90:10444-8 (1993).

Outras modalidades da presente invenção incluem conduzir uma ou mais entre as etapas de seleção *in vitro*. O componente selecionado, por exemplo, sintetase e/ou tRNA, depois pode ser introduzido em uma célula para o uso na incorporação *in vivo* de um aminoácido não natural.

Detalhes adicionais para produzir O-RS e alterar a especificidade de substrato da sintetase podem ser encontrados no Pedido de Patente U.S. 10/126.931 intitulado "Methods and Compositions for the Production of Orthogonal tRNA-Aminoacyl tRNA Synthetase Pairs;" e, USSN 10/825.867 intitulado "EXPANDING THE EUKARIOTIC GENETIC CODE", os quais são incorporados como referência neste relatório. Detalhes adicionais para produzir O-RS podem ser encontrados em Hamano-Takaku *et al.*, (2000) A mutant *Escherichia coli* Tyrosyl-tRNA Synthetase Utilizes the Unnatural Amino Acid Azatyrosine More Efficiently than Tyrosine, *Journal of Biological Chemistry*, 275(51):40324-40328; Kiga *et al.* (2002), An engineered *Escherichia coli* tyrosyl-tRNA synthetase for site-specific incorporation of an unnatural amino acid into proteins in eukariotic translation and its application in a wheat germ cell-free system, *PNAS* 99(15): 9715-9723; e, Francklyn *et al.*, (2002), Aminoacyl-tRNA synthetases: Versatile players in the changing theater of translation; *RNA*, 8:1363-1372, cada um dos quais é incorporado como referência neste relatório.

FONTE E ORGANISMOS HOSPEDEIROS

Os componentes de tradução da presente invenção são tipicamente derivados de organismos não eucarióticos. Por exemplo, o O-tRNA ortogonal pode ser derivado de um organismo não eucariótico, por exemplo, uma archaeobactéria, tal como *Methanococcus jannaschii*, *Methanobacterium thermoautotrophicum*, *Halobacterium*, tal como *Haloferax volcanii* e *Halobacterium* espécie NRC-1, *Archaeoglobus fulgidus*, *Pyrococcus furiosus*, *Pyrococcus horikoshii*, *Aeuropyrum pernix* ou semelhantes, ou uma eubactéria, tal como *Escherichia coli*, *Thermus thermophilus*, *Bacillus stearothermophilus* ou semelhantes, enquanto a O-RS

ortogonal pode ser derivada de um organismo não eucariótico, por exemplo, *Methanobacterium thermoautotrophicum*, *Halobacterium*, tal como *Haloferax volcanii* e *Halobacterium* espécie *NRC-1*, *Archaeoglobus fulgidus*, *Pyrococcus furiosus*, *Pyrococcus horikoshii*, *Aeuropyrum pernix* ou semelhantes, ou uma eubactéria, tal como *Escherichia coli*, *Thermus thermophilus*, *Bacillus stearothermophilus* ou semelhantes. Em uma modalidade, fontes eucarióticas também podem ser usadas, incluindo, mas não limitadas às plantas, algas, protistas, fungos, leveduras, animais (por exemplo, mamíferos, insetos, artrópodes, etc.) ou semelhantes.

Os componentes individuais de um par de O-tRNA/O-RS podem ser derivados do mesmo organismo ou de organismos diferentes. Em uma modalidade, o par de O-tRNA/O-RS é derivado do mesmo organismo. Alternativamente, o O-tRNA e a O-RS do par de O-tRNA/O-RS são derivados de organismos diferentes. Por exemplo, o O-tRNA pode ser derivado, por exemplo, de *Halobacterium sp NRC-1*, e a O-RS pode ser derivada, por exemplo, de *Methanobacterium thermoautotrophicum*.

O O-tRNA, O-RS ou par de O-tRNA/O-RS podem ser selecionados ou triados *in vivo* ou *in vitro* e/ou usados em uma célula, por exemplo, uma célula não eucariótica (tal como célula de *E. coli*), ou uma célula eucariótica, para produzir um polipeptídeo com um aminoácido selecionado (por exemplo, um aminoácido não natural). Uma célula não eucariótica pode ser derivada de uma variedade de fontes, tais como o domínio filogenético Archaea, incluindo, mas não limitado a *Methanococcus jannaschii*, *Methanobacterium thermoautotrophicum*, *Halobacterium*, tal como *Haloferax volcanii* e *Halobacterium* espécie *NRC-1*, *Archaeoglobus fulgidus*, *Pyrococcus furiosus*, *Pyrococcus horikoshii*, *Aeuropyrum pernix* ou semelhantes, ou pode pertencer ao domínio filogenético Eubacteria (incluindo, mas não limitado a *Escherichia coli*, *Thermus thermophilus*, *Bacillus stearothermophilus*, *Pseudomonas fluorescens*, *Pseudomonas aeruginosa*, *Pseudomonas putida*, etc.) ou semelhantes. Uma célula eucariótica pode ser derivada de uma variedade de fontes, incluindo, mas não limitada a uma planta (por exemplo, planta complexa, tal como monocotiledôneas ou dicotiledôneas), alga, protista, fungo, levedura (incluindo, mas não limitada a *Saccharomyces cerevisiae*), animal (incluindo, mas não limitado a um mamífero, inseto, artrópode, etc.) ou semelhantes. Composições de células com componentes de tradução da presente invenção também são uma característica da presente invenção. Veja também USSN 10/825.867 intitulado "Expanding the Eukariotic Genetic Code" para triagem de O-tRNA e/ou O-RS em uma espécie para o uso em uma outra espécie.

Para expressar um polipeptídeo de interesse com um aminoácido selecionado em uma célula hospedeira, uma pessoa habilitada pode subclonar polinucleotídeos que codificam um polipeptídeo de interesse em um vetor de expressão que contém um promotor para direcionar a transcrição, um terminador de transcrição/tradução e, se for um ácido nucleico

que codifica uma proteína, um sítio de ligação de ribossomo para iniciação de tradução. Promotores bacterianos adequados são bem conhecidos na técnica e descritos, por exemplo, em Sambrook *et al.* e Ausubel *et al.*

Sistemas de expressão bacterianos para expressar um polipeptídeo de interesse estão disponíveis em, incluindo, mas não limitados a *E. coli*, *Bacillus sp.*, *Pseudomonas fluorescens*, *Pseudomonas aeruginosa*, *Pseudomonas putida* e *Salmonella* (Palva *et al.*, Gene 22:229-235 (1983); Mosbach *et al.*, Nature 302:543-545 (1983)). Kits para tais sistemas de expressão são comercialmente disponíveis. Sistemas de expressão eucarióticos para células de mamíferos, leveduras e insetos são bem conhecidos na técnica e também são comercialmente disponíveis.

Um tRNA e/ou RS da presente invenção e/ou um polipeptídeo de interesse podem ser utilizados e/ou expressados em qualquer número de sistemas de expressão adequados, incluindo, por exemplo, levedura, células de insetos, células de mamíferos e bactérias. Uma descrição de sistemas de expressão exemplares é fornecida abaixo.

Levedura Conforme usado neste relatório, o termo “levedura” inclui qualquer uma entre as várias leveduras capazes de expressar um polipeptídeo de interesse. Tais leveduras incluem, mas não são limitadas a leveduras ascospórogenas (*Endomycetales*), leveduras basidiospórogenas e leveduras pertencentes ao grupo Fungi imperfecti (*Blastomycetes*). As leveduras ascospórogenas são divididas em duas famílias, *Spermophthoraceae* e *Saccharomycetaceae*. A última é compreendida de quatro subfamílias, *Schizosaccharomycoidae* (por exemplo, gênero *Schizosaccharomyces*), *Nadsonioideae*, *Lipomycoideae* e *Saccharomycoideae* (por exemplo, gêneros *Pichia*, *Kluyveromyces* e *Saccharomyces*). As leveduras basidiospórogenas incluem os gêneros *Leucosporidium*, *Rhodosporeidium*, *Sporidiobolus*, *Filobasidium* e *Filobosidiella*. Leveduras pertencentes ao grupo Fungi Imperfecti (*Blastomycetes*) são divididas em duas famílias, *Sporobolomycetaceae* (por exemplo, gêneros *Sporobolomyces* e *Bullera*) e *Cryptococcaceae* (por exemplo, gênero *Candida*).

De particular interesse para o uso com a presente invenção são as espécies dos gêneros *Pichia*, *Kluyveromyces*, *Saccharomyces*, *Schizosaccharomyces*, *Hansenula*, *Torulopsis* e *Candida*, incluindo, mas não limitadas a *P. pastoris*, *P. guillerimondii*, *S. cerevisiae*, *S. carlsbergensis*, *S. diastaticus*, *S. douglasii*, *S. kluyveri*, *S. norbensis*, *S. oviformis*, *K. lactis*, *K. fragilis*, *C. albicans*, *C. maltosa* e *H. polymorpha*. Leveduras são geralmente disponíveis a partir de uma variedade de fontes, incluindo, mas não limitadas ao Yeast Genetic Stock Center, Department of Biophysics and Medical Physics, University of California (Berkeley, CA), e à American Type Culture Collection (“ATCC”) (Manassas, VA).

O termo “hospedeiro de levedura” ou “célula hospedeira de levedura” inclui a levedura que pode ser, ou foi, usada como um recipiente para vetores recombinantes ou outro DNA de transferência. O termo inclui a progênie da célula hospedeira de levedura original

que recebeu os vetores recombinantes ou outro DNA de transferência. Deve ser entendido que a progênie de uma célula parental única pode não ser necessária e completamente idêntica em morfologia ou em complemento de DNA genômico ou total ao precursor original, devido à mutação accidental ou deliberada. A progênie da célula parental que é suficientemente similar ao precursor a ser caracterizado pela propriedade relevante, tal como a presença de uma sequência de nucleotídeos que codifica um polipeptídeo de interesse, é incluída na progênie intencionada por esta definição.

A expressão e vetores de transformação, incluindo réplicons extracromossômicos ou vetores de integração, foram desenvolvidos para transformação em diversos hospedeiros de levedura. Por exemplo, vetores de expressão foram desenvolvidos para *S. cerevisiae* (Sikorski *et al.*, GENETIC (1989) 122:19; Ito *et al.*, J. BACTERIOL. (1983) 153:163; Hinnen *et al.*, PROC. NATL. ACAD. SCI. USA (1978) 75:1929); *C. albicans* (Kurtz *et al.*, MOL. CELL. BIOL. (1986) 6: 142); *C. maltosa* (Kunze *et al.*, J. BASIC MICROBIOL. (1985) 25: 141); *H. polymorpha* (Gleeson *et al.*, J. GEN. MICROBIOL. (1986) 132:3459; Roggenkamp *et al.*, MOL. GENETIC AND GENOMICS (1986) 202:302); *K. fragilis* (Das *et al.*, J. BACTERIOL. (1984) 158:1165); *K. lactis* (De Louvencourt *et al.*, J. BACTERIOL. (1983) 154:737; Van den Berg *et al.*, BIOTECHNOLOGY (NY) (1990) 8:135); *P. guillerimondii* (Kunze *et al.*, J. BASIC MICROBIOL. (1985) 25:141); *P. pastoris* (Patentes U.S. N^{os} 5.324.639; 4.929.555; e 4.837.148; Cregg *et al.*, MOL. CELL. BIOL. (1985) 5:3376); *Schizosaccharomyces pombe* (Beach *et al.*, NATURE (1982) 300:706); e *Y. lipolytica*; *A. nidulans* (Ballance *et al.*, BIOCHEM. BIOPHYS. RES. COMMUN. (1983) 112:284-89; Tilburn *et al.*, GENE (1983) 26:205-221; e Yelton *et al.*, PROC. NATL. ACAD. SCI. USA (1984) 81:1470-74); *A. niger* (Kelly e Hynes, EMBO J. (1985) 4:475-479); *T. reesia* (EP 0 244 234); e fungos filamentosos, tais como, por exemplo, *Neurospora*, *Penicillium*, *Tolypocladium* (WO 91/00357), incorporados como referência neste relatório.

Sequências controle para vetores de levedura são conhecidas àqueles de habilidade comum na técnica e incluem, mas não são limitadas a regiões promotoras a partir de genes, tais como álcool desidrogenase (ADH) (EP 0 284 044); enolase; glicoquinase; glicose-6-fosfato isomerase; gliceraldeído-3-fosfato-desidrogenase (GAP ou GAPDH); hexoquinase; fosfofrutoquinase; 3-fosfoglicerato mutase; e piruvato quinase (PyK) (EP 0 329 203). O gene PHO5 de levedura, que codifica ácido fosfatase, também pode fornecer sequências promotoras úteis (Miyanojara *et al.*, PROC. NATL. ACAD. SCI. USA (1983) 80:1). Outras sequências promotoras adequadas para o uso com hospedeiros de levedura podem incluir os promotores para 3-fosfoglicerato quinase (Hitzeman *et al.*, J. BIOL. CHEM. (1980) 255:12073); e outras enzimas glicolíticas, tais como piruvato decarboxilase, triosefosfato isomerase e fosfoglicose isomerase (Holland *et al.* BIOCHEMISTRY (1978) 17:4900; Hess *et al.*, J. ADV. ENZYME REG. (1969) 7:149). Promotores de levedura induzíveis tendo a

vantagem adicional de transcrição controlada por condições de crescimento podem incluir as regiões promotoras para álcool desidrogenase 2; isocitocromo C; ácido fosfatase; metalotioneína; gliceraldeído-3-fosfato desidrogenase; enzimas degradantes associadas ao metabolismo de nitrogênio; e enzimas responsáveis pela utilização de maltose e galactose. Vetores e promotores adequados para o uso na expressão da levedura são ainda descritos no EP 0 073 657.

Acentuadores de levedura também podem ser usados com promotores de levedura. Além disso, promotores sintéticos também podem funcionar como promotores de levedura. Por exemplo, as sequências de ativação a montante (UAS) de um promotor de levedura podem ser unidas com a região de ativação de transcrição de um outro promotor de levedura, criando um promotor híbrido sintético. Exemplos de tais promotores híbridos incluem a sequência regulatória de ADH ligada à região de ativação de transcrição de GAP. Veja Patentes U.S. Nºs 4.880.734 e 4.876.197, que são incorporadas como referência neste relatório. Outros exemplos de promotores híbridos incluem promotores que consistem das sequências regulatórias dos genes ADH2, GAL4, GAL10 ou PHO5, combinadas com a região de ativação transcricional de um gene de enzima glicolítica, tal como GAP ou PyK. Veja EP 0 164 556. Além disso, um promotor de levedura pode incluir promotores que ocorrem naturalmente de origem que não levedura que têm a capacidade de se ligar à RNA polimerase de levedura e iniciar a transcrição.

Outros elementos controle que podem compreender parte dos vetores de expressão de levedura incluem terminadores, por exemplo, a partir de GAPDH ou dos genes enolase (Holland *et al.*, J. BIOL. CHEM. (1981) 256:1385). Além disso, a origem de replicação a partir da origem de plasmídeo 2 μ é adequada para levedura. Um gene de seleção adequado para o uso em levedura é o gene *trp1* presente no plasmídeo de levedura. Veja Tschumper *et al.*, GENE (1980) 10:157; Kingsman *et al.*, GENE (1979) 7:141. O gene *trp1* fornece um marcador de seleção para uma cepa mutante de levedura desprovida da capacidade de crescimento em triptofano. Similarmente, cepas de levedura deficientes de Leu2 (ATCC 20.622 ou 38.626) são complementadas por plasmídeos conhecidos que conduzem o gene *Leu2*.

Métodos de introduzir DNA exógeno em hospedeiros de levedura são conhecidos àqueles de habilidade comum na técnica e, tipicamente, incluem, mas não são limitados à transformação de esferoplastos ou de células hospedeiras de levedura intactas tratadas com cátions alcalinos. Por exemplo, a transformação de levedura pode ser realizada de acordo com o método descrito em Hsiao *et al.*, PROC. NATL. ACAD. SCI. USA (1979) 76:3829 e Van Solingen *et al.*, J. BACT. (1977) 130:946. Entretanto, outros métodos para introduzir DNA em células, tais como por injeção nuclear, eletroporação ou fusão de protoplastos, também podem ser usados, conforme descrito, no geral, em SAMBROOK *ET AL.*,

MOLECULAR CLONING: A LAB. MANUAL (2001). As células hospedeiras de levedura depois podem ser cultivadas usando técnicas padrão conhecidas àqueles de habilidade comum no ramo.

Outros métodos para expressar proteínas heterólogas em células hospedeiras de levedura são conhecidos àqueles de habilidade comum na técnica. Veja, no geral, Publicação de Patente U.S. Nº 20020055169, Patentes U.S. Nºs 6.361.969; 6.312.923; 6.183.985; 6.083.723; 6.017.731; 5.674.706; 5.629.203; 5.602.034; e 5.089.398; Patentes Reexaminadas U.S. Nºs RE37.343 e RE35.749; Pedidos de Patente Publicados PCT WO 99/07862; WO 98/37208; e WO 98/26080; Pedidos de Patente Europeus EP 0 946 736; EP 0 732 403; EP 0 480 480; WO 90/10277; EP 0 340 986; EP 0 329 203; EP 0 324 274; e EP 0 164 556. Veja também Gellissen *et al.*, ANTONIE VAN LEEUWENHOEK (1992) 62(1-2): 79-93; Romanos *et al.*, Yeast (1992) 8(6):423-488; Goeddel, METHODS IN ENZYMOLOGY (1990) 185:3-7, incorporados como referência neste relatório.

As cepas de hospedeiro de levedura podem ser cultivadas em fermentadores durante o estágio de amplificação usando métodos de fermentação em lote de alimentação padrão conhecidos àqueles de habilidade comum na técnica. Os métodos de fermentação podem ser adaptados para incluir diferenças em uma via de utilização de carbono de hospedeiro de levedura particular ou modo de controle de expressão. Por exemplo, a fermentação de um hospedeiro de levedura *Saccharomyces* pode exigir uma única alimentação de glicose, fonte de nitrogênio complexa (por exemplo, hidrolisados de caseína), e múltipla suplementação de vitaminas. Ao contrário, a levedura *P. pastoris* metilotrófica pode exigir glicerol, metanol e alimentações de mineral traço, porém apenas sais de amônio simples (nitrogênio) para crescimento e expressão ideais. Veja, por exemplo, Patente U.S. Nº 5.324.639; Elliott *et al.*, J. PROTEIN CHEM. (1990) 9:95; e Fieschko *et al.*, BIOTECH. BIOENG. (1987) 29:1113, incorporados como referência neste relatório.

Tais métodos de fermentação, entretanto, podem ter certas características comuns independentes da cepa de hospedeiro de levedura utilizada. Por exemplo, um nutriente limitante de crescimento, tipicamente carbono, pode ser adicionado ao fermentador durante a fase de amplificação para permitir crescimento máximo. Além disso, métodos de fermentação geralmente utilizam um meio de fermentação designado para conter quantidades adequadas de carbono, nitrogênio, sais basais, fósforo e outros nutrientes secundários (vitaminas, minerais traço e sais, etc.). Exemplos de meios de fermentação adequados para o uso com *Pichia* são descritos nas Patentes U.S. Nºs 5.324.639 e 5.231.178, que são incorporadas como referência neste relatório.

Células de Inseto Infectadas por Baculovírus O termo “hospedeiro de inseto” ou “célula hospedeira de inseto” refere-se a um inseto que pode ser, ou foi, usado como um recipiente para vetores recombinantes ou outro DNA de transferência. O termo inclui a progênie

da célula hospedeira de inseto original que foi transfectada. Deve ser entendido que a progênie de uma célula parental única pode não ser necessária e completamente idêntica em morfologia ou em complemento de DNA genômico ou total ao precursor original, devido à mutação accidental ou deliberada. A progênie da célula parental que é suficientemente similar ao precursor a ser caracterizado pela propriedade relevante, tal como a presença de uma sequência de nucleotídeos que codifica um polipeptídeo de interesse, é incluída na progênie intencionada por esta definição.

A seleção de células de inseto adequadas para expressão de um polipeptídeo de interesse é conhecida àqueles de habilidade comum na técnica. Várias espécies de inseto são bem descritas na técnica e são comercialmente disponíveis, incluindo *Aedes aegypti*, *Bombyx mori*, *Drosophila melanogaster*, *Spodoptera frugiperda* e *Trichoplusia ni*. Na seleção de hospedeiros de inseto para expressão, hospedeiros adequados podem incluir aqueles mostrados para ter, *inter alia*, capacidade de secreção satisfatória, baixa atividade proteolítica, e robustez global. Insetos são geralmente disponíveis a partir de uma variedade de fontes, incluindo, mas não limitadas ao Insect Genetic Stock Center, Department of Biophysics and Medical Physics, University of California (Berkeley, CA); e à American Type Culture Collection ("ATCC") (Manassas, VA).

Geralmente, os componentes de um sistema de expressão de inseto infectado por baculovírus incluem um vetor de transferência, usualmente um plasmídeo bacteriano, que contém um fragmento do genoma de baculovírus e um sítio de restrição conveniente para a inserção do gene heterólogo a ser expressado; um baculovírus do tipo selvagem com sequências homólogas ao fragmento específico do baculovírus no vetor de transferência (isto permite a recombinação homóloga do gene heterólogo no genoma de baculovírus); e células hospedeiras de inseto e meios de crescimento apropriados. Os materiais, métodos e técnicas usados em vetores de construção, células de transfecção, placas de coleta, células de crescimento em cultura e semelhantes são conhecidos na técnica e manuais que descrevem estas técnicas são disponíveis.

Depois de inserir o gene heterólogo no vetor de transferência, o vetor e o genoma do tipo selvagem viral são transfectados em uma célula hospedeira de inseto onde o vetor e genoma viral se recombinam. O vírus recombinante acondicionado é expressado e placas recombinantes são identificadas e purificadas. Materiais e métodos para sistemas de expressão celular de baculovírus/insetos são comercialmente disponíveis na forma de kit, por exemplo, da Invitrogen Corp. (Carlsbad, CA). Estas técnicas são geralmente conhecidas àqueles de habilidade comum no ramo e completamente descritas em SUMMERS E SMITH, TEXAS AGRICULTURAL EXPERIMENT STATION BULLETIN Nº 1555 (1987), incorporado como referência neste relatório. Veja também, RICHARDSON, 39 METHODS IN MOLECULAR BIOLOGY: BACULOVIRUS EXPRESSION PROTOCOLS (1995); AUSUBEL

ET AL., CURRENT PROTOCOLS IN MOLECULAR BIOLOGY 16.9-16.11 (1994); KLING E POSSEE, THE BACULOVIRUS SYSTEM: A LABORATORY GUIDE (1992); e O'REILLY ET AL., BACULOVIRUS EXPRESSION VECTORS: A LABORATORY MANUAL (1992).

De fato, a produção de várias proteínas heterólogas usando sistemas de expressão celular de baculovírus/insetos é conhecida àqueles de habilidade comum na técnica. Veja, por exemplo, Patentes U.S. Nºs 6.368.825; 6.342.216; 6.338.846; 6.261.805; 6.245.528; 6.225.060; 6.183.987; 6.168.932; 6.126.944; 6.096.304; 6.013.433; 5.965.393; 5.939.285; 5.891.676; 5.871.986; 5.861.279; 5.858.368; 5.843.733; 5.762.939; 5.753.220; 5.605.827; 5.583.023; 5.571.709; 5.516.657; 5.290.686; WO 02/06305; WO 01/90390; WO 01/27301; WO 01/05956; WO 00/55345; WO 00/20032; WO 99/51721; WO 99/45130; WO 99/31257; WO 99/10515; WO 99/09193; WO 97/26332; WO 96/29400; WO 96/25496; WO 96/06161; WO 95/20672; WO 93/03173; WO 92/16619; WO 92/02628; WO 92/01801; WO 90/14428; WO 90/10078; WO 90/02566; WO 90/02186; WO 90/01556; WO 89/01038; WO 89/01037; WO 88/07082, os quais são incorporados como referência neste relatório.

Vetores que são úteis em sistema de expressão celular de baculovírus/insetos são conhecidos na técnica e incluem, por exemplo, vetores de expressão e transferência de insetos derivados do baculovírus vírus de poli-hedrose nuclear de Autographacalifornica (AcNPV), que é um vetor de expressão viral independente de auxiliares. Vetores de expressão viral derivados deste sistema usualmente usam o promotor do gene poliedrina viral forte para conduzir a expressão de genes heterólogos. Veja, no geral, O'Reilly ET AL., BACULOVIRUS EXPRESSION VECTORS: A LABORATORY MANUAL (1992).

Antes de inserir o gene estranho no genoma de baculovírus, os componentes acima descritos, compreendendo uma sequência promotora, líder (se desejado) e codificante de interesse, e sequência de terminação de transcrição, são tipicamente construídos em um constructo de transcolocação intermediário (vetor de transferência). Os constructos de transcolocação intermediários são frequentemente mantidos em um réplicon, tal como um elemento extracromossômico (por exemplo, plasmídeos) capaz de manutenção estável em um hospedeiro, tal como bactérias. O réplicon terá um sistema de replicação, permitindo assim que ele seja mantido em um hospedeiro adequado para clonagem e amplificação. Mais especificamente, o plasmídeo pode conter o sinal de poliadenilação de poliedrina (Miller, ANN. REV. MICROBIOL. (1988) 42:177) e um gene resistente à ampicilina procariótico (*amp*) e origem de replicação para seleção e propagação em *E. coli*.

Um vetor de transferência comumente usado para introduzir genes estranhos em AcNPV é pAc373. Muitos outros vetores, conhecidos àqueles de habilidade na técnica, também foram designados, incluindo, por exemplo, pVL985, que altera o códon de partida de poliedrina de ATG para ATT, e que introduz 32 pares de base no sítio de clonagem BamHI a jusante de ATT. Veja Luckow e Summers, VIROLOGY 170:31 (1989). Outros vetores co-

mercialmente disponíveis incluem, por exemplo, PBlueBac4.5/V5-His; pBlueBacHis2; pMel-Bac; pBlueBac4.5 (Invitrogen Corp., Carlsbad, CA).

Depois da inserção do gene heterólogo, o vetor de transferência e genoma baculoviral do tipo selvagem são cotransfectados em uma célula hospedeira de inseto. Métodos para introduzir DNA heterólogo no sítio desejado nos baculovírus são conhecidos na técnica. Veja SUMMERS E SMITH, TEXAS AGRICULTURAL EXPERIMENT STATION BULLETIN Nº 1555 (1987); Smith *et al.*, MOL. CELL. BIOL. (1983) 3:2156; Luckow e Summers, VIROLOGY (1989) 170:31. Por exemplo, a inserção pode ser em um gene, tal como o gene poliedrina, por recombinação cruzada dupla homóloga; a inserção também pode ser em um sítio de enzima de restrição construído no gene do baculovírus desejado. Veja Miller *et al.*, BIOESSAYS (1989) 11(4):91.

A transfecção pode ser realizada por eletroporação. Veja TROTTER E WOOD, 39 METHODS IN MOLECULAR BIOLOGY (1995); Mann e King, J. GEN. VIROL. (1989) 70:3501. Alternativamente, os lipossomas podem ser usados para transfectar as células de inseto com o vetor de expressão recombinante e o baculovírus. Veja, por exemplo, Liebman *et al.*, BIOTECHNIQUES (1999) 26(1):36; Graves *et al.*, BIOCHEMISTRY (1998) 37:6050; Nomura *et al.*, J. BIOL. CHEM. (1998) 273(22):13570; Schmidt *et al.*, PROTEIN EXPRESSION AND PURIFICATION (1998) 12:323; Siffert *et al.*, NATURE GENETICS (1998) 18:45; TILKINS ET AL., CELL BIOLOGY: A LABORATORY HANDBOOK 145-154 (1998); Cai *et al.*, PROTEIN EXPRESSION AND PURIFICATION (1997) 10:263; Dolphin *et al.*, NATURE GENETICS (1997) 17:491; Kost *et al.*, GENE (1997) 190:139; Jakobsson *et al.*, J. BIOL. CHEM. (1996) 271:22203; Rowles *et al.*, J. BIOL. CHEM. (1996) 271(37):22376; Reverey *et al.*, J. BIOL. CHEM. (1996) 271(39): 23607-10; Stanley *et al.*, J. BIOL. CHEM. (1995) 270:4121; Sisk *et al.*, J. VIROL. (1994) 68(2):766; e Peng *et al.*, BIOTECHNIQUES (1993) 14(2):274. Lipossomas comercialmente disponíveis incluem, por exemplo, Cellfectin® e Lipofectin® (Invitrogen, Corp., Carlsbad, CA). Além disso, a transfecção por fosfato de cálcio pode ser usada. Veja TROTTER E WOOD, 39 METHODS IN MOLECULAR BIOLOGY (1995); Kitts, NAR (1990) 18(19):5667; e Mann e King, J. GEN. VIROL. (1989) 70:3501.

Vetores de expressão de baculovírus usualmente contêm um promotor de baculovírus. Um promotor de baculovírus é qualquer sequência de DNA capaz de se ligar a uma RNA polimerase de baculovírus e iniciar a transcrição (3') a jusante de uma sequência codificante (por exemplo, gene estrutural) no mRNA. Um promotor terá uma região de iniciação de transcrição que é usualmente colocada próxima à extremidade 5' da sequência codificante. Esta região de iniciação de transcrição tipicamente inclui um sítio de ligação de RNA polimerase e um sítio de iniciação de transcrição. Um promotor de baculovírus também pode ter um segundo domínio chamado acentuador, que, se presente, é usualmente distal ao gene estrutural. Além disso, a expressão pode ser regulada ou constitutiva.

Genes estruturais, abundantemente transcritos tardiamente no ciclo de infecção, fornecem sequências promotoras particularmente úteis. Exemplos incluem sequências derivadas do gene que codifica a proteína poliedro viral (FRIESEN *ET AL.*, *The Regulation of Baculovirus Gene Expression* in THE MOLECULAR BIOLOGY OF BACULOVIRUS (1986); EP 0 127 839 e 0 155 476) e o gene que codifica a proteína p10 (Vlak *et al.*, J. GEN. VIROL. (1988) 69:765).

O vetor de expressão de baculovírus recém formado é acondicionado em um baculovírus recombinante infeccioso e placas subsequentemente cultivadas podem ser purificadas por técnicas conhecidas àqueles de habilidade comum no ramo. Veja Miller *et al.*, BIOESSAYS (1989) 11(4):91; SUMMERS E SMITH, TEXAS AGRICULTURAL EXPERIMENT STATION BULLETIN Nº 1555 (1987).

Vetores de expressão de baculovírus recombinantes foram desenvolvidos para infecção em várias células de insetos. Por exemplo, os baculovírus recombinantes foram desenvolvidos, *inter alia*, para *Aedes aegypti* (ATCC Nº CCL-125), *Bombyx mori* (ATCC Nº CRL-8910), *Drosophila melanogaster* (ATCC Nº 1963), *Spodoptera frugiperda* e *Trichoplusia ni*. Veja Wright, NATURE (1986) 321:718; Carbonell *et al.*, J. VIROL. (1985) 56:153; Smith *et al.*, MOL. CELL. BIOL. (1983) 3:2156. Veja, no geral, Fraser *et al.*, IN VITRO CELL. DEV. BIOL. (1989) 25:225. Mais especificamente, as linhagens celulares usadas para sistemas de vetor de expressão de baculovírus comumente incluem, mas não são limitadas a Sf9 (*Spodoptera frugiperda*) (ATCC Nº CRL-1711), Sf21 (*Spodoptera frugiperda*) (Invitrogen Corp., Cat. Nº 11497-013 (Carlsbad, CA)), Tri-368 (*Trichoplusia ni*) e High-Five™ BTI-TN-5B1-4 (*Trichoplusia ni*).

Células e meios de cultura são comercialmente disponíveis para expressão direta e fusão de polipeptídeos heterólogos em um baculovírus/expressão, e a tecnologia de cultura celular é geralmente conhecida àqueles de habilidade comum na técnica.

E. Coli, espécies de Pseudomonas e outros Procariontes Técnicas de expressão bacteriana são conhecidas àqueles de habilidade comum no ramo. Uma ampla variedade de vetores está disponível para o uso em hospedeiros bacterianos. Os vetores podem ser vetores de cópia única ou de baixa ou alta multicópia. Vetores podem servir para clonagem e/ou expressão. Tendo em vista a ampla literatura com respeito aos vetores, disponibilidade comercial de muitos vetores e ainda manuais descrevendo vetores e seus mapas e características de restrição, nenhum debate extensivo é exigido neste relatório. Como é bem conhecido, os vetores normalmente envolvem marcadores que permitem a seleção, em que tais marcadores podem fornecer resistência ao agente citotóxico, prototrofia ou imunidade. Frequentemente, uma pluralidade de marcadores está presente, a qual fornece características diferentes.

Um promotor bacteriano é qualquer sequência de DNA capaz de se ligar à RNA po-

limerase bacteriana e iniciar a transcrição (3') a jusante de uma sequência codificante (por exemplo, gene estrutural) no mRNA. Um promotor terá uma região de iniciação de transcrição que é usualmente colocada próxima à extremidade 5' da sequência codificante. Esta região de iniciação de transcrição tipicamente inclui um sítio de ligação de RNA polimerase e um sítio de iniciação de transcrição. Um promotor bacteriano também pode ter um segundo domínio chamado operador que pode se sobrepor a um sítio de ligação de RNA polimerase adjacente no qual a síntese de RNA é iniciada. O operador permite transcrição negativa regulada (induzível), conforme uma proteína repressora de gene pode se ligar ao operador e, desse modo, inibir a transcrição de um gene específico. A expressão constitutiva pode ocorrer na ausência de elementos regulatórios negativos, tais como o operador. Além disso, a regulação positiva pode ser obtida por uma sequência de ligação de proteína ativadora de gene, que, se presente, é usualmente próxima (5') à sequência de ligação de RNA polimerase. Um exemplo de uma proteína ativadora de gene é a proteína ativadora de catabólito (CAP), que ajuda a iniciar a transcrição do operon lac em *Escherichia coli* (*E. coli*) [Raibaud *et al.*, ANNU. REV. GENET. (1984) 18:173]. A expressão regulada, portanto, pode ser positiva ou negativa, desse modo, acentuando ou reduzindo a transcrição.

Sequências que codificam enzimas da via metabólica fornecem, particularmente, sequências promotoras úteis. Exemplos incluem sequências promotoras derivadas de enzimas metabolizadoras de açúcar, tais como galactose, lactose (lac) [Chang *et al.*, NATURE (1977) 198: 1056] e maltose. Exemplos adicionais incluem sequências promotoras derivadas de enzimas biossintéticas, tais como triptofano (trp) [Goeddel *et al.*, NUC. ACIDS RES. (1980) 8:4057; Yelverton *et al.*, NUCL. ACIDS RES. (1981) 9:731; Pat. U.S. Nº 4.738.921; Pub. EP Nºs 036 776 e 121 775, os quais são incorporado como referência neste relatório]. O sistema promotor de β -galactosidase (bla) [Weissmann (1981) "The cloning of interferon and other mistakes." Em Interferon 3 (Ed. I. Gresser)], sistemas de promotor do lambda PL do bacteriófago [Shimatake *et al.*, NATURE (1981) 292: 128] e T5 [Pat. U.S. Nº 4.689.406, os quais são incorporados como referência neste relatório] também fornecem sequências promotoras úteis. Promotores fortes, tais como o promotor T7, podem ser usados para induzir o polipeptídeo de interesse em níveis altos. Exemplos de tais vetores são conhecidos àqueles de habilidade comum na técnica e incluem a série pET29 da Novagen, e os vetores pPOP descritos no WO99/05297, que é incorporado como referência neste relatório. Tais sistemas de expressão produzem níveis altos de polipeptídeo no hospedeiro sem comprometer os parâmetros de viabilidade ou crescimento da célula hospedeira. pET19 (Novagen) é um outro vetor conhecido na técnica.

Além disso, os promotores sintéticos que não ocorrem na natureza também funcionam como promotores bacterianos. Por exemplo, as sequências de ativação de transcrição de um promotor bacteriano ou bacteriófago podem ser unidas com as sequências do operon

de um outro promotor bacteriano ou bacteriófago, criando um promotor híbrido sintético [Pat. U.S. Nº 4.551.433, que é incorporada como referência neste relatório]. Por exemplo, o promotor tac é um promotor trp-lac híbrido compreendido das sequência do promotor trp e do operon lac que são reguladas pelo repressor lac [Amann *et al.*, GENE (1983) 25:167; de Boer *et al.*, PROC. NATL. ACAD. SCI. (1983) 80:21]. Além disso, um promotor bacteriano pode incluir promotores que ocorrem naturalmente de origem não bacteriana que têm a capacidade de se ligar à RNA polimerase bacteriana e iniciar a transcrição. Um promotor que ocorre naturalmente de origem não bacteriana também pode ser ligado a uma RNA polimerase compatível para produzir níveis altos de expressão de alguns genes em procariontes. O sistema T7 RNA polimerase/promotor do bacteriófago é um exemplo de um sistema de promotor ligado [Studier *et al.*, J. MOL. BIOL. (1986) 189:113; Tabor *et al.*, Proc. Natl. Acad. Sci. (1985) 82: 1074]. Além disso, um promotor híbrido também pode ser compreendido de um promotor de bacteriófago e uma região de operador de *E. coli* (Pub. EP Nº 267 851).

Além de uma sequência promotora funcional, um sítio de ligação de ribossomo eficaz também é útil para a expressão de genes estranhos em procariontes. Em *E. coli*, o sítio de ligação de ribossomo é chamado de sequência de Shine-Dalgarno (SD) e inclui um códon de iniciação (ATG) e uma sequência de 3 a 9 nucleotídeos no comprimento de 3 a 11 nucleotídeos localizados a montante do códon de iniciação [Shine *et al.*, NATURE (1975) 254:34]. A sequência de SD promove a ligação do mRNA ao ribossomo pelo pareamento de bases entre a sequência de SD e o 3' do rRNA 16S de *E. coli* [Steitz *et al.* "Genetic signals and nucleotide sequences in messenger RNA", Em Biological Regulation and Development: Gene Expression (Ed. R. F. Goldberger, 1979)]. Para expressar genes eucarióticos e procariontes com sítio de ligação de ribossomo fraco [Sambrook *et al.* "Expression of cloned genes in Escherichia coli", Molecular Cloning: A Laboratory Manual, 1989].

O termo "hospedeiro bacteriano" ou "célula hospedeira bacteriana" refere-se a uma bactéria que pode ser, ou foi, usada como um recipiente para vetores recombinantes ou outro DNA de transferência. O termo inclui a progênie da célula hospedeira bacteriana original que foi transfectada. Deve ser entendido que a progênie de uma célula parental única pode não ser necessária e completamente idêntica em morfologia ou em complemento de DNA genômico ou total ao precursor original, devido à mutação accidental ou deliberada. A progênie da célula parental que é suficientemente similar ao precursor a ser caracterizado pela propriedade relevante, tal como a presença de uma sequência de nucleotídeos que codifica um polipeptídeo de interesse, é incluída na progênie intencionada por esta definição.

A seleção de bactérias hospedeiras adequadas para a expressão de polipeptídeos é conhecida àqueles de habilidade comum na técnica. Na seleção de hospedeiros bacterianos para expressão, hospedeiros adequados podem incluir aqueles mostrados para ter, *inter alia*, capacidade de formação de corpo de inclusão satisfatória, baixa atividade proteolítica e

robustez global. Hospedeiros bacterianos são geralmente disponíveis a partir de uma variedade de fontes, incluindo, mas não limitadas ao Bacterial Genetic Stock Center, Department of Biophysics and Medical Physics, University of California (Berkeley, CA); e à American Type Culture Collection ("ATCC") (Manassas, VA). A fermentação industrial/farmacêutica geralmente usa bactérias derivadas de cepas K (por exemplo, W3110) ou de bactérias derivadas de cepas B (por exemplo, BL21). Estas cepas são particularmente úteis, pois seus parâmetros de crescimento são extremamente bem conhecidos e robustos. Além disso, estas cepas não são patogênicas, o que é comercialmente importante por razões de segurança e ambientais. Outros exemplos de hospedeiros de *E. coli* adequados incluem, mas não são limitados a cepas de BL21, DH10B ou derivados das mesmas. Em uma outra modalidade dos métodos da presente invenção, o hospedeiro de *E. coli* é uma cepa negativa para protease incluindo, mas não limitada a OMP- e LON-. A cepa celular hospedeira pode ser uma espécie de *Pseudomonas*, incluindo, mas não limitada a *Pseudomonas fluorescens*, *Pseudomonas aeruginosa* e *Pseudomonas putida*. *Pseudomonas fluorescens* biovar 1, designada cepa MB101, é conhecida ser útil para a produção recombinante e está disponível para processos de produção de proteína terapêutica. Exemplos de um sistema de expressão de *Pseudomonas* incluem o sistema disponível da Dow Chemical Company como uma cepa hospedeira (Midland, MI disponível na World Wide Web em dow.com). As Patentes U.S. Nºs 4.755.465 e 4.859.600, as quais são incorporadas como referência neste relatório, descrevem o uso de cepas de *Pseudomonas* como uma célula hospedeira para a produção de hGH.

Uma vez que uma cepa celular hospedeira recombinante foi estabelecida (isto é, o constructo de expressão foi introduzido na célula hospedeira e as células hospedeiras com o constructo de expressão apropriado são isoladas), a cepa celular hospedeira recombinante é cultivada sob condições apropriadas para a produção do polipeptídeo de interesse. Como será evidente a uma pessoa de habilidade na técnica, o método de cultura da cepa celular hospedeira recombinante será dependente da natureza do constructo de expressão utilizado e da identidade da célula hospedeira. Cepas hospedeiras recombinantes são normalmente cultivadas usando métodos que são bem conhecidos na técnica. Células hospedeiras recombinantes são tipicamente cultivadas em meio líquido contendo fontes assimiláveis de carbono, nitrogênio e sais inorgânicos e, opcionalmente, contendo vitaminas, aminoácidos, fatores de crescimento e outros suplementos de cultura proteínáceos conhecidos àqueles de habilidade comum na técnica. Meios líquidos para a cultura de células hospedeiras podem opcionalmente conter antibióticos ou antifúngicos para prevenir o crescimento de microorganismos e/ou compostos indesejáveis incluindo, mas não limitados a antibióticos, para selecionar células hospedeiras contendo o vetor de expressão.

Vários métodos bem conhecidos de introduzir ácidos nucleicos alvo em células são

disponíveis e qualquer um dos mesmos pode ser usado na invenção. Estes incluem: fusão das células recipientes com protoplastos bacterianos contendo o DNA, eletroporação, bombardeamento por projétil e infecção com vetores virais (debatidos abaixo), etc. As células bacterianas podem ser usadas para amplificar o número de plasmídeos contendo construtos de DNA desta invenção. As bactérias são cultivadas em fase log e os plasmídeos dentro das bactérias podem ser isolados por uma variedade de métodos conhecidos na técnica (veja, por exemplo, Sambrook). Além disso, kits são comercialmente disponíveis para a purificação de plasmídeos a partir de bactérias, (veja, por exemplo, EasyPrep™, FlexiPrep™, da Pharmacia Biotech; StrataClean™ da Stratagene; e QIAprep™ da Qiagen). Os plasmídeos isolados e purificados depois são manipulados para produzir outros plasmídeos, usados para transfectar células ou incorporados em vetores relacionados para infectar organismos. Vetores típicos contêm terminadores de transcrição e tradução, sequências de iniciação de transcrição e tradução e promotores úteis para a regulação da expressão do ácido nucleico alvo particular. Os vetores opcionalmente compreendem cassetes de expressão genéricos contendo pelo menos uma sequência terminadora independente, sequências que permitem a replicação do cassete em eucariontes, procariontes ou ambos (incluindo, mas não limitadas aos vetores *shuttle*) e marcadores de seleção para sistemas procarióticos e eucarióticos. Vetores são adequados para replicação e integração em procariontes, eucariontes ou ambos. Veja, Gillam & Smith, *Gene* 8:81 (1979); Roberts, *et al.*, *Nature*, 328:731 (1987); Schneider, E., *et al.*, *Protein Expr. Purif.* 6(1): 10-14 (1995); Ausubel, Sambrook, Berger (*supra*). Um catálogo de bactérias e bacteriófagos úteis para clonagem é fornecido, por exemplo, pela ATCC, por exemplo, O ATCC Catalogue of Bacteria and Bacteriophage (1992) Gherna *et al.* (eds) publicado pela ATCC. Procedimentos básicos adicionais para sequenciamento, clonagem e outros aspectos da biologia molecular e considerações teóricas essenciais também são encontrados em Watson *et al.* (1992) *Recombinant DNA Second Edition* Scientific American Books, NY. Além disso, essencialmente, qualquer ácido nucleico (e, na prática, qualquer ácido nucleico marcado, se padrão ou não padrão) pode ser customizado ou na forma padrão a partir de qualquer variedade de fontes comerciais, tais como a Midland Certified Reagent Company (Midland, TX disponível na World Wide Web em m-crc.com), The Great American Gene Company (Ramona, CA disponível na World Wide Web em genco.com), ExpressGen Inc. (Chicago, IL disponível na World Wide Web em express-gen.com), Operon Technologies Inc. (Alameda, CA), entre outros.

Células hospedeiras recombinantes podem ser cultivadas em formatos em lotes ou contínuos, com coleta celular (no caso onde o polipeptídeo de interesse se acumula intracelularmente) ou coleta do sobrenadante de cultura em formatos em lotes ou contínuos. Para a produção em células hospedeiras procarióticas, lotes de cultura e coleta celular são preferidos.

CÓDONS SELETORES

Códons seletores da presente invenção expandem a estrutura de códon genético do mecanismo biossintético da proteína. Por exemplo, um códon seletor inclui, por exemplo, um códon único de três bases, um códon sem sentido, tal como um códon de parada, incluindo, mas não limitado a um códon âmbar (UAG), um códon ocre ou um códon opala (UGA),
 5 um códon não natural, um códon de quatro bases (ou mais), um códon raro ou semelhantes. Vários códons seletores podem ser introduzidos em um gene ou polinucleotídeo desejado, por exemplo, um ou mais, dois ou mais, três ou mais, etc.

Códons seletores da invenção expandem a estrutura de códon genético do mecanismo biossintético da proteína. Por exemplo, um códon seletor inclui, mas não é limitado a
 10 um códon único de três bases, um códon sem sentido, tal como um códon de parada, incluindo, mas não limitado a um códon âmbar (UAG), um códon ocre ou um códon opala (UGA), um códon não natural, um códon de quatro ou mais bases, um códon raro ou semelhantes. É evidente àqueles de habilidade comum na técnica que existe uma ampla faixa no número
 15 de códons seletores que podem ser introduzidos em um gene ou polinucleotídeo desejado, incluindo, mas não limitados a um ou mais, dois ou mais, três ou mais, 4, 5, 6, 7, 8, 9, 10 ou mais em um polinucleotídeo único que codifica o antígeno ou organismo total. Como exemplos não limitantes de organismos totais geneticamente modificados usando a presente invenção, MAP e *E. coli*.

Em uma modalidade, os métodos envolvem o uso de um códon seletor que é um códon de parada para a incorporação de um aminoácido selecionado, por exemplo, um aminoácido não natural, *in vivo*. Por exemplo, um O-tRNA é produzido, o qual reconhece o códon de parada e é aminoacilado por uma O-RS com um aminoácido selecionado. Este O-tRNA não é reconhecido pelas aminoacil-tRNA sintetases do hospedeiro que ocorrem naturalmente. A mutagênese sítio-dirigida convencional pode ser usada para introduzir o códon de parada no sítio de interesse em um polipeptídeo de interesse. Veja, por exemplo, Sayers, J.R., *et al.* (1988), *5'-3' Exonucleases in phosphorothioate-based oligonucleotide-directed mutagenesis*. Nucleic Acids Res. 16:791-802. Quando a O-RS, O-tRNA e o ácido nucleico que codifica um polipeptídeo de interesse são combinados, por exemplo, *in vivo*, o aminoácido selecionado é incorporado em resposta ao códon de parada para fornecer um polipeptídeo contendo o aminoácido selecionado, por exemplo, um aminoácido não natural, na posição específica. Em uma modalidade da presente invenção, um códon de parada usado como um códon seletor é um códon âmbar, UAG, e/ou um códon opala, UGA. Por exemplo, veja SEQ ID NO.: 6 para um exemplo de um O-tRNA que reconhece um códon âmbar, e
 30 veja SEQ ID NO.: 7 para um exemplo de um O-tRNA que reconhece um códon opala. Um código genético em que UAG e UGA são usados como um códon seletor pode codificar 22 aminoácidos, preservando o códon sem sentido ocre, UAA, que é o sinal de terminação

mais abundante.

A incorporação de aminoácidos selecionados, por exemplo, aminoácidos não naturais, *in vivo*, pode ser feita sem perturbação significativa da célula hospedeira. Por exemplo, em células não eucarióticas, tais como *Escherichia coli*, pois a eficácia de supressão para o códon UAG depende da competição entre o O-tRNA, por exemplo, o tRNA supressor âmbar, e o fator de liberação 1 (RF1) (que se liga ao códon UAG e inicia a liberação do peptídeo crescente a partir do ribossomo), a eficácia de supressão pode ser modulada, por exemplo, aumentando-se o nível de expressão do O-tRNA, por exemplo, do tRNA supressor, ou usando-se uma cepa deficiente de RF1. Em células eucarióticas, pois a eficácia de supressão para o códon UAG depende da competição entre o O-tRNA, por exemplo, o tRNA supressor âmbar, e um fator de liberação eucariótico (por exemplo, eRF) (que se liga a um códon de parada e inicia a liberação do peptídeo crescente a partir do ribossomo), a eficácia de supressão pode ser modulada, por exemplo, aumentando-se o nível de expressão do O-tRNA, por exemplo, do tRNA supressor.

Aminoácidos não naturais também podem ser codificados com códons raros. Por exemplo, quando a concentração de arginina em uma reação de síntese de proteína *in vitro* é reduzida, o códon arginina raro, AGG, provou ser eficaz para a inserção de Ala através de um tRNA sintético acilado com alanina. Veja, por exemplo, Ma *et al.*, Biochemistry, 32:7939 (1993). Neste caso, o tRNA sintético compete com o tRNA^{Arg} que ocorre naturalmente, que existe como uma espécie secundária em *Escherichia coli*. Alguns organismos não usam todos os códons tripletos. Um códon AGA não determinado em *Micrococcus luteus* foi utilizado para a inserção de aminoácidos em um extrato de transcrição/tradução *in vitro*. Veja, por exemplo, Kowal e Oliver, *Nucl. Acid. Res.*, 25:4685 (1997). Os componentes da presente invenção podem ser gerados para usar estes códons raros *in vivo*.

Os códons seletores também compreendem códons estendidos, por exemplo, códons de quatro ou mais bases, tais como, quatro, cinco, seis ou mais códons de base. Exemplos de códons de quatro bases incluem, mas não são limitados a AGGA, CUAG, UAGA, CCCU e semelhantes. Exemplos de códons de cinco base incluem, mas não são limitados a AGGAC, CCCCUC, CCCUC, CUAGA, CUACU, UAGGC e semelhantes. Uma característica pode incluir usar códons estendidos com base na supressão de mudança de fase de leitura. Os códons de quatro ou mais bases podem inserir, por exemplo, um ou múltiplos aminoácidos selecionados, incluindo, mas não limitados a aminoácidos não naturais, na mesma proteína. Por exemplo, na presença de O-tRNAs modificados, por exemplo, tRNAs supressores de mudança de fase de leitura especial, com alças anticódon, por exemplo, com uma sequência CU(X)_n XXXAA (onde n = 1), o códon de quatro ou mais bases é lido como aminoácido único. Por exemplo, veja SEQ ID NOs.: 6, 12 do PCT/US04/22061 para O-tRNAs que reconhecem um códon de quatro bases. Em outras modalidades, as al-

ças anticódon podem decodificar, por exemplo, pelo menos um códon de quatro bases, pelo menos um códon de cinco bases ou pelo menos um códon de seis bases ou mais. Visto que existem 256 códons de quatro bases possíveis, múltiplos aminoácidos não naturais podem ser codificados na mesma célula usando um códon de quatro ou mais bases. Veja, Anderson *et al.*, (2002) *Exploring the Limits of Codon and Anticodon Size*, Chemistry and Biology, 9:237-244; Magliery, (2001) *Expanding the Genetic Code: Selection of Efficient Suppressors of Four-base Codons and Identification of "Shifty" Four-base Codons with a Library Approach in Escherichia coli*, J. Mol. Biol. 307: 755-769.

Por exemplo, códons de quatro bases foram usados para incorporar aminoácidos não naturais em proteínas usando métodos biossintéticos *in vitro*. Veja, por exemplo, Ma *et al.*, (1993) Biochemistry, 32:7939; e Hohsaka *et al.*, (1999) J. Am. Chem. Soc. 121:34. CGGG e AGGU foram usados para incorporar simultaneamente 2-naftilalanina e um NBD derivado de lisina em estreptavidina *in vitro* com dois tRNAs supressores de mudança de fase de leitura quimicamente acilados. Veja, por exemplo, Hohsaka *et al.*, (1999) J. Am. Chem. Soc. 121:12194. Em um estudo *in vivo*, Moore *et al.* examinaram a capacidade de derivados de tRNA^{Leu} com anticódon NCUA suprimirem códons UAGN (N pode ser U, A, G ou C), e descobriram que o quadrupeto UAGA pode ser decodificado por um tRNA^{Leu} com um anticódon UCUA com uma eficácia de 13 a 26 % com pouca decodificação no quadro 0 ou -1. Veja, Moore *et al.*, (2000) J. Mol. Biol. 298:195. Em uma modalidade, códons estendidos com base em códons raros ou códons sem sentido podem ser usados na invenção, os quais podem reduzir perda de sentido de leitura ininterrupta e supressão de mudança de fase de leitura em outros sítios indesejados.

Para um sistema fornecido, um códon seletor também pode incluir um dos códons de três bases naturais, onde o sistema endógeno não usa (ou raramente usa) o códon de base natural. Por exemplo, isto inclui um sistema que é desprovido de um tRNA que reconhece o códon de três bases natural e/ou um sistema onde o códon de três bases é um códon raro.

Códons seletores incluem opcionalmente pares de base não naturais. Estes pares de base não naturais expandem adicionalmente o alfabeto genético existente. Um par de base extra aumenta o número de códons tripletos de 64 para 125. As propriedades dos terceiros pares de base incluem pareamento de base estável e seletivo, incorporação enzimática eficaz no DNA com alta fidelidade por uma polimerase e a extensão do *primer* contínua eficaz depois da síntese do par de base nascente não natural. As descrições de pares de base não naturais que podem ser adaptadas para os métodos e composições incluem, por exemplo, Hirao, *et al.*, (2002) *An unnatural base pair for incorporating amino acid analogues into protein*, Nature Biotechnology. 20:177-182. Veja, também, Wu, Y., *et al.*, (2002) J. Am. Chem. Soc. 124: 14626-14630. Outras publicações relevantes são listadas abaixo.

Para a utilização *in vivo*, o nucleosídeo não natural é permeável à membrana e é fosforilado para formar o trifosfato correspondente. Além disso, a informação genética aumentada é estável e não é destruída por enzimas celulares. Esforços prévios de Benner e colaboradores tomam vantagem dos padrões de ligação de hidrogênio que são diferentes daqueles em pares de Watson-Crick canônicos, o exemplo mais notável é o par iso-C:iso-G. Veja, por exemplo, Switzer *et al.*, (1989) J. Am. Chem. Soc., 111:8322; e Piccirilli *et al.*, (1990) Nature, 343:33; Kool, (2000) Curr. Opin. Chem. Biol., 4:602. Estas bases, no geral, são pareadas de forma imperfeita, em algum grau, com bases naturais e não podem ser enzimaticamente replicadas. Kool e colaboradores demonstraram que interações de acondicionamento hidrofóbicas entre bases podem substituir a ligação de hidrogênio para conduzir a formação do par de base. Veja, Kool, (2000) Curr. Opin. Chem. Biol., 4:602; e Guckian e Kool, (1998) Angew. Chem. Int. Ed. Engl., 36, 2825. Em um esforço para desenvolver um par de base não natural que satisfaz todas as exigências acima, Schultz, Romesberg e colaboradores sintetizaram e estudaram sistematicamente uma série de bases hidrofóbicas não naturais. Descobriu-se que um autopar PICS:PICS é mais estável do que pares de base naturais, e pode ser eficazmente incorporado no DNA pelo fragmento Klenow da DNA polimerase I de *Escherichia coli* (KF). Veja, por exemplo, McMinn *et al.*, (1999) J. Am. Chem. Soc., 121: 11585-6; e Ogawa *et al.*, (2000) J. Am. Chem. Soc., 122:3274. Um autopar 3MN:3MN pode ser sintetizado por KF com eficácia e seletividade suficientes para a função biológica. Veja, por exemplo, Ogawa *et al.*, (2000) J. Am. Chem. Soc., 122:8803. Entretanto, estas bases atuam como um terminador de cadeia para replicação adicional. Uma DNA polimerase mutante foi recentemente desenvolvida, a qual pode ser usada para replicar o autopar PICS. Além disso, um autopar 7AI pode ser replicado. Veja, por exemplo, Tae *et al.*, (2001) J. Am. Chem. Soc., 123:7439. Um novo par de metalobases, Dipic:Py, também foi desenvolvido, o qual forma um par estável sob ligação de Cu(II). Veja, Meggers *et al.*, (2000) J. Am. Chem. Soc., 122:10714. Pelo fato de que os códons estendidos e os códons não naturais são intrinsecamente ortogonais aos códons naturais, os métodos da presente invenção podem tomar vantagem desta propriedade para gerar tRNAs ortogonais para os mesmos.

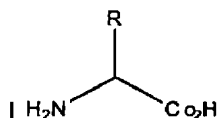
Um sistema de atalho de tradução também pode ser usado para incorporar um aminoácido selecionado, por exemplo, um aminoácido não natural, em um polipeptídeo desejado. Em um sistema de atalho de tradução, uma sequência grande é inserida em um gene, mas não é traduzida em proteína. A sequência contém uma estrutura que serve como um sinal para induzir o ribossomo a saltar sobre a sequência e retomar a tradução a jusante da inserção.

AMINOÁCIDOS SELECIONADOS E NÃO NATURAIS

Conforme usado neste relatório, um aminoácido selecionado refere-se a qualquer

aminoácido que ocorre naturalmente ou aminoácido não natural desejado. Um aminoácido que ocorre naturalmente inclui qualquer um entre os vinte alfa-aminoácidos geneticamente codificados: alanina, arginina, asparagina, ácido aspártico, cisteína, glutamina, ácido glutâmico, glicina, histidina, isoleucina, leucina, lisina, metionina, fenilalanina, prolina, serina, treonina, triptofano, tirosina e valina. Em uma modalidade, o aminoácido selecionado é incorporado em uma cadeia polipeptídica crescente com alta fidelidade, por exemplo, mais de cerca de 70 % de eficácia para um códon seletor fornecido, mais de 75 % de eficácia para um códon seletor fornecido, mais de cerca de 80 % de eficácia para um códon seletor fornecido, mais de cerca de 85 % de eficácia para um códon seletor fornecido, mais de cerca de 90 % de eficácia para um códon seletor fornecido, mais de cerca de 95 % de eficácia para um códon seletor fornecido ou mais de cerca de 99 % de eficácia ou mais para um códon seletor fornecido.

Conforme usado neste relatório, um aminoácido não natural refere-se a qualquer aminoácido, aminoácido modificado ou análogo de aminoácido, exceto a selenocisteína e/ou a pirrolisina, e os vinte alfa-aminoácidos geneticamente codificados seguintes: alanina, arginina, asparagina, ácido aspártico, cisteína, glutamina, ácido glutâmico, glicina, histidina, isoleucina, leucina, lisina, metionina, fenilalanina, prolina, serina, treonina, triptofano, tirosina, valina. A estrutura genérica de um alfa-aminoácido é ilustrada pela Fórmula I:

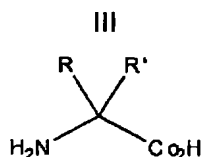
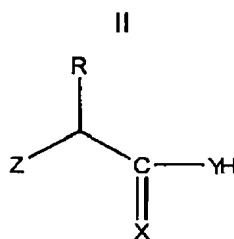


Um aminoácido não natural é tipicamente qualquer estrutura que tem a Fórmula I, em que o grupo R é qualquer substituinte, exceto um dos usados nos vinte aminoácidos naturais. Veja, por exemplo, *Biochemistry* de L. Stryer, 3ª ed. 1988, Freeman and Company, Nova Iorque, para as estruturas dos vinte aminoácidos naturais. Note que, os aminoácidos não naturais da presente invenção podem ser compostos que ocorrem naturalmente, exceto os vinte alfa-aminoácidos acima.

Pelo fato de os aminoácidos não naturais da presente invenção diferirem tipicamente dos aminoácidos naturais apenas na estrutura da cadeia lateral, os aminoácidos não naturais formam ligações amida com outros aminoácidos, incluindo, mas não limitados aos naturais ou não naturais, da mesma maneira pela qual elas são formadas em proteínas que ocorrem naturalmente. Entretanto, os aminoácidos não naturais têm grupos de cadeias laterais que os distinguem dos aminoácidos naturais. Por exemplo, R na Fórmula I pode compreender um alquil-, aril-, acil-, ceto-, azido-, hidroxil-, hidrazina-, ciano-, halo-, hidrazida, alquenila, alquinila, éter, tiol, seleno, sulfonila, borato, boronato, fosfo, fosfono, fosfina, heterocíclico, enona, imina, aldeído, éster, tioácido, hidroxilamina, amina e semelhantes, ou

qualquer combinação dos mesmos. Outros aminoácidos que não ocorrem naturalmente incluem, mas não são limitados aos aminoácidos que compreendem um reticulador fotoativável, aminoácidos marcados por spin, aminoácidos fluorescentes, aminoácidos de ligação ao metal, aminoácidos que contêm metal, aminoácidos radioativos, aminoácidos com novos grupos funcionais, aminoácidos que interagem com outras moléculas covalente ou não covalentemente, aminoácidos *photocaged* e/ou fotoisomerizáveis, aminoácidos contendo biotina ou análogo de biotina, aminoácidos glicosilados, tais como uma serina substituída por açúcar, outros aminoácidos modificados por carboidrato, aminoácidos contendo ceto, aminoácidos que compreendem polietilenoglicol ou poliéter, aminoácidos substituídos com átomo pesado, aminoácidos quimicamente cliváveis e/ou fotocliváveis, aminoácidos com cadeias laterais alongadas quando comparadas as de aminoácidos naturais, incluindo, mas não limitados a poliéteres ou hidrocarbonetos de cadeia longa, incluindo, mas não limitada a maior do que cerca de 5 ou maior do que cerca de 10 carbonos, aminoácidos contendo açúcar ligado ao carbono, aminoácidos redox-ativos, aminoácidos contendo aminotioácidos e aminoácidos contendo uma ou mais porções tóxicas. Veja também as Publicações de Pedido de Patente U.S. 2003/0082575 e 2003/0108885, que são incorporadas como referência neste relatório. Os aminoácidos não naturais podem ter um reticulador fotoativável que é usado, por exemplo, para ligar uma proteína a um suporte sólido. Os aminoácidos não naturais podem ter uma porção sacarídeo ligada à cadeia lateral do aminoácido.

Além de os aminoácidos não naturais que contêm novas cadeias laterais, os aminoácidos não naturais também compreendem opcionalmente estruturas de cadeia principal modificadas, por exemplo, conforme ilustrado pelas estruturas das Fórmulas II e III:



em que Z tipicamente compreende OH, NH₂, SH, NH-R' ou S-R'; X e Y, que podem ser os mesmos ou diferentes, tipicamente compreendem S ou O, e R e R', que são opcionalmente os mesmos ou diferentes, são tipicamente selecionados a partir da mesma lista de constituintes para o grupo R descrita acima para os aminoácidos não naturais que têm a

Fórmula I, assim como hidrogênio. Por exemplo, os aminoácidos não naturais podem compreender substituições no grupo amino ou carboxila, conforme ilustrado pelas Fórmulas II e III. Os aminoácidos não naturais desse tipo incluem, mas não são limitados a ácidos α -hidróxi, α -tioácidos, α -aminotiocarboxilatos, por exemplo, com cadeias laterais que correspondem a cadeias laterais dos vinte aminoácidos naturais comuns ou não naturais. Além disso, as substituições do α -carbono incluem opcionalmente L, D ou α - α -aminoácidos dissubstituídos, tais como D-glutamato, D-alanina, D-metil-O-tirosina, ácido aminobutírico e semelhantes. Outras alternativas estruturais incluem aminoácidos cíclicos, tais como análogos de prolina, assim como análogos de prolina com 3, 4, 6, 7, 8 e 9 membros no anéis, β e γ aminoácidos, tais como β -alanina substituída e ácido γ -aminobutírico.

Muitos aminoácidos não naturais são baseados em aminoácidos naturais, tais como tirosina, glutamina, fenilalanina e semelhantes. Os análogos de tirosina incluem tirosinas para-substituídas, tirosinas orto-substituídas e tirosinas meta-substituídas, em que a tirosina substituída compreende um grupo ceto (incluindo, mas não limitado a um grupo acetila), um grupo benzoila, um grupo amino, uma hidrazina, uma hidroxiamina, um grupo tiol, um grupo carbóxi, um grupo isopropila, um grupo metila, um hidrocarboneto C₆-C₂₀ de cadeia reta ou ramificada, um hidrocarboneto saturado ou não saturado, um grupo O-metila, um grupo poliéter, um grupo nitro ou semelhantes. Além disso, os anéis de arila multiplamente substituídos também são contemplados. Os análogos de glutamina incluem, mas não são limitados a derivados de α -hidróxi, derivados γ -substituídos, derivados cíclicos e derivados de glutamina amido-substituídos. Os análogos de fenilalanina exemplares incluem, mas não são limitados a fenilalaninas para-substituídas, fenilalaninas orto-substituídas e fenilalaninas meta-substituídas, em que o substituinte compreende um grupo hidróxi, um grupo metóxi, um grupo metila, um grupo alila, um aldeído, uma azida, um iodo, um bromo, um grupo ceto (incluindo, mas não limitado a um grupo acetila) ou semelhantes. Os exemplos específicos de aminoácidos não naturais incluem, mas não são limitados a p-acetil-L-fenilalanina, p-propargil-fenilalanina, O-metil-L-tirosina, L-3-(2-naftil)alanina, 3-metilfenilalanina, O-4-alil-L-tirosina, tri-O-acetil-GlcNAc β -serina, L-Dopa, fenilalanina fluorada, isopropil-L-fenilalanina, p-azido-L-fenilalanina, p-acil-L-fenilalanina, p-benzoil-L-fenilalanina, L-fosfosserina, fosfonosserina, fosfotirosina, p-iodo-fenilalanina, p-bromofenilalanina, p-amino-L-fenilalanina, isopropil-L-fenilalanina e p-propargilóxi-fenilalanina e semelhantes. Os exemplos de estruturas de uma variedade de aminoácidos não naturais são fornecidos, por exemplo, no WO 2002/085623 intitulado "In vivo incorporation of unnatural amino acids", que é incorporado como referência neste relatório. Veja também Kiick *et al.*, (2002) *Incorporation of azides into recombinant proteins for chemoselective modification by the Staudinger ligation*, PNAS 99:19-24, que é incorporado como referência neste relatório, para análogos de metionina adicionais.

Um aminoácido não natural incorporado em um polipeptídeo no terminal amino pode ser composto de um grupo R que é qualquer substituinte, exceto um usado nos vinte aminoácidos naturais e um segundo grupo reativo a partir do grupo NH_2 normalmente presente em α -aminoácidos (veja a Fórmula I). Um aminoácido não natural similar pode ser incorporado no terminal carbóxi com um segundo grupo reativo diferente do grupo COOH normalmente presente em α -aminoácidos (veja a Fórmula I).

Os aminoácidos não naturais da invenção podem ser selecionados ou designados para fornecer características indisponíveis nos vinte aminoácidos naturais. Por exemplo, o aminoácido não natural pode ser opcionalmente designado ou selecionado para modificar as propriedades biológicas de uma proteína, por exemplo, na qual eles são incorporados. Por exemplo, as propriedades seguintes podem ser opcionalmente modificadas pela inclusão de um aminoácido não natural em uma proteína: toxicidade, biodistribuição, solubilidade, estabilidade, por exemplo, térmica, hidrolítica, oxidativa, resistência à degradação enzimática e semelhantes, facilidade de purificação e processamento, propriedades estruturais, propriedades espectroscópicas, propriedades químicas e/ou fotoquímicas, atividade catalítica, potencial redox, meia-vida, habilidade para reagir com outras moléculas, por exemplo, covalente ou não covalentemente, e semelhantes.

As estruturas de uma variedade de aminoácidos não naturais são fornecidas, por exemplo, nas Figuras 16, 17, 18, 19, 26 e 29 do WO 2002/085923 intitulado "In vivo incorporation of unnatural amino acids", que é incorporado como referência neste relatório. Os exemplos não se destinam a ser limitantes de qualquer modo a aminoácidos que possam ser aderidos a um tRNA da presente invenção.

Uma vantagem de um aminoácido não natural é que ele apresenta porções químicas adicionais que podem ser usadas para adicionar moléculas adicionais. Estas modificações podem ser feitas *in vivo* em uma célula eucariótica ou não eucariótica, ou em *in vitro*. Assim, em certas modalidades, a modificação pós-tradução se dá através do aminoácido não natural. Um aminoácido não natural em um polipeptídeo pode ser usado para ligar uma outra molécula ao polipeptídeo, incluindo, mas não limitado a um marcador, um pigmento, um polímero, um polímero solúvel em água, um derivado de polietilenoglicol, um fotorreticulador, um radionuclídeo, um composto citotóxico, um fármaco, um marcador de afinidade, um marcador de fotoafinidade, um composto reativo, uma resina, uma segunda proteína ou polipeptídeo ou análogo de polipeptídeo, um anticorpo ou fragmento de anticorpo, um quelante de metal, um cofator, um ácido graxo, um carboidrato, um polinucleotídeo, um DNA, um RNA, um polinucleotídeo antissentido, um sacarídeo, um dendrímero solúvel em água, uma ciclodextrina, um ácido ribonucleico inibitório, um biomaterial, uma nanopartícula, um marcador de *spin*, um fluoróforo, uma porção contendo metal, uma porção radioativa, um novo grupo funcional, um grupo que interage covalente ou não covalentemente com outras

moléculas, uma porção *photocaged*, uma porção excitável por radiação actínica, uma porção fotoisomerizável, biotina, um derivado de biotina, um análogo de biotina, uma porção que incorpora um átomo pesado, um grupo quimicamente clivável, um grupo fotoclivável, uma cadeia lateral alongada, um açúcar ligado ao carbono, um agente redox-ativo, um amino

5 no tioácido, uma porção tóxica, uma porção isotopicamente marcada, uma sonda biofísica, um grupo fosforescente, um grupo quimioluminescente, um grupo eletrodensso, um grupo magnético, um grupo de intercalação, um cromóforo, um agente de transferência energética, um agente biologicamente ativo, um marcador detectável, uma pequena molécula, um ponto quântico, um nanotransmissor ou qualquer combinação dos citados acima ou qualquer outro

10 composto ou substância desejável, compreendendo um segundo grupo reativo a pelo menos um aminoácido não natural compreendendo um primeiro grupo reativo utilizando metodologia química que é conhecida por uma pessoa de habilidade comum na técnica adequada aos grupos reativos particulares.

Por exemplo, a modificação pós-tradução pode ocorrer através de uma reação nucleofílica-eletrofílica. Atualmente, a maioria das reações usadas para a modificação seletiva

15 de proteínas envolve a formação de ligação covalente entre parceiros de reação nucleofílica e eletrofílica, incluindo, mas não limitadas à reação de α -halocetonas com cadeias laterais de histidina ou cisteína. A seletividade nesses casos é determinada pelo número e acessibilidade dos resíduos nucleofílicos na proteína. Nas proteínas da invenção, outras reações

20 mais seletivas podem ser usadas, tais como a reação de um ceto-aminoácido não natural com hidrazidas ou compostos amino-óxi, *in vitro* e *in vivo*. Veja, por exemplo, Cornish, *et al.*, (1996) J. Am. Chem. Soc., 118:8150-8151; Mahal, *et al.*, (1997) Science, 276:1125-1128; Wang, *et al.*, (2001) Science, 292:498-500; Chin *et al.*, (2002) J. Am. Chem. Soc., 124:9026-9027; Chin *et al.*, (2002) Proc. Natl. Acad. Sci., 99: 11020-11024; Wang, *et al.*, (2003) Proc.

25 Natl. Acad. Sci., 100:56-61; Zhang *et al.*, (2003) Biochemistry, 42:6735-6746; e Chin *et al.*, (2003) Science, 301:964-7, os quais são incorporados como referência neste relatório. Isto permite a marcação seletiva de, virtualmente, qualquer proteína com uma série de reagentes, incluindo fluoróforos, agentes reticuladores, derivados de sacarídeo e moléculas citotóxicas. Veja também a Patente U.S. Nº 6.927.042 intitulada "Glycoprotein synthesis", que é

30 incorporada como referência neste relatório. As modificações pós-tradução através de um azido-aminoácido também podem ser feitas através da ligação de Staudinger (incluindo, mas não limitada com reagentes de triarilfosfina). Veja, por exemplo, Kiick *et al.*, (2002) *Incorporation of azides into recombinant proteins for chemoselective modification by the Staudinger ligation*, PNAS 99:19-24.

35 Síntese Química de Aminoácidos Não Naturais

Muitos dos aminoácidos não naturais são comercialmente disponíveis, por exemplo, pela Sigma-Aldrich (St. Louis, MO, USA), Novabiochem (uma divisão de EMD Biosciences,

Darmstadt, Alemanha), ou Peptech (Burlington, MA, USA). Aqueles que não são comercialmente disponíveis são opcionalmente sintetizados, conforme fornecido neste relatório ou usando métodos padrão conhecidos àqueles de habilidade na técnica. Para as técnicas de síntese orgânica, veja, por exemplo, Organic Chemistry de Fessenden e Fessenden, (1982, Segunda Edição, Willard Grant Press, Boston Mass.); Advanced Organic Chemistry de March (Terceira Edição, 1985, Wiley e Sons, Nova Iorque); e Advanced Organic Chemistry de Carey e Sundberg (Terceira Edição, Partes A e B, 1990, Plenum Press, Nova Iorque). As publicações adicionais que descrevem a síntese de aminoácidos não naturais incluem, por exemplo, WO 2002/085923 intitulado "In vivo incorporation of Unnatural Amino Acids"; Matsoukas *et al.*, (1995) J. Med. Chem., 38, 4660-4669; King, F. E. & Kidd, D. A. A. (1949) "A New Synthesis of Glutamine and of γ -Dipeptides of Glutamic Acid from Phthylated Intermediates" J. Chem. Soc., 3315-3319; Friedman, O. M. & Chatterji, R. (1959) "Synthesis of Derivatives of Glutamine as Model Substrates for Anti-Tumor Agents" J. Am. Chem. Soc. 81, 3750-3752; Craig, J. C. *et al.*, (1988) "Absolute Configuration of the Enantiomers of 7-Chloro-4 [[4-(diethylamino)-1-methylbutyl]amino]quinoline (Chloroquine)" J. Org. Chem. 53, 1167-1170; Azoulay, M., Vilmont, M. & Frappier, F. (1991) "Glutamine analogues as Potential Antimalarials" Eur. J. Med. Chem. 26, 201-5; Koskinen, A. M. P. & Rapoport, H. (1989) "Synthesis of 4-Substituted Prolines as Conformationally Constrained Amino Acid Analogues" J. Org. Chem. 54, 1859-1866; Christie, B. D. & Rapoport, H. (1985) "Synthesis of Optically Pure Pivalates from L-Asparagine. Application to the Total Synthesis of (+)-Apovincamine through Amino Acid Decarbonylation and Iminium Ion Cyclization" J. Org. Chem. 1989:1859-1866; Barton *et al.*, (1987) "Synthesis of Novel α -Amino-Acids and Derivatives Using Radical Chemistry: Synthesis of L- and D- α -Amino-Adipic Acids, L- α -aminopimelic Acid and Appropriate Unsaturated Derivatives" Tetrahedron 43:4297-4308; e Subasinghe *et al.*, (1992) *Quisqualic acid analogues: synthesis of beta-heterocyclic 2-aminopropanoic acid derivatives and their activity at a novel quisqualate-sensitized site.* J. Med. Chem. 35: 4602-7. Veja também a Publicação de Patente U.S. Nº US 2004/0198637, intitulada "Protein Arrays", que é incorporada como referência.

Absorção Celular de Aminoácidos Não Naturais

A absorção de aminoácidos não naturais por uma célula é um assunto que é tipicamente considerado quando se projeta e seleciona aminoácidos não naturais, por exemplo, para a incorporação em uma proteína. Por exemplo, a alta densidade de carga de α -aminoácidos sugere que estes compostos provavelmente não são permeáveis à célula. Os aminoácidos naturais são obtidos em uma célula por intermédio de uma coleta de sistemas de transporte com base em proteína. Uma rápida triagem pode ser feita, a qual avalia quais aminoácidos não naturais, se algum, são obtidos pelas células. Veja, por exemplo, os ensaios de toxicidade, por exemplo, na Publicação de Patente U.S. Nº US 2004/0198637, intitula-

da “Protein Arrays”, que é incorporada como referência neste relatório, e Liu, D. R. & Schultz, P. G. (1999) *Progress toward the evolution of an organism with an expanded genetic code* PNAS United States 96:4780-4785. Embora a absorção seja facilmente analisada por vários ensaios, uma alternativa para projetar aminoácidos não naturais que são sujeitos às

5 vias de absorção celular é fornecer vias biossintéticas para criar aminoácidos *in vivo*.

Biossíntese de Aminoácidos Não Naturais

Muitas vias biossintéticas já existem em células para a produção de aminoácidos e outros compostos. Enquanto um método de biossíntese para um aminoácido não natural particular pode não existir na natureza em uma célula, a invenção fornece tais métodos. Por

10 exemplo, as vias biossintéticas para os aminoácidos não naturais são criadas opcionalmente na célula hospedeira pela adição de novas enzimas ou pela modificação de vias existentes da célula hospedeira. As novas enzimas adicionais são opcionalmente enzimas que ocorrem naturalmente ou enzimas artificialmente desenvolvidas. Por exemplo, a biossíntese da p-aminofenilalanina (conforme apresentado em um exemplo no WO 2002/085923, intitulado

15 “In vivo incorporation of unnatural amino acids”) baseia-se na adição de uma combinação de enzimas conhecidas a partir de outros organismos. Os genes para estas enzimas podem ser introduzidos em uma célula pela transformação da célula com um plasmídeo compreendendo os genes. Os genes, quando expressados na célula, fornecem uma via enzimática para sintetizar o composto desejado. Os exemplos dos tipos de enzimas que são opcionalmente

20 adicionadas são fornecidos nos exemplos abaixo. As sequências de enzimas adicionais são encontradas, por exemplo, no Genbank. As enzimas artificialmente desenvolvidas também são opcionalmente adicionadas em uma célula da mesma maneira. Dessa maneira, o mecanismo celular e recursos de uma célula são manipulados para produzir aminoácidos não naturais.

25 Uma variedade de métodos está disponível para produzir novas enzimas para uso em vias biossintéticas ou para o desenvolvimento de vias existentes. Por exemplo, a recombinação recursiva, por exemplo, conforme desenvolvido por Maxygen, Inc. (disponível na *World Wide Web* em maxygen.com), é opcionalmente usada para desenvolver novas enzimas e vias. Veja, por exemplo, Stemmer (1994), *Rapid evolution of a protein in vitro by DNA*

30 *shuffling* Nature 370(4):389-391; e Stemmer, (1994), *DNA shuffling by random fragmentation and reassembly: In vitro recombination for molecular evolution*, Proc. Nat. Acad. Sci. USA., 91:10747-10751. De maneira similar, DesignPath™, desenvolvido pela Genencor (disponível na *World Wide Web* em genencor.com) é opcionalmente usado para a engenharia da via metabólica, por exemplo, para projetar uma via para criar O-metil-L-tirosina em uma célula.

35 Esta tecnologia reconstrói vias existentes em organismos hospedeiros usando uma combinação de novos genes, incluindo, mas não limitados àqueles identificados através da genômica funcional e evolução e projeto moleculares. A Diversa Corporation (disponível na *World*

Wide Web em *diversa.com*) também fornece a tecnologia para triagem rápida de bibliotecas de genes e vias de genes para criar novas vias.

Tipicamente, o aminoácido não natural produzido com uma via biossintética construída da presente invenção é produzido em uma concentração suficiente para uma biossíntese proteica eficaz, por exemplo, uma quantidade celular natural, mas não a um grau que afete significativamente a concentração de outros aminoácidos ou que leve os recursos celulares à exaustão. As concentrações típicas produzidas *in vivo* dessa maneira são de cerca de 10 mM até cerca de 0,05 mM. Uma vez que uma célula é transformada com um plasmídeo compreendendo os genes para produzir enzimas desejadas para uma via específica e um aminoácido não natural é criado, seleções *in vivo* são opcionalmente usadas para otimizar adicionalmente a produção do aminoácido não natural tanto para a síntese de proteína ribossomal quanto para o crescimento celular.

SEQUÊNCIA DE ÁCIDOS NUCLEICOS E POLIPEPTÍDEOS E VARIANTES

Conforme descrito acima e abaixo, a invenção fornece sequências de polinucleotídeos de ácido nucleico e sequências de aminoácidos polipeptídicas, por exemplo, de tRNAs e RSs e, por exemplo, as composições e os métodos que compreendem as ditas sequências. Os exemplos das ditas sequências, por exemplo, de tRNAs e RSs, são divulgados. Entretanto, uma pessoa de habilidade na técnica avaliará que a invenção não está limitada àquelas sequências divulgadas neste relatório, tal como nos Exemplos. Uma pessoa habilitada avaliará que a invenção também fornece muitas sequências relacionadas e não relacionadas com as funções descritas neste relatório, por exemplo, codificando um O-tRNA ou uma O-RS.

A invenção fornece polipeptídeos (O-RSs) e polinucleotídeos, por exemplo, O-tRNA, polinucleotídeos que codificam O-RSs ou porções das mesmas, oligonucleotídeos usados para isolar clones de aminoacil-tRNA sintetase, etc. Polinucleotídeos da presente invenção incluem aqueles que codificam proteínas ou polipeptídeos de interesse da invenção com um ou mais códons seletores. Além disso, os polinucleotídeos da presente invenção incluem, por exemplo, um polinucleotídeo que compreende uma sequência de nucleotídeos, conforme apresentado em qualquer na SEQ ID NO.: 1, 2, 3; um polinucleotídeo que é complementar a, ou uma variação conservativa da mesma. Similarmente, um ácido nucleico que hibridiza um polinucleotídeo indicado acima sob condições altamente severas sobre, substancialmente, o comprimento total do ácido nucleico é um polinucleotídeo da presente invenção.

Em certas modalidades, um vetor (por exemplo, um plasmídeo, um cosmídeo, um fago, uma bactéria, um vírus, um polinucleotídeo puro, um polinucleotídeo conjugado, etc.) compreende um polinucleotídeo da presente invenção. Em uma modalidade, o vetor é um vetor de expressão. Em uma outra modalidade, o vetor de expressão inclui um promotor

ligado, de maneira operável, a um ou mais dos polinucleotídeos da presente invenção. Em uma outra modalidade, uma célula compreende um vetor que inclui um polinucleotídeo da presente invenção.

5 Uma pessoa habilitada também avaliará que muitas variantes das sequências divulgadas estão incluídas na invenção. Por exemplo, as variações conservativas das sequências divulgadas que produzem uma sequência funcionalmente idêntica são incluídas na invenção. As variantes das sequências de polinucleotídeos de ácido nucleico, em que as variantes se hibridizam com pelo menos uma sequência divulgada, são consideradas incluídas na invenção. As subsequências únicas das sequências divulgadas neste relatório, conforme determinado, por exemplo, por técnicas de comparação de sequências padrão, também estão incluídas na invenção.

Variações conservativas

Devido à degeneração do código genético, as “substituições silenciosas” (isto é, substituições em uma sequência de ácidos nucleicos as quais não resultam em uma alteração em um polipeptídeo codificado) são uma característica implícita de toda sequência de ácidos nucleicos que codifica um aminoácido. De maneira similar, as “substituições conservativas de aminoácidos”, em que um ou poucos aminoácidos em uma sequência de aminoácidos são substituídos com aminoácidos diferentes com propriedades altamente similares, também são prontamente identificadas como altamente similares a um constructo divulgado.

15 20 Tais variações conservativas de cada sequência divulgada são uma característica da presente invenção.

As “variações conservativas” de uma sequência de ácidos nucleicos particular referem-se àqueles ácidos nucleicos que codificam sequências de aminoácidos idênticas ou essencialmente idênticas, ou onde o ácido nucleico não codifica uma sequência de aminoácidos, às sequências essencialmente idênticas. Uma pessoa de habilidade comum na técnica reconhecerá que substituições, deleções ou adições individuais, que alteram, adicionam ou deletam um único aminoácido ou uma pequena porcentagem de aminoácidos em uma sequência codificada, são “variações conservativamente modificadas” ou “variantes conservativamente modificadas” onde as alterações resultam na deleção de um aminoácido, na

25 30 adição de um aminoácido ou na substituição de um aminoácido por um aminoácido quimicamente similar. Assim, as “variações conservativas” de uma sequência de polipeptídeos listada da presente invenção incluem substituições de uma pequena porcentagem, tipicamente menos que 5 %, mais tipicamente menos que 4 %, 2 % ou 1 %, dos aminoácidos da sequência de polipeptídeos, por um aminoácido conservativamente selecionado do mesmo grupo de substituição conservativa. A adição de sequências que não alteram a atividade codificada de uma molécula de ácido nucleico, tal como a adição de uma sequência não funcional, é uma variação conservativa do ácido nucleico básico.

35

As tabelas de substituições conservativas que fornecem aminoácidos funcionalmente similares são conhecidas àqueles de habilidade comum na técnica. Cada um dos oito grupos seguintes contém aminoácidos que são substituições conservativas entre si:

- 1) Alanina (A), Glicina (G);
 - 5 2) Ácido Aspártico (D), Ácido Glutâmico (E);
 - 3) Asparagina (N), Glutamina (Q);
 - 4) Arginina (R), Lisina (K);
 - 5) Isoleucina (I), Leucina (L), Metionina (M), Valina (V);
 - 6) Fenilalanina (F), Tirosina (Y), Triptofano (W);
 - 10 7) Serina (S), Treonina (T); e
 - 8) Cisteína (C), Metionina (M)
- (veja, por exemplo, Creighton, *Proteins: Structures and Molecular Properties* (W H Freeman & Co.; 2ª edição (Dezembro, 1993))

Hibridização de Ácidos Nucleicos

- 15 A hibridização comparativa pode ser usada para identificar ácidos nucleicos da presente invenção, tal como SEQ ID NO.: 1 a 3, incluindo variações conservativas de ácidos nucleicos da presente invenção, e este método de hibridização comparativa é um método preferido para distinguir ácidos nucleicos da presente invenção. Além disso, os ácidos nucleicos alvo que se hibridizam com os ácidos nucleicos representados pela SEQ ID NO.: 1 a 3
- 20 sob condições de alta, ultra-alta e/ou ultra-ultra-alta estringências são uma característica da presente invenção. Os exemplos de tais ácidos nucleicos incluem aqueles com uma ou poucas substituições de ácido nucleico silenciosas ou conservativas se comparadas a uma sequência de ácidos nucleicos fornecida.

- Diz-se que um ácido nucleico teste se hibridiza especificamente com uma sonda de
- 25 ácido nucleico quando ele se hibridiza pelo menos $1/2$ à sonda quanto ao alvo complementar de ligação perfeita, isto é, com uma razão sinal/ruído de pelo menos $1/2$ tão alta quanto a da hibridização da sonda ao alvo sob condições nas quais a sonda de ligação perfeita se liga ao alvo complementar de ligação perfeita com uma razão sinal/ruído que é pelo menos cerca de 5x-10x tão alta quanto aquela observada para a hibridização de qualquer um dos ácidos nucleicos alvo não ligados.
- 30 Os ácidos nucleicos se “hibridizam” quando eles se associam, tipicamente em solução. Os ácidos nucleicos se hibridizam devido a uma variedade de forças físico-químicas bem caracterizadas, tais como ligação de hidrogênio, exclusão de solventes, empilhamento de base (*base stacking*) e semelhantes. O termo “condições de hibridização severas” refere-

- 35 se às condições de baixa força iônica e alta temperatura, conforme é conhecido na técnica. Tipicamente, sob condições severas, uma sonda se hibridizará a sua subsequência alvo em uma mistura complexa de ácido nucleico (incluindo, mas não limitada a DNA ou RNA celular

total ou de biblioteca), mas não se hibridiza a outras sequências na mistura complexa. Um guia extensivo para a hibridização de ácidos nucleicos é encontrado em Tijssen (1993) *Laboratory Techniques in Biochemistry and Molecular Biology-Hybridization with Nucleic Acid Probes* parte I capítulo 2, "Overview of principles of hybridization and the strategy of nucleic acid probe assays", (Elsevier, Nova Iorque), assim como em Ausubel, *et al.*, *Current Protocols in Molecular Biology* (1995). Hames e Higgins (1995) Gene Probes 1 IRL impresso em Oxford University Press, Oxford, England (Hames e Higgins 1) e Hames e Higgins (1995) Gene Probes 2 IRL impresso em Oxford University Press, Oxford, England (Hames e Higgins 2) e fornece detalhes sobre a síntese, marcação, detecção e quantificação de DNA e RNA, incluindo oligonucleotídeos. Geralmente, as condições severas são selecionadas em aproximadamente 5 a 10 °C inferiores ao ponto de fusão térmico (T_m) para a sequência específica em um pH de força iônica definida. O T_m é a temperatura (sob força iônica, pH e concentração nucleica definidos) na qual 50 % das sondas complementares ao alvo hibridizam à sequência alvo em equilíbrio (como as sequências alvo estão presentes em excesso, no T_m , 50 % das sondas são ocupadas em equilíbrio). Condições severas podem ser aquelas em que a concentração de sal é menor do que cerca de 1,0 M de íon sódio, tipicamente cerca de 0,01 a 1,0 M de concentração de íon sódio (ou outros sais) ao pH 7,0 a 8,3 e a temperatura é pelo menos cerca de 30 °C para sondas curtas (incluindo, mas não limitadas entre 10 a 50 nucleotídeos) e pelo menos cerca de 60 °C para sondas longas (incluindo, mas não limitadas a mais do que 50 nucleotídeos). Condições severas também podem ser obtidas com a adição de agentes desestabilizantes, tais como formamida. Para a hibridização seletiva ou específica, um sinal positivo pode ser pelo menos duas vezes o ruído, opcionalmente 10 vezes a hibridização de ruído. Condições de hibridização severas exemplares podem ser as seguintes: 50 % de formamida, SSC 5X e 1 % de SDS, incubação a 42 °C ou SSC 5X, 1 % de SDS, incubação a 65 °C, com lavagem em SSC 0,2X e 0,1 % de SDS a 65 °C. Tais lavagens podem ser realizadas durante 5, 15, 30, 60, 120 minutos ou mais.

Um exemplo de condições de hibridização severas para a hibridização de ácidos nucleicos complementares, os quais têm mais de 100 resíduos complementares em um filtro em um *Southern* ou *Northern blot*, é formalina 50 % com 1 mg de heparina a 42 °C, com a hibridização realizada durante a noite. Um exemplo de condições de lavagem severas é uma lavagem com SSC 0,2X a 65 °C durante 15 minutos (veja, Sambrook, *et al.*, *Molecular Cloning: A Laboratory Manual* (3ª ed., 2001) para uma descrição do tampão SSC). Frequentemente, a lavagem de alta estrigência é precedida por uma lavagem de baixa estrigência para a remoção do sinal de fundo da sonda. Uma lavagem de baixa estrigência exemplar é SSC 2X a 40 °C durante 15 minutos. Em geral, uma razão sinal/ruído de 5x (ou mais alta) do que aquela observada para uma sonda não relacionada no ensaio de hibridização particular indica a detecção de uma hibridização específica.

As “condições de lavagem de hibridização severas”, no contexto de experimentos de hibridização de ácidos nucleicos, tais como hibridizações de *Southern* e de *Northern* são dependentes de sequência, e são diferentes sob parâmetros ambientais diferentes. As sequências mais longas se hibridizam especificamente em altas temperaturas. Um guia exten-

5 sivo para a hibridização de ácidos nucleicos é encontrado em Tijssen (1993), *supra*, e em Hames e Higgins 1 e 2, *supra*. A hibridização severa e as condições de lavagem podem ser facilmente determinadas empiricamente para qualquer ácido nucleico de teste. Por exemplo, na determinação da hibridização altamente severa e das condições de lavagem, a hibridiza-

10 ção e as condições de lavagem são aumentadas gradualmente (por exemplo, aumentando-se a temperatura, diminuindo-se a concentração de sal, aumentando-se a concentração de detergente e/ou aumentando-se a concentração de solventes orgânicos, tais como a formali-

15 na, na hibridização ou na lavagem), até que um grupo selecionado de critérios seja satisfeito. Por exemplo, as condições de lavagem e hibridização são gradualmente aumentadas até que uma sonda se ligue a um alvo complementar de ligação perfeita com uma razão sinal/ruído que seja pelo menos 5x tão alta quanto aquela observada para a hibridização da sonda com um alvo não ligado.

As condições “muito severas” são selecionadas para serem iguais ao ponto de fusão térmico do DNA (T_m) para uma sonda particular. O T_m é a temperatura (sob força iônica e pH definidos) na qual 50 % da sequência de teste se hibridiza com uma sonda de ligação

20 perfeita. Para as finalidades da presente invenção, geralmente, condições de lavagem e hibridização “altamente severas” são selecionadas em aproximadamente 5 °C inferiores ao T_m para a sequência específica sob uma força iônica e pH definidos.

As condições de lavagem e hibridização de “ultra-alta estringência” são aquelas nas quais a estringência das condições de lavagem e hibridização é aumentada até que a razão

25 sinal/ruído para a ligação da sonda ao ácido nucleico alvo complementar de ligação perfeita seja pelo menos 10x tão alta quanto aquela observada para a hibridização com qualquer um dos ácidos nucleicos alvo não ligados. Um ácido nucleico alvo que hibridiza a uma sonda sob tais condições, com uma razão sinal/ruído de pelo menos $\frac{1}{2}$ daquela do ácido nucleico alvo complementar de ligação perfeita, que ele se liga à sonda sob condições de ultra-alta

30 estringência.

De maneira similar, mesmo os níveis mais altos de estringência podem ser determinados aumentando-se gradualmente as condições de hibridização e/ou lavagem do ensaio de hibridização relevante. Por exemplo, aqueles nos quais a estringência das condições de hibridização e lavagem é aumentada até que a razão sinal/ruído para ligação da sonda

35 ao ácido nucleico alvo complementar de ligação perfeita seja pelo menos 10x, 20x, 50x, 100x, 500x ou mais tão alta quanto aquelas observadas para a hibridização de qualquer um dos ácidos nucleicos alvo não ligados. Um ácido nucleico alvo que hibridiza a uma sonda

sob tais condições, com uma razão sinal/ruído de pelo menos $\frac{1}{2}$ daquela do ácido nucleico alvo complementar de ligação perfeita, que ele se liga à sonda sob condições de ultra-ultra-alta estringência.

- Os ácidos nucleicos que não hibridizam entre si sob condições severas ainda serão consideravelmente idênticos se os polipeptídeos que eles codificam são consideravelmente idênticos. Isto ocorre, por exemplo, quando uma cópia de um ácido nucleico é criada usando a degeneração de códons máxima permitida pelo código genético.

Subsequências únicas

- Em um aspecto, a invenção fornece um ácido nucleico que compreende uma subsequência única em um ácido nucleico selecionado a partir das sequências de O-tRNAs e O-RSs divulgadas neste relatório. A subsequência única é única em comparação a um ácido nucleico que corresponde a qualquer sequência de ácidos nucleicos de O-tRNA ou O-RS conhecida. Um alinhamento pode ser realizado usando, por exemplo, BLAST ajustado aos parâmetros pré-determinados. Qualquer subsequência única é útil, por exemplo, como uma sonda para identificar os ácidos nucleicos da presente invenção.

De maneira similar, a invenção inclui um polipeptídeo que compreende uma subsequência única em um polipeptídeo selecionado a partir das sequências de O-RSs divulgadas neste relatório. Neste ponto, a subsequência única é única em comparação a um polipeptídeo que corresponde a qualquer sequência de polipeptídeos conhecida.

- A invenção também fornece ácidos nucleicos alvo que se hibridizam sob condições severas com um oligonucleotídeo codificante único que codifica uma subsequência única em um polipeptídeo selecionado a partir das sequências de O-RSs, em que a subsequência única é única em comparação a um polipeptídeo que corresponde a qualquer um dos polipeptídeos controle (por exemplo, sequências parentais a partir das quais sintetases da invenção foram derivadas, por exemplo, por mutação). As sequências únicas são determinadas, conforme observado acima.

Comparação, identidade e homologia de sequências

- Os termos “idêntico” ou porcentagem de “identidade”, no contexto de duas ou mais sequências de polipeptídeos ou de ácidos nucleicos, referem-se a duas ou mais sequências ou subsequências que são as mesmas ou têm uma porcentagem especificada de resíduos de aminoácidos ou nucleotídeos que são os mesmos, quando comparados e alinhados para uma correspondência máxima, em uma janela de comparação, ou região designada, conforme medido usando um dos algoritmos de comparação de sequências descritos abaixo (ou outros algoritmos disponíveis a pessoas habilitadas na técnica) ou por alinhamento manual e inspeção visual.

O termo “substancialmente idêntico”, no contexto de dois ácidos nucleicos ou polipeptídeos (por exemplo, DNAs que codificam um O-tRNA ou uma O-RS ou a sequência de

aminoácidos de uma O-RS), refere-se a duas ou mais sequências ou subsequências que possuem pelo menos cerca de 60 %, cerca de 80 %, cerca de 90 a 95 %, cerca de 98 %, cerca de 99 % ou mais de identidade entre nucleotídeos ou entre resíduos de aminoácidos, quando comparados e alinhados para uma correspondência máxima em uma janela de comparação, ou região designada, conforme medido usando um algoritmo de comparação de sequências (ou outros algoritmos disponíveis a pessoas de habilidade comum da técnica) ou por alinhamento manual e inspeção visual. Tais sequências “substancialmente idênticas” são tipicamente consideradas como “homólogas”, sem referência à ancestralidade real. A “identidade substancial” pode existir em uma região das sequências que tem pelo menos cerca de 50 resíduos em comprimento, uma região de pelo menos cerca de 100 resíduos, ou uma região de pelo menos cerca de 150 resíduos, ou por todo o comprimento das duas sequências a serem comparadas.

Para comparação de sequências e determinação de homologia, tipicamente, uma sequência atua como uma sequência de referência a qual sequências de teste são comparadas. Quando se usa um algoritmo de comparação de sequências, sequências de teste e de referência são inseridas em um computador, coordenadas de subsequência são designadas, se necessário, e parâmetros do programa de algoritmo de sequência são designados. O algoritmo de comparação de sequência então calcula a porcentagem de identidade de sequência para a(s) sequência(s) de teste em relação à sequência de referência, com base nos parâmetros do programa designado.

Os métodos de alinhamento de sequências para comparação são conhecidos àqueles de habilidade comum na técnica. O alinhamento ideal de sequências para comparação pode ser conduzido pelo algoritmo de homologia local de Smith e Waterman (1970) *Adv. Appl. Math.* 2:482c, pelo algoritmo de alinhamento de homologia de Needleman e Wunsch (1970) *J. Mol. Biol.* 48:443, pela pesquisa para o método de similaridade de Pearson e Lipman (1988) *Proc. Nat'l. Acad. Sci. USA* 85:2444, por implementações computadorizadas destes algoritmos (GAP, BESTFIT, FASTA e TFASTA no Wisconsin Genetics Software Package, Genetics Computer Group, 575 Science Dr., Madison, WI), ou por alinhamento manual e inspeção visual (veja, por exemplo, Ausubel *et al.*, *Current Protocols in Molecular Biology* (1995 suplemento)).

Um exemplo de um algoritmo que é adequado para determinar a porcentagem de identidade de sequência e similaridade de sequência são os algoritmos BLAST e BLAST 2.0, que são descritos em Altschul *et al.* (1997) *Nuc. Acids Res.* 25:3389-3402, e Altschul *et al.* (1990) *J. Mol. Biol.* 215:403-410 (1990). O software para realizar análises BLAST está publicamente disponível através do National Center for Biotechnology Information disponível na World Wide Web em ncbi.nlm.nih.gov. Este algoritmo envolve inicialmente a identificação de pares de sequência de alta pontuação (HSPs) pela identificação de palavras curtas de

comprimento W na sequência de teste, que se liga ou satisfaz algumas pontuações de limiar valoradas positivamente T quando alinhadas com uma palavra do mesmo comprimento em uma sequência de base de dados. T é referido como o limiar de pontuação de palavra da vizinhança (Altschul et al, *supra*). Estes sucessos de palavras da vizinhança iniciais atuam como origens para o início das buscas para encontrar HSPs maiores do que os contêm.

Os sucessos de palavras são então prolongados em ambos os sentidos ao longo de cada sequência até o ponto em que a pontuação cumulativa do alinhamento possa ser aumentada. As pontuações cumulativas são calculadas usando, para sequências de nucleotídeos, os parâmetros M (pontuação de recompensa para um par de resíduos pareados; sempre >0) e N (pontuação de penalidade para resíduos mal pareados; sempre <0). Para as sequências de aminoácidos, uma matriz de pontuação é usada para calcular a pontuação cumulativa. A extensão dos sucessos de palavras em cada direção é detida quando: a pontuação cumulativa do alinhamento diminui pela quantidade de X a partir do seu valor máximo obtido; a pontuação cumulativa vai para zero ou abaixo, devido ao acúmulo de um ou mais alinhamentos de resíduos de pontuação negativa; ou o fim de cada sequência é atingido. Os parâmetros do algoritmo BLAST W, T e X determinam a sensibilidade e velocidade do alinhamento. O programa BLASTN (para sequências de nucleotídeos) usa como padrão um comprimento de palavra (W) de 11, uma expectativa (E) de 10, um corte de 100, M = 5, N = -4, e uma comparação de ambas as fitas. Para sequências de aminoácidos, o programa BLASTP usa como padrão um comprimento de palavra (W) de 3, uma expectativa (E) de 10, e alinhamentos (B) de matriz de pontuação BLOSUM62 (veja Henikoff & Henikoff (1992) Proc. Nat. Acad. Sci. USA 89:10915) de 50, expectativa (E) de 10, M = 5, N = -4, e uma comparação de ambas as fitas. O algoritmo de BLAST é tipicamente realizado com o filtro de “baixa complexidade” desativado.

Além de calcular a porcentagem de identidade de sequência, o algoritmo BLAST também realiza uma análise estatística da similaridade entre duas sequências (veja, por exemplo, Karlin e Altschul (1993) Proc. Natl. Acad. Sci. USA 90:5873-5787). Uma medição de similaridade fornecida pelo algoritmo BLAST é a menor probabilidade de soma (P(N)), que fornece uma indicação da probabilidade pela qual uma comparação entre duas sequências de nucleotídeos ou aminoácidos ocorreria por acaso. Por exemplo, um ácido nucleico é considerado similar a uma sequência de referência se a menor probabilidade de soma em uma comparação do ácido nucleico de teste ao ácido nucleico de referência é menor do que cerca de 0,2, ou menor do que cerca de 0,01, ou menor do que cerca de 0,001.

Mutagênese e Outras Técnicas de Biologia Molecular

Os polinucleotídeos e polipeptídeos da presente invenção e usados na invenção podem ser manipulados usando técnicas de biologia molecular. Uma sequência de nucleotídeos pode ser convenientemente modificada por mutagênese sítio-dirigida, de acordo com

métodos convencionais. Alternativamente, a sequência de nucleotídeos pode ser preparada por síntese química, incluindo, mas não limitada pelo uso de um sintetizador oligonucleotídico, em que os oligonucleotídeos são designados com base na sequência de aminoácidos do polipeptídeo desejado e de preferência, selecionando aqueles códons que são favorecidos na célula hospedeira, em que o polipeptídeo recombinante será produzido. Por exemplo, vários oligonucleotídeos pequenos que codificam porções do polipeptídeo desejado podem ser sintetizados e montados por PCR, ligação ou reação em cadeia de ligação. Veja, por exemplo, Barany *et al.*, *Proc. Natl. Acad. Sci.* 88:189-193 (1991); Patente U.S. 6.521.427, os quais são incorporados como referência neste relatório.

Esta invenção utiliza técnicas de rotina no campo da genética recombinante. Os textos básicos que divulgam métodos gerais de uso nesta invenção incluem Sambrook *et al.*, *Molecular Cloning, A Laboratory Manual* (3ª Ed. 2001); Kriegler, *Gene Transfer and Expression: A Laboratory Manual* (1990); e *Current Protocols in Molecular Biology* (Ausubel, *et al.*, eds., 1994).

Os textos gerais que descrevem técnicas de biologia molecular incluem Berger e Kimmel, Guide to Molecular Cloning Techniques, Methods in Enzymology volume 152 Academic Press, Inc., San Diego, CA (Berger); Sambrook *et al.*, Molecular Cloning - A Laboratory Manual (3ª Ed.) Vol. 1-3, Cold Spring Harbor Laboratory, Cold Spring Harbor, Nova Iorque, 2001 ("Sambrook") e Current Protocols in Molecular Biology, F. M. Ausubel *et al.*, eds., Current Protocols, a joint venture between Greene Publishing Associates, Inc. e John Wiley & Sons, Inc., (suplementado em 1999) ("Ausubel"). Estes textos descrevem a mutagenese, o uso de vetores, promotores e muitos outros tópicos relevantes relacionados, por exemplo, à criação de genes ou polinucleotídeos que incluem códons seletores para a produção de proteínas que incluem aminoácidos selecionados (por exemplo, aminoácidos não naturais), tRNAs ortogonais, sintetases ortogonais e pares dos mesmos.

Vários tipos de mutagenese são usados na invenção para uma variedade de propósitos, incluindo, mas não limitada à produção de novas sintetases ou tRNAs, modificação de moléculas de tRNA, produção de bibliotecas de tRNAs, modificação de tRNA, produção de bibliotecas de tRNAs, modificação de moléculas de RS, produção de bibliotecas de sintetases, produzir códons seletores, inserção de códons seletores que codificam um aminoácido selecionado em uma proteína ou polipeptídeo de interesse. Eles incluem, mas não são limitados à mutagenese sítio-dirigida, mutagenese de ponto aleatória, recombinação homóloga, embaralhamento do DNA ou outros métodos de mutagenese recursivos, construção quimérica, mutagenese usando moldes que contém uracila, mutagenese dirigida por oligonucleotídeo, mutagenese de DNA modificado por fosforotioato, mutagenese usando DNA duplex com interrupções ou semelhantes ou qualquer combinação dos mesmos. Métodos adequados adicionais incluem reparo de erro de pareamento pontual, mutagenese usando cepas

hospedeiras deficientes em reparo, seleção por restrição e purificação por restrição, mutagênese por deleção, mutagênese por síntese gênica total, reparo de quebra de dupla fita e semelhantes. A mutagênese, incluindo, mas não limitadas as que envolvem construções quiméricas, também está incluída na presente invenção. Em uma modalidade, a mutagênese pode ser guiada pela informação conhecida sobre a molécula que ocorre naturalmente ou sobre a molécula que ocorre naturalmente modificada ou alterada, incluindo, mas não limitada à sequência, comparações de sequências, propriedades físicas, estrutura secundária, terciária ou quaternária, estrutura cristalina ou semelhantes.

Os textos e exemplos encontrados neste relatório descrevem estes procedimentos.

- 10 A informação adicional é encontrada nas seguintes publicações e referências citadas a este respeito: Ling *et al.*, *Approaches to DNA mutagenesis: an overview*, Anal Biochem. 254(2):157-178 (1997); Dale *et al.*, *Oligonucleotide-directed random mutagenesis using the phosphorothioate method*, Methods Mol. Biol. 57:369-374 (1996); Smith, *In vitro mutagenesis*, Ann. Rev. Genet. Methods Mol. Biol. 19:423-462 (1985); Botstein & Shortle, *Strategies and applications of in vitro mutagenesis*, Science 229:1193-1201 (1985); Carter, *Site-directed mutagenesis*, Biochem. J. 237:1-7 (1986); Kunkel, *The efficiency of oligonucleotide directed mutagenesis*, in Nucleic Acids & Molecular Biology (Eckstein, F. e Lilley, D. M. J. Eds., Springer Verlag, Berlin) (1987); Kunkel, *Rapid and efficient site-specific mutagenesis without phenotypic selection*, Proc. Natl. Acad. Sci. USA 82:488-492 (1985); Kunkel *et al.*, *Rapid and efficient site-specific mutagenesis without phenotypic selection*, Methods in Enzymol. 154:367-382 (1987); Bass *et al.*, *Mutant Trp repressors with new DNA-binding specificities*, Science 242:240-245 (1988); Zoller & Smith, *Oligonucleotide-directed mutagenesis using M13-derived vectors: an efficient and general procedure for the production of point mutations in any DNA fragment*, Nucleic Acids Res. 10:6487-6500 (1982); Zoller & Smith, *Oligonucleotide-directed mutagenesis of DNA fragments cloned into P13 vectors*, Methods in Enzymol. 100:468-500 (1983); Zoller & Smith, *Oligonucleotide-directed mutagenesis: a simple method using two oligonucleotide primers and a single-stranded DNA template*, Methods in Enzymol. 154:329-350 (1987); Taylor *et al.*, *The use of phosphorothioate-modified DNA in restriction enzyme reactions to prepare nicked DNA*, Nucl. Acids Res. 13:8749-8764 (1985); Taylor *et al.*, *The rapid generation of oligonucleotide-directed mutations at high frequency using phosphorothioate-modified DNA*, Nucl. Acids Res. 13:8765-8785 (1985); Nakamaye & Eckstein, *Inhibition of restriction endonuclease Nci I cleavage by phosphorothioate groups and its application to oligonucleotide-directed mutagenesis*, Nucl. Acids Res. 14:9679-9698 (1986); Sayers *et al.*, *5'-3' Exonucleases in phosphorothioate-based oligonucleotide-directed mutagenesis*, Nucleic Acids Res., 16(3):791-802 (1988); Sayers *et al.*, *Strand specific cleavage of phosphorothioate-containing DNA by reaction with restriction endonucleases in the presence of ethidium bromide*, (1988) Nucleic Acids Res., 16:803-814; Kramer *et al.*, *The*

gapped duplex DNA approach to oligonucleotide-directed mutation construction, Nucleic Acids Res., 12:9441-9456 (1984); Kramer & Fritz *Oligonucleotide-directed construction of mutations via gapped duplex DNA*, Methods in Enzymol. 154:350-367 (1987); Kramer *et al.*, *Improved enzymatic in vitro reactions in the gapped duplex DNA approach to oligonucleotide-directed construction of mutations*, Nucleic Acids Res., 16:7207 (1988); Fritz *et al.*, *Oligonucleotide-directed construction of mutations: a gapped duplex DNA procedure without enzymatic reactions in vitro*, Nucleic Acids Res., 16:6987-6999 (1988); Kramer *et al.*, *Different base/base mismatches are corrected with different efficiencies by the methyl-directed DNA mismatch repair system of E. Coli*, Cell 38:879-887 (1984); Carter *et al.*, *Improved oligonucleotide site-directed mutagenesis using M13 vectors*, Nucleic Acids Res., 13:4431-4443 (1985), Carter, *Improved oligonucleotide-directed mutagenesis using M13 vectors*, Methods in Enzymol. 154:382-403 (1987); Eghtedarzadeh & Henikoff, *Use of oligonucleotides to generate large deletions*, Nucl. Acids Res. 14:5115 (1986); Wells *et al.*, *Importance of hydrogen-bond formation in stabilizing the transition state of subtilisin*, Phil. Trans. R. Soc. Lond. A 317:415-423 (1986); Nambiar *et al.*, *Total synthesis and cloning of a gene coding for the ribonuclease S protein*, Science 223:1299-1301 (1984); Sakmar e Khorana, *Total synthesis and expression of a gene for the alpha-subunit of bovine rod outer segment guanine nucleotide-binding protein (transducin)*, Nucl. Acids Res. 14:6361-6372 (1988); Wells *et al.*, *Cassette mutagenesis: an efficient method for generation of multiple mutations at defined sites*, Gene 34:315-323 (1985); Grundström *et al.*, *Oligonucleotide-directed mutagenesis by microscale 'shot-gun' gene synthesis*, Nucl. Acids Res. 13:3305-3316 (1985); Mandecki, *Oligonucleotidedirected double-strand break repair in plasmids of Escherichia coli: a method for site-specific mutagenesis*, Proc. Natl. Acad. Sci. USA, 83:7177-7181 (1986); Arnold, *Protein engineering for unusual environments*, Current Opinion in Biotechnology 4:450-455 (1993); Sieber, *et al.*, Nature Biotechnology, 19:456-460 (2001); W. P. C. Stemmer, Nature 370:389-91 (1994); e I. A. Lorimer, I. Pastan, Nucl. Acids Res. 23:3067-8 (1995). Os detalhes adicionais em muitos desses métodos podem ser encontrados em Methods in Enzymology Volume 154, que também descreve controles úteis para problemas *trouble-shooting* com vários métodos de mutagênese.

Os oligonucleotídeos, por exemplo, para o uso em mutagênese da presente invenção, por exemplo, modificando bibliotecas de tRNAs ou sintetases, ou alterando tRNAs ou RRs, são tipicamente quimicamente sintetizados, de acordo com o método de triéster de fosforamidita de fase sólida descrito por Beaucage e Caruthers, Tetrahedron Letts. 22(20):1859-1862 (1981), por exemplo, usando um sintetizador automático, conforme descrito em Needham-VanDevanter *et al.*, Nucleic Acids Res. 12:6159-6168 (1984).

Além disso, essencialmente, qualquer ácido nucleico pode ser customizado ou fabricado sob medida de qualquer dentre uma variedade de fontes comerciais, tal como The

Midland Certified Reagent Company (mcrc@oligos.com), The Great American Gene Company (www.genco.com), ExpressGen Inc. (www.expressgen.com), Operon Technologies Inc. (Alameda, Calif.), entre outras.

A invenção também refere-se às células hospedeiras eucarióticas, células hospedeiras não eucarióticas e organismos para a incorporação *in vivo* de um aminoácido não natural por intermédio de pares de tRNA/RS ortogonais. As células hospedeiras são geneticamente construídas (incluindo, mas não limitadas às transformadas, transduzidas ou transfectadas) com os polinucleotídeos da presente invenção ou constructos que incluem um polinucleotídeo da presente invenção, incluindo, mas não limitados a um vetor da presente invenção, que, por exemplo, pode ser um vetor de clonagem ou um vetor de expressão. Por exemplo, as regiões codificantes para o tRNA ortogonal, a tRNA sintetase ortogonal e a proteína a ser derivada são, de maneira operável, ligadas aos elementos de controle da expressão gênica que são funcionais na célula hospedeira desejada. O vetor, por exemplo, pode estar na forma de um plasmídeo, um cosmídeo, um fago, uma bactéria, um vírus, um polinucleotídeo puro ou um polinucleotídeo conjugado. Os vetores são introduzidos em células e/ou microorganismos por métodos padrão que incluem eletroporação (Fromm *et al.*, (1985) Proc. Natl. Acad. Sci. USA 82, 5824), infecção por vetores virais, penetração balística de alta velocidade por pequenas partículas com o ácido nucleico ou dentro da matriz de pequenas microesferas ou partículas ou sobre a superfície (Klein *et al.*, (1987) *Nature* 327, 70-73) e/ou semelhantes.

Muitos métodos bem conhecidos para introduzir ácidos nucleicos alvo em células são disponíveis, qualquer um dos quais pode ser usado na invenção. Estes incluem: fusão das células receptoras com protoplastos bacterianos contendo o DNA, eletroporação, bombardeamento de projétil e infecção com vetores virais (debatida adicionalmente abaixo), etc. As células bacterianas podem ser usadas para amplificar o número de plasmídeos contendo constructos de DNA desta invenção. As bactérias são cultivadas até a fase log e os plasmídeos nas bactérias podem ser isolados por uma variedade de métodos conhecidos na técnica (veja, por exemplo, Sambrook). Além disso, os kits são comercialmente disponíveis para a purificação de plasmídeos de bactérias, (veja, por exemplo, EasyPrep™, FlexiPrep™, ambos da Pharmacia Biotech; StrataClean™ da Stratagene; e QIAprep™ da Qiagen). Os plasmídeos isolados e purificados depois são manipulados para produzir outros plasmídeos, usados para transfectar células ou incorporados em vetores relacionados para infectar organismos. Vetores típicos contêm terminadores de transcrição e tradução, sequências de iniciação de transcrição e tradução e promotores úteis para a regulação da expressão do ácido nucleico alvo particular. Os vetores opcionalmente compreendem cassetes de expressão genéricos contendo pelo menos uma sequência terminadora independente, sequências que permitem a replicação do cassete em eucariontes, procariontes ou ambos (incluindo, mas

não limitadas aos vetores *shuttle*) e marcadores de seleção para sistemas procarióticos e eucarióticos. Vetores são adequados para replicação e integração em procariontes, eucariotes ou ambos. Veja, Gillam & Smith, *Gene* 8:81 (1979); Roberts, *et al.*, *Nature*, 328:731 (1987); Schneider, E., *et al.*, *Protein Expr. Purif.* 6(1): 10-14 (1995); Ausubel, Sambrook, Berger (*supra*). Um catálogo de bactérias e bacteriófagos úteis para clonagem é fornecido, por exemplo, pela ATCC, por exemplo, O ATCC Catalogue of Bacteria and Bacteriophage (1992) Gherna *et al.* (eds) publicado pela ATCC. Procedimentos básicos adicionais para sequenciamento, clonagem e outros aspectos da biologia molecular e considerações teóricas essenciais também são encontrados em Watson *et al.* (1992) *Recombinant DNA Second Edition* Scientific American Books, NY. Além disso, essencialmente, qualquer ácido nucleico (e, na prática, qualquer ácido nucleico marcado, se padrão ou não padrão) pode ser customizado ou na forma padrão a partir de qualquer variedade de fontes comerciais, tais como a Midland Certified Reagent Company (Midland, TX disponível na World Wide Web em mcrc.com), The Great American Gene Company (Ramona, CA disponível na World Wide Web em genco.com), ExpressGen Inc. (Chicago, IL disponível na World Wide Web em expressgen.com), Operon Technologies Inc. (Alameda, CA), entre outros.

As células hospedeiras construídas podem ser cultivadas em meios nutrientes convencionais modificados, conforme apropriado, para tais atividades como, por exemplo, etapas de triagem, ativação de promotores ou seleção de transformantes. Estas células podem opcionalmente ser cultivadas em organismos transgênicos. Outras referências úteis, por exemplo, para cultura e isolamento de células (por exemplo, para isolamento subsequente de ácido nucleico) incluem Freshney (1994) Culture of Animal Cells, a Manual of Basic Technique, terceira edição, Wiley-Liss, Nova Iorque e as referências citadas neste relatório; Payne *et al.* (1992) Plant Cell and Tissue Culture in Liquid Systems John Wiley & Sons, Inc. Nova Iorque, NY; Gamborg e Phillips (eds) (1995) Plant Cell Tissue and Organ Culture; Fundamental Methods Springer Lab Manual, Springer-Verlag (Berlin Heidelberg Nova Iorque) e Atlas e Parks (eds) The Handbook of Microbiological Media (1993) CRC Press, Boca Raton, FL.

A habilidade para incorporar aminoácidos não naturais diretamente em proteínas *in vivo* oferece uma ampla variedade de vantagens incluindo, mas não limitadas aos altos rendimentos de proteínas mutantes, facilidade técnica, o potencial para estudar as proteínas mutantes em células ou, possivelmente, em organismos vivos, e o uso dessas proteínas mutantes em tratamentos terapêuticos e usos diagnósticos. A habilidade para incluir aminoácidos não naturais com vários tamanhos, acidez, nucleofilicidades, hidrofobicidades e outras propriedades em proteínas pode expandir enormemente a habilidade para manipular racional e sistematicamente as estruturas de proteínas, tanto para investigar a função proteica quanto para criar novas proteínas ou organismos com novas propriedades.

PROTEÍNAS E POLIPEPTÍDEOS DE INTERESSE

A incorporação de um aminoácido não natural pode ser feita para uma variedade de propósitos, incluindo, mas não limitados às mudanças sob medida na estrutura proteica e/ou função, mudança de tamanho, acidez, nucleofilicidade, ligação de hidrogênio, hidrofobicidade, acessibilidade de sítios alvo de proteases, alveijamento a uma porção (incluindo, mas não limitado a, para um arranjo de proteínas), adição de uma molécula biologicamente ativa, adesão de um polímero, adesão de um radionuclídeo, modulação de meia-vida sérica, modulação de penetração tecidual (por exemplo, tumores), modulação de transporte ativo, modulação tecidual, especificidade ou distribuição à célula ou órgão, modulação da imunogenicidade, modulação da resistência à protease, etc. Proteínas, que incluem um aminoácido não natural, podem ter novas propriedades catalíticas ou biofísicas melhoradas ou ainda inteiramente renovadas. Por exemplo, as seguintes propriedades são opcionalmente modificadas pela inclusão de um aminoácido não natural em uma proteína: toxicidade, biodistribuição, propriedades estruturais, propriedades espectroscópicas, propriedades químicas e/ou fotoquímicas, habilidade catalítica, meia vida (incluindo, mas não limitada à meia-vida sérica), habilidade para reagir com outras moléculas, incluindo, mas não limitada a, covalente ou não covalentemente, e semelhantes. As composições, incluindo proteínas que incluem pelo menos um aminoácido não natural, são úteis, incluindo, mas não limitadas para novas terapias, diagnósticos, enzimas catalíticas, enzimas industriais, proteínas de ligação (incluindo, mas não limitadas a anticorpos) e, incluindo, mas não limitadas ao estudo de estrutura e função proteicas. Veja, por exemplo, Dougherty, (2000) *Unnatural Amino Acids as Probes of Protein Structure and Function*, Current Opinion in Chemical Biology, 4:645-652.

Uma proteína pode ter pelo menos um, pelo menos dois, pelo menos três, pelo menos quatro, pelo menos cinco, pelo menos seis, pelo menos sete, pelo menos oito, pelo menos nove ou pelo menos dez ou mais aminoácidos não naturais. Os aminoácidos não naturais podem ser os mesmos ou diferentes e podem estar em 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 ou mais sítios diferentes na proteína que compreende 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 ou mais aminoácidos não naturais diferentes. Uma proteína pode ter pelo menos um, mas menos que todos, de um aminoácido particular, presente na proteína, é substituído por um aminoácido não natural. Para uma proteína fornecida com mais de um aminoácido não natural, os aminoácidos não naturais podem ser idênticos ou diferentes (a proteína pode incluir dois ou mais tipos diferentes de aminoácidos não naturais ou pode incluir dois dos mesmos aminoácidos não naturais). Para uma proteína fornecida com mais de dois aminoácidos não naturais, os aminoácidos não naturais podem ser os mesmos, diferentes ou uma combinação de um aminoácido não natural múltiplo do mesmo tipo com pelo menos um aminoácido não natural diferente.

Através da produção de proteínas ou polipeptídeos de interesse com pelo menos

um aminoácido não natural em células eucarióticas, proteínas ou polipeptídeos incluirão tipicamente modificações pós-tradução eucarióticas. Em certas modalidades, uma proteína inclui pelo menos um aminoácido não natural e pelo menos uma modificação pós-tradução que é feita *in vivo* por uma célula eucariótica, onde a modificação pós-tradução não é feita por uma célula procariótica. Por exemplo, a modificação pós-tradução inclui acetilação, acilação, modificação lipídica, palmitoilação, adição de palmitato, fosforilação, modificação por ligação de glicolípido, glicosilação e semelhantes. Ainda em um outro aspecto, a modificação pós-tradução inclui processamento proteolítico de precursores (incluindo, mas não limitados a precursor de calcitonina, precursor peptídico relacionado ao gene de calcitonina, hormônio pré-pró-paratireoide, pré-pró-insulina, pró-insulina, pré-pró-opiomelanocortina, pró-opiomelanocortina e semelhantes), montagem em uma proteína de subunidades múltiplas ou montagem macromolecular, tradução em um outro sítio na célula (incluindo, mas não limitado a organelas, tais como o retículo endoplasmático, o complexo de Golgi, o núcleo, lisossomos, peroxissomos, mitocôndria, cloroplastos, vacúolos, etc., ou através de via secretória). Em certas modalidades, a proteína compreende uma sequência de secreção ou localização, um marcador de epítopo, um marcador FLAG, um marcador de poli-histidina, uma fusão de GST ou semelhantes.

Os métodos para produzir uma proteína em uma célula com um aminoácido selecionado em uma posição especificada também são uma característica da presente invenção. Por exemplo, um método inclui cultivar, em um meio apropriado, uma célula, em que a célula compreende um ácido nucleico que compreende pelo menos um códon seletor e codifica uma proteína; e fornecer o aminoácido selecionado; em que a célula compreende adicionalmente: um tRNA ortogonal (O-tRNA) que funciona na célula e reconhece o códon seletor; e uma aminoacil-tRNA sintetase ortogonal (O-RS) que, de preferência, aminoacila o O-tRNA com o aminoácido selecionado. Tipicamente, o O-tRNA compreende atividade de supressão na presença de uma sintetase cognata em resposta a um códon seletor. Uma proteína produzida por este método também é uma característica da presente invenção.

As composições da presente invenção e composições feitas pelos métodos da presente invenção estão opcionalmente em uma célula. Os pares de O-tRNA/O-RS ou componentes individuais da invenção depois podem ser usados em um mecanismo de tradução do sistema hospedeiro, o que resulta em um aminoácido selecionado, por exemplo, um aminoácido não natural, sendo incorporado em uma proteína. Pedidos de Patente USSN 10/825.867, intitulado "Expanding the Eukaryotic Genetic Code"; e 10/126.927, intitulado "In vivo incorporation of unnatural amino acids" descrevem este processo e são incorporados como referência neste relatório. Por exemplo, quando um par de OtRNA/O-RS é introduzido em um hospedeiro, por exemplo, um *E.scherichia coli*, o par permite a incorporação *in vivo* de aminoácido selecionado, tal como um aminoácido não natural, por exemplo, um aminoá-

cido sintético, tal como um derivado de um aminoácido leucina, que pode ser adicionado de maneira exógena ao meio de crescimento, em uma proteína, em resposta a um códon seletor. Opcionalmente, as composições da presente invenção podem estar em um sistema de tradução *in vitro* ou em sistema(s) *in vivo*.

5 Qualquer proteína (ou porção da mesma) que inclui um aminoácido selecionado, por exemplo, um aminoácido não natural, (e qualquer ácido nucleico codificante correspondente, por exemplo, que inclui um ou mais códons seletores) pode ser produzida usando as composições e métodos neste relatório. Qualquer polipeptídeo é adequado para a incorporação de um ou mais aminoácidos selecionados. Nenhuma tentativa é feita para identificar
10 as centenas de milhares de proteínas conhecidas, qualquer uma das quais pode ser modificada para incluir um ou mais aminoácidos não naturais, por exemplo, pela adequação de quaisquer métodos de mutação disponíveis para incluir um ou mais códons seletores apropriados em um sistema de tradução relevante. Os repositórios de sequências comuns para proteínas conhecidas incluem GenBank EMBL, DDBJ e o NCBI. Outros repositórios podem
15 ser facilmente identificados por busca na Internet.

 Tipicamente, as proteínas são, por exemplo, pelo menos 60 %, pelo menos 70 %, pelo menos 75 %, pelo menos 80 %, pelo menos 90 %, pelo menos 95 %, pelo menos 99 % ou mais idênticas a qualquer proteína disponível (por exemplo, uma proteína terapêutica, uma proteína diagnóstica, uma enzima industrial, ou porção das mesmas e semelhantes), e
20 elas compreendem um ou mais aminoácidos selecionados. Os exemplos de proteínas terapêuticas, diagnósticas e outras que podem ser modificadas para compreender um ou mais aminoácidos selecionados, por exemplo, um aminoácido não natural, podem ser encontrados, mas não estão limitados àqueles em USSN 10/825.867, intitulado "Expanding the Eukaryotic Genetic Code"; e Pedido de Patente U.S. Número de Série 10/126.927, intitulado "In
25 vivo incorporation of unnatural amino acids".

 Em certas modalidades, a proteína ou polipeptídeo de interesse (ou porção do mesmo) nos métodos e/ou composições da presente invenção é codificado por um ácido nucleico. Tipicamente, o ácido nucleico possui pelo menos um códon seletor, pelo menos dois códons seletores, pelo menos três códons seletores, pelo menos quatro códons seletores,
30 res, pelo menos cinco códons seletores, pelo menos seis códons seletores, pelo menos sete códons seletores, pelo menos oito códons seletores, pelo menos nove códons seletores, dez ou mais códons seletores.

 Os genes que codificam proteínas ou polipeptídeos de interesse podem ser mutagenizados usando métodos bem conhecidos para uma pessoa habilitada na técnica e descritos neste relatório em "Mutagênese e Outras Técnicas de Biologia Molecular" para incluir,
35 por exemplo, um ou mais códons seletores para a incorporação de um aminoácido selecionado, por exemplo, um aminoácido não natural. Por exemplo, um ácido nucleico para uma

proteína de interesse é mutagenizado para incluir um ou mais códons seletores, permitindo a inserção de um ou mais aminoácidos selecionados, por exemplo, um aminoácido não natural. A invenção inclui qualquer variante, por exemplo, mutante, versões de qualquer proteína, por exemplo, incluindo pelo menos um aminoácido selecionado. De maneira similar, a invenção também inclui ácidos nucleicos correspondentes, isto é, qualquer ácido nucleico com um ou mais códons seletores que codificam um ou mais aminoácidos selecionados.

Para preparar uma proteína que inclui um aminoácido selecionado, uma pessoa habilitada pode usar células hospedeiras e organismos que estão adaptados para a incorporação *in vivo* do aminoácido selecionado por intermédio de pares de tRNA/RS ortogonais. As células hospedeiras são geneticamente construídas (por exemplo, transformadas, transduzidas ou transfectadas) com um ou mais vetores que expressam o tRNA ortogonal, a tRNA sintetase ortogonal e um vetor que codifica a proteína a ser derivatizada. Cada um desses componentes pode estar no mesmo vetor, ou cada um pode estar em um vetor separado, dois componentes podem estar em um vetor e o terceiro componente em um segundo vetor. O vetor pode estar, por exemplo, na forma de um plasmídeo, um cosmídeo, um fago, uma bactéria, um vírus, um polinucleotídeo puro ou um polinucleotídeo conjugado.

SISTEMAS ALTERNADOS

Muitas estratégias foram utilizadas para introduzir aminoácidos não naturais em proteínas em células hospedeiras não recombinantes, células hospedeiras mutagenizadas ou em sistemas livres de célula. A derivatização de aminoácidos com cadeias laterais reativas, tais como Lys, Cys e Tyr, resultou na conversão de lisina a N²-acetil-lisina. A síntese química também fornece um método simples para incorporar aminoácidos não naturais. Com o desenvolvimento recente de ligação enzimática e ligação química nativa de fragmentos peptídicos, é possível preparar proteínas maiores. Veja, por exemplo, P. E. Dawson e S. B. H. Kent Ann. Ver. Biochem. 69:923 (2000). Ligação peptídica química e ligação química nativa são descritas na Patente U.S. Nº 6.184.344, Publicação de Patente U.S. Nº 2004/0138412, Publicação de Patente U.S. Nº 2003/0208046, WO 02/098902 e WO 03/042235, os quais são incorporados como referência neste relatório. Um método geral na biossíntese *in vitro*, no qual um tRNA supressor quimicamente acilado com o aminoácido não natural desejado é adicionado a um extrato *in vitro* capaz de suportar a biossíntese proteica, foi usado para incorporar sítio-especificamente mais de 100 aminoácidos não naturais em uma variedade de proteínas virtualmente de qualquer tamanho. Veja, por exemplo, V. W. Cornish, D. Mendel e P. G. Schultz, Angew. Chem. Int. Ed. Engl. 1995, 34:621; C. J. Noren, S. J. Anthony-Cahill, M. C. Griffith, P. G. Schultz, *A general method for site-specific incorporation of unnatural amino acids into proteins*, Science 244:182-188 (1989); e J. D. Bain, C. G. Glabe, T. A. Dix, A. R. Chamberlin, E. S. Diala, *Biosynthetic site-specific incorporation of non-natural amino acid into a polypeptide*, J. Am. Chem. Soc. 111:8013-8014 (1989). Uma

ampla faixa de grupos funcionais foi introduzida em proteínas para estudos de estabilidade de proteína, dobra da proteína, mecanismo enzimático e transdução de sinal.

Um método *in vivo*, denominado incorporação por pressão seletiva, foi desenvolvido para explorar a promiscuidade de sintetases do tipo selvagem. Veja, por exemplo, N. Budisa, C. Minks, S. Alefelder, W. Wenger, F. M. Dong, L. Moroder e R. Huber, FASEB J. 13:41 (1999). Uma cepa auxotrófica, na qual a via metabólica relevante que fornece à célula um aminoácido natural particular é desligada, é cultivada em meios mínimos contendo concentrações limitadas do aminoácido natural, enquanto a transcrição do gene alvo é reprimida. No início da fase de crescimento estacionário, o aminoácido natural é depletado e substituído com o análogo de aminoácido não natural. A indução de expressão da proteína recombinante resulta no acúmulo de uma proteína contendo o análogo não natural. Por exemplo, usando esta estratégia, o, m e p-fluorfenilalaninas foram incorporadas em proteínas e exibem dois ombros (*shoulders*) característicos no espectro de UV que podem ser facilmente identificados, veja, por exemplo, C. Minks, R. Huber, L. Moroder e N. Budisa, Anal. Biochem., 284:29 (2000); trifluormetionina foi usada para substituir a metionina em lisozima de bacteriófago T4 para estudar a sua interação com ligantes de quito-oligossacarídeos por RMN de ^{19}F , veja, por exemplo, H. Duewel, E. Daub, V. Robinson e J. f. Honek, Biochemistry, 36:3404 (1997); e trifluorleucina foi incorporada no lugar da leucina, resultando no aumento da estabilidade térmica e química de uma proteína de zíper de leucina. Veja, por exemplo, Y. Tang, G. Ghirlanda, W. A. Petka, T. Nakajima, W. F. DeGrado e D. A. Tirrell, Angew. Chem. Int. Ed. Engl., 40:1494 (2001). Além disso, selenometionina e telurometionina são incorporadas em várias proteínas recombinantes para facilitar a solução de fases em cristalografia de raios-X. Veja, por exemplo, W. A. Hendrickson, J. R. Horton e D. M. Lemas-ter, EMBO J., 9:1665 (1990); J. O. Boles, K. Lewinski, M. Kunkle, J. D. Odon, B. Dunlap, L. Lebioda e M. Hatada, Nat. Struct. Biol., 1:283 (1994); N. Budisa, B. Steipe, P. Demange, C. Eckerskon, J. Kellermann e R. Huber, Eur. J. Biochem., 230:788 (1995); e N. Budisa, W. Karnbrock, S. Steinbacher, A. Humm, L. Prade, T. Neuefeind, L. Moroder e R. Huber, J. Mol. Biol., 270:616 (1997). Os análogos de metionina com funcionalidade alqueno ou alquino também foram eficazmente incorporados, permitindo a modificação adicional de proteínas por meios químicos. Veja, por exemplo, J. C. M. van Hest e D. A. Tirrell, FEBS Lett., 428:68 (1998); J. C. M. van Hest, K. L. Kiick e D. A. Tirrell, J. Am. Chem. Soc., 122:1282 (2000); e K. L. Kiick e D. A. Tirrell, Tetrahedron, 56:9487 (2000); Patente U.S. Nº 6.586.207, Publicação de Patente U.S. 2002/0042097, os quais são incorporados como referência neste relatório.

O sucesso desse método depende do reconhecimento dos análogos de aminoácidos não naturais por aminoacil-tRNA sintetases, que, em geral, exigem alta seletividade para garantir a fidelidade da tradução proteica. Uma maneira para expandir o âmbito desse

método é relaxar a especificidade do substrato de aminoacil-tRNA sintetases, o que foi obtido em um número limitado de casos. Por exemplo, a substituição de Ala²⁹⁴ por Gly na fenilalanil-tRNA sintetase (PheRS) de *Escherichia coli* aumenta o tamanho da fenda de ligação de substrato e resulta na acilação de tRNA^{Phe} por p-Cl-fenilalanina (p-Cl-Phe). Veja, M. Ibba, P. Kast e H. Hennecke, Biochemistry, 33:7107 (1994). Uma cepa de *Escherichia coli* que abriga esta PheRS mutante permite a incorporação de p-Cl-fenilalanina ou p-Br-fenilalanina no lugar da fenilalanina. Veja, por exemplo, M. Ibba e H. Hennecke, FEBS Lett., 364:272 (1995); e N. Sharma, R. Furter, P. Kast e D. A Tirrell, FEBS Lett., 467:37 (2000). De maneira similar, uma mutação de ponto Phe130Ser próxima ao sítio de ligação a aminoácido da tiro-

10 sil-tRNA sintetase de *Escherichia coli* permitiu que a azatirosina fosse mais eficazmente incorporada em relação à tirosina. Veja, F. Hamano-Takaku, T. Iwama, S. Saito-Yano, K. Takaku, Y. Monden, M. Kitabatake, D. Soll e S. Nishimura, J. Biol. Chem., 275:40324 (2000).

Uma outra estratégia para incorporar aminoácidos não naturais em proteínas *in vivo* é modificar sintetases que têm mecanismos de revisão. Estas sintetases não podem discriminar e, portanto, ativar aminoácidos que são estruturalmente similares aos aminoácidos naturais cognatos. Este erro é corrigido em um sítio separado, que desacula o aminoácido mal carregado do tRNA para manter a fidelidade da tradução proteica. Se a atividade de revisão da sintetase é desabilitada, análogos estruturais que são mal ativados podem escapar da função de edição e serem incorporados. Esta abordagem foi recentemente demonstrada com a valil-tRNA sintetase (ValRS). Veja, V. Doring, H. D. Mootz, L. A. Nangle, T. L. Hendrickson, V. de Crecy-Lagard, P. Schimmel e P. Marliere, Science, 292:501 (2001). A ValRS pode aminoacilar de forma errônea tRNA^{Val} com Cys, Thr ou aminobutirato (Abu); estes aminoácidos não cognatos são subsequentemente hidrolisados pelo domínio de edição. Após a mutagênese aleatória do cromossomo de *Escherichia coli*, foi selecionada uma

15 cepa mutante de *Escherichia coli* que tinha uma mutação no sítio de edição de ValRS. Esta ValRS defectiva na edição carrega incorretamente tRNA^{Val} com Cys. Pelo fato de Abu se assemelhar estericamente a Cys (o grupo -SH de Cys é substituído com -CH₃ em Abu), a ValRS mutante também incorpora Abu em proteínas quando esta cepa mutante de *Escherichia coli* é cultivada na presença de Abu. As análises por espectrometria de massas mostram que cerca de 24 % das valinas são substituídas por Abu em cada posição de valina na proteína nativa.

20

25

30

A síntese em fase sólida e métodos semissintéticos também permitiram a síntese de várias proteínas contendo novos aminoácidos. Por exemplo, veja as seguintes publicações e referências citadas a este respeito, que as seguintes: Crick, F. J. C. Barrett, L. Brenner, S. Watts-Tobin, R. *General nature of the genetic code for proteins*. Nature, 192:1227-1232 (1961); Hofmann, K., Bohn, H. *Studies on polypeptides. XXXVI. The effect of pyrazole-imidazole replacements on the S-protein activating potency of an S-peptide fragment*. J. Am.

35

Chem. 88(24) 5914-5919 (1966); Kaiser, E. T. *Synthetic approaches to biologically active peptides and proteins including enzymes*, Acc Chem Res, 22:47-54 (1989); Nakatsuka, T., Sasaki, T., Kaiser, E. T. *Peptide segment coupling catalyzed by the semisynthetic enzyme thiosubtilisin*, J. Am. Chem. Soc., 3808-3810 (1987); Schnolzer, M., Kent, S B H. *Constructing proteins by dovetailing unprotected synthetic peptides: backbone-engineered HIV protease*, Science, 256(5054):221-225 (1992); Chaiken, I. M. *Semisynthetic peptides and proteins*. CRC Crit Rev Biochem, 11(3):255-301 (1981); Offord, R. E. *Protein engineering by chemical means?* Protein Eng. 1(3):151-157 (1987); e Jackson, D. Y., Burnier, J., Quan, C., Stanley, M., Tom, J., Wells, J. A. *A Designed Peptide Ligase for Total Synthesis of Ribonuclease A with Unnatural Catalytic Residues*, Science, 266(5183):243 (1994).

A modificação química foi usada para introduzir uma variedade de cadeias laterais não naturais, incluindo cofatores, marcadores de *spin* e oligonucleotídeos em proteínas *in vitro*. Veja, por exemplo, Corey, D. R., Schultz, P. G. *Generations of a hybrid sequence-specific single-stranded deoxyribonuclease*, Science, 238(4832):1401-1403 (1987); Kaiser, E. T., Lawrence D. S., Rokita, S. E. *The chemical modification of enzymatic specificity*, Annu Rev Biochem, 54:565-595 (1985); Kaiser, E. T., Lawrence, D. S. *Chemical mutation of enzyme active sites*, Science, 226(4674):505-511 (1984); Neet, K. E., Nanci A., Koshland, D. E. *Properties of thiol-subtilisin*, J. Biol. Chem., 243(24):6392-6401 (1968); Polgar, L. B., M. L. *A new enzyme containing a synthetically formed active site. Thiol-subtilisin* J. Am. Chem. Soc., 3153-3154 (1966); e Pollack, S. J., Nakayama, G. Schultz, P. G. *Introduction of nucleophiles and spectroscopic probes into antibody combining sites*, Science, 242(4881):1038-1040 (1988).

Definindo Polipeptídeos através de Imunorreatividade

Pelo fato de os polipeptídeos da presente invenção fornecerem uma variedade de novas sequências de polipeptídeos (por exemplo, compreendendo aminoácidos selecionados (por exemplo, aminoácidos não naturais), no caso de proteínas sintetizadas no sistema de tradução neste relatório, ou, por exemplo, no caso de novas sintetases, novas sequências de aminoácidos padrão), os polipeptídeos também fornecem novas características estruturais que podem ser reconhecidas, por exemplo, em ensaios imunológicos. A produção de antissoro, o qual se liga especificamente aos polipeptídeos da presente invenção, assim como os polipeptídeos que são ligados por tal antissoro, são uma característica da invenção. O termo "anticorpo", conforme usado neste relatório, inclui, mas não é limitado a um polipeptídeo consideravelmente codificado por um gene de imunoglobulina ou genes de imunoglobulina, ou fragmentos dos mesmos, que se ligam e reconhecem especificamente um analito (antígeno). Os exemplos incluem anticorpos policlonais, monoclonais, quiméricos e de cadeia única e semelhantes. Os fragmentos de imunoglobulinas, incluindo fragmentos Fab e fragmentos produzidos por uma biblioteca de expressão, incluindo exibição de fago,

também são incluídos no termo “anticorpo”, conforme usado neste relatório. Veja, por exemplo, Paul (1999) Fundamental Immunology, 4^a Ed., Raven Press, Nova Iorque, para estrutura e terminologia de anticorpos.

Por exemplo, a invenção inclui RSs e proteínas preparadas utilizando tRNAs e/ou
5 RSs da presente invenção que se ligam especificamente a ou que são especificamente imunorreativas a um anticorpo ou antissoro gerado contra um imunógeno compreendendo uma sequência de aminoácidos. Para eliminar a reatividade cruzada com outros homólogos, o anticorpo ou antissoro é subtraído com a proteína disponível, tal como o polipeptídeo do tipo selvagem, por exemplo, os polipeptídeo “controle”. Quando a proteína selvagem corresponde a um ácido nucleico, um polipeptídeo codificado pelo ácido nucleico é gerado e usado
10 para propósitos de subtração de anticorpo/antissoro.

Em um formato típico, o imunoensaio usa um antissoro que foi recrutado contra um ou mais polipeptídeos ou uma subsequência considerável dos mesmos (isto é, pelo menos 30 % da sequência completa fornecida). O grupo de imunógenos polipeptídicos potenciais
15 derivados da proteína são coletivamente referidos abaixo como “os polipeptídeos imunogênicos”. O antissoro resultante é opcionalmente selecionado para ter baixa reatividade cruzada contra os homólogos da sintetase controle e qualquer de tal reatividade cruzada é removida, por exemplo, por imunoabsorção, com um ou mais dos homólogos controle, antes do uso do antissoro policlonal no imunoensaio.

20 De modo a produzir antissoro para uso em um ensaio imune, um ou mais dos polipeptídeos imunogênicos é produzido e purificado, conforme descrito neste relatório. Por exemplo, uma proteína recombinante pode ser produzida em uma célula recombinante. Uma cepa obtida por cruzamento endogâmico de camundongos (usada neste ensaio, pois os resultados são mais reprodutíveis devido à identidade genética virtual dos camundongos) é
25 imunizada com a(s) proteína(s) imunogênica(s) em combinação com um adjuvante padrão, como adjuvante de Freund, e um protocolo padrão de imunização de camundongo (veja, por exemplo, Harlow e Lane (1988) Antibodies, A Laboratory Manual, Cold Spring Harbor Publications, Nova Iorque, para uma descrição padrão de produção de anticorpo, formatos de imunoensaios e condições que podem ser usadas para determinar imunorreatividade específica.
30

As referências e debate adicionais de anticorpos também são encontrados neste relatório e podem ser aplicados neste relatório para definir polipeptídeos por imunorreatividade. Senão, um ou mais polipeptídeos sintéticos ou recombinantes derivados das sequências divulgadas neste relatório são conjugados a uma proteína de transporte e usados como um
35 imunógeno.

Os soros policlonais são coletados e titulados contra o polipeptídeo imunogênico no imunoensaio, por exemplo, um imunoensaio em fase sólida com uma ou mais das proteínas

imunogênicas imobilizadas em um suporte sólido. O antissoro policlonal com um título de 10^6 ou mais é selecionados, reservado e subtraído com os polipeptídeos de sintetase controle para produzir antissoro policlonal, titulado, reservado e subtraído.

O antissoro policlonal, titulado, reservado e subtraído é testado quanto à reatividade cruzada contra os homólogos controle em um imunoensaio comparativo. Neste ensaio comparativo, as condições de ligação discriminatórias são determinadas para o antissoro policlonal, titulado e subtraído que resulta em uma razão sinal/ruído pelo menos cerca de 5 a 10 vezes maior para ligação do antissoro policlonal titulado à proteína imunogênica em comparação à ligação aos homólogos de sintetase controle. Isto é, a estringência da reação de ligação é ajustada pela adição de competidores não específicos, tais como albumina ou leite em pó desnatado, e/ou pelo ajuste de condições salinas, temperatura e/ou semelhantes. Estas condições de ligação são usadas em ensaio subsequentes para determinar se um polipeptídeo de teste (um polipeptídeo comparado aos polipeptídeos imunogênicos e/ou aos polipeptídeos controle) é especificamente ligado pelo antissoro policlonal subtraído.

Em um outro exemplo, os imunoensaios no formato de ligação competitiva são usados para a detecção de um polipeptídeo de teste. Por exemplo, conforme observado, os anticorpos que reagem cruzadamente são removidos da mistura de antissoro reservada por imunoabsorção com os polipeptídeos controle. O(s) polipeptídeo(s) imunogênico(s) depois são imobilizados em um suporte sólido que é exposto ao antissoro reservado subtraído. As proteínas de teste são adicionadas ao ensaio para competir pela ligação ao antissoro subtraído reservado. A habilidade da(s) proteína(s) de teste em competir pela ligação ao antissoro subtraído reservado em comparação à(s) proteína(s) imobilizada(s) é comparada à habilidade do(s) polipeptídeo(s) imunogênico(s) adicionado(s) ao ensaio em competir pela ligação (os polipeptídeos imunogênicos competem eficazmente com os polipeptídeos imunogênicos imobilizados pela ligação ao antissoro reservado). A reatividade cruzada percentual para a proteína de teste é calculada, usando cálculos padrão.

Em um ensaio paralelo, a habilidade das proteínas controle em competir pela ligação ao antissoro subtraído reservado é opcionalmente determinada em comparação à habilidade do(s) polipeptídeo(s) imunogênico(s) competirem pela ligação ao antissoro. De novo, a reatividade cruzada percentual para os polipeptídeos controle é calculada, usando cálculos padrão. Quando a reatividade cruzada percentual é pelo menos 5 a 10x tão alta quanto para os polipeptídeos de teste em comparação aos polipeptídeos controle e ou quando a ligação dos polipeptídeos de teste está aproximadamente na faixa da ligação do polipeptídeo imunogênico, diz-se que os polipeptídeos de teste se ligam especificamente ao antissoro subtraído reservado.

Em geral, o antissoro imunoabsorvido e reservado pode ser usado em um imunoensaio de ligação competitiva, conforme descrito neste relatório, para comparar qualquer

polipeptídeo de teste ao(s) polipeptídeo(s) imunogênico(s) e/ou controle. De modo a fazer esta comparação, os polipeptídeos imunogênicos, teste e controle são avaliados em uma ampla gama de concentrações e quantidade de cada polipeptídeo exigida para inibir 50 % da ligação do antissoro subtraído, por exemplo, para um controle imobilizado, proteína de teste ou imunogênica é determinada usando técnicas padrão. Se a quantidade do polipeptídeo de teste requerido para a ligação no ensaio competitivo é menor do que duas vezes a quantidade do polipeptídeo imunogênico que é exigido, então se diz que o polipeptídeo de teste se liga especificamente a um anticorpo gerado para a proteína imunogênica, a quantidade é pelo menos cerca de 5 a 10x tão alta quanto para o polipeptídeo controle.

Como uma determinação adicional de especificidade, o antissoro reservado é opcional e completamente imunoabsorvido com o(s) polipeptídeo(s) imunogênico(s) (ao invés dos polipeptídeos controle) até que pouca ou nenhuma ligação do antissoro reservado subtraído por polipeptídeo imunogênico para o(s) polipeptídeo(s) imunogênico(s) usado(s) na aminoabsorção é detectável. Este antissoro completamente imunoabsorvido é então testado quanto à reatividade com o polipeptídeo de teste. Se pouca ou nenhuma reatividade é observada (isto é, não mais que 2x a razão sinal/ruído observada para ligação do antissoro completamente imunoabsorvido ao polipeptídeo imunogênico), então o polipeptídeo de teste é especificamente ligado ao antissoro elicitado pela proteína imunogênica.

Os detalhes adicionais sobre proteínas, anticorpos, antissoro, etc. podem ser encontrados em USSN 10/825.867 intitulado "Expanding the Eukaryotic Genetic Code"; WO 2002/085923, intitulado "In vivo incorporation of unnatural amino acids"; Patente U.S. Nº 6.927.042 intitulada "Glycoprotein synthesis"; e Publicação de Patente U.S. Nº US 2004/0198637, intitulada "Protein Arrays", os quais são incorporados como referência neste relatório.

KITS

Os kits também são uma característica da invenção. Por exemplo, é fornecido um kit para produzir uma proteína que compreende pelo menos um aminoácido selecionado, por exemplo, um aminoácido não natural, em uma célula, onde o kit inclui um recipiente contendo uma sequência de polinucleotídeos que codifica um O-tRNA e/ou um O-tRNA e/ou sequência de polinucleotídeos que codifica uma O-RS e/ou uma O-RS. Em uma modalidade, o kit inclui adicionalmente pelo menos um aminoácido selecionado. Em uma outra modalidade, o kit inclui um tRNA aminoacilado da invenção. Em uma outra modalidade, o kit compreende adicionalmente materiais de instrução para produzir a proteína.

Um exemplo adicional é um kit para produzir uma proteína que compreende pelo menos um aminoácido selecionado, por exemplo, aminoácido não natural, em um sistema de tradução livre de células, onde o kit inclui um recipiente contendo uma sequência de polinucleotídeos que codifica um O-tRNA e/ou um O-tRNA e/ou uma sequência de polinucleotí-

deos que codifica uma O-RS e/ou uma O-RS. Em uma modalidade, o kit inclui adicionalmente um aminoácido selecionado. Em uma outra modalidade, o kit inclui um tRNA aminoacilado da invenção. Em uma outra modalidade, o kit compreende adicionalmente materiais de instrução para produzir a proteína.

5 EXEMPLOS

Os seguintes exemplos são oferecidos para ilustrar, mas não para limitar a invenção reivindicada. Uma pessoa habilitada reconhecerá uma variedade de parâmetros não críticos que devem ser alterados sem divergir do escopo da invenção reivindicada.

Exemplo 1: Seleção de Aminoacil-tRNA Sintetase contra Para-Acetil Fenilalanina

10 Duas bibliotecas de DNA foram triadas para aminoacil-tRNA sintetases contra para-acetil fenilalanina, um aminoácido não naturalmente codificado. Estas bibliotecas consistem em seis mutações no gene da tirosil tRNA sintetase de *Methanococcus jannaschi* no plasmídeo pBK.

15 Foi realizado procedimento de seleção que consistiu de cinco rodadas alternadas de seleção, três positivas, duas negativas. As bibliotecas foram combinadas em uma razão de 1:1 e eletroporadas na linhagem celular positiva (GeneHog com plasmídeo de seleção positiva, pREP) e plaqueadas em placas com meios mínimos (GMML) com antibióticos apropriados e o aminoácido não naturalmente codificado para-acetil fenilalanina (pAF). As placas foram incubadas a 37 °C durante cerca de 40 horas, em que as células foram recuperadas por *scraping*. O DNA foi extraído usando um procedimento de Qiagen Mini-Prep e
20 então foi purificado em gel de agarose para isolar o DNA de plasmídeo da biblioteca.

 Este DNA foi então eletroporado em linhagem celular negativa (GeneHog com plasmídeo de seleção negativa derivado de pBAD). Estes transformantes foram plaqueados em placas de LB com antibiótico apropriado sem o aminoácido não naturalmente codificado
25 (pAF). Após cerca de 17 horas, estas células foram recuperadas por *scraping* e o DNA de plasmídeo foi purificado usando o procedimento de Qiagen Mini-Prep e então foi purificado em gel de agarose.

 As rodadas subsequentes de seleção foram feitas utilizando o mesmo método de eletroporação, plaqueamento, recuperação e purificação de DNA. Na última (quinta) rodada
30 de seleção, as diluições seriadas foram feitas das células de seleção positiva transformadas que foram plaqueadas em placas com meios mínimos. As colônias individuais foram então retiradas e cultivadas em um bloco de 96 poços durante a noite. Este bloco foi então replicado em placas com meios mínimos com concentrações variadas de cloranfenicol (o antibiótico de seleção positiva) com e sem aminoácido não natural pAF. Após cerca de 40 horas de
35 crescimento a 37 °C, as placas foram visualmente comparadas para determinar quais colônias cresceram na maior concentração de cloranfenicol, mas que não cresceram ou cresceram de forma insatisfatória na ausência do aminoácido não naturalmente codificado pAF. As

colônias que preencheram estes critérios foram cultivadas durante a noite. O DNA foi isolado das culturas por Mini-Prep e purificação em gel de agarose e foi sequenciado.

A partir desta seleção por pAF, foram encontrados 13 clones tendo sequências únicas de aminoácido e foram submetidos à caracterização adicional para determinar a fidelidade e processividade da pAF-tRNA sintetase.

Para caracterizar estas sintetases, supressões âmbar em pequena escala foram realizadas para mostrar que o aminoácido não naturalmente codificado pAF foi incorporado em um polipeptídeo e os resultados foram visualizados por SDS-PAGE. Uma única colônia foi retirada e cultivada durante a noite em caldo LB, que foi então usado para inocular 50 mL de LB. Estas células foram cultivadas em uma OD de 0,3 a 0,4, ponto em que alíquotas de 1,5 mL foram tomadas como pontos de pré-indução e a cultura foi dividida em dois frascos. Um mM de pAF foi adicionado a uma divisão e ambas foram cultivadas durante 30 minutos. Após 30 minutos de crescimento, ambas as culturas (pAF +/-) foram induzidas com 0,2 % de L-Arabinose e cultivadas durante 4,5 horas e a OD₆₀₀ foi registrada. As alíquotas de 1,5 mL foram tomadas dos frascos de pAF +/- para a análise através de SDS-PAGE.

As alíquotas de 1,5 mL (pré-indução, pAF +, pAF -) foram centrifugadas e 10.000 xg durante 10 minutos para sedimentar as células. As células depois foram colocadas em suspensão em quantidades proporcionais de Bacterial Protein Extraction Reagent (BPER, Pierce) relativas as suas OD₆₀₀ no momento da recuperação. A DNase I foi adicionada às células lisadas e incubada a 4 °C durante 20 minutos. As amostras depois foram combinadas com um agente redutor e pigmentos de carregamento e corridas em um gel Bis-Tris 4 a 12 % em tampão MES durante 30 minutos. O gel foi lavado em H₂O DI duas vezes durante 10 minutos e pigmentado com pigmento azul de coomassie. As bandas de pAF +/- foram comparadas para a fidelidade da pAF-tRNA RS para resultar na incorporação de pAF e a banda pAF + foi comparada à pAF-tRNA RS previamente selecionada.

Para verificar a processividade das RSs, o mesmo procedimento foi realizado com um plasmídeo contendo mioglobina C-H6 S4am (S4am-Myo). O S4am-Myo depois foi purificado por IMAC e enviado para sequenciamento de proteína para determinar a quantidade de incorporação de pAF.

Das pAF-tRNA RSs identificadas desta seleção foi encontrada uma sintetase (E9) que incorpora pAF eficazmente, com eficácia maior do que 95 % de incorporação de pAF em S4am-Myo. A incorporação foi determinada por sequenciamento de aminoácidos, enquanto a processividade foi mostrada por comparação de bandas de proteína em géis de SDS-PAGE. A sequência de nucleotídeos para E9 é mostrada na SEQ ID NO.: 4 e a sequência de aminoácidos de E9 é mostrada na SEQ ID NO: 5.

Um mutante adicional com atividade similar a E9 foi identificado e tem a sequência de aminoácidos mostrada na SEQ ID NO 17.

Exemplo 2: Mutagênese de tRNA

Três mutantes foram gerados de tRNA J17. A sequência de DNA de J17 do tipo selvagem é mostrada na SEQ ID NO: 8 e nas Publicações de Patente U.S. Nºs 2003/0108885 como SEQ ID NO: 1 e US 2003/0082575 como SEQ ID NO: 1 (Pedido de Patente U.S. Nºs de Série 10/126.931 e 10/126.927, respectivamente), ambos dos quais são incorporados como referência neste relatório. tRNA de J17 tem um par oscilante U51:G63 no tronco TΨC, conforme mostrado na Figura 1.

Três mutantes J17 (F12, F13 e F14) foram gerados para produzir pares de bases de Watson-Crick nas posições 51 e 63 do tronco TΨC. A mutagênese foi realizada por PCR de sobreposição e os constructos finais foram clonados em sítios de EcoRI e NdeI em um plasmídeo pET19 compreendendo a sequência de polinucleotídeos que codifica a aminoacil tRNA sintetase E9 (SEQ ID NO.: 4) e a sequência de polinucleotídeos que codifica o hormônio de crescimento humano (hGH) com uma substituição do códon âmbar (SEQ ID NO.: 16). A expressão de hGH ocorreu sob o controle do promotor de T7.

Foram gerados dois fragmentos por PCR de sobreposição. O primeiro fragmento foi obtido por extensão de *primers*. A sequência do *forward primer* usada para gerar cada um dos três mutantes foi:

GTAACGCTGAATTCCCGGCGGTAGTTCAGCAGGGCAGAACGGCGGACTCTAA
ATCCGCAT GGCGC (Ftam11; SEQ ID NO: 9).

Para gerar o mutante F12 (51C:63G), o *primer* reverso seguinte foi usado:

GATCTGCAGTGGTCCGGCGGGCCGGATTGTAACCGGCGCCATGCGGATTAG
AGTCCGCC GTTCTGC (Ftam12; SEQ ID NO: 10).

Para gerar o mutante F13 (51U:63A), o *primer* reverso seguinte foi usado:

GATCTGCAGTGGTCCGGCGGGCTGGATTGTAACCGAGCGCCATGCGGATTAG
AGTCCGCC GTTCTGC (Ftam13; SEQ ID NO: 11).

Para gerar o mutante F14 (51A:63U), o *primer* reverso seguinte foi usado:

GATCTGCAGTGGTCCGGCGGGCAGGATTGTAACCTGCGCCATGCGGATTAG
AGTCCGCC GTTCTGC (Ftam14; SEQ ID NO: 12).

Para gerar o segundo fragmento, o plasmídeo pET19 J17 E9 hGH compreendendo a sequência de polinucleotídeos para tRNA J17 (SEQ ID NO: 8), a sequência de polinucleotídeos codificando a tRNA sintetase E9 (SEQ ID NO: 4) e a sequência de polinucleotídeos codificando o hormônio de crescimento humano com uma substituição de códon âmbar (SEQ ID NO: 16) foram usadas como molde para amplificação com o par de *primers* seguinte:

CGCCGGACCACTGCAGATCCTTAGCGAAAGCTAAGGATTTTTTTTAAGC (*forward primer*, Ftam15; SEQ ID NO: 13) e CAAATTCGTCCATATGGGATTCC (Ftam16; SEQ ID NO: 14). O *forward primer* foi usado para prolongar a sequência a partir da extremidade

3' de tRNA até o sítio Nde I do plasmídeo. O produto resultante foi purificado em gel.

A etapa final de PCR de sobreposição envolveu o *forward primer* GTAACGCTGAATTCCCGGCG (Ftam17; SEQ ID NO: 15), *primer reverso* Ftam16 (SEQ ID NO: 14), o primeiro fragmento e o segundo fragmento. Os produtos montados foram digeridos com EcoR I e Nde I e ligados no plasmídeo pET19 J17 E9 hGH digerido com EcoR I e Nde I. A sequência de cada constructo foi confirmada por sequenciamento e as sequências de DNA para cada um dos tRNAs mutantes de J17 são mostradas como SEQ ID NO: 1 (F12), SEQ ID NO: 2 (F13), e SEQ ID NO: 3 (F14). Os tRNAs foram nomeados após seus *primers* reversos correspondentes.

Expressão de proteína

Os plasmídeos codificando os tRNAs (J17, F12, F13 ou F14) foram transformados em células hospedeiras bacterianas de *E. coli* cepa 1 e cepa 2 por meios químicos e plaqueados em placas de LB ágar com 50 pg/mL de carbenicilina. As placas foram incubadas a 37 °C durante a noite. Para cada tRNA, uma única colônia foi retirada para começar uma cultura durante a noite a 37 °C em 1 mL de 2xYT com 50 pg/mL de carbenicilina. Este 1 mL de cultura foi usado para inocular duas culturas de 2xYT com 50 pg/mL de carbenicilina a 37 °C. Uma cultura de 10 mL foi suplementada com 4 mM de *para*-acetilfenilalanina. Na OD₆₀₀ = 0,7, a expressão de hGH foi induzida com 0,4 mM de IPTG. Após cultivo das células a 37 °C durante 4 horas com 250 rpm, as células foram coletadas por centrifugação a 5000 x g durante 5 minutos. As células foram lisadas com reagente B-PER (Pierce, Rockford, IL), suplementadas com 5 pg/mL de DNase I. O lisato celular total foi analisado através de SDS-PAGE 4 a 12 %.

A Figura 2 mostra uma análise de lisatos celulares de cepa 1 de *E. coli* através de SDS-PAGE. A supressão de um códon seletor no hormônio de crescimento humano foi realizada usando J17 ou tRNA mutante (F12, F13, F14) de J17 e a aminoacil tRNA sintetase E9. As células que hospedam mutantes J17 cresceram mais devagar em relação às células que hospedam J17. Nenhum produto de hGh de comprimento total foi observado por SDS-PAGE para os mutantes de tRNA na ausência de *para*-acetilfenilalanina 4 mM. Na presença de *para*-acetilfenilalanina 4 mM, o produto de comprimento total foi produzido com cada um dos mutantes de tRNA, demonstrando que estes pares de tRNA mutante-RS E9 são ortogonais ao mecanismo de *E. coli*. Com base em SDS-PAGE, o rendimento de hGH suprimido dos mutantes de J17 foi de aproximadamente 1,5 a 2 vezes maior do que aquele de J17 na cepa 1 de *E. coli*.

Um mutante de J17, F13, foi adicionalmente testado em linhagem celular bacteriana de cepa 2 de *E. coli* para supressão âmbar, conforme mostrado na Figura 3. Na cepa 2 de *E. coli*, a expressão, assim como os rendimentos de supressão âmbar foram reduzidos em relação aqueles da cepa 1 de *E. coli*. Na ausência de *para*-acetilfenilalanina, nenhum produ-

to de hGH de comprimento total foi observado por SDS-PAGE. Na presença de *para*-acetilfenilalanina 4 mM, hGH de comprimento total foi observado para ambos os tRNAs. Com base em SDS-PAGE, o rendimento de hGH suprimido de F13 foi cerca de três vezes maior do que aquele de J17.

- 5 Uma corrida de fermentação comparando J17 e F13 foi realizada com um volume final de aproximadamente 1,5 L. O plasmídeo que codifica o tRNA J17 e o plasmídeo que codifica tRNA de F13 foram transformados na cepa 1 de *E. coli*. A densidade celular final para cada um foi de aproximadamente 190 g de células úmidas/L. O título de hGH foi de 347 mg/L para o clone J17 e 542 mg/L para o clone F13.

10

TABELA 1

SEQ ID NO:	Marcação	SEQUÊNCIA
1	DNA F12	CCGGCGGTAGTTCAGCAGGGCAGAACGGCGGACTCTAAATCC G CATGGCGCCGGTTCAAATCCGGCCCGCCGGACCA
2	DNA F13	CCGGCGGTAGTTCAGCAGGGCAGAACGGCGGACTCTAAATCC G CATGGCGCTGGTTCAAATCCAGCCCGCCGGACCA
3	DNA F14	CCGGCGGTAGTTCAGCAGGGCAGAACGGCGGACTCTAAATCC G CATGGCGCAGGTTCAAATCCTGCCCGCCGGACCA
4	Ácido nucleico E9 RS	ATGGACGAATTTGAAATGATAAAGAGAAACACATCTGAAATTAT CAGCGAGGAAGAGTTAAGAGAGGTTTTAAAAAAGATGAAAA TCTGCTGTTATAGGTTTTGAACCAAGTGGTAAAATACATTTAGG GCATTATCTCCAAATAAAAAAGATGATTGATTTACAAAATGCTG GATTTGATATAATTATATATTTGGCTGATTTACACGCCTATTTAA ACCAGAAAGGAGAGTTGGATGAGATTAGAAAAATAGGAGATTA TAACAAAAAAGTTTTTGAAGCAATGGGGTTAAAGGCAAAATAT GTTTATGGAAGTGAACATGGTCTTGATAAGGATTATACACTGAA TGTCTATAGATTGGCTTTAAAACTACCTTAAAAAGAGCAAGAA GGAGTATGGAACCTTAGCAAGAGAGGATGAAAATCCAAAGGT TGCTGAAGTTATCTATCCAATAATGCAGGTTAATGGGATTCATT ATGAGGGCGTTGATGTTGCAGTTGGAGGGATGGAGCAGAGAA A AATACACATGTTAGCAAGGGAGCTTTTACCAAAAAAGGTTGTTT GTATTCACAACCCTGTCTTAACGGGTTTGGATGGAGAAGGAAA G

		ATGAGTTCTTCAAAAGGGAATTTTATAGCTGTTGATGACTCTCC A GAAGAGATTAGGGCTAAGATAAAGAAAGCATACTGCCCAGCTG GAGTTGTTGAAGGAAATCCAATAATGGAGATAGCTAAATACTTC CITGAATATCCTTTAACCATAAAAAGGCCAGAAAAATTTGGIGG AGATTTGACAGTTAATAGCTATGAGGAGTTAGAGAGTTTATTTA AAAATAAGGAATTGCATCCAATGGATTTAAAAAATGCTGTAGCT GAAGAACTTATAAAGATTTTAGAGCCAATTAGAAAGAGATTATA A
5	Aminoácido E9 RS	MDEFEMIKRNTSEIIEEELREVLKKDEKSAVIGFEPGKIHLGHYL Q IKKMIDLQNAGFDIIYLADLHAYLNQKGELDEIRKIGDYNKKVFEA MGLKAKYVYGSEHGLDKDYTLNVYRLALKTTTLKRARRSMELIAR EDENPKVAEVIYPIMQVNGIHYEGVDVAVGGMEQRKIHMLARELL PKKVVCIHNPVLTGLDGEGKMSSSKGNFIAVDDSPPEMAKIKKAY CPAGVVEGNPIMEIAKYFLEYPLTIKRPEKFGGDLTVNSYEELESFL KNKELHPMDLKNVAEELIKILEPIRKRL
6	HL(TAG)3 tRNA DNA	CCCAGGGTAGCCAAGCTCGGCCAACGGCGACGGACTCTAAAT CC GTTCTCGTAGGAGTTCGAGGGTTCGAATCCCTTCCCTGGGACC A
7	HL(TGA)1 tRNA DNA	GCGGGGGTTGCCGAGCCTGGCCAAAGGCGCCGGACTTCAAAT CC GGTCCCGTAGGGGTTCGGGGTTCAAATCCCCGCCCCCGCAC CA
8	J17 <i>M. jannas- chii</i> mtRNA ^{Tyr} CUA DNA	CCGGCGGTAGTTCAGCAGGGCAGAACGGCGGACTCTAAATCC G CATGGCGCTGGTTCAAATCCGGGCCCGCCGGACCA
9	Primer FTam11	GTAACGCTGAATTCCTGGCGGTAGTTCAGCAGGGCAGAACGG C GGAATCTAAATCCGCATGGCGC
10	Primer FTam12	GATCTGCAGTGGTCCGGCGGGCCGGATTTGAACCGGCGCCAT GC GGATTTAGAGTCCGCCGTTCTGC
11	Primer	GATCTGCAGTGGTCCGGCGGGCTGGATTTGAACCGCGCCAT

	Ftam13	GCCATGC GGATTTAGAGTCCGCCGTTCTGC
12	Primer FTam14	GATCTGCAGTGGTCCGGCGGGCAGGATTTGAACCTGCGCCAT GC GGATTTAGAGTCCGCCGTTCTGC
13	Primer FTam15	CGCCGGACCACTGCAGATCCTTAGCGAAAGCTAAGGATTTTTT T TAAGC
14	Primer FTam16	CAAATTCGTCCATATGGGATTCC
15	Primer FTam17	GTAACGCTGAATTCCCGGCG
16	hGH (DNA)	ATGGGCCACCACCACCACCACCACTTCCCAACCATTCCCTTAT CC AGGCTTTTTGACAACGCTATGCTCCGCGCCCATCGTCTGCACC A GCTGGCCTTTGACACCTACCAGGAGTTTGAAGAAGCCTAGATC C CAAAGGAACAGAAGTATTCATTCTGCAGAACCCCCAGACCTC C CTCTGTTTCTCAGAGTCTATTCCGACACCCTCCAACAGGGAGG A AACACAACAGAAATCCAACCTAGAGCTGCTCCGCATCTCCCTG C TGCTCATCCAGTCGTGGCTGGAGCCCGTGCAGTTCCTCAGGA GT GTCTTCGCCAACAGCCTGGTGTACGGCGCCTCTGACAGCAAC GT CTATGACCTCCTAAAGGACCTAGAGGAAGGCATCCAAACGCTG A TGGGGAGGCTGGAAGATGGCAGCCCCCGGACTGGGCAGATCT T CAAGCAGACCTACAGCAAGTTCGACACAACTCACACAACGAT GACGCACTACTCAAGAACTACGGGCTGCTCTACTGCTTCAGGA A GGACATGGACAAGGTCGAGACATTCCTGCGCATCGTGCAGTG CC

		GCTCTGTGGAGGGCAGCTGTGGCTTCTAA
17	Mutante D286R de E9	MDEFEMIKRN TSEIIEEEEL REVLKKDEKS AVIGFEPGK IHLGHYLQIK KMIDLQNAGF DIIYLADLH AYLNQKGELD E1RKIGDYNK KVFEAMGLKA KYVYGSEHGL DKDYTLNVYR LALKTTLKRA RRSMEIARE DENPKVAEVI YPIMQVNGIH YEGVDVAVGG MEQRKTEIMLA RELLPKKVVC IHNPVLTGLD GEGKMSSSKG NFIIVDDSP EIRAK1KKAY CPAGVVEGNP IMEIAKYFLE YPLT1KRPEK FGDLTVNSY EELESLEFKNK ELHPMRLKNA VAEELIKILE PIRKRL
18	Primer 1 de Y31	gtattaccctgttatccctatggcgaaattcagaatgattctc
19	Primer 2 de Y31	cagaaaccgcgtgtgttc <u>ct</u> aaacaatgtaatcattac
20	Primer 3 de Y31	gtaatgattacattgtt <u>ag</u> gaacaacacgcggttctg
21	Primer 4 de Y31	gagagcgggtgttgcgtactg
22	Primer 5 de Y31	attaccctgttatccctagacgctcagtggaacgaaaactcacg
23	Primer 6 de Y31	tagggataacagggttaataacaatttcagggtg
24	Primer 7 de Y31	ggcagctgtcgaaagtaccg
25	Primer 8 de Y31	<u>ct</u> acggttcgatcatctccagctagggataacagggttaat
26	Primer 1 de N51	gtattaccctgttatccctatggcgaaattcagaatgattctc
27	Primer 2 de	gattacattgtt <u>ag</u> gaacaacacgcggttctg

	N51	
28	Primer 3 de N51	cagaaaccgctgtgttc <u>ta</u> aacaatgtaatc
29	Primer 4 de N51	Gagagcgggtagcgtactg
30	Primer 5 de N51	attaccctgltatcccta gacgctcagtggaacgaaaactcacg
31	Primer 6 de N51	tagggataacagggtataacaatttcaggtg
32	Primer 7 de N51	ggtcagcttgcgaaagtaccg
33	Primer 8 de N51	<u>cta</u> aacaatgtaatctagggataacagggtaat
34	MAP Y31 Âmbar Mutante A12	atggctatctcaatcaagaccccagaagatatcgaaaaaatgcgcgtcgctggcc gactggctgccgaagtgcaggagatgatcgaaaccgtatgtaaacccgggcgtcag <u>caccggcgagctggatcgcatctgtaatgattacattgttaatgaacaacacgcg</u> gtttctgcctgcctcggctatcacggctatccgaaatccggttgcattcttatta atgaagtgggtgccacgggtatcccgacgatgctaagctgctgaaagatggcga tatcgtaacattgatgtcaccgtaataaagatggttccacggcgataacctg aaaatgtttatcgctggaagccgaccatcatggcgaaacgtctgtgccgatca cgcaagaaagcctgtacctggcgctacgcatggtaaaaccaggcattaatctgcg cgaaatcgggtgcggcgattcagaaattgtcgaagcagaaggcttctccgtcgtt cgtgaatattgcggacacgggtattggtcgggcttccatgaagaaccgcaggtgc tgactatgactcccgtgaaaccaacgtcgtactgaaacctgggatgacgttcac catcgagccaatggtcaacgcgggtaaaaaagagatccgcaccatgaaagatggc tgacggtaaaaaccaaagatcgacgcttgcacacaaatgagcatactattg tggtgactgataacggctgcgaaattctgacgctacgcaaggatgacaccatccc ggcgataatctgcacgacgaataa
35	MAP N51 Âmbar Mutante D2	atggctatctcaatcaagaccccagaagatatcgaaaaaatgcgcgtcgctggcc gactggctgccgaagtgcaggagatgatcgaaaccgtatgtaaacccgggcgtcag caccggcgagctggatcgcatctgtaatgattacattgtt <u>ag</u> gaacaacacgcg

	gtttctgcctgcctcggtatcacggctatccgaaatccgtttgcatctctatta atgaagtgggtgtgccacggtatcccggacgatgctaagctgctgaaagatggcga tatcgtaacattgatgtcacgtaataaagatggttccacggcgataacctcg aaaatgtttatcgtaggtaagccgaccatcatggggaacgtctgtgccgcatca cgcaagaaagcctgtacctggcgctacgcatggtaaaaccaggcattaatctgcg cgaaatcgggtgcggcgattcagaaatttgcgaagcagaaggcttctccgtcgtt cgtgaatattgcggacacggtattggtcgcggcttccatgaagaaccgcaggtgc tgcactatgactcccgtgaaaccaacgtcgtactgaaacctgggatgacgttcac catcgagccaatggtcaacgcgggtaaaaaagagatccgcaccatgaaagatggc tggacggtaaaaaccaagatcgcagcttgtctgcacaatatgagcatactattg tggtagtgataacggctgcgaaattctgacgctacgcaaggatgacaccatccc ggcgataatctcgcacgacgaataa
--	--

Exemplo 3: Vacina Viva da Ambrix para *Mycobacterium avium* subespécie *paratuberculosis*

Mycobacterium avium subespécie *paratuberculosis* (MAP) é uma bactéria patogênica que causa doença de Johne no gado bovino e outros ruminantes. Um estudo recente

5 estimou que 21 % dos rebanhos leiteiros dos Estados Unidos são infectados, resultando em perdas econômicas consideráveis à indústria de laticínios, totalizando mais do que \$200 milhões por ano. A doença de Johne apenas pode ser tratada com uma combinação de anti-

10 tibióticos, tais como Rifabutina e um macrolídeo, tal como Claritromicina. Os regimes de tratamento podem durar anos. Uma vacina de MAP eficaz é fundamental, pois as vacinas para paratuberculose comercialmente disponíveis não são completamente eficazes na prevenção da doença.

Conexão Entre MAP e Doença de Crohn Humana

A doença de Crohn afeta entre 400.000 a 600.000 pessoas na América do Norte. As estimativas de prevalência para o Norte da Europa variaram de 27 a 48 em 100.000.

15 Suspeita-se que MAP seja um agente causador da doença de Crohn em seres humanos, mas esta conexão é controversa. Partindo do pressuposto que MAP é um agente causador da doença de Crohn, a Giaconda pretende comercializar uma combinação de Rifabutina, claritromicina e clofazimina (Myoconda) como uma terapia medicamentosa potencial para a

20 doença de Crohn. Com uma resposta clínica de quase 95 % do experimento clínico inPhase II, o experimento Fase IIIa demonstrou um aperfeiçoamento estatisticamente significativo na obtenção de remissão em 16 semanas em comparação à terapia convencional. Entretanto, Myoconda® não apresentou resultados estatisticamente significantes na manutenção da remissão depois que a terapia foi concluída neste experimento.

Atualmente, as vacinas disponíveis oferecidas são vacinas vivas atenuadas (cepa

316 F), com base em células totais mortas (cepa 18) e também existe um experimento de vacina de subunidade em andamento, entretanto, as vacinas atuais apenas retardam a eliminação fecal e progressão à doença clínica, mas não protegem contra a infecção (Expert Rev. Vaccines 7(6) 2008). Portanto, procura-se uma vacina que seja racionalmente atenuada.

Devido ao fato de que a sequência genômica completa de MAP foi publicada (PNAS 2005 12344-12349) e a expressão do gene estranho em MAP foi realizada, a deleção específica do gene e mutação sítio-específica são agora ferramentas disponíveis para MAP (Applied and Environmental Microbiology 2008, 1687-1695). A vacina da Ambrx será um organismo total geneticamente modificado MAP dependente de aminoácido não natural que será administrada como uma vacina para prevenir infecção por MAP em animais. Genes essenciais são selecionados e depois são transformados para incluir uma ou mais mutações âmbar sítio-específicas. O MAP é caracterizado através da cultura celular com e sem o aminoácido não natural presente no meio de cultura, taxas de reversão são caracterizadas e o organismo total.

Exemplo 4: Criação de Supressão Competente/Células Lambda Red

Uma cepa de *E. coli* foi preparada. A cepa de *E. coli* é capaz de supressão âmbar e Recombinação Lambda Red. *E. coli* W3110 foram preparadas contendo pKD46n mostrado na Figura 5 e depois foram transformadas por eletroporação com o plasmídeo pVK10/camR mostrado na Figura 4. Transformantes positivos foram selecionados em placas de LB ágar contendo ampicilina 50 µg/mL e cloranfenicol 34 µg/mL cultivadas durante 48 horas a 30 °C.

Células eletrocompetentes contendo os dois plasmídeos foram preparadas através do crescimento de uma cultura de 500 mL contendo os dois plasmídeos em meio LB contendo ampicilina 50 µg/mL, cloranfenicol 34 µg/mL e arabinose 0,2 % em uma OD600 final de 0,5. As células depois foram lavadas três vezes em glicerol 10 % e recolocadas em suspensão em 500 µL de glicerol 10 %. Aliquotas de 60 µL foram armazenadas a -80 °C para uso posterior.

Exemplo 5: Criação de Metiona Amino Peptidase Letal Condicional

Cepas de *E. coli* com metiona amino peptidase condicionalmente letal foram geradas. Uma mutação âmbar foi introduzida em um dos dois sítios no gene de metiona amino peptidase, Y31 ou N51. Para integrar estes mutantes no cromossomo de *E. coli*, uma parte linear de DNA recombinante foi gerada por sobreposição da PCR, conforme mostrado no diagrama na Figura 6. Primeiro, usando os *primers* 1 e 2 (SEQ ID NO: 18 ou 26), a extremidade 5' do gene de metiona amino peptidase foi clonada (produto A da PCR), terminando no aminoácido selecionado (Y31 ou N51) que foi convertido do códon nativo ao códon âmbar. A extremidade 3' do gene de metiona amino peptidase foi clonada, iniciando com a mutação âmbar, usando os *primers* 3 (SEQ ID NO: 20 ou 28) e 4 (SEQ ID NO: 21 ou 29) (produto B

da PCR). Os dois produtos de reação se sobrepõem na sequência que flanqueia a mutação âmbar, e estes produtos foram amplificados com os *primers* 1 (SEQ ID NO: 18 ou 26) e 4 (SEQ ID NO: 21 ou 29) para fornecer o produto C, o gene de metionina amino peptidase completo com uma substituição no códon âmbar na posição 31 ou 51.

Para criar o molde de mutagênese completo, um marcador de resistência à Canamicina flanqueado por I-sceI residente em sítios de reconhecimento da endonuclease foi amplificado usando os *primers* 5 (SEQ ID NO: 22 ou 30) e 6 (SEQ ID NO: 23 ou 31) no DNA a partir de uma cepa de *E. coli* previamente gerada (produto D da PCR). Os *primers* 7 (SEQ ID NO: 24 ou 32) e 8 (SEQ ID NO: 25 ou 33) foram usados para amplificar a sequência genômica a montante da metionina amino peptidase e da extremidade 5' da metionina amino peptidase, para incluir a remoção do marcador de resistência à canamicina e recombinação homóloga (produto E da PCR). O molde de DNA final para a transformação foi criado amplificando-se os fragmentos E, D e C usando os *primers* 7 (SEQ ID NO: 24 ou 32) e 4 (SEQ ID NO: 21 ou 29).

Exemplo 7: Transformação e Seleção de Clones de Metionina Amino Peptidase Letal Condicional

O molde de transformação gerado por sobreposição de PCR foi usado para transformar células positivas de pKD46, pVK10-camR eletrocompetentes. Transformantes positivos foram selecionados através do crescimento de colônias em placas de LB ágar contendo Ampicilina 50 µg/mL, Canamicina 50 µg/mL e pAF 2 mM durante 48 a 72 horas a 30 °C. Para testar o crescimento dependente de pAF, 96 colônias para cada sítio foram cultivadas em 1 mL de LB contendo Ampicilina 50 µg/mL, Canamicina 50 µg/mL e pAF 1 mM durante 72 horas a 30 °C, também mostrado nas Figuras 7, 8 e 9. Estes clones depois foram plaquedados em réplica em LB ágar contendo Ampicilina 50 µg/mL, Canamicina 50 µg/mL e pAF 2 mM e LB ágar contendo Ampicilina 50 µg/mL e Canamicina 50 µg/mL sem pAF. Estas placas em réplica foram cultivadas durante 72 horas a 30 °C e depois foram contadas quanto ao crescimento. Os clones que exibiram crescimento apenas na presença de pAF 2 mM foram selecionados para análise adicional.

Exemplo 8: Seleção e Maturação do Clone

Os clones que foram positivos quanto à dependência de pAF foram selecionados para estudo adicional.

Estes clones foram Y31 A12, Y31 C5, Y31 E3, N51 D2 e N51 H9. Os clones foram novamente triados individualmente e todos cresceram apenas em meio contendo pAF. Todos os clones foram sequenciados quanto à presença de um códon âmbar (tag) no sítio apropriado: Y31 ou N51. Todos os cinco apresentaram um códon âmbar na localização apropriada. Os clones Y31 A12 e N51 D2 depois foram isolados e o plasmídeo pKD46 foi removido através de tratamento por calor repetido a 37 °C.

Uma vez que pKD46 e resistência à ampicilina foram removidos, as células foram transformadas com plasmídeo contendo a enzima I-SceI. Depois do tratamento com esta enzima, vários clones foram selecionados e testados quanto à resistência à canamicina. Os clones que não foram resistentes à canamicina foram resequenciados e congelados em estoques de glicerol 10 %. Estes clones serão os clones de metionina amino peptidase dependentes de pAF finais.

Em algumas modalidades da presente invenção, as taxas de reversão desejadas, esquematicamente consideradas na Figura 11, serão menores do que 10^{-9} de frequência de reversão com uma única troca. Os clones Y31 e N51 foram testados e as taxas de reversão são mostradas na Tabela 2.

TABELA 2

Clone	Teste 1	Teste 2
Y31	$2,78 \times 10^{-8}$	$<3 \times 10^{-8}$ (nada detectado)
N51	$6,82 \times 10^{-8}$	$<2 \times 10^{-8}$ (nada detectado)

TABELA 3

Genes Essenciais em <i>E. coli</i>		
Alt. Símbolo	Símbolo	Produto de Gene
b0026	ileS	Isoleucil-tRNA sintetase (EC 6.1.1.5)
b0027	lspA	Peptidase de sinal de lipoproteína (EC 3.4.23.36)
b0029	ispH*	Proteína IspH (lytB)
b0083	ftsL	Proteína de divisão celular ftsL
b0084	ftsI	Precursor da peptidoglicano sintetase ftsI
b0085	murE	UDP-N-acetilmuramoilalanil-D-glutamato-2,6-diaminopimelato ligase (EC 6.3.2.13)
b0088	murD	UDP-N-acetilmuramoilalanina--D-glutamato ligase (EC 6.3.2.9)
b0089	ftsW	Proteína de divisão celular ftsW
b0090	murG	UDP-N-acetilglucosamina-N-acetilmuramil-(Pentapeptídeo) pirofosforil-undecaprenol N-acetilglucosamina transferase (EC 2.4.1.)
b0091	murC	UDP-N-acetilmuramato-alanina ligase (EC 6.3.2.8)
b0093	ftsQ	Proteína de divisão celular ftsQ
b0094	ftsA	Proteína de divisão celular ftsA
b0095	ftsZ	Proteína de divisão celular ftsZ
b0098	secA	Subunidade secA da pré-proteína translocase

b0168	map	Metionina aminopeptidase (EC 3.4.11.18)
b0169	rpsB	30S proteína ribossomal S2
b0171	pyrH	Uridilato quinase (EC 2.7.4.-)
b0172	frr	Fator de reciclagem do ribossomo
b0174	uppS	Undecaprenil pirofosfato sintetase (EC 2.5.1.31)
b0175	cdsA	Fosfatidato citidililtransferase (EC 2.7.7.41)
b0180	fabZ	(3R)-Hidroximiristoil-[proteína de transporte de acila] desidratase (EC 4.2.1.-)
b0181	lpxA	Acil-[proteína de transporte de acila]-UDP-N-acetilglucosamina O-aciltransferase (EC 2.3.1.129)
b0182	lpxB	Lipídeo-A-dissacarídeo sintase (EC 2.4.1.182)
b0184	dnaE	Subunidade alfa da DNA polimerase III (EC 2.7.7.7)
b0185	accA	Subunidade alfa da acetil-coenzima A carboxilase carboxil transferase (EC 6.4.1.2)
b0194	proS	Prolil-tRNA sintetase (EC 6.1.1.15)
b0416	nusB	proteína B da substância de utilização de N
b0420	dxs	1-Desóxi-D-xilulose 5-fosfato sintase (EC 4.1.3.37)
b0421	ispA	Geraniltranstransferase (EC 2.5.1.10)
b0470	dnaX	Subunidade tau de DNA polimerase III (EC 2.7.7.7)
b0474	adk	Adenilato quinase (EC 2.7.4.3)
b0526	cysS	Cisteinil-tRNA sintetase (EC 6.1.1.16)
b0634	mrdB	Proteína determinante na forma de bastão rodA
b0635	mrdA	Proteína de ligação à penicilina 2
b0640	holA	Subunidade delta de DNA polimerase III (EC 2.7.7.7)
b0642	leuS	Leucil-tRNA sintetase (EC 6.1.1.4)
b0680	ginS	Glutaminil-tRNA sintetase (EC 6.1.1.18)
b0893	serS	Seril-tRNA sintetase (EC 6.1.1.11)
b0914	msbA	Proteína de ligação a ATP de transporte provável msbA
b0918	kdsB	3-Desóxi-mano-octulosonato citidililtransferase (EC 2.7.7.38)
b0930	asnS	Asparaginil-tRNA sintetase (EC 6.1.1.22)
b0954	fabA	3-Hidroxidecanoil[proteína de transporte de acila] desidratase (EC 4.2.1.60)
b1084	rne	Ribonuclease E (EC 3.1.4.-)
b1092	fabD	Malonil CoA-proteína de transporte de acila transacilase (EC 2.3.1.39)
b1093	fabG	3-Oxoacil-[proteína de transporte de acila] redutase (EC 1.1.1.100)

b1098	tmk	Timidilato quinase (EC 2.7.4.9)
b1099	holB	Subunidade delta' de DNA polimerase III (EC 2.7.7.7)
b1116	lolC	Proteína transmembrana do sistema de liberação de lipoproteína lolC
b1117	lolD	proteína de ligação a ATP do sistema de lipoproteína lolD
b1118	lolE	Proteína transmembrana do sistema de liberação de lipoproteína lolE
b1133	trmU	tRNA (5-metilaminometil-2-tiouridilato)-metiltransferase (EC 2.1.1.61)
b1207	prsA	Ribose-fosfato pirofosfoquinase (EC 2.7.6.1)
b1274	topA	DNA topoisomerase I (EC 5.99.1.2)
b1288	fabI	Enoil-[proteína de transporte de acila] redutase [NADH] (EC 1.3.1.9)
b1714	pheS	Cadeia alfa da fenilalanil-tRNA sintetase (EC 6.1.1.20)
b1716	rplT	50S proteína ribossomal L20
b1718	infC	Fator de iniciação de tradução IF-3
b1740	nadE	NAD(+) sintetase dependente de NH(3) (EC 6.3.5.1)
b1866	aspS	Aspartil-tRNA sintetase (EC 6.1.1.12)
b1876	argS	Arginil-tRNA sintetase (EC 6.1.1.19)
b2185	rplY	50S proteína ribossomal L25
b2234	nrdA	Cadeia alfa da ribonucleosídeo-difosfato redutase 1 (EC 1.17.4.1)
b2235	nrdB	Cadeia beta da ribonucleosídeo-difosfato redutase 1 (EC 1.17.4.1)
b2323	fabB	3-Oxoacil-[proteína de transporte de acila] sintase I (EC 2.3.1.41)
b2400	gltX	Glutamil-tRNA sintetase (EC 6.1.1.17)
b2411	ligA	DNA ligase (EC 6.5.1.2)
b2412	zipA	Proteína de divisão celular zipA
b2496	yfgE	Proteína hipotética yfgE
b2514	hisS	Histidil-tRNA sintetase (EC 6.1.1.21)
b2563	acpS	Holo-[proteína de transporte de acila] sintase (EC 2.7.8.7)
b2606	rplS	50S proteína ribossomal L19
b2607	trmD	tRNA (Guanina-N(1)-)-metiltransferase (EC 2.1.1.31)
b2609	rpsP	30S proteína ribossomal S16
b2610	ffh	Proteína da partícula de reconhecimento do sinal
b2614	grpE	Proteína GrpE
b3018	plsC	1-Acil-sn-glicerol-3-fosfato aciltransferase (EC 2.3.1.51)
b3019	parC	Subunidade A da topoisomerase IV (EC 5.99.1.-)
b3030	parE	Subunidade B da topoisomerase IV (EC 5.99.1.-)
b3041	ribB	3,4-di-Hidróxi-2-butanona 4-fosfato sintase
b3067	rpoD	Fator sigma da RNA polimerase rpoD
b3165	rpsO	30S proteína ribossomal S15

b3183	yhbZ	Proteína de ligação a GTP hipotética yhbZ
b3230	rpsI	30S proteína ribossomal S9
b3231	rplM	50S proteína ribossomal L13
b3287	def	Peptídeo deformilase (EC 3.5.1.88)
b3298	rpsM	30S proteína ribossomal S13
b3300	secY	Subunidade secY da pré-proteína translocase
b3301	rplO	50S proteína ribossomal L15
b3303	rpsE	30S proteína ribossomal S5
b3304	rplR	50S proteína ribossomal L18
b3305	rplF	50S proteína ribossomal L6
b3307	rpsN	30S proteína ribossomal S14
b3309	rplX	50S proteína ribossomal L24
b3310	rplN	50S proteína ribossomal L14
b3311	rpsQ	30S proteína ribossomal S17
b3316	rpsS	30S proteína ribossomal S19
b3317	rplB	50S proteína ribossomal L2
b3318	rplW	50S proteína ribossomal L23
b3319	rplD	50S proteína ribossomal L4
b3320	rplC	50S proteína ribossomal L3
b3321	rpsJ	30S proteína ribossomal S10
b3340	fusA	Fator de elongação G
b3342	rpsL	30S proteína ribossomal S12
b3384	trpS	Triptofanil-tRNA sintetase (EC 6.1.1.2)
b3464	ftsY	Proteína de divisão celular ftsY
b3559	glyS	Cadeia beta da glicil-tRNA sintetase (EC 6.1.1.14)
b3633	kdtA	3-Desóxi-D-mano-ácido octulosônico-transferase (EC 2.-.-.-)
b3634	coaD	Fosfopanteteína adenililtransferase (EC 2.7.7.3)
b3640	dut	Desoxiuridina 5'-trifosfato nucleotídeo-hidrolase (EC 3.6.1.23)
b3648	gmK	Guanilato quinase (EC 2.7.4.8)
b3702	dnaA	Proteína do iniciador de replicação cromossômica dnaA
b3730	glmU	Proteína glmU bifuncional
b3865	engB	Proteína de ligação a GTP provável engB
b3967	murI	Glutamato racemase (EC 5.1.1.3)
b3985	rplJ	50S proteína ribossomal L10
b3986	rplL	50S proteína ribossomal L7/L12
b3987	rpoB	Cadeia beta da RNA polimerase dirigida ao DNA (EC 2.7.7.6)

b3988	rpoC	Cadeia beta' da RNA polimerase dirigida ao DNA (EC 2.7.7.6)
b4052	dnaB	DNA helicase replicativa (EC 3.6.1.-)
b4142	groS	Chaperonina 10 kDa
b4143	groL	Chaperonina 60 kDa
b4147	efp	Fator de elongação P
b4200	rpsF	30S proteína ribossomal S6
b4226	ppa	Pirofosfatase inorgânica (EC 3.6.1.1)
b4258	valS	Valil-tRNA sintetase (EC 6.1.1.9)
b4361	dnaC	Proteína de replicação de DNA dnaC
b4362	dnaT	Proteína primossomal I

Exemplo 9:

Cepa de origem do mesmo tipo e passagem, conforme usado para vacina contra poliomielite (oral), são transformadas para incluir mutações âmbar em mais do que um gene essencial, e cultura.

5 Cultura Celular:

- i) Substratos celulares diferentes para a propagação de vírus incluem linhagem celular diploide humana, células de rim de macaco, linhagem celular Vero ou qualquer outra cultura celular adequada, ii) Produção de células em larga escala: 1. Preparação do banco celular de trabalho do fabricante (MWCB): MWCB, no caso de culturas celulares primárias (células derivadas de tecido normal e armazenadas sob congelamento em temperatura menor do que 70° C.) ou a partir de quaisquer outras, permitiu que a linhagem celular contínua seja preparada a partir das células e as células reservadas depois da subcultura em série dentro do número especificado de culturas celulares. 2. Qualquer um dos sistemas seguintes pode ser usado, o qual é sempre adequado, para o aumento gradual no volume de células. Frascos rotatórios: O crescimento em frascos rotatórios difere daquele em culturas estacionárias na distribuição de células em superfície de vidro e em densidade celular máxima atingível. Culturas em frasco rotatório não se deterioram menos rapidamente, conforme elas podem tolerar uma densidade quase duas vezes maior do que aquela das culturas estacionárias. As células de mamíferos cultivadas em cultura de monocamada se adaptarão melhor à técnica de frascos rotatórios. Preparações celulares: Estas são pilhas de bandejas de cultura que compartilham uma entrada comum e orifício de saída. Estas fornecem células com uma superfície de crescimento tão grande quanto dentro do frasco rotatório, mas ocupam menos espaço dentro do incubador. Estas são ideais para culturas celulares aderentes, têm baixa contaminação, risco e são compactas. Sistema de transporte Roux flasks Micro 3. Amostras para culturas celulares controle são separadamente incubadas durante pelo menos duas semanas e estas são examinadas quanto à evidência de alterações citopáticas.

Propagação do Vírus da Poliomielite em Culturas Celulares:

i) Sistema de lotes de origens: a) Nível de passagem como segue para a Vacina contra a Poliomielite Oral OR b) Seu próximo nível de passagem, conforme derivado para suspensão em massa monovalente que pode ser usada para a fabricação da vacina contra a poliomielite oral. 1. A vacina é fabricada na base do sistema de lotes de origens virais. Lotes de origens virais são uma quantidade de vírus processada junta e de composição uniforme preparada a partir dos lotes de origem. 2. A subcultura do vírus de origem não deve ser feita mais do que 10 vezes, contadas a partir de um lote de origem usado para a produção da vacina em que o laboratório original e testes de campo foram feitos.

Coleta Única e Reservatórios Monovalentes:

O vírus é propagado em culturas celulares e colhido a partir de culturas celulares derivadas de um lote único de células e processado. Isto é conhecido como coleta única. As coletas únicas de um tipo de suspensão viral são processadas ao mesmo tempo para obter forma bruta de reservatório monovalente.

Filtração ou Clarificação do Reservatório Monovalente:

A suspensão viral bruta de cada reservatório monovalente é gradualmente purificada através de filtros de porosidade decrescente. O propósito da etapa de filtração é remover material particulado e outras substâncias que podem afetar o processo de inativação, tal como agregados tendem a aumentar em repouso.

Concentração do Reservatório Monovalente:

Cada reservatório monovalente filtrado é concentrado por Ultra-filtração.

Purificação do Reservatório Monovalente:

A cromatografia com DEAE sepharose/coluna de DNA-ase imobilizada/coluna de imunoabsorção é utilizada para a purificação do reservatório monovalente. O processo de purificação reduz sistematicamente o nível de DNA celular a partir daquele da coleta de vírus inicial por um fator de pelo menos 108. A concentração de D-antígeno é determinada por ELISA e exibição de imunogenicidade comparável em ratos. Consequentemente, a potência é ajustada. Os reservatórios monovalentes do Poliovírus são combinados para formar o produto em massa trivalente final.

Formulação e Fabricação da Vacina:

A massa purificada é filtrada, preferivelmente através de filtro de 0,22 micron ou 0,45 micron.

O vírus monovalente transformado continua na presença do aminoácido não natural e é opcionalmente misturado com estabilizador e/ou preservante para a preparação da vacina monovalente, bivalente, trivalente ou multivalente.

Teste da Eficácia da Vacina:

A eficácia da vacina fabricada acima em comparação à vacina de referência esta-

belecida mostrou-se similar no método *in vitro* e/ou *in vivo*.

Vários outros testes com base nos métodos padrão aceitos foram realizados na vacina fabricada acima, tal como determinação de pH, teste de esterilidade, teste quanto à transformação eficaz (por exemplo, divisão e verificação de morte na ausência do aminoácido não natural), estudo da toxicidade anormal, medição de volume, determinação de potência *in vitro*, teste da endotoxina, teste de identidade, estimativa do teor de nitrogênio da proteína e determinação do teor de estabilizador e teor de formaldeído para garantir sua efetividade, segurança e potência.

A potência *in vitro* da SIPV fabricada acima foi determinada usando ELISA Sanduíche. Dois anticorpos primários recrutados em espécies diferentes contra o vírus da poliomielite foram utilizados para o mesmo e um segundo anticorpo conjugado com HRP é novamente usado no segundo anticorpo primário. Os resultados foram analisados por comparação com padrão de referência.

O teste de identidade para a vacina foi realizado usando ELISA direto. O antígeno foi revestido em placas de ELISA de 96 poços e soro anti-polio tipo-específico é utilizado para determinar a presença de todos os três tipos de vírus da poliomielite na vacina.

Embora a invenção antecedente tenha sido descrita em alguns detalhes para fins de clareza e entendimento, será evidente para uma pessoa habilitada na técnica a partir de uma leitura desta divulgação que várias alterações na forma e detalhes podem ser feitas sem divergir do verdadeiro escopo da presente invenção. Por exemplo, todas as técnicas e aparelhos descritos acima podem ser usados em várias combinações. Todas as publicações, patentes, pedidos de patentes e/ou outros documentos citados neste pedido são incorporados como referência na sua totalidade para todos os fins na mesma extensão em que cada publicação individual, patente, pedido de patente e/ou outro documento foram indicados individualmente para serem incorporados como referência para todos os fins.

REIVINDICAÇÕES

1. Célula de *Mycobacterium avium* subespécie *paratuberculosis* (MAP), **CHARACTERIZADA** pelo fato de que um aminoácido não natural foi incorporado na sequência de aminoácidos de um produto de gene essencial.

5 2. MAP, de acordo com a reivindicação 1, **CHARACTERIZADO** pelo fato de que o aminoácido não natural é incorporado em um produto de gene de MAP que é exigido para replicação.

10 3. MAP, de acordo com a reivindicação 1, **CHARACTERIZADO** pelo fato de que mais do que um aminoácido não natural é incorporado em mais do que um produto de gene essencial.

 4. Célula de *E. coli*, **CHARACTERIZADA** pelo fato de que um aminoácido não natural foi incorporado em um produto de gene essencial.

 5. Célula bacteriana, **CHARACTERIZADA** pelo fato de que um aminoácido não natural foi incorporado em um produto de gene essencial.

15 6. Célula, **CHARACTERIZADA** pelo fato de que um aminoácido não natural foi incorporado em um produto de gene essencial.

Figura 1

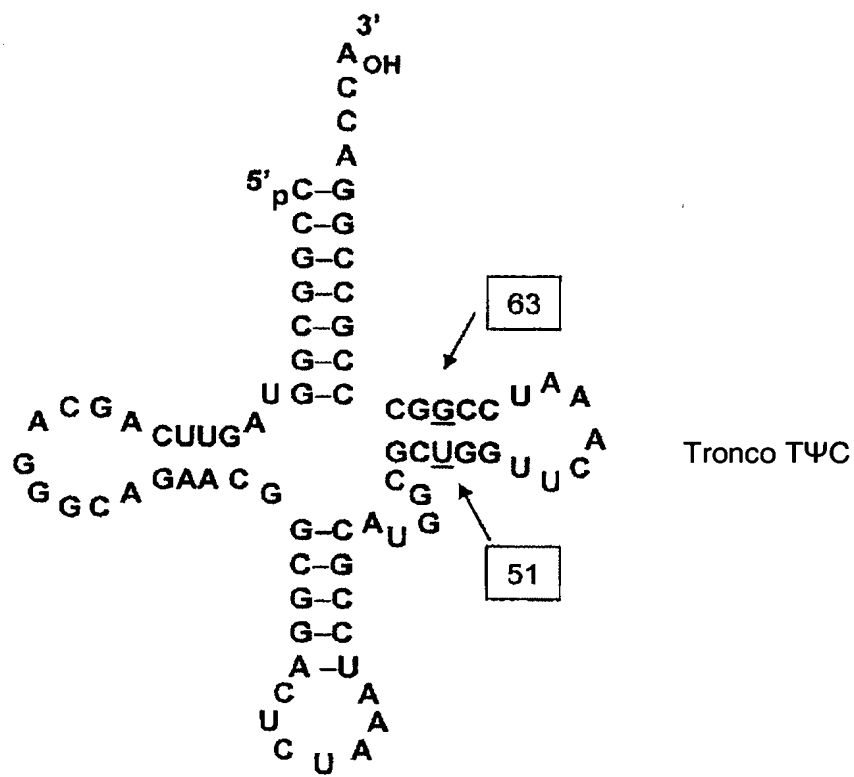


Figura 2

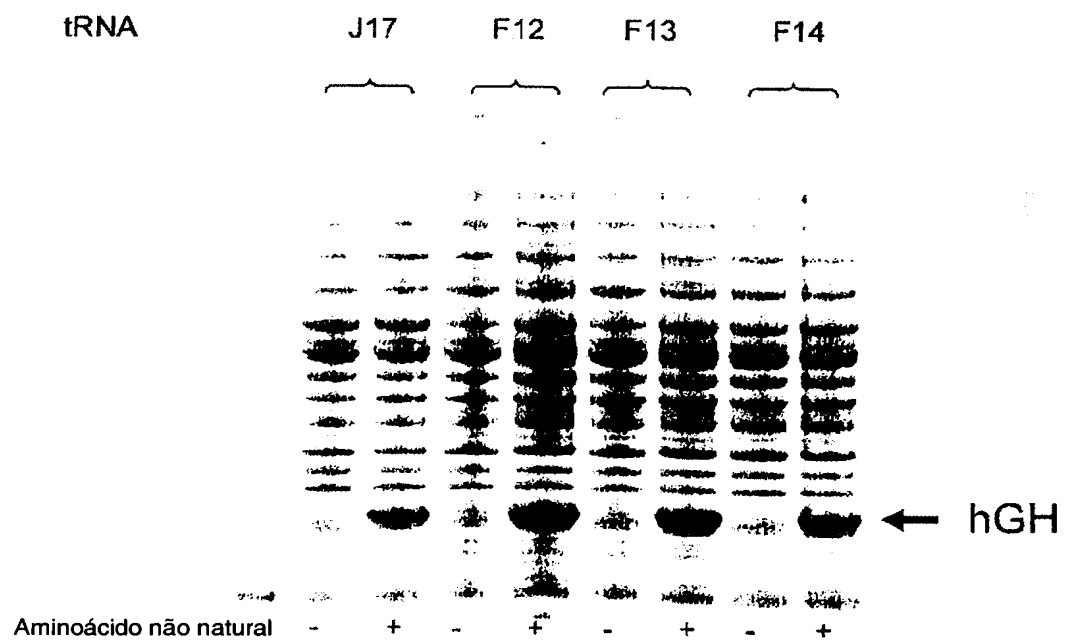


Figura 3

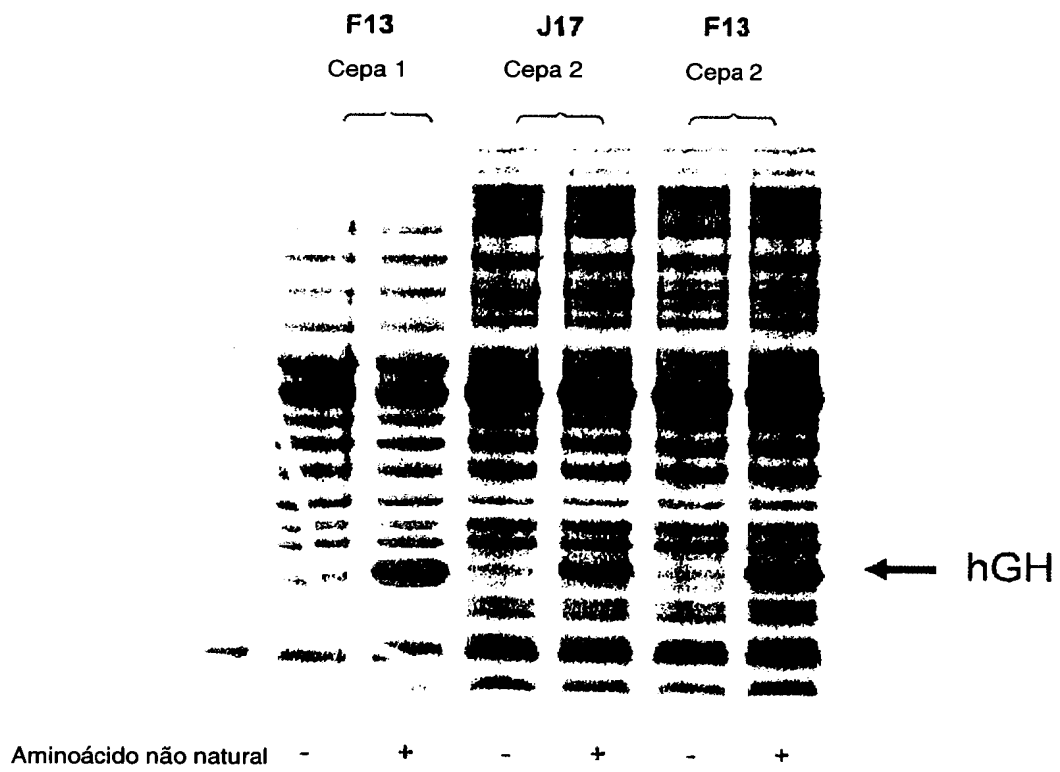


Figura 4

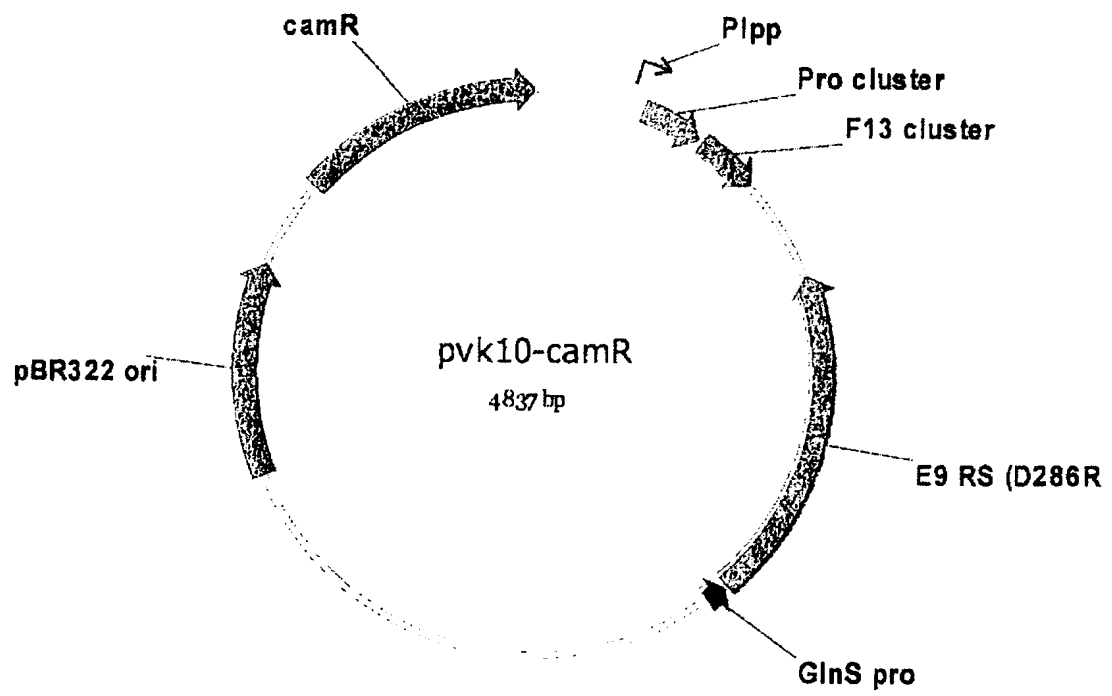


Figura 5

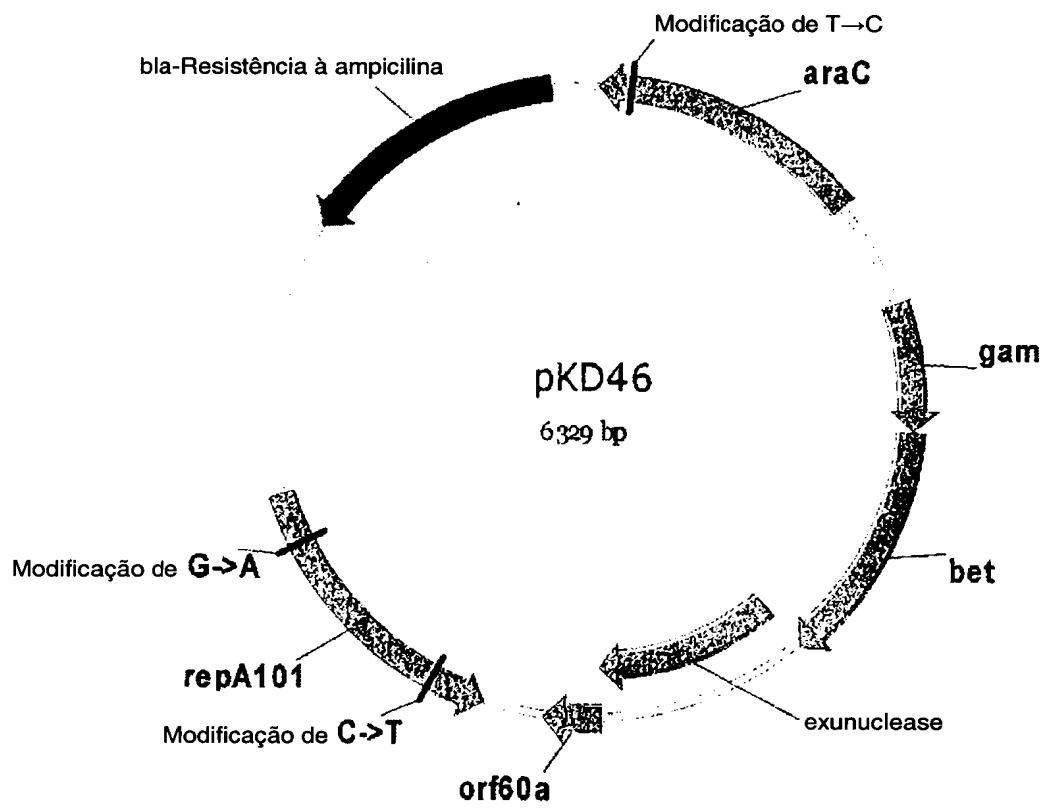


Figura 6

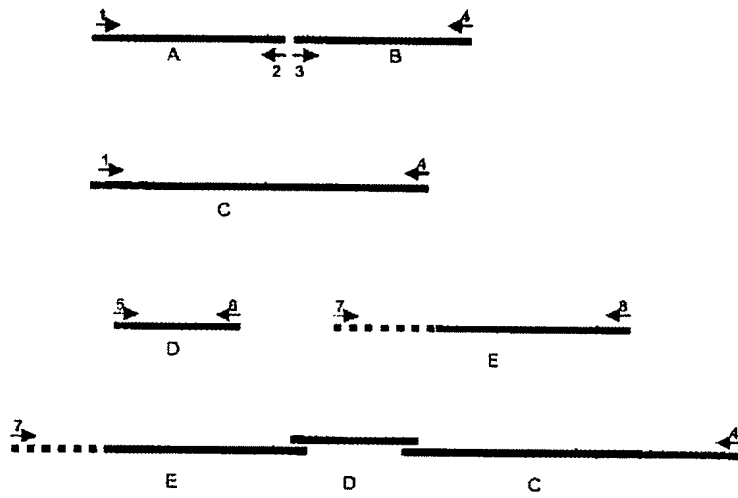


Figura 7

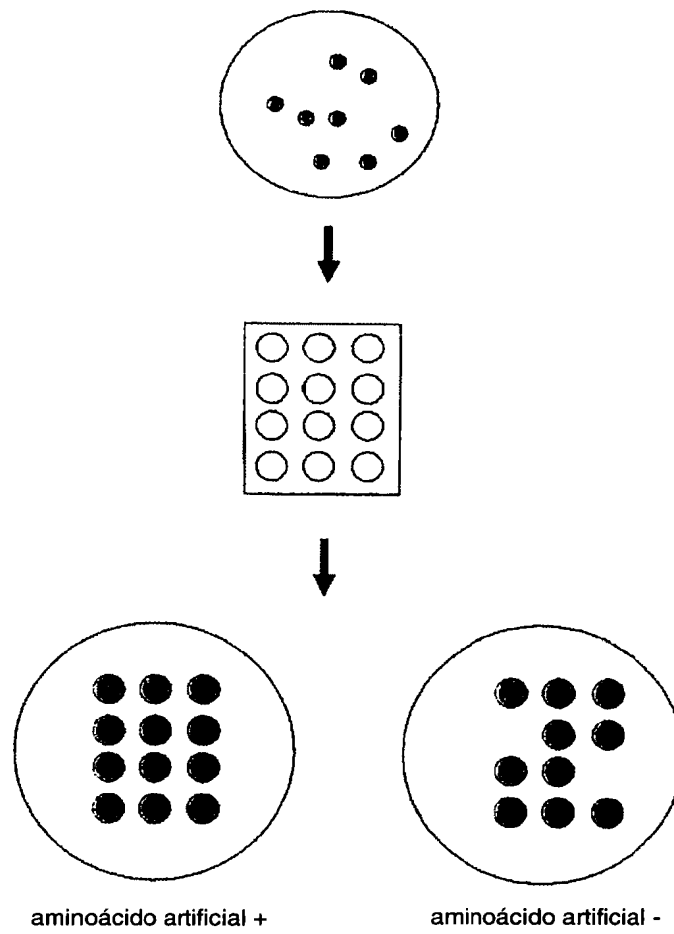
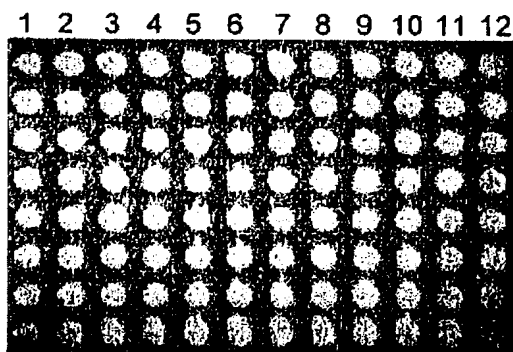
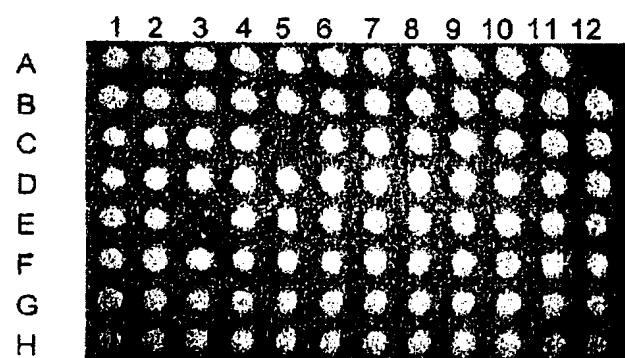


Figura 8

Y31- 1^a Triagem

Y31 MAP
LB agar
50ug/mL Amp
50ug/mL Kan
+ 2mM pAF



Y31 MAP
LB agar
50ug/mL Amp
50ug/mL Kan

Figura 9

N51- 1ª Triagem

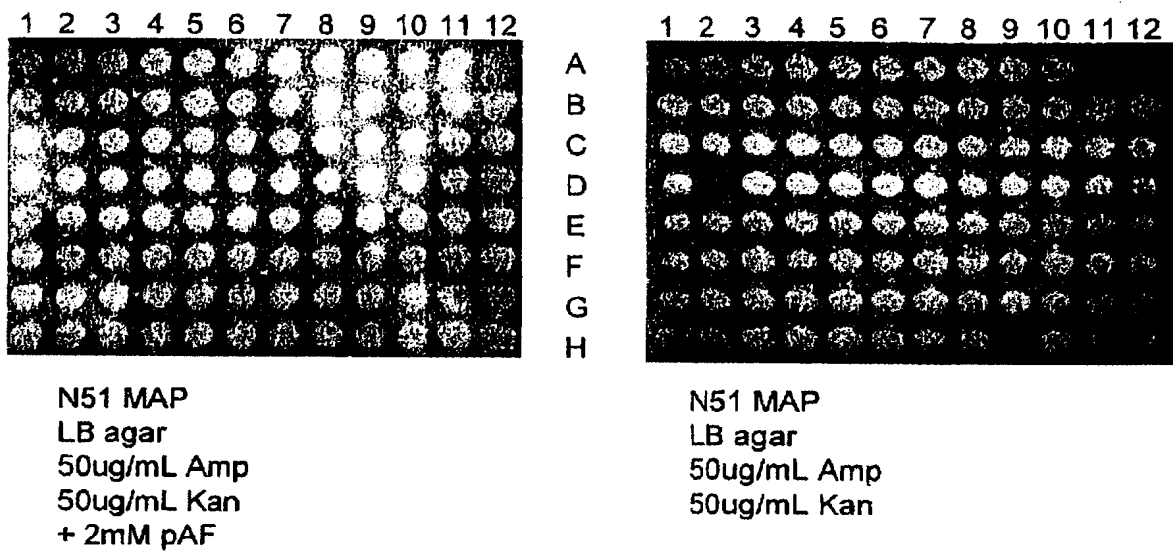


Figura 10A

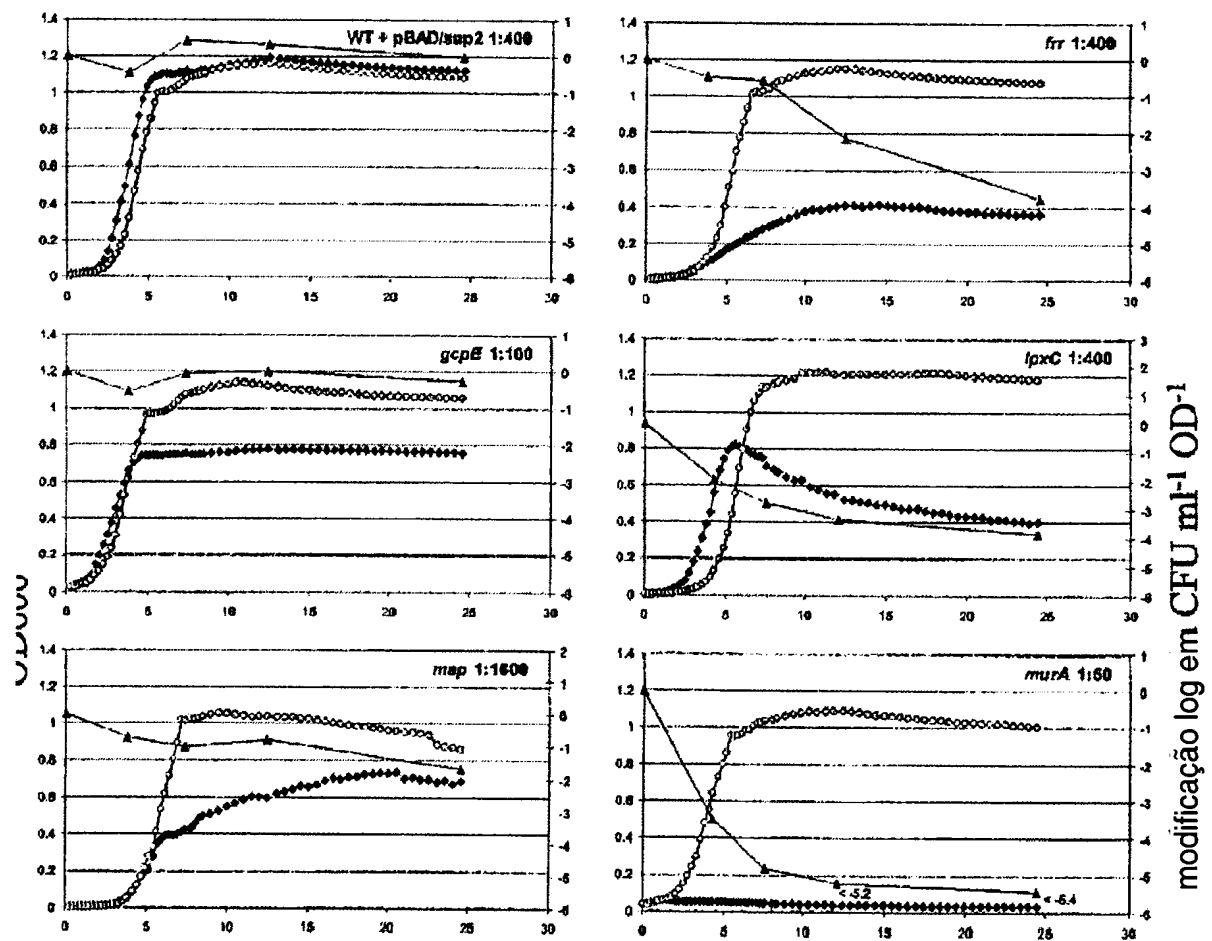


Figura 10B

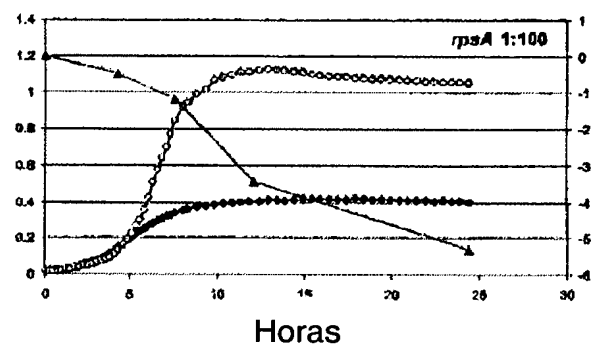
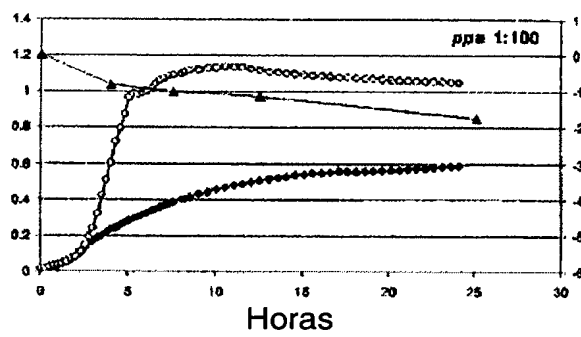


Figura 11

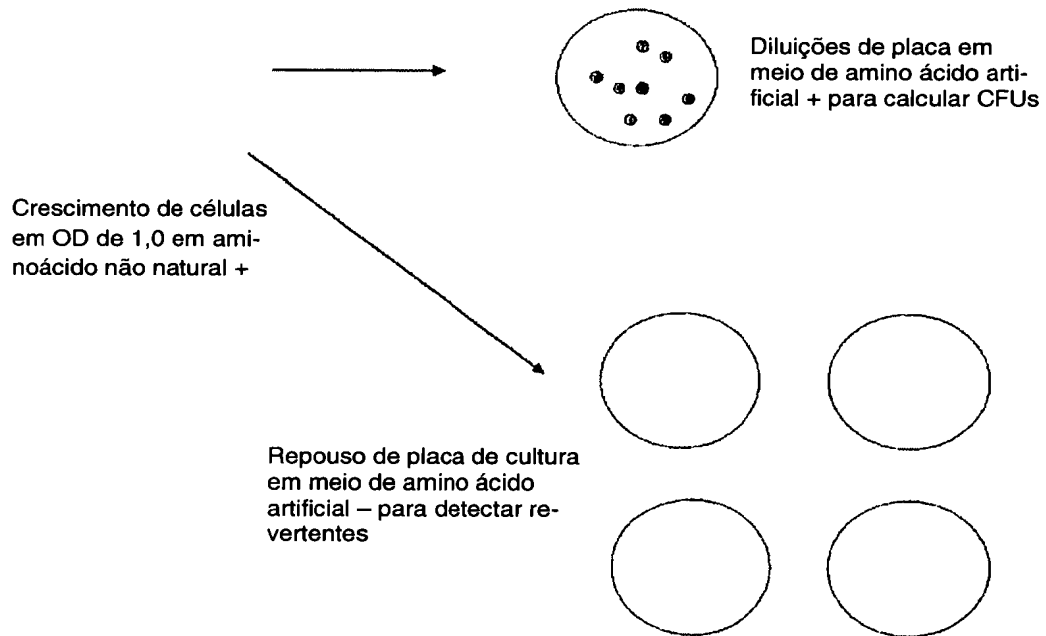


Figura 12

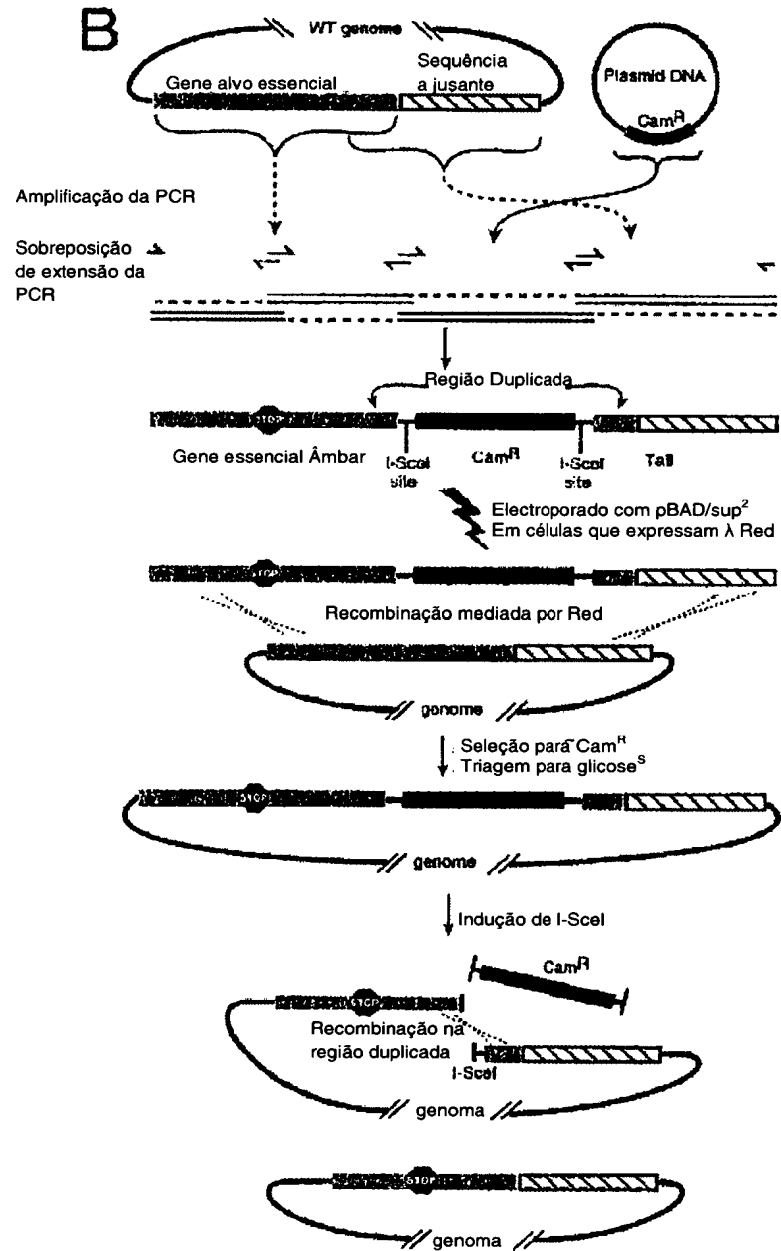


Figura 13

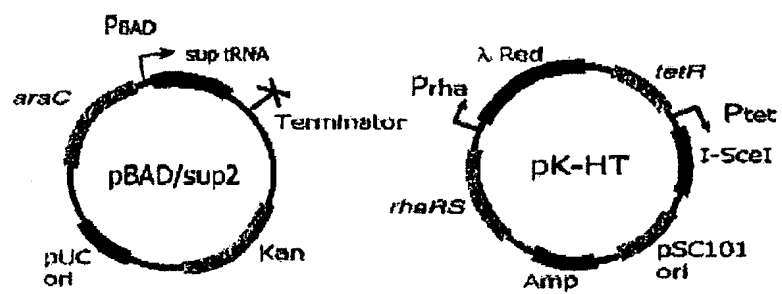


Figura 14

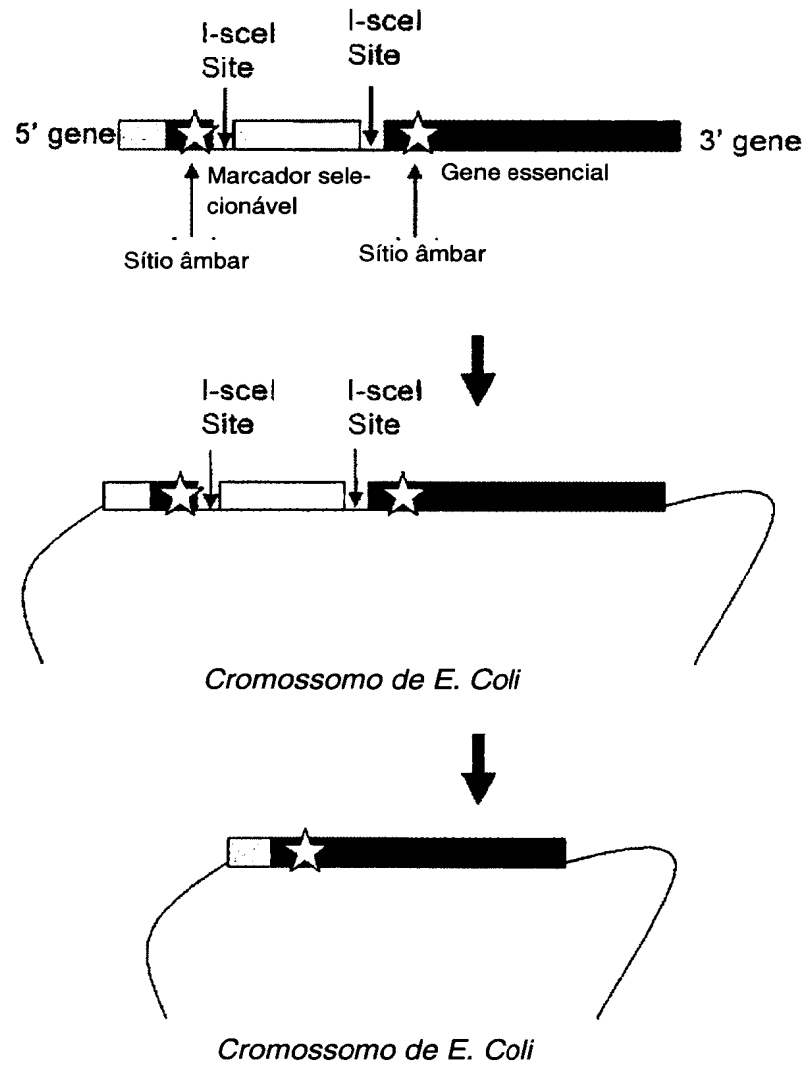


Figura 15

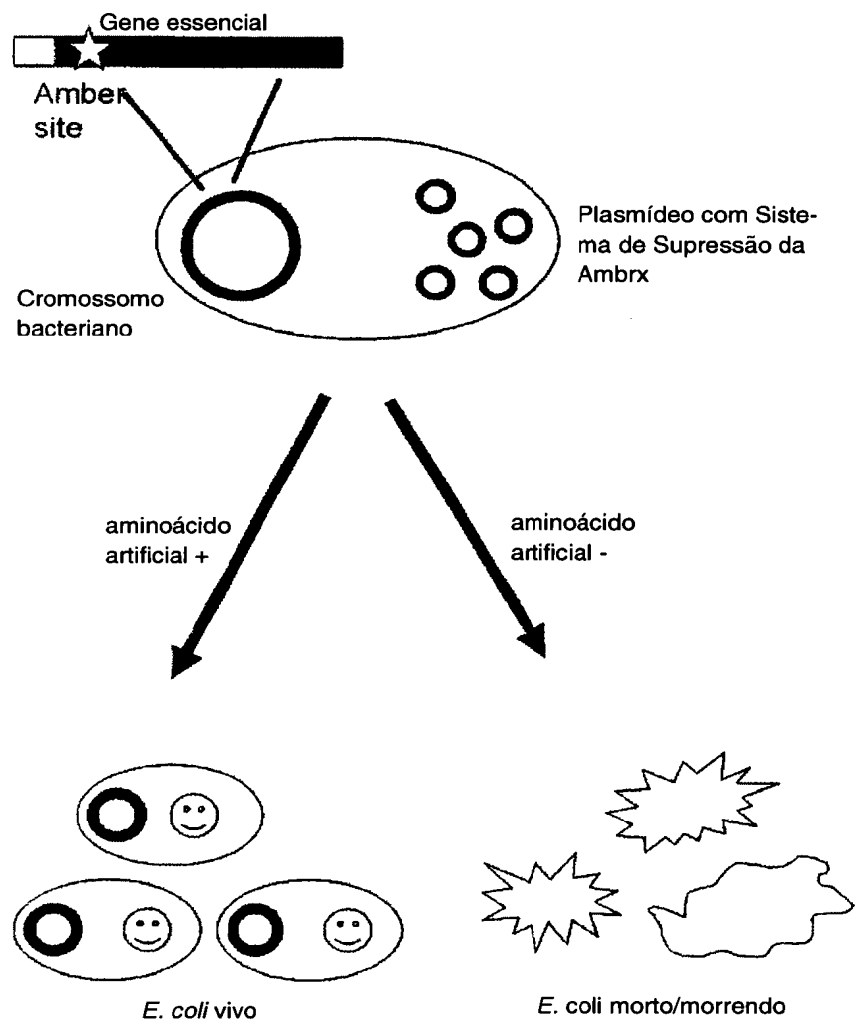


Figura 16

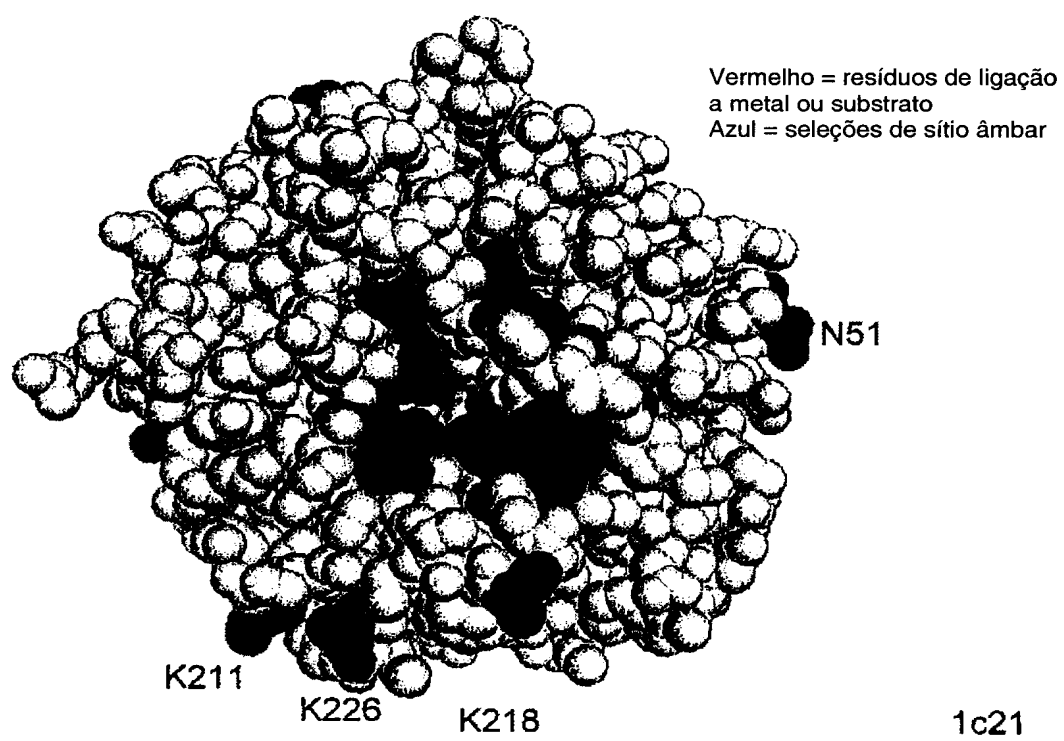
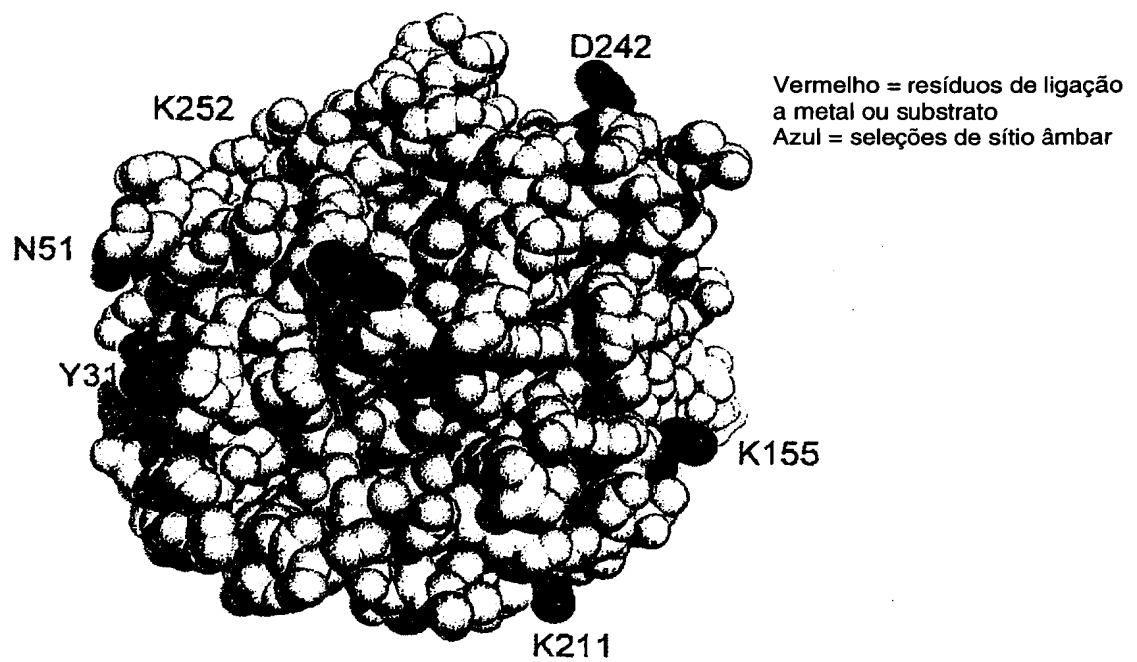


Figura 17



1c21

RESUMO

“MICRO-ORGANISMOS E VACINAS DEPENDENTES DE REPLICAÇÃO ATRAVÉS DO USO DE AMINOÁCIDOS NÃO NATURAIS”

5 Composições e métodos para produzir vacinas, incluindo métodos em que vacinas de organismo total são fornecidas com capacidades de replicação limitadas, desse modo, aumentando a segurança e eficácia da vacina, através do uso de aminoácido não naturais ou não naturalmente codificados.

LISTAGEM DE SEQUÊNCIAS

<110> Tian, Feng
Hehli, Brad

5

<120> Microorganismos e Vacinas Dependentes de Replicação Através do Uso
de Aminoácidos Não Naturais

<130> AMBX-0157.00PCT

10

<150> 61/100.688

<151> 26-09-2008

<160> 35

15

<170> PatentIn version 3.4

<210> 1

<211> 77

20

<212> DNA

<213> Artificial Sequence

<220>

<223> tRNA mutante derivado de tRNA Methanococcus jannaschii

25

<400> 1

ccggcggttag ttcagcaggg cagaacggcg gactctaaat ccgcatggcg ccggttcaaa 60

tccggcccgc cggacca

77

30

<210> 2

<211> 77

<212> DNA

35

<213> Artificial Sequence

<220>

<223> tRNA mutante derivado de tRNA Methanococcus jannaschii

<400> 2

ccggcggttag ttcagcaggg cagaacggcg gactctaaat ccgcatggcg ctggttcaaa 60

5

tccagcccg cggacca

77

<210> 3

10 <211> 77

<212> DNA

<213> Artificial Sequence

<220>

15 <223> tRNA mutante derivado de tRNA Methanococcus jannaschii

<400> 3

ccggcggttag ttcagcaggg cagaacggcg gactctaaat ccgcatggcg caggttcaaa 60

20

tcctgcccgc cggacca

77

<210> 4

<211> 921

25 <212> DNA

<213> Artificial Sequence

<220>

30 <223> Sintetase mutante derivada de sintetase de Methanococcus
jannaschii

<400> 4

atggacgaat tgaaatgat aaagagaaac acatctgaaa ttatcagcga ggaagagtta 60

35

agagagggtt taaaaaaaga tgaaaaatct gctgttatag gtttgaacc aagtggtaaa 120

atacatttag ggcattatct ccaaataaaa aagatgattg attacaaaa tgctggattt 180

gatataatta tatattggc tgattacac gcctatttaa accagaaagg agagtggat 240

gagattagaa aaataggaga ttataacaaa aaagttttg aagcaatggg gttaaaggca 300

5

aaatatgtt atggaagtga acatggtctt gataaggatt atacactgaa tgtctataga 360

ttggctttaa aaactacctt aaaaagagca agaaggagta tggaacttat agcaagagag 420

10 gatgaaaatc caaaggttgc tgaagttatc tatccaataa tgcaggttaa tgggattcat 480

tatgagggcg ttgatgttc agttggaggg atggagcaga gaaaaatata catgttagca 540

agggagcttt taccaaaaaa ggttgtttg atcacaacc ctgtctaac gggtttggat 600

15

ggagaaggaa agatgagtc ttcaaaaggg aattttatag ctgtgatga ctctccagaa 660

gagattaggg ctaagataaa gaaagcatac tgcccagctg gagttgtga aggaaatcca 720

20 ataatggaga tagctaaata ctccttgaa ttcctttaa ccataaaaag gccagaaaaa 780

tttggaggag attgacagt taatagctat gaggagtag agagttatt taaaaataag 840

gaattgcac caatggatt aaaaaatgct gtagctgaag aactataaa gatttagag 900

25

ccaattagaa agagattata a 921

<210> 5

30 <211> 306

<212> PRT

<213> Artificial Sequence

<220>

35 <223> Sintetase mutante derivada de sintetase de Methanococcus
jannaschii

<400> 5

Met Asp Glu Phe Glu Met Ile Lys Arg Asn Thr Ser Glu Ile Ile Ser

1 5 10 15

5

Glu Glu Glu Leu Arg Glu Val Leu Lys Lys Asp Glu Lys Ser Ala Val

20 25 30

10

Ile Gly Phe Glu Pro Ser Gly Lys Ile His Leu Gly His Tyr Leu Gln

35 40 45

15 Ile Lys Lys Met Ile Asp Leu Gln Asn Ala Gly Phe Asp Ile Ile Ile

50 55 60

Tyr Leu Ala Asp Leu His Ala Tyr Leu Asn Gln Lys Gly Glu Leu Asp

20 65 70 75 80

Glu Ile Arg Lys Ile Gly Asp Tyr Asn Lys Lys Val Phe Glu Ala Met

85 90 95

25

Gly Leu Lys Ala Lys Tyr Val Tyr Gly Ser Glu His Gly Leu Asp Lys

100 105 110

30

Asp Tyr Thr Leu Asn Val Tyr Arg Leu Ala Leu Lys Thr Thr Leu Lys

115 120 125

35 Arg Ala Arg Arg Ser Met Glu Leu Ile Ala Arg Glu Asp Glu Asn Pro

130 135 140

Lys Val Ala Glu Val Ile Tyr Pro Ile Met Gln Val Asn Gly Ile His
 145 150 155 160

5

Tyr Glu Gly Val Asp Val Ala Val Gly Gly Met Glu Gln Arg Lys Ile
 165 170 175

10 His Met Leu Ala Arg Glu Leu Leu Pro Lys Lys Val Val Cys Ile His
 180 185 190

15 Asn Pro Val Leu Thr Gly Leu Asp Gly Glu Gly Lys Met Ser Ser Ser
 195 200 205

20 Lys Gly Asn Phe Ile Ala Val Asp Asp Ser Pro Glu Glu Ile Arg Ala
 210 215 220

Lys Ile Lys Lys Ala Tyr Cys Pro Ala Gly Val Val Glu Gly Asn Pro
 225 230 235 240

25

Ile Met Glu Ile Ala Lys Tyr Phe Leu Glu Tyr Pro Leu Thr Ile Lys
 245 250 255

30 Arg Pro Glu Lys Phe Gly Gly Asp Leu Thr Val Asn Ser Tyr Glu Glu
 260 265 270

35 Leu Glu Ser Leu Phe Lys Asn Lys Glu Leu His Pro Met Asp Leu Lys
 275 280 285

Asn Ala Val Ala Glu Glu Leu Ile Lys Ile Leu Glu Pro Ile Arg Lys
 290 295 300

5 Arg Leu
 305

<210> 6

10 <211> 88

<212> DNA

<213> Halobacterium sp. NRC-1

<400> 6

15 cccagggtag ccaagctcgg ccaacggcga cggactctaa atccgtctc gtaggagttc 60

gagggttcga atccctccc tgggacca 88

20 <210> 7

<211> 88

<212> DNA

<213> Artificial Sequence

25 <220>

<223> tRNA ortogonal que reconhece um códon opala

<400> 7

gcgggggttg ccgagcctgg ccaaaggcgc cggactcaa atccggtccc gtaggggttc 60

30

cggggttcaa atccccgccc ccgcacca 88

<210> 8

35 <211> 77

<212> DNA

<213> Methanococcus jannaschii

<400> 8

ccggcggtag ttcagcaggg cagaacggcg gactctaaat ccgcatggcg ctggttcaaa 60

5 tccggccccg cggacca

77

<210> 9

<211> 65

10 <212> DNA

<213> Artificial Sequence

<220>

<223> Primer de oligonucleotídeo sintético

15

<400> 9

gtaacgctga attcccggcg gtagtcagc agggcagaac ggcggactct aaatccgcat 60

ggcgc

65

20

<210> 10

<211> 67

<212> DNA

25 <213> Artificial Sequence

<220>

<223> Primer de oligonucleotídeo sintético

30 <400> 10

gatctgcagt ggtccggcgg gccggatttg aaccggcgcc atgcggattt agagtccgcc 60

gttctgc

67

35

<210> 11

<211> 67

<212> DNA

<213> Artificial Sequence

<220>

5 <223> Primer de oligonucleotídeo sintético

<400> 11

gatctgcagt ggtccggcgg gctggatttg aaccagcgcc atgcggattt agagtccgcc 60

10 gttctgc 67

<210> 12

<211> 67

15 <212> DNA

<213> Artificial Sequence

<220>

<223> Primer de oligonucleotídeo sintético

20

<400> 12

gatctgcagt ggtccggcgg gcaggatttg aacctgcgcc atgcggattt agagtccgcc 60

gttctgc 67

25

<210> 13

<211> 49

<212> DNA

30 <213> Artificial Sequence

<220>

<223> Primer de oligonucleotídeo sintético

35

<400> 13

cgccggacca ctgcagatcc ttagcgaaag ctaaggattt ttttaagc 49

<210> 14

<211> 23

<212> DNA

5 <213> Artificial Sequence

<220>

<223> Primer de oligonucleotídeo sintético

10 <400> 14

caaattcgtc catatgggat tcc

23

<210> 15

15 <211> 20

<212> DNA

<213> Artificial Sequence

<220>

20 <223> Primer de oligonucleotídeo sintético

<400> 15

gtaacgctga attcccggcg

20

25

<210> 16

<211> 600

<212> DNA

<213> Homo sapiens

30

<400> 16

atggggccacc accaccacca ccacttcca accattccct tatccaggct tttgacaac 60

gctatgctcc gcgcccacg tctgcaccag ctggcctttg acacctacca ggagttgaa 120

35

gaagcctaga tcccaaagga acagaagtat tcattcctgc agaaccacca gacctccctc 180

tgtttctcag agtctattcc gacaccctcc aacagggagg aaacacaaca gaaatccaac 240

ctagagctgc tccgcatctc cctgctgctc atccagtcgt ggctggagcc cgtgcagttc 300

5 ctcaggagtg tcttcgcca cagcctggtg tacggcgcct ctgacagcaa cgtctatgac 360

ctcctaaagg acctagagga aggcattcaa acgctgatgg ggaggctgga agatggcagc 420

ccccggactg ggcagatctt caagcagacc tacagcaagt tcgacacaaa ctcacacaac 480

10

gatgacgcac tactcaagaa ctacgggctg ctctactgct tcaggaagga catggacaag 540

gtcgagacat tctgcgcat cgtgcagtgc cgctctgtgg agggcagctg tggcttctaa 600

15

<210> 17

<211> 306

<212> PRT

<213> Artificial Sequence

20

<220>

<223> Sintetase mutante derivada de sintetase de Methanococcus
jannaschii

25

<400> 17

Met Asp Glu Phe Glu Met Ile Lys Arg Asn Thr Ser Glu Ile Ile Ser

1 5 10 15

30

Glu Glu Glu Leu Arg Glu Val Leu Lys Lys Asp Glu Lys Ser Ala Val

20 25 30

35

Ile Gly Phe Glu Pro Ser Gly Lys Ile His Leu Gly His Tyr Leu Gln

35 40 45

Ile Lys Lys Met Ile Asp Leu Gln Asn Ala Gly Phe Asp Ile Ile Ile

50 55 60

5

Tyr Leu Ala Asp Leu His Ala Tyr Leu Asn Gln Lys Gly Glu Leu Asp

65 70 75 80

10

Glu Ile Arg Lys Ile Gly Asp Tyr Asn Lys Lys Val Phe Glu Ala Met

85 90 95

15 Gly Leu Lys Ala Lys Tyr Val Tyr Gly Ser Glu His Gly Leu Asp Lys

100 105 110

Asp Tyr Thr Leu Asn Val Tyr Arg Leu Ala Leu Lys Thr Thr Leu Lys

20 115 120 125

Arg Ala Arg Arg Ser Met Glu Leu Ile Ala Arg Glu Asp Glu Asn Pro

130 135 140

25

Lys Val Ala Glu Val Ile Tyr Pro Ile Met Gln Val Asn Gly Ile His

145 150 155 160

30

Tyr Glu Gly Val Asp Val Ala Val Gly Gly Met Glu Gln Arg Lys Ile

165 170 175

35 His Met Leu Ala Arg Glu Leu Leu Pro Lys Lys Val Val Cys Ile His

180 185 190

Asn Pro Val Leu Thr Gly Leu Asp Gly Glu Gly Lys Met Ser Ser Ser
 195 200 205

5

Lys Gly Asn Phe Ile Ala Val Asp Asp Ser Pro Glu Glu Ile Arg Ala
 210 215 220

10 Lys Ile Lys Lys Ala Tyr Cys Pro Ala Gly Val Val Glu Gly Asn Pro
 225 230 235 240

Ile Met Glu Ile Ala Lys Tyr Phe Leu Glu Tyr Pro Leu Thr Ile Lys
 15 245 250 255

Arg Pro Glu Lys Phe Gly Gly Asp Leu Thr Val Asn Ser Tyr Glu Glu
 260 265 270

20

Leu Glu Ser Leu Phe Lys Asn Lys Glu Leu His Pro Met Arg Leu Lys
 275 280 285

25

Asn Ala Val Ala Glu Glu Leu Ile Lys Ile Leu Glu Pro Ile Arg Lys
 290 295 300

30 Arg Leu
 305

<210> 18

35 <211> 43

<212> DNA

<213> Artificial

<220>

<223> Primer 1 de Y31

5 <400> 18

gtattaccct gttatcccta tggcgaaatt cagaatgatt ctc 43

<210> 19

10 <211> 39

<212> DNA

<213> Artificial

<220>

15 <223> Primer 2 de Y31

<400> 19

cagaaaccgc gtgtgttcc taaacaatgt aatcattac 39

20

<210> 20

<211> 39

<212> DNA

<213> Artificial

25

<220>

<223> Primer 3 de Y31

<400> 20

30 gtaatgatta cattgttag gaacaacacg cggtttctg 39

<210> 21

<211> 21

35 <212> DNA

<213> Artificial

<220>

<223> Primer 4 de Y31

<400> 21

5 gagagcgggtg ttgcgtact g

21

<210> 22

<211> 44

10 <212> DNA

<213> Artificial

<220>

<223> Primer 5 de Y31

15

<400> 22

attaccctgt tatccctaga cgctcagtg aacgaaaact cacg

44

20 <210> 23

<211> 31

<212> DNA

<213> Artificial

25 <220>

<223> Primer 6 de Y31

<400> 23

tagggataac agggaatac aattcaggt g

31

30

<210> 24

<211> 22

<212> DNA

35 <213> Artificial

<220>

<223> Primer 7 de Y31

<400> 24

ggtcagcttg tcgaaagtac cg

22

5

<210> 25

<211> 40

<212> DNA

10 <213> Artificial

<220>

<223> Primer 8 de Y31

15 <400> 25

ctacgggtcg atcatctcca gctagggata acagggtaat

40

<210> 26

20 <211> 43

<212> DNA

<213> Artificial

<220>

25 <223> Primer 1 de N51

<400> 26

gtattaccct gttatcccta tggcgaaatt cagaatgatt ctc

43

30

<210> 27

<211> 34

<212> DNA

<213> Artificial

35

<220>

<223> Primer 2 de N51

<400> 27

gattacattg ttaggaaca acacgcggtt tctg

34

5

<210> 28

<211> 34

<212> DNA

<213> Artificial

10

<220>

<223> Primer 3 de N51

<400> 28

15 cagaaaccgc gtgtgttcc taaacaatgt aatc

34

<210> 29

<211> 21

20 <212> DNA

<213> Artificial

<220>

<223> Primer 4 de N51

25

<400> 29

gagagcgggtg ttgcgtact g

21

30 <210> 30

<211> 44

<212> DNA

<213> Artificial

35 <220>

<223> Primer 5 de N51

<400> 30

attaccctgt tatccctaga cgctcagtg aacgaaaact cacg 44

5 <210> 31

<211> 31

<212> DNA

<213> Artificial

10 <220>

<223> Primer 6 de N51

<400> 31

tagggataac agggaatac aatttcaggt g 31

15

<210> 32

<211> 22

<212> DNA

20 <213> Artificial

<220>

<223> Primer 7 de N51

25 <400> 32

ggtcagcttg tcgaaagtac cg 22

<210> 33

30 <211> 33

<212> DNA

<213> Artificial

<220>

35 <223> Primer 8 de N51

<400> 33

ctaaacaatg taatctaggg ataacagggt aat 33

<210> 34

5 <211> 795

<212> DNA

<213> Artificial

<220>

10 <223> MAP Y31 Mutante Âmbar A12

<400> 34

atggctatct caatcaagac cccagaagat atcgaaaaaa tgcgcgtcgc tggccgactg 60

15 gctgccgaag tgctggagat gatcgaaccg taggttaaac cgggcgtcag caccggcgcag 120

ctggatcgca tctgtaatga ttacattgtt aatgaacaac acgcgggttc tgcctgcctc 180

ggctatcacg gctatccgaa atccgtttgc atctctatta atgaagtggg gtgccacggg 240

20

atccccggacg atgctaagct gctgaaagat ggcgatatcg ttaacattga tgcaccgta 300

atcaaagatg gttccacgg cgatacctcg aaaatgttta tcgtcggtaa gccgaccatc 360

25 atgggcgaac gtctgtgccg catcacgcaa gaaagcctgt acctggcgct acgcatggta 420

aaaccaggca ttaatctgcg cgaaatcggg gcggcgattc agaaatttgt cgaagcagaa 480

ggcttctccg tcgttcgtga atattgcgga cacggtattg gtcgcggctt ccatgaagaa 540

30

ccgcaggtgc tgcactatga ctcccgtgaa accaacgtcg tactgaaacc tgggatgacg 600

ttcaccatcg agccaatggg caacgcgggt aaaaaagaga tccgcacat gaaagatggc 660

35 tggacggtaa aaaccaaaga tcgcagcttg tctgcacaat atgagcatac tattgtggtg 720

actgataacg gctgcgaaat tctgacgcta cgcaaggatg acaccatccc ggcgataatc 780

tcgcacgacg aataa

795

5 <210> 35
 <211> 795
 <212> DNA
 <213> Artificial

10 <220>
 <223> MAP N51 Mutante Âmbiar D2

<400> 35

atggctatct caatcaagac cccagaagat atcgaaaaaa tgcgcgtcgc tggccgactg 60

15 gctgccgaag tgctggagat gatcgaaccg tatgttaaac cgggcgtcag caccggcgag 120

ctggatcgca tctgtaatga ttacattgtt taggaacaac acgcggtttc tgctgcctc 180

20 ggctatcacg gctatccgaa atccgtttgc atctctatta atgaagtggg gtgccacggg 240

atcccggacg atgctaagct gctgaaagat ggcgatatcg ttaacattga tgcaccgta 300

atcaaagatg gttccacgg cgatacctcg aaaatgttta tcgtcggtaa gccgaccatc 360

25 atgggcgaac gtctgtgccg catcacgcaa gaaagcctgt acctggcgct acgcatggta 420

aaaccaggca ttaatctgcg cgaaatcggg gcggcgattc agaaatttgc cgaagcagaa 480

30 ggcttctccg tcgttcgtga atattgcgga cacggtattg gtcgcggctt ccatgaagaa 540

ccgcagggtc tgcactatga ctcccgtgaa accaacgtcg tactgaaacc tgggatgacg 600

ttcaccatcg agccaatggg caacgcgggt aaaaaagaga tccgcacccat gaaagatggc 660

35 tggacggtaa aaaccaaaga tcgcagcttg tctgcacaat atgagcatac tattgtggtg 720

actgataacg gctgcgaaat tctgacgcta cgcaaggatg acaccatccc ggcgataatc 780

tcgcacgacg aataa

795