

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

(21) **PI0709935-5 A2**

(22) Data de Depósito: 07/03/2007  
(43) Data da Publicação: 02/08/2011  
(RPI 2117)



(51) *Int.Cl.:*  
C12N 15/74 2006.01

(54) Título: **SISTEMAS PARA A EXPRESSÃO DE COMPONENTES ORTOGONAIS DE TRADUÇÃO EM CÉLULAS HOSPEDEIRAS EUBACTERIANAS**

(30) Prioridade Unionista: 09/03/2006 US 60/780.973, 17/03/2006 US 60/783.497, 29/10/2006 US 60/855.336

(73) Titular(es): The Scripps Research Institute

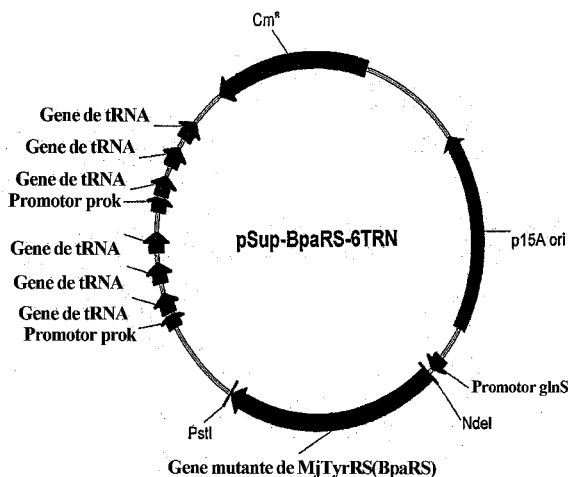
(72) Inventor(es): Peter G. Schultz, Youngha Ryu

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

(86) Pedido Internacional: PCT US2007005914 de 07/03/2007

(87) Publicação Internacional: WO 2007/103490 de 13/09/2007

(57) Resumo: SISTEMAS PARA A EXPRESSÃO DE COMPONENTES ORTOGONAIS DE TRADUÇÃO EM CÉLULAS HOSPEDEIRAS EUBACTERIANAS A invenção refere-se a composições e métodos para a produção in vivo de polipeptídeos que compreendem um ou mais aminoácidos não-naturais. Especificamente, a invenção fornece sistemas plasmídicos para a expressão eubacteriana eficiente de polipeptídeos que compreendem um ou mais aminoácidos não-naturais em posições geneticamente programadas.





"SISTEMAS PARA A EXPRESSÃO DE COMPONENTES  
ORTOGONAIS DE TRADUÇÃO EM CÉLULAS HOSPEDEIRAS EUBACTERIANAS"

REFERÊNCIA CRUZADA A PEDIDOS DE PATENTE AFINS

Este pedido de patente reivindica prioridade para  
5 e o benefício do pedido de patente provisório nº de série US  
60/780.973, depositado em 9 de março de 2006; pedido de  
patente provisório nº de série US 60/783.497, depositado em  
17 de março de 2006; e pedido de patente provisório nº de  
série US 60/855.336, depositado em 29 de outubro de 2006,  
10 cujos teores são aqui incorporados como referência em sua  
totalidade para todos propósitos.

DECLARAÇÃO QUANTO A DIREITOS ÀS INVENÇÕES FEITAS  
SOB PESQUISA E DESENVOLVIMENTO PATROCINADOS PELO GOVERNO  
FEDERAL

15 A invenção foi realizada com o patrocínio do  
Governo dos Estados Unidos da América do Norte, sob o  
contrato nº ER46051 do Ministério da energia. O Governo dos  
Estados Unidos pode ter certos direitos à invenção.

CAMPO TÉCNICO DA INVENÇÃO

20 O estudo da estrutura e função de proteínas tem se  
baseado historicamente nas propriedades químicas que estão  
disponíveis, usando os grupos R dos aminoácidos de  
ocorrência natural. Infelizmente, cada organismo conhecido,  
desde bactérias até os seres humanos, codifica os mesmos  
25 aminoácidos comuns com as raras exceções de selenocisteína  
(vide, por exemplo, Bock et al., *Molecular Biology* 5:515-20  
(2002)) e pirrolisina (vide, por exemplo, Srinivasan et al.,  
*Science* 296:1459-62 (1991)). Esta seleção limitada de grupos

R restringiu o estudo da estrutura e função de proteínas, onde os estudos estão confinados pelas propriedades químicas dos aminoácidos de ocorrência natural.

Foi desenvolvido um método genérico para a  
5 incorporação específica de sítio *in vivo* de aminoácidos não-naturais quimicamente diversos com propriedades físico-químicas e biológicas inusitadas em proteínas em organismos procarióticos e eucarióticos (Wang et al., *Science* 292:498-500 (2001); Chin et al., *Science* 301:964-967 (2003); e Wang  
10 e Schultz, *Angew. Chem. Int. Ed.* 44:34-66 (2005)). Este método se baseia em um par singular de códon-tRNA e a aminoacil-tRNA sintetase correspondente (aaRs ou simplesmente RS) para cada aminoácido não-natural que funciona eficientemente na tradução de proteínas, mas não sofre reação cruzada com  
15 qualquer um dos tRNAs, RSs, aminoácidos ou códons endógenos no organismo hospedeiro (isto é, ele deve ser ortogonal). O uso desses pares tRNA-RS ortogonais tornou possível codificar geneticamente um grande número de aminoácidos estruturalmente diversos, incluindo aqueles com química  
20 inusitada (Wang et al., *Proc. Natl. Acad. Sci. USA* 100:56-61 (2003)) e reatividade fotoquímica (Chin et al., *Proc. Natl. Acad. Sci. USA* 99:11020-11024 (2002); Wu et al., *J. Am. Chem. Soc.* 126:14306-14307 (2004)), bem como glicosilados (Zhang et al., *Science* 303:371-373 (2004)), fluorescentes  
25 (Wang e Schultz, *Angew. Chem. Int. Ed.* 44:34-66 (2005)), ligação de metais (Wang e Schultz, *Angew. Chem. Int. Ed.* 44:34-66 (2005) e aminoácidos ativos em oxirredução (Alfonta et al., *J. Am. Chem. Soc.* 125:14662-14663 (2003)). Um par

MjtRNA-Tyr(CUA)-MjTyrRS mutante específico foi particularmente útil para codificar novos aminoácidos em *E. coli* (Wang e Schultz, *Chem. Biol.* 8:883-890 (2001)).

Entretanto, apesar do sucesso desta técnica em  
5 incorporar uma série diversa de aminoácidos não-naturais *in vivo*, a eficiência do sistema de expressão para a produção de proteínas mutantes que contêm aminoácidos não-naturais, que não foi otimizada, e a eficiência da supressão do sistema ortogonal para superar o códon seletor pode ser  
10 deficiente. Há uma necessidade nessas técnicas de desenvolver reagentes para melhorar a eficiência da supressão de sistemas ortogonais de tradução. A invenção aqui descrita preenche estas e outras necessidades, como ficará evidente depois da revisão da descrição que se segue.

15 SUMÁRIO DA INVENÇÃO

A invenção fornece sistemas de vetores de expressão aperfeiçoados para a expressão bacteriana eficiente de proteínas mutantes que compreendem um ou mais aminoácidos não-naturais em sítios específicos codificados  
20 geneticamente por códons seletores (códons *nonsense* âmbar). Estes sistemas resultam em uma eficiência significativamente melhor na incorporação de aminoácidos não-naturais em proteínas em eubactérias (por exemplo, *E. coli*). As características inusitadas dos vetores de expressão da  
25 invenção são amplamente compatíveis com uma série de esqueletos de vetores de expressão em *E. coli* e cepas de *E. coli*, e são facilmente adaptadas para a expressão de outras proteínas ou tRNAs de interesse, além da expressão de amino-

acyl-tRNA sintetases ortogonais ou tRNAs supressores ortogonais.

Esta tecnologia de tradução ortogonal utilizada pela invenção é conhecida nessas técnicas. Entretanto, a invenção não está limitada em qualquer aspecto aos componentes específicos da tradução ortogonal que são usados (isto é, a amino-acyl-tRNA sintetase específica ou o tRNA supressor ortogonal específico). Na realidade, a invenção fornece composições e métodos aperfeiçoados que encontram amplo uso em sistemas de vetores de expressão bacterianos que não estão limitados à expressão de amino-acyl-tRNA sintetases ou RNAs supressores ortogonais.

Em alguns aspectos, a invenção fornece várias construções de ácidos nucleicos. Estas modalidades incluem, por exemplo:

Construção A: Construções que têm seqüências de nucleotídeos promotoras e terminadoras derivadas a partir de um gene de tRNA de prolina em *Escherichia coli* e uma seqüência de nucleotídeos expressável, onde as seqüências promotora e terminadora são ambas ligadas operacionalmente à seqüência de nucleotídeos expressável, e onde a seqüência de nucleotídeos expressável é heteróloga às seqüências de nucleotídeos promotora e extermadora.

Construção B: As construções que têm uma seqüência de nucleotídeos promotora que é um promotor de *E. coli glnS* modificado que tem a seqüência de SEQ ID NO: 13 e uma seqüência de nucleotídeos expressável, onde a seqüência de nucleotídeos promotora de *E. coli glnS* modificada está

ligada operacionalmente à sequência de nucleotídeos expressável.

Em alguns aspectos, as características das construções A e B são usadas no mesmo vetor, e onde as  
5 sequências de nucleotídeos expressáveis são diferentes.

Construção C: Construções que têm uma sequência de nucleotídeos que codifica um tRNA ortogonal (O-tRNA) e uma sequência de nucleotídeos que codifica uma amino-acil-tRNA sintetase ortogonal (O-RS), onde a O-RD amino-acila  
10 preferencialmente o O-tRNA com um aminoácido não-natural.

Construção D: Construções que têm um operon policistrônico que compreende uma pluralidade de sequências de nucleotídeos do gene de tRNA, onde pelo menos um gene de tRNA é separado de pelo menos um gene de tRNA adjacente por  
15 um ligante de polinucleotídeos heterólogos derivado de um ligante de polinucleotídeos de ocorrência natural de um operon de tRNA de ocorrência natural.

Em algumas modalidades, no caso em que as construções A, B e /c são usadas simultaneamente, a  
20 sequência de nucleotídeos expressável em A codifica o O-tRNA, e onde a sequência de nucleotídeos expressável em B codifica a O-RS.

Em algumas modalidades dessas construções, o gene do tRNA de prolina em *E. coli* é selecionado entre os genes  
25 *proK*, *ProL* e *proM* de *E. coli*. Em algumas modalidades, as sequências promotora e terminadora do tRNA de prolina de *E. coli* derivadas das sequências promotora e terminadora de *E. coli proK* fornecidas em SEQ ID NOS: 32 (promotora) e 33

(terminadora), respectivamente. Em algumas modalidades, a sequência de nucleotídeos expressável na construção A é um operon policistrônico que compreende uma pluralidade de uma ou mais sequências de nucleotídeos. Em algumas modalidades,

5 a sequência de nucleotídeos expressável na construção A codifica um tRNA, por exemplo, onde o tRNA é derivado de um ou mais tRNA de *Archaea*, ou onde o tRNA é codificado por uma sequência de nucleotídeos de SEQ ID NO: 1 (MjtRNA-Tyr(CUA)). Em algumas modalidades, a sequência de nucleotídeos

10 expressável é um operon policistrônico que compreende uma pluralidade das sequências de nucleotídeos de SEQ ID NO: 1 (MjtRNA-Tyr(CUA)). Em alguns aspectos, a sequência de nucleotídeos expressável é uma pluralidade de operons policistrônicos.

15 Quando a construção B é usada, a sequência de nucleotídeos expressável pode codificar um polipeptídeo, por exemplo, uma amino-acil-tRNA sintetase. Em alguns aspectos, a sequência de nucleotídeos expressável da construção B codifica uma amino-acil-tRNA sintetase ortogonal (O-RS),

20 onde a O-RS preferencialmente amino-acila um O-tRNA correspondente com um aminoácido não-natural. Em alguns aspectos, a O-RS é derivada de uma amino-acil-tRNA sintetase de *Methanococcus jannaschii*, por exemplo, uma tirosil-tRNA sintetase de *Methanococcus jannaschii*. Nestes aspectos,

25 opcionalmente, a O-RS tem uma substituição de ácido aspártico para arginina na posição do aminoácido 286 ou em uma posição análoga à posição 286, em relação à sequência de aminoácidos da tirosil-tRNA sintetase de *Methanococcus*

*jannaschii* do tipo selvagem, fornecida em SEQ ID NO: 2 (Mj-tRNA-Tyr RS do tipo selvagem).

Estas construções da invenção podem ser usadas em uma célula hospedeira, por exemplo, uma célula hospedeira eubacteriana, tal como uma célula de *E. coli*.

Quando a construção D é usada, o operon policistrônico pode compreender uma pluralidade de ligantes de polinucleotídeos heterólogos idênticos. Em alguns destes casos, pelo menos dois dos ligantes de polinucleotídeos heterólogos são diferentes. Este operon policistrônico de D pode compreender uma pluralidade de ligantes de polinucleotídeos heterólogos. Em algumas modalidades, o ligante de polinucleotídeos heterólogos de D pode compreender um nucleotídeo de timidina no terminal 5', um nucleotídeo de adenosina no terminal 3', ou ambos, um nucleotídeo de timidina no terminal 5' e um nucleotídeo de adenosina no terminal 3'. O ligante de polinucleotídeos heterólogos pode ser derivado do ligante de polinucleotídeos de ocorrência natural localizado entre os genes de tRNA de *Escherichia coli* selecionados entre: *valU* e *valX*; *ileT* e *alaT*; *serV* e *argV*; *valV* e *valW*; *glyT* e *thrT*; *metT* e *leuW*; *glnW* e *metU*; *hisR* e *leuT*; *glnU* e *glnW*; *leuP* e *leuV*; *glnV* e *glnX*; *alaW* e *alaX*; *ileU* e *alaU*; *ileV* e *alaV*; *metU* e *glnV*; *glyW* e *cysT*; *argX* e *hisR*; e *argY* e *argZ*. Em alguns aspectos, o ligante de polinucleotídeos heterólogos de D é derivado da sequência de nucleotídeos de SEQ ID NO: 14 (ligante *valU/valX*) ou 15 (ligante *ileT/alaT*).

Em outros aspectos, a invenção fornece sistemas de



tradução para a expressão de um polipeptídeo de interesse, compreendendo pelo menos um aminoácido não-natural em uma posição especificada. Estes sistemas da invenção contêm:

(a) um aminoácido não-natural;

5 (b) uma construção de ácido nucléico, sendo que a construção compreende uma seqüência de nucleotídeos que codifica um tRNA ortogonal (O-tRNA) e uma seqüência de nucleotídeos que codifica uma amino-acil-tRNA sintetase ortogonal (O-RS), onde a O-RS preferencialmente amino-acila  
10 o O-tRNA com o aminoácido não-natural; e

(c) um polinucleotídeo que codifica o polinucleotídeo de interesse, sendo que o polinucleotídeo compreende pelo menos um códon seletor que é reconhecido pelo O-tRNA, onde a posição do códon seletor no  
15 polinucleotídeo controla a posição especificada do aminoácido não-natural no polipeptídeo de interesse depois da expressão do polinucleotídeo, para produzir o polipeptídeo.

Nestes sistemas da invenção, a construção de ácido  
20 nucléico pode compreender qualquer uma ou mais características listadas abaixo:

(1) as seqüências de nucleotídeos promotora e terminadora derivadas de um gene do tRNA de *Escherichia coli*, onde as seqüências promotora e terminadora estão,  
25 ambas, ligadas operacionalmente à seqüência de nucleotídeos que compreende ou codifica o O-tRNA, e onde o o-tRNA é heterólogo às seqüências de nucleotídeos promotora e terminadora;

(2) uma sequência de nucleotídeos correspondente a um promotor de *E. coli glnS* modificado que tem uma sequência de SEQ ID NO: 13, onde a sequência de nucleotídeos de *E. coli glnS* modificada está ligada operacionalmente à  
5 sequência de nucleotídeos que codifica a O-RS; ou

(3) um operon policistrônico que compreende uma pluralidade das sequências de nucleotídeos do gene de O-tRNA, onde pelo menos um gene de O-tRNA é separado do pelo menos um gene de O-tRNA adjacente por um ligante de  
10 polinucleotídeos heterólogos derivado de um ligante de polinucleotídeos de ocorrência natural a partir de um operon de tRNA de ocorrência natural.

Nestes sistemas, o gene do tRNA de prolina de *E. coli* pode ser selecionado entre os genes *proK*, *proL* e *proM*  
15 de *E. coli*. As sequências promotora e terminadora do tRNA de prolina de *E. coli* podem ser derivadas a partir das sequências promotora e terminadora de *E. coli proK* fornecida em SEQ ID NOS: 32 (promotora) e 33 (terminadora), respectivamente. O operon policistrônico pode compreender  
20 uma pluralidade de ligantes de polinucleotídeos heterólogos idênticos, por exemplo, onde pelo menos dois dos ligantes de polinucleotídeos heterólogos são diferentes. Nestes sistemas, o ligante de polinucleotídeos heterólogos pode compreender um nucleotídeo de timidina no terminal 5', ou um  
25 nucleotídeo de adenosina no terminal 3', ou ambos, um nucleotídeo de timidina no terminal 5' e um nucleotídeo de adenosina no terminal 3'. O ligante de polinucleotídeos heterólogos pode ser derivado de um ligante de

polinucleotídeos de ocorrência natural localizado entre os genes do tRNA de *Escherichia coli* endógenos selecionados entre: *valU* e *valX*; *ileT* e *alaT*; *serV* e *argV*; *valV* e *valW*; *glyT* e *thrT*; *metT* e *leuW*; *glnW* e *metU*; *hisR* e *leuT*; *glnU* e  
 5 *glnW*; *leuP* e *leuV*; *glnV* e *glnX*; *alaW* e *alaX*; *ileU* e *alaU*; *ileV* e *alaV*; *metU* e *glnV*; *glyW* e *cysT*; *argX* e *hisR*; e *argY* e *argZ*. Em alguns aspectos, o ligante de polinucleotídeos heterólogos é derivado da sequência de nucleotídeos de SEQ ID NO: 14 (ligante *valU/valX*) ou 15 (ligante *ileT/alaT*).

10               Nestes sistemas, o O-tRNA pode ser derivado a partir de um ou mais tRNA de *Archaea*. Em alguns aspectos dos sistemas, a sequência de nucleotídeos que codifica um O-tRNA é um operon policistrônico que compreende uma pluralidade de sequência de nucleotídeos que codificam um O-tRNA. Em alguns  
 15 aspectos, a sequência de nucleotídeos que codifica um O-tRNA compreende uma sequência de nucleotídeos de SEQ ID NO: 1 (*MjtRNA-Tyr(CUA)*). Em algumas modalidades, a sequência de nucleotídeos expressável é um operon policistrônico que compreende uma pluralidade das sequências de nucleotídeos de  
 20 SEQ ID NO: 1 (*MjtRNA-Tyr(CUA)*). Em outros aspectos do sistema, a sequência de nucleotídeos que codifica um O-tRNA é um operon policistrônico que compreende uma pluralidade da sequência de nucleotídeos de SEQ ID NO : 1 (*MjtRNA-Tyr(CUA)*).

25               Em outros aspectos, o sistema de tradução pode utilizar uma O-RS derivada de uma amino-acil-tRNA sintetase de *Methanococcus jannaschii*, por exemplo, uma tirosil-tRNA sintetase de *Methanococcus jannaschii*. Em alguns aspectos,

esta O-RS tem uma substituição de ácido aspártico para arginina na posição do aminoácido 286 ou em uma posição análoga à posição 286, em relação à sequência de aminoácidos da tirosil-tRNA sintetase de *Methanococcus jannaschii* do tipo selvagem fornecida em SEQ ID NO: 2 (Mj-tRNA-Tyr RS do tipo selvagem).

Em alguns aspectos destes sistemas, os componentes ficam contidos em uma célula hospedeira, por exemplo, uma célula hospedeira eubacteriana, tal como uma célula de *E. coli*.

Em outros aspectos, a invenção fornece métodos para produzir, em uma célula hospedeira, um polipeptídeo de interesse que compreende um aminoácido não-natural em uma posição específica. Estes métodos têm as seguintes etapas:

(a) disponibilizar: (i) um aminoácido não-natural; (ii) uma construção de ácido nucléico que compreende uma sequência de nucleotídeos que codifica um tRNA ortogonal (O-tRNA) e uma sequência de nucleotídeos que codifica uma amino-acil-tRNA sintetase ortogonal (O-RS), onde a O-RS preferencialmente amino-acila o O-tRNA com o aminoácido não-natural; e (iii) um polinucleotídeo que codifica o polipeptídeo de interesse, sendo que o polinucleotídeo compreende pelo menos um códon seletor que é reconhecido pelo O-tRNA, e onde a posição do códon seletor se correlaciona com a posição especificada do aminoácido não-natural no polipeptídeo de interesse; (iv) uma célula hospedeira que compreende (i), (ii) e (iii); e

(b) desenvolver a célula hospedeira; e

(c) incorporar o aminoácido não-natural na posição especificada no polipeptídeo durante a tradução do polipeptídeo na célula hospedeira, produzindo desta forma o polipeptídeo de interesse que compreende o aminoácido não-natural na posição especificada.

Nestes métodos, o O-tRNA pode ser derivado de um ou mais tRNA de *Archaea*. A construção usada nos métodos pode utilizar um operon policistrônico que compreende uma pluralidade de seqüências de nucleotídeos que codificam uma ou mais espécies de O-tRNA. O O-tRNA pode compreender uma seqüência de nucleotídeos de SEQ ID NO: 1 (*MjtRNA-Tyr(CUA)*), ou alternativamente, um operon policistrônico que compreende uma pluralidade das seqüências de nucleotídeos de SEQ ID NO: 1 (*MjtRNA-Tyr(CUA)*). Em alguns aspectos deste método, a O-RS pode ser derivada de uma amino-acil-tRNA sintetase de *Methanococcus jannaschii*. Em alguns aspectos, a O-RS tem uma substituição de ácido aspártico para arginina na posição do aminoácido 286 ou em uma posição análoga à posição 286, em relação à seqüência de aminoácidos da tirosil-tRNA sintetase de *Methanococcus jannaschii* do tipo selvagem fornecida em SEQ ID NO: 2 (*Mj-tRNA-Tyr RS* do tipo selvagem).

Estes métodos para produzir os polipeptídeos podem utilizar uma série de construções de ácidos nucléicos, onde as construções podem ser qualquer uma entre:

(I) seqüências de nucleotídeos promotoras e terminadoras derivadas de um gene do tRNA de prolina de *Escherichia coli*, onde as seqüências promotora e terminadora estão, ambas, ligadas operacionalmente à seqüência de

nucleotídeos que codifica o O-tRNA, e onde a sequência de nucleotídeos que codifica o O-tRNA é heteróloga às sequências promotora e terminadora;

(II) uma sequência de nucleotídeos promotora  
5 correspondente a um promotor de *E. coli glnS* modificado que tem uma sequência de nucleotídeos de SEQ ID NO: 13, onde a sequência de nucleotídeos de *E. coli glnS* modificada está ligada operacionalmente à sequência de nucleotídeos que codifica O-RS; e

10 (III) um operon policistrônico que compreende uma pluralidade de sequências de nucleotídeos do gene de O-tRNA, onde pelo menos um gene é separado de pelo menos um gene de O-tRNA adjacente por um ligante de polinucleotídeos heterólogos derivado de um ligante de  
15 polinucleotídeos de ocorrência natural de um operon de tRNA de ocorrência natural.

Quando (I) é usado, o tRNA de prolina de *E. coli* pode ser selecionado entre o tRNA de *E. coli proK*, *proL* e *proM*. Quando *E. coli proK* é usado, SEQ ID NO: 32 (promotora)  
20 e 33 (terminadora) podem ser utilizadas.

Quando um operon policistrônico de (III) é usado, uma pluralidade de ligantes de polinucleotídeos heterólogos pode ser utilizada, alternativamente quando pelo menos dois dos ligantes de polinucleotídeos heterólogos são diferentes.  
25 Em alguns aspectos, o operon policistrônico de (III) usa um nucleotídeo de timidina no terminal 5', um nucleotídeo de adenosina no terminal 3', ou ambos, um nucleotídeo de timidina no terminal 5' e um nucleotídeo de adenosina no

terminal 3'. Além disso, o ligante de polinucleotídeos heterólogos pode ser derivado do ligante de polinucleotídeos de ocorrência natural localizado entre os genes do tRNA de *Escherichia coli* endógenos selecionados entre: *valU* e *valX*;  
 5 *ileT* e *alaT*; *serV* e *argV*; *valV* e *valW*; *glyT* e *thrT*; *metT* e *leuW*; *glnW* e *metU*; *hisR* e *leuT*; *glnU* e *glnW*; *leuP* e *leuV*; *glnV* e *glnX*; *alaW* e *alaX*; *ileU* e *alaU*; *ileV* e *alaV*; *metU* e *glnV*; *glyW* e *cysT*; *argX* e *hisR*; e *argY* e *argZ*. Particularmente, a sequência de nucleotídeos de SEQ ID NO:  
 10 14 (ligante *valU/valX*) ou 15 (ligante *ileT/alaT*) pode ser usada. Em alguns aspectos, a célula hospedeira é uma célula hospedeira eubacteriana, por exemplo, uma célula hospedeira de *Escherichia coli*.

Em outros aspectos, a invenção fornece métodos  
 15 para produzir proteínas que têm aminoácidos não-naturais. Estes métodos usam sistemas de vetores inusitados da invenção, tornando possível a produção aperfeiçoada das proteínas que contêm os aminoácidos não-naturais. Estes métodos para produzir um polipeptídeo de interesse que  
 20 compreende um aminoácido não-natural em uma posição especificada são realizados em uma célula hospedeira, onde as etapas do método incluem: (a) disponibilizar:

- (i) um aminoácido não-natural;
- (ii) uma construção de ácido nucléico que  
 25 compreende uma sequência de nucleotídeos que codifica um tRNA ortogonal (O-tRNA);
- (iii) uma construção de ácido nucléico que compreende uma sequência de nucleotídeos que codifica uma

amino-acil-tRNA sintetase ortogonal (O-RS), onde a O-RS preferencialmente amino-acila o O-tRNA com o aminoácido não-natural;

(iv) um polinucleotídeo que codifica o polipeptídeo de interesse, sendo que o polinucleotídeo compreende pelo menos um códon seletor que é reconhecido pelo O-tRNA, e onde a posição do códon seletor se correlaciona com a posição especificada do aminoácido não-natural no polipeptídeo de interesse; e

(v) uma célula hospedeira que compreende (i), (ii), (iii) e (iv).

Estes métodos requerem que as construções de ácidos nucleicos de (i) e (ii) compreendam coletivamente pelo menos uma das três seguintes características:

(I) seqüências de nucleotídeos promotoras e terminadoras derivadas de um gene do tRNA de *Escherichia coli*, onde as seqüências promotoras e terminadoras estão, ambas, ligadas operacionalmente à seqüência de nucleotídeos que codifica o O-tRNA, e onde a seqüência de nucleotídeos que codifica o O-tRNA é heteróloga às seqüências promotoras e terminadoras;

(II) uma seqüência de nucleotídeos promotora correspondente a um promotor de *E. coli glnS* modificado, tendo uma seqüência de nucleotídeos de SEQ ID NO: 13, onde a seqüência de nucleotídeos de *E. coli glnS* modificada está ligada operacionalmente à seqüência de nucleotídeos que codifica a O-RS; e

(III) um operon policistrônico que compreende



uma pluralidade de seqüências de nucleotídeos do gene de O-tRNA, onde pelo menos um gene de O-tRNA é separado do pelo menos um gene de O-tRNA adjacente por um ligante de polinucleotídeos heterólogos derivado de um ligante de polinucleotídeos de ocorrência natural de um operon de tRNA de ocorrência natural.

O método é realizado (b) desenvolvendo a célula hospedeira; e (c) incorporando o aminoácido não-natural na posição especificada no polipeptídeo durante a tradução do polipeptídeo na célula hospedeira, produzindo desta forma o polipeptídeo de interesse que compreende o aminoácido não-natural na posição especificada.

Opcionalmente, a seqüência de nucleotídeos que codifica um O-tRNA é derivada de um ou mais tRNA de *Archaea*. O O-tRNA pode ser colocado em um operon policistrônico que compreende uma pluralidade de seqüências de nucleotídeos que codificam uma ou mais espécies de O-tRNA. Em alguns aspectos, a seqüência de nucleotídeos que codifica um O-tRNA compreende uma seqüência de nucleotídeos de SEQ ID NO: 1 (*MjtRNA-Tyr(CUA)*). Um operon policistrônico pode opcionalmente utilizar uma pluralidade de seqüência de nucleotídeos de SEQ ID NO: 1 (*MjtRNA-Tyr(CUA)*).

Em outros aspectos, a O-RS é derivada de uma amino-acil-tRNA sintetase de *Methanococcus jannaschii*, tal como uma tirosil-tRNA sintetase de *Methanococcus jannaschii*. Opcionalmente, a O-RS tem uma substituição de ácido aspártico para arginina na posição do aminoácido 286 ou em uma posição análoga à posição 286, em relação à seqüência de

aminoácidos de tirosil-tRNA sintetase de *Methanococcus jannaschii* do tipo selvagem. Fornecida em SEQ ID NO: 2 (Mj-tRNA-Tyr RS do tipo selvagem).

Em outras modalidades destes métodos, a construção  
 5 de ácido nucléico de (I) compreende um tRNA de prolina de *E. coli* selecionado entre o tRNA de *E. coli proK*, *proL* e *proM*. Além disso, a construção de (I) pode usar as seqüências promotora e terminadora de *E. coli proK* fornecida em SEQ ID NOS: 32 (promotora) e 33 (terminadora), respectivamente.

10 Em outras modalidades do método, a construção de ácido nucléico de (III) usa uma pluralidade de ligantes de polinucleotídeos heterólogos idênticos. Opcionalmente, pelo menos dois dos ligantes de polinucleotídeos heterólogos são diferentes.

15 Opcionalmente, o operon policistrônico de (III) nestes métodos usa um ligante de polinucleotídeos heterólogos que compreende um nucleotídeo de timidina no terminal 5', um nucleotídeo de adenosina no terminal 3', ou ambos, um nucleotídeo de timidina no terminal 5' e um  
 20 nucleotídeo de adenosina no terminal 3'. O ligante de polinucleotídeos heterólogos aqui utilizado pode ser derivado de um ligante de polinucleotídeos de ocorrência natural localizado entre os genes do tRNA endógenos de *Escherichia coli* selecionados entre: *valU* e *valX*; *ileT* e *alaT*; *serV* e *argV*; *valV* e *valW*; *glyT* e *thrT*; *metT* e *leuW*; *glnW* e *metU*; *hisR* e *leuT*; *glnU* e *glnW*; *leuP* e *leuV*; *glnV* e *glnX*; *alaW* e *alaX*; *ileU* e *alaU*; *ileV* e *alaV*; *metU* e *glnV*; *glyW* e *cysT*; *argX* e *hisR*; e *argY* e *argZ*. Por exemplo, o  
 25

ligante de polinucleotídeos heterólogos pode ser derivado da sequência de nucleotídeos de SEQ ID NO: 14 (ligante *valU/valX*) ou 15 (ligante *ileT/alaT*). Opcionalmente, estes métodos são realizados em uma célula hospedeira eubacteriana, tal como *Escherichia coli*.

Em outros aspectos, a invenção fornece também sistemas de tradução para a expressão de um polipeptídeo de interesse que tem pelo menos um aminoácido não-natural em uma posição especificada. Essencialmente, estes sistemas incluem:

- (a) um aminoácido não-natural;
- (b) uma construção de ácido nucléico que compreende uma sequência de nucleotídeos que codifica um tRNA ortogonal (O-tRNA), e uma sequência de nucleotídeos que codifica uma amino-acyl-tRNA sintetase ortogonal (O-RS), onde a O-RS preferencialmente amino-acila o O-tRNA com o aminoácido não-natural; e
- (c) um polinucleotídeo que codifica o polipeptídeo de interesse, sendo que o polinucleotídeo compreende pelo menos um códon seletor que é reconhecido pelo O-tRNA, onde a posição do códon seletor no polinucleotídeo controla a posição especificada do aminoácido não-natural no polipeptídeo de interesse depois da expressão do polinucleotídeo para produzir o polipeptídeo.

Opcionalmente, estes componentes do sistema estão integrados em uma célula hospedeira.

Em ainda outras modalidades, a invenção fornece

métodos para produzir um polipeptídeo de interesse que tem um aminoácido não-natural em uma posição especificada. Essencialmente, estes métodos usam as etapas de:

(i) um aminoácido não-natural;

5           (ii) uma construção de ácido nucléico que compreende uma seqüência de nucleotídeos que codifica um tRNA ortogonal (O-tRNA), e uma seqüência de nucleotídeos que codifica uma amino-acil-tRNA sintetase ortogonal (O-RS), onde a O-RS preferencialmente amino-acila o O-tRNA com o  
10 aminoácido não-natural; e

          (iii) um polinucleotídeo que codifica o polipeptídeo de interesse, sendo que o polinucleotídeo compreende pelo menos um códon seletor que é reconhecido pelo O-tRNA, e onde a posição do códon seletor se  
15 correlaciona com a posição especificada do aminoácido não-natural no polipeptídeo de interesse; e

          (iv) uma célula hospedeira que compreende (i), (ii), e (iii).

Estes métodos requerem ainda (b) desenvolver a  
20 célula hospedeira; e (c) incorporar o aminoácido não-natural na posição especificada no polipeptídeo durante a tradução do polipeptídeo na célula hospedeira, produzindo desta forma o polipeptídeo de interesse que compreende o aminoácido não-natural na posição especificada.

25           Definições

Antes de descrever a invenção detalhadamente, deve-se entender que esta invenção não está limitada aos sistemas biológicos específicos, os quais podem,

evidentemente, variar. Deve-se entender também que a terminologia aqui utilizada é com o propósito de descrever apenas modalidades específicas, e não pretende ser limitativa. Como utilizadas neste relatório descritivo e nas reivindicações apensadas, as formas singulares "o", "a", "um" e "uma" incluem os plurais referentes a menos que o teor dite claramente o contrário. Assim sendo, por exemplo, a referência a "uma célula" inclui combinações de duas ou mais células; a referência a "um polinucleotídeo" inclui, como uma questão prática, muitas cópias deste polinucleotídeo.

A menos que aqui e abaixo definido e no restante do relatório descritivo, todos os termos técnicos e científicos aqui utilizados têm o mesmo significado que aquele comumente entendido pelos versados nas técnicas às quais a invenção pertence. Ao descrever e reivindicar a presente invenção, a terminologia que se segue será utilizada de acordo com as definições enunciadas abaixo.

Ortogonal: Como aqui utilizado, o termo "ortogonal" refere-se a uma molécula (por exemplo, um tRNA ortogonal (O-tRNA) e/ou uma amino-acil-tRNA sintetase ortogonal (O-RS)) que funciona com componentes endógenos de uma célula com eficiência reduzida em comparação com uma molécula correspondente que é endógena para a célula ou sistema de tradução, ou que deixa de funcionar com componentes endógenos da célula. No contexto de tRNA's e amino-acil-tRNA sintetases, ortogonal refere-se a uma incapacidade ou eficiência reduzida, por exemplo, menos do

que 20% de eficiência, menos do que 10% de eficiência, menos do que 5% de eficiência, ou menos do que 1% de eficiência, de um tRNA ortogonal para funcionar com uma tRNA sintetase endógena em comparação com um tRNA endógeno para funcionar com a tRNA sintetase endógena, ou de uma amino-acil-tRNA sintetase ortogonal para funcionar com um tRNA endógeno em comparação com uma tRNA sintetase endógena para funcionar com o tRNA endógeno. A molécula ortogonal carece de uma molécula complementar endógena funcionalmente normal na célula. Por exemplo, um tRNA ortogonal em uma célula é amino-acetilado por qualquer RS endógena da célula com eficiência reduzida ou mesmo zero, em comparação com uma amino-acilação de um tRNA endógeno pela RS endógena. Em outro exemplo, uma RS ortogonal amino-acila qualquer tRNA endógeno de uma célula de interesse com eficiência reduzida ou mesmo zero, em comparação com a amino-acilação do tRNA endógeno por uma RS endógena. Uma segunda molécula ortogonal pode ser introduzida dentro da célula, que funciona com a primeira molécula ortogonal. Por exemplo, um par tRNA/RS ortogonal inclui componentes complementares introduzidos que funcionam juntos na célula com uma eficiência (por exemplo, uma eficiência de 45%, uma eficiência de 50%, uma eficiência de 60%, uma eficiência de 70%, uma eficiência de 75%, uma eficiência de 80%, uma eficiência de 90%, uma eficiência de 95%, ou uma eficiência de 99% ou mais) em comparação com aquela de um controle, por exemplo, um par tRNA/RS endógeno correspondente, ou um par ortogonal ativo (por exemplo, um par tRNA/RS tirosil ortogonal).

Tirosil-tRNA ortogonal: Como aqui utilizado, um tirosil-tRNA ortogonal (tirosil-O-tRNA) é um tRNA que é ortogonal a um sistema de tradução de interesse, onde o tRNA é: (1) idêntico ou substancialmente similar a um tirosil-tRNA de ocorrência natural, (2) derivado de um tirosil-tRNA de ocorrência natural por mutagênese natural ou artificial, (3) derivado por qualquer processo que leva em consideração uma seqüência de uma seqüência de tirosil-tRNA do tipo selvagem ou mutante de (1) ou (2), (4) homólogo a um tirosil-tRNA do tipo selvagem ou mutante, (5) homólogo a qualquer tRNA exemplificativo que é designado como um substrato para uma tirosil-tRNA sintetase, por exemplo, sintetases de SEQ ID NOS: 2, 4, 6, 8 ou 10, (6) uma variante conservativa de qualquer tRNA exemplificativo que é designado como um substrato para uma tirosil-tRNA sintetase, por exemplo, o O-tRNA de SEQ ID NO: 1. O tirosil-tRNA pode existir carregado com um aminoácido, ou em um estado não-carregado. Deve-se entender também que um "tirosil-O-tRNA" é opcionalmente carregado (amino-acilado) por uma sintetase cognata com um aminoácido diferente de tirosina ou leucina, respectivamente, por exemplo, com um aminoácido não-natural. Na realidade, deve-se avaliar que um tirosil-tRNA da invenção é usado vantajosamente para inserir essencialmente qualquer aminoácido, seja ele natural ou artificial, em um polipeptídeo em crescimento, durante a tradução, em resposta a um códon seletor.

Tirosil-aminoácido sintetase ortogonal: Como aqui utilizado, uma tirosil-aminoácido sintetase ortogonal

(tirosil-O-RS) é uma enzima que de preferência amino-acila o tirosil-O-tRNA com um aminoácido em um sistema de tradução de interesse. O aminoácido que a tirosil-O-RS carrega no tirosil-O-tRNA pode ser qualquer aminoácido, seja ele  
5 natural, não-natural ou artificial, e não é aqui limitado. A sintetase é opcionalmente a mesma ou homóloga a uma tirosil-aminoácido sintetase de ocorrência natural, ou igual ou homóloga a uma sintetase designada como uma O-RS em, por exemplo, SEQ ID NOS: 4, 6, 8 ou 10. Por exemplo, a O-RS pode  
10 ser uma variante conservativa de uma tirosil-O-RS de, por exemplo, SE ID NOS: 4, 6, 8 ou 10 e/ou pode ser pelo menos 50%, 60%, 70%, 80%, 90%, 95%, 98%, 99%, ou mais, idêntica em seqüência a uma O-RS de, por exemplo, SEQ ID NOS: 4, 6, 8 ou 10.

15           Cognato: O termo "cognato" refere-se a componentes que funcionam em conjunto, por exemplo, um tRNA ortogonal e uma amino-acil-tRNA sintetase ortogonal. Os componentes podem ser referidos também como sendo complementares.

          Preferencialmente amino-acila: Como aqui utilizado  
20 com relação a sistemas de tradução ortogonal, uma O-RS "preferencialmente amino-acila" um O-tRNA cognato quando a O-RS carrega o O-tRNA com um aminoácido mais eficientemente do que ela carrega qualquer tRNA endógeno em um sistema de expressão. Isto é, quando o O-tRNA e qualquer dado tRNA  
25 endógeno estão presentes em um sistema de tradução em razões molares aproximadamente iguais, a O-RS carregará o O-tRNA mais freqüentemente do que ele carregará o tRNA endógeno. De preferência, a razão relativa do O-tRNA carregado pela O-RS



para o tRNA endógeno carregado pela O-RS é alta, resultando de preferência na O-RS carregar o O-tRNA exclusivamente, ou quase que exclusivamente, quando o O-tRNA e o tRNA endógeno estão presentes em concentrações molares iguais no sistema de tradução. A razão relativa entre O-tRNA e tRNA endógeno que está carregado pela O-RS, quando o O-tRNA e a O-RS estão presentes em concentrações molares iguais, é maior do que 1:1, de preferência pelo menos cerca de 2:1, mais preferivelmente 5:1, ainda mais preferivelmente 10:1, mais preferivelmente ainda 20:1, ainda mais preferivelmente 50:1, mais preferivelmente ainda 75:1, muito mais preferivelmente 95:1, 98:1, 99:1, 100:1, 500:1, 1.000:1, 5.000:1 ou mais.

A O-RS "preferencialmente amino-acila um O-tRNA com um aminoácido não-natural" quando (a) a O-RS preferencialmente amino-acila o O-tRNA em comparação com um tRNA endógeno, e (b) quando esta amino-acilação é específica para o aminoácido não-natural em comparação com a amino-acilação do O-tRNA pela O-RS com qualquer aminoácido natural. Isto é, quando os aminoácidos não-naturais e naturais estão presentes em quantidades molares iguais em um sistema de tradução que compreende a O-RS e O-tRNA, a O-RS carregará o O-tRNA com o aminoácido não-natural mais freqüentemente do que com o aminoácido natural. De preferência, a razão relativa do O-tRNA carregado com o aminoácido não-natural para o O-tRNA carregado com o aminoácido natural é alta. Mais preferivelmente, a O-RS carrega o O-tRNA exclusivamente, ou quase que exclusivamente, com o aminoácido não-natural. A razão

relativa entre o carregamento do O-tRNA com o aminoácido não-natural e o carregamento do O-tRNA com o aminoácido natural, quando os aminoácidos não-natural de natural estão presentes no sistema de tradução em concentrações molares iguais, é maior do que 1:1, de preferência pelo menos cerca de 2:1, mais preferivelmente 5:1, ainda mais preferivelmente 10:1, mais preferivelmente ainda 20:1, ainda mais preferivelmente 50:1, mais preferivelmente ainda 75:1, muito mais preferivelmente 95:1, 98:1, 99:1, 100:1, 500:1, 1.000:1, 5.000:1 ou mais.

Códon seletor: O termo "códon seletor" refere-se aos códons reconhecidos pelo O-tRNA no processo de tradução e não são reconhecidos por um tRNA endógeno. A alça anticódon do O-tRNA reconhece o códon seletor no RNAm e incorpora seu aminoácido, por exemplo, um aminoácido não-natural, neste sítio no polipeptídeo. Os códons seletores podem incluir, por exemplo, códons *nonsense*, tais como códons finalizadores, por exemplo, códons âmbar, ocre e opala; códons de quatro ou mais bases; códons raros; códons derivados de pares de bases naturais ou não-naturais, e/ou similares

tRNA supressor: Um tRNA supressor é um tRNA que altera a leitura de um RNA mensageiro (RNAm) em um dado sistema de tradução, por exemplo, disponibilizando um mecanismo para incorporar um aminoácido em uma cadeia de polipeptídeo em resposta a um códon seletor. Por exemplo, um tRNA supressor pode ler através de, por exemplo, um códon finalizador (por exemplo, um códon âmbar, ocre ou opala), um

códon de quatro bases, um códon raro, etc.

Atividade de supressão: Como aqui utilizado, o termo "atividade de supressão" refere-se genericamente à capacidade de um tRNA (por exemplo, um tRNA supressor) 5 permitir a leitura da tradução através de um códon (por exemplo, um códon seletor que é um códon âmbar ou um códon de 4 ou mais bases) que resultaria de outra forma na terminação da tradução ou em uma tradução mal feita (por exemplo, deslocamento de estrutura). A atividade de 10 supressão de um tRNA supressor pode ser expressada como uma porcentagem da atividade de leitura da tradução observada em comparação com um segundo tRNA supressor, ou comparada com um sistema de controle, por exemplo, um sistema de controle sem uma O-RS.

15 A presente invenção fornece vários métodos pelos quais a atividade de supressão pode ser quantificada. A supressão percentual de um tRNA e O-RS específico contra um códon seletor (por exemplo, um códon âmbar) de interesse refere-se à porcentagem de atividade de um dado marcador do 20 teste expressado (por exemplo, LacZ), que inclui um códon seletor, em um ácido nucléico que codifica o marcador do teste expressado, em um sistema de tradução de interesse, onde o sistema de tradução de interesse inclui uma O-RS e um O-tRNA, em comparação com uma construção de controle 25 positivo, onde o controle positivo carece do O-tRNA, a O-RD e o códon seletor. Assim sendo, por exemplo, caso uma construção de marcador de controle positivo ativo que carece de um códon seletor tenha uma atividade observada de X em um

5      dado sistema de tradução, em unidades relevantes para o  
ensaio do marcador em questão, então a supressão percentual  
de uma construção em teste que compreende o códon seletor é  
a porcentagem de X que a construção do marcador do teste  
10      apresenta sob condições ambientais essencialmente iguais sob  
as quais o marcador de controle positivo foi expressado,  
exceto que a construção do marcador do teste é expressada em  
um sistema de tradução que inclui também o O-tRNA e a O-RS.  
Tipicamente, os sistema de tradução que expressa o marcador  
15      do teste inclui também um aminoácido que é reconhecido pelo  
O-tRNA e/ou pela O-RS. Opcionalmente, a medição da supressão  
percentual pode ser refinada por comparação do marcador do  
teste com uma construção de marcador de controle "de fundo"  
ou "negativo", que inclui o mesmo códon seletor que o  
20      marcador do teste, mas em um sistema que não inclui o O-  
tRNA, a O-RS e/ou o aminoácido relevante reconhecido pelo O-  
tRNA e/ou O-RS. Este controle negativo é útil em medições de  
supressão percentual para normalização, a fim de explicar os  
efeitos de sinais de fundo a partir do marcador no sistema  
de tradução de interesse.

25      A eficiência da supressão pode ser determinada por  
qualquer um entre inúmeros ensaios conhecidos nessas  
técnicas. Por exemplo, pode ser usado um ensaio de  $\beta$ -  
galactosidase repórter, por exemplo, um peptídeo *lacZ*  
derivado (onde a construção tem um códon seletor na  
seqüência de ácidos nucleicos de *lacZ*) é introduzido dentro  
das células a partir de um organismo apropriado (por  
exemplo, um organismo no qual os componentes ortogonais

podem ser usados), junto com o plasmídeo que compreende um O-tRNA da invenção. Uma sintetase cognata também introduzida (seja como um polipeptídeo ou como um polinucleotídeo que codifica a sintetase cognata quando expressado). As células  
5 são desenvolvidas no meio até uma densidade desejada, por exemplo, até uma  $DO_{600}$  de cerca de 0,5, e os ensaios de  $\beta$ -galactosidase são realizados, por exemplo, usando o Kit de Ensaio de  $\beta$ -Galactosidase BetaFluor™ (Novagen). A supressão percentual pode ser calculada como a porcentagem de  
10 atividade para uma amostra em relação a um controle comparável, por exemplo, o valor observado a partir da construção de *LacZ* derivada, onde a construção tem um códon *sense* correspondente na posição desejada em vez de um códon seletor.

15            Sistema de tradução: O termo "sistema de tradução" refere-se aos componentes que incorporam um aminoácido dentro de uma cadeia de polipeptídeo (proteína) em crescimento. Os componentes de um sistema de tradução podem incluir, por exemplo, ribossomas, tRNA's, sintetases, RNAm,  
20 e similares. O O-tRNA e/ou as O-RS's da invenção podem ser adicionados ao sistema de tradução *in vitro* ou *in vivo*, ou podem ser parte de um sistema de tradução *in vitro* ou *in vivo*, por exemplo, em uma célula não-eucariótica, por exemplo, uma eubactéria (tal como *E. coli*), ou em uma célula  
25 eucariótica, por exemplo, uma célula de levedura, uma célula de mamífero, uma célula vegetal, uma célula de alga, uma célula de fungo, uma célula de inseto, e/ou similares.

Aminoácido não-natural: Como aqui utilizado, o

termo "aminoácido não-natural" refere-se a qualquer aminoácido, aminoácido, aminoácido modificado e/ou análogo de aminoácido, que não é um dos 20 aminoácidos de ocorrência natural comuns, ou selenocisteína ou pirrolisina. Por exemplo, os aminoácidos não-naturais p-benzoil-L-fenilalanina (Bpa), para-acetil-L-fenilalanina (pAcPhe), para-azido-L-fenilalanina (pAzPhe) e para-iodo-L-fenilalanina (pIPhe) encontram uso na invenção.

Em resposta a: Como aqui utilizado, o termo "em resposta a" refere-se ao processo no qual um O-tRNA da invenção reconhece um códon seletor e medeia a incorporação do aminoácido não-natural, que é acoplado ao RNA, dentro da cadeia de polipeptídeo em crescimento.

Polipeptídeo: Um polipeptídeo é qualquer oligômero de aminoácidos (naturais ou não-naturais, ou uma combinação deles), de qualquer comprimento, tipicamente, porém não exclusivamente, unido por ligações peptídicas covalentes. Um polipeptídeo pode ser de qualquer origem, por exemplo, polipeptídeo de ocorrência natural, um polipeptídeo produzido por técnicas recombinantes de genética molecular, um polipeptídeo de uma célula ou sistema de tradução, ou um polipeptídeo produzido por meios sintéticos isentos de células. Um polipeptídeo caracteriza-se por sua seqüência de aminoácidos, por exemplo, a estrutura principal dos seus aminoácidos componentes. Como aqui utilizado, o termo "seqüência de aminoácidos" de um polipeptídeo não se limita a seqüências de comprimento inteiro, mas podem ser seqüências parciais ou completas. Além disso, não se

pretende que um polipeptídeo esteja limitado por possuir ou não possuir qualquer atividade biológica específica. Como aqui utilizado, o termo "proteína" é sinônimo de polipeptídeo. O termo "peptídeo" refere-se a um polipeptídeo  
5 pequeno, por exemplo, porém sem limitações, um comprimento entre 2 e 25 aminoácidos.

Variante conservativa: Como aqui utilizado, o termo "variante conservativa", no contexto de um componente de tradução, refere-se a um componente de tradução, por  
10 exemplo, um O-tRNA variante conservativo ou uma O-RS variante conservativa, que funcionalmente se comporta de forma similar a um componente da base, de tal modo que a variante conservativa seja similar, por exemplo, a um O-tRNA ou O-RS, tendo variações na sequência em comparação com um  
15 O-tRNA ou O-RS referencial. Por exemplo, uma O-RS, ou uma variante conservativa desta O-RS, amino-acilará um O-tRNA cognato com um aminoácido não-natural, por exemplo, um aminoácido que compreende um grupamento N-acetilgalatosamina. Neste exemplo, a O-RS e a variante  
20 conservativa desta O-RS não têm as mesmas seqüências de aminoácidos. A variante conservativa pode ter, por exemplo, uma variação, duas variações, três variações, quatro variações, ou cinco ou mais variações na sequência, desde que a variante conservativa ainda seja complementar ao O-  
25 tRNA ou O-RS correspondente.

Em algumas modalidades, uma O-RS variante conservativa compreende uma ou mais substituições conservativas de aminoácidos em comparação com a O-RS a

partir da qual ela foi derivada. Em algumas modalidades, uma O-RS variante conservativa compreende uma ou mais substituições conservativas de aminoácidos em comparação com a O-RS a partir da qual ela foi derivada, e além disso, retém a atividade biológica da O-RS; por exemplo, uma O-RS variante conservativa que retém pelo menos 10% da atividade da molécula O-RS originária a partir da qual ela foi derivada, ou alternativamente, pelo menos 20%, pelo menos 30%, ou pelo menos 40%. Em algumas modalidades preferidas, a O-RS variante conservativa retém pelo menos 50% da atividade biológica da molécula de O-RS originária a partir da qual ela foi derivada. As substituições conservativas de aminoácidos de uma O-RS variante conservativa podem ocorrer em qualquer domínio da O-RS, incluindo a bolsa de ligação de aminoácido.

Polinucleotídeo ou ácido nucléico: Os termos "ácido nucléico", "seqüência de ácidos nucléicos", "seqüência de nucleotídeos", "oligonucleotídeo", "polinucleotídeo" ou "molécula de ácido nucléico" ou termos similares referem-se a oligômeros de bases ligados tipicamente por um esqueleto de açúcar/fosfato, tais como oligonucleotídeos ou polinucleotídeos, e fragmentos ou partes deles, e ao DNA ou RNA de origem genômica ou sintética, que pode ser de filamento único ou filamento duplo, e representam um filamento *sense* ou *antisense*. Os termos "ácido nucléico", "polinucleotídeo", e "nucleotídeo" incluem também especificamente ácidos nucléicos constituídos de bases diferentes das cinco bases de ocorrência biológica



(isto é, adenina, guanina, timina, citosina e uracila), e incluem também ácidos nucleicos que têm estruturas de esqueletos não-naturais, tais como moléculas de PNA.

As moléculas de ácidos nucleicos (por exemplo, DNA  
5 ou RNA) são ditas como tendo "extremidades 5'" e  
"extremidades 3'" porque os mononucleotídeos não reagidos  
para produzir oligonucleotídeos ou polinucleotídeos de tal  
modo que o 5'-fosfato de um anel pentose do mononucleotídeo  
fique anexado ao oxigênio de 3' do seu vizinho em uma  
10 direção por intermédio de uma ligação fosfodiéster. Assim  
sendo, um polinucleotídeo terá tipicamente uma "extremidade  
5'" que compreende um 5'-fosfato e uma "extremidade 3'" que  
compreende o oxigênio de 3'. Uma seqüência de  
polinucleotídeo, mesmo se interna a um ácido nucléico maior,  
15 também pode ser dita como tendo direcionalidade 5' e 3'. Em  
uma molécula de DNA linear ou circular, os distintos  
elementos são referidos como estando a "montante" ou 5' da  
"jusante" ou elementos 3'. Esta terminologia reflete o fato  
de que a transcrição prossegue de uma maneira 5' até 3' ao  
20 longo do filamento de DNA.

Gene: Como aqui utilizado, o termo "gene" refere-se  
mais genericamente a uma combinação de elementos  
polinucleotídeos, os quais, quando ligados operacionalmente  
de uma maneira nativa ou recombinante, produzem algum  
25 produto ou função. O termo "gene" deve ser interpretado  
amplamente neste caso, englobando as formas RNAm, DNAm, RNAm  
e DNA genômico de um gene. Em alguns casos, o gene é  
herdável. Em alguns aspectos, os genes compreendem

seqüências codificadoras (por exemplo, "estrutura de leitura aberta" ou "região codificadora") necessárias para a produção de um polipeptídeo, enquanto que em outros aspectos, os genes não codificam um polipeptídeo. Os  
5 exemplos de genes que não codificam polipeptídeos incluem genes de RNA ribossômico (rRNA) e genes de RNA de transferência (tRNA).

O termo "gene" pode englobar opcionalmente seqüências reguladoras não-codificadoras que residem em um  
10 locus genético. Por exemplo, além de uma região codificadora de um ácido nucléico, o termo "gene" engloba também as seqüências de nucleotídeos transcritas do rRNA de comprimento inteiro adjacente às extremidades 5' e 3' da região codificadora. Estas regiões não-codificadoras têm um  
15 tamanho variável, e tipicamente se estendem em ambas extremidades, 5' e 3' da região codificadora. As seqüências que estão localizadas a 5' e 3' da região codificadora e ficam contidas no rRNA são referidas como seqüências não-traduzidas 5' e 3' (5' UT e 3' UT). 5' e 3' UT podem ter papéis  
20 reguladores, inclusive iniciação da tradução, clivagem pós-translacional e poliadenilação. O termo "gene" engloba as formas rRNA, DNAc e genômica de um gene.

Em alguns aspectos, a forma genômica ou clone genômico de um gene inclui as seqüências do rRNA transcrito,  
25 bem como outras seqüências não-transcritas que ficam fora do transcrito. As regiões reguladoras que ficam fora da unidade de transcrição do rRNA são denominadas algumas vezes "seqüências flanqueadoras 5' ou 3'". Uma forma genômica

funcional de um gene contém tipicamente elementos reguladores necessários para a regulação da transcrição. Por exemplo, o termo "promotor" é utilizado usualmente para descrever uma região do DNA, tipicamente, mas não  
5 exclusivamente, a 5' do sítio da iniciação da transcrição, suficiente para conferir uma iniciação precisa da transcrição. Em algumas modalidades, um promotor é constitutivamente ativo, enquanto que em modalidades alternativas, o promotor é condicionalmente ativo (por  
10 exemplo, quando a transcrição é iniciada apenas sob certas condições fisiológicas). Em procariotos, a atividade de um promotor pode ser modulada por uma sequência "operadora" adjacente. Em algumas modalidades, a região flanqueadora 3' contém sequências adicionais que regulam a terminação da  
15 transcrição, algumas vezes denominadas sequências "terminadoras". Genericamente, o termo "elemento regulador" refere-se a qualquer elemento genético que controla algum aspecto da expressão de sequências de ácidos nucleicos.

Sequência de nucleotídeos expressável: Como aqui  
20 utilizado, o termo "sequência de nucleotídeos expressável" refere-se a qualquer sequência de nucleotídeos que é capaz de ser transcrita (por exemplo, transcrita por uma RNA polimerase dependente de DNA) para gerar um transcrito. A sequência do transcrito não está limitada, e pode ser  
25 codificadora de proteínas (por exemplo, pode codificar uma amino-acil-tRNA sintetase) ou pode ser não-codificadora de proteínas (por exemplo, codificar uma molécula de tRNA).

Ligado operacionalmente: Como aqui utilizados, os

termos "em combinação operável", "em ordem operável", "ligado operacionalmente", "unido operacionalmente", e termos similares, quando utilizados com relação a ácidos nucleicos, referem-se à ligação de seqüências de ácidos nucleicos colocadas em relações funcionais entre si. Por exemplo, uma seqüência promotora, estrutura de leitura aberta e seqüência terminadora ligadas operacionalmente resultam na produção precisa de uma molécula de RNA. Em alguns aspectos, elementos de ácidos nucleicos ligados operacionalmente resultam na transcrição de uma estrutura de leitura aberta, e ulteriormente, na produção de um polipeptídeo (isto é, a expressão da estrutura de leitura aberta).

Operon: Como aqui utilizado, o termo "operon" refere-se a uma unidade genética (por exemplo, uma região cromossômica) que controla a expressão gênica em procariotos. Um operon compreende tipicamente um ou mais genes que codificam um ou mais polipeptídeos ou RNA's, e a região (ou regiões) reguladora adjacente que controla a transcrição dos genes. A região reguladora compreende tipicamente um promotor e um operador. A região codificadora de um gene procariótico é denominada historicamente um "cístron". Os operons que contêm múltiplos cístrons são denominados "policistrônicos". Os genes em um operon policistrônico são tipicamente relacionados em função e são tipicamente co-transcritos como uma única unidade e expressados de uma maneira coordenada.

Construção: Como aqui utilizado, o termo

"construção" é utilizado com relação a qualquer polinucleotídeo ou outra molécula que pode transferir um segmento ou segmentos de ácidos nucleicos para dentro de uma célula. O termo "vetor" ou "veículo" é usado algumas vezes  
5 de forma intercambiável com "vetor". Um vetor compreende opcionalmente partes que medeiam a propagação e manipulação de vetores (por exemplo, seqüências necessárias para replicação, genes que conferem resistência a fármacos ou antibióticos, um sítio de múltipla clonagem, elementos  
10 promotores/intensificadores ligados operacionalmente que permitem a expressão de um gene clonado, etc.). Os vetores são freqüentemente derivados a partir de plasmídeos, bacteriófagos, ou vírus vegetais ou animais. Um "vetor de clonagem" ou "vetor-ponte (*shuttle*)" ou "vetor de  
15 subclonagem" contém partes ligadas operacionalmente que facilitam etapas de subclonagem (por exemplo, um sítio de múltipla clonagem que contém múltiplos sítios de endonucleases de restrição).

Vetor de expressão: O termo "vetor de expressão",  
20 como aqui utilizado, refere-se a um vetor recombinante que compreende seqüências de polinucleotídeos ligadas operacionalmente, que facilitam a expressão de uma seqüência codificadora em um organismo hospedeiro específico (por exemplo, vetor de expressão bacteriana). As seqüências de  
25 polinucleotídeos que facilitam a expressão em procariotos incluem tipicamente, por exemplo, um promotor, seqüências de terminação da transcrição, isto é, seqüências terminadoras, um operador (opcional), e um sítio de ligação de ribossomas,

freqüentemente junto com outras seqüências.

Codificação: Como aqui utilizado, o termo "codificação" refere-se a qualquer processo pelo qual as informações em uma macromolécula polimérica ou seqüência *string* são usadas para direcionar a produção de uma segunda molécula ou seqüência *string* que é diferente da primeira molécula ou seqüência *string*. Como aqui utilizado, o termo é usado genericamente e pode ter uma série de aplicações. Em alguns aspectos, o termo "codificação" descreve o processo de replicação de DNA semiconservativa, onde um filamento de uma molécula de DNA de filamento duplo é usado como um modelo para codificar um filamento *sister* por uma DNA polimerase dependente de DNA.

Em outro aspecto, o termo "codificação" refere-se a qualquer processo pelo qual as informações em uma molécula são usadas para direcionar a produção de uma segunda molécula que tem uma natureza química diferente da primeira molécula. Por exemplo, uma molécula de DNA pode codificar uma molécula de RNA (por exemplo, pelo processo de transcrição que incorpora uma enzima RNA polimerase dependente de DNA). Além disso, uma molécula de RNA pode codificar um polipeptídeo, como no processo de tradução. Quando utilizado para descrever o processo de tradução, o termo "codificar" se estende também para o códon triplete que codifica um aminoácido. Em alguns aspectos, uma molécula de RNA pode codificar uma molécula de DNA, por exemplo, pelo processo de transcrição reversa que incorpora uma DNA polimerase dependente de RNA. Em outro aspecto, uma molécula

de DNA pode codificar um polipeptídeo, onde deve-se entender que "codificar", como utilizado neste caso, incorpora os processos de transcrição e tradução.

Heterólogo: Como aqui utilizados, os termos "heterólogo" ou "exógeno", como aplicados a polinucleotídeos ou polipeptídeos, referem-se a moléculas que foram rearranjadas ou supridas artificialmente para um sistema biológico, e não estão em uma configuração nativa (por exemplo, com relação à seqüência, posição genômica ou arranjo de partes) ou não são nativas para esse sistema biológico específico. Os termos indicam que o material relevante se originou de uma fonte que não a fonte de ocorrência natural, ou referem-se a moléculas que têm uma configuração, localização genética ou arranjo de partes não-natural. Os termos "exógeno" e "heterólogo" são utilizados algumas vezes de forma intercambiável com "recombinante".

Recombinante: O termo "recombinante" com relação a um ácido nucléico ou polipeptídeo indica que o material (por exemplo, um ácido nucléico, gene, polinucleotídeo, polipeptídeo, etc., recombinante) foi alterado por intervenção humana. Genericamente, o arranjo de partes de uma molécula recombinante não é uma configuração nativa, ou a seqüência primária do polinucleotídeo ou polipeptídeo recombinante foi manipulada de alguma forma. A alteração para produzir o material recombinante pode ser realizada no material no seu ambiente ou estado natural, ou fora dele. Por exemplo, um ácido nucléico de ocorrência natural torna-se um ácido nucléico recombinante caso ele seja alterado, ou

caso ele seja transcrito do DNA que foi alterado, por meio de intervenção humana realizada dentro da célula a partir da qual ele se origina. Uma estrutura de leitura aberta de sequência gênica é recombinante caso essa sequência de nucleotídeos tenha sido removida do seu contexto natural e clonada para qualquer tipo de vetor de ácido nucléico artificial. O termo "recombinante" pode se referir também a um organismo que hospeda material recombinante. Os protocolos e reagentes para produzir moléculas recombinantes, especialmente ácidos nucléicos recombinantes, são comuns e rotineiros nessas técnicas (vide, por exemplo, Maniatis *et al.* (editores), "Molecular Cloning: A Laboratory Manual", Cold Spring Harbor Laboratory Press, NY (1982); Sambrook *et al.* (editores), "Molecular Cloning: A Laboratory Manual", Segunda Edição, Volumes 1-3, Cold Spring Harbor Laboratory Press, NY (1989); e Ausubel *et al.* (editores), "Current Protocols in Molecular Biology", Volumes 1-4, John Wiley & Sons, New York (1994)).

Nativa ou endógena: Em contraste com uma molécula heteróloga ou exógena, uma molécula "nativa" ou "endógena" é nativa para o sistema biológico, espécie ou cromossoma em estudo. Um gene "nativo" ou "endógeno" é um gene que não contém elementos de ácidos nucléicos codificados por fontes que não o cromossoma no qual ele é normalmente encontrado na natureza. Um gene, transcrito ou polipeptídeo endógeno é codificado por seu locus cromossômico natural, e não fornecido artificialmente para a célula.

Célula hospedeira: O termo "célula hospedeira"



refere-se tipicamente a uma célula que contém um ácido nucléico heterólogo, tal como um vetor, e sustenta a replicação e/ou expressão do ácido nucléico. As células hospedeiras podem ser células procarióticas, tal como *E. coli*, ou células eucarióticas, tais como células de leveduras, insetos, anfíbios ou mamíferos. De preferência, as células hospedeiras são células vegetais. No contexto da invenção, uma célula hospedeira particularmente preferida é uma célula hospedeira de soja.

10                   Eucarioto: Como aqui utilizado, o termo "eucarioto" refere-se a organismos pertencentes ao Reino *Eucarya*. Os eucariotos são genericamente distinguíveis de procariotos por sua organização tipicamente multicelular (porém não exclusivamente multicelular, por exemplo, 15 levedura), pela presença de núcleo ligado à membrana e por outras organelas ligadas à membrana, material genético linear (isto é, cromossomas lineares), ausência de operons, presença de íntrons, capeamento de mensagens e poli-A RNAm, e outras características bioquímicas, tal como uma estrutura 20 ribossômica peculiar. Os organismos eucarióticos incluem, por exemplo, animais (por exemplo, mamíferos, insetos, répteis, aves, etc.), ciliados, plantas (por exemplo, monocotilédones, dicotilédones, algas, etc.), fungos, leveduras, flagelados, microsporídios, protistas, etc.

25                   Procarioto: Como aqui utilizado, o termo "procarioto" refere-se a organismos pertencentes ao Reino *Monera* (também denominado *Procarya*). Os organismos procarióticos são genericamente distinguíveis de eucariotos

por sua organização unicelular, reprodução assexuada por brotamento ou fissão, falta de um núcleo ligado à membrana ou outras organelas ligadas à membrana, cromossoma circular, presença de operons, ausência de íntrons, capeamento de  
 5 mensagens e poli-A RNAm, e outras características bioquímicas, tal como uma estrutura ribossômica peculiar. Os *Prokarya* incluem os sub-reinos *Eubateria* e *Archaea* (algumas vezes denominados "*Archaeobacteria*"). As cianobactérias (as algas azuis esverdeadas) e os micoplasmas recebem algumas  
 10 vezes classificações separadas sob o Reino *Monera*.

Bactérias: Como aqui utilizados, os termos "bactérias" e "eubactérias" referem-se a organismos procarióticos que são distinguíveis das *Archaea*. Similarmente, *Archaea* refere-se a procariotos que são  
 15 distinguíveis de eubactérias. *Eubateria* e *Archaea* podem ser distinguidas por inúmeros critérios morfológicos e bioquímicos. Por exemplo, as diferenças em seqüências do RNA ribossômico, estrutura de RNA polimerase, presença ou ausência de íntrons, sensibilidade a antibióticos, presença  
 20 ou ausência de peptidoglicanos da parede celular e outros componentes da parede celular, estruturas ramificadas versus não-ramificadas de lipídeos da membrana, e presença/ausência de histonas e proteínas semelhantes a histonas, são usadas para designar um organismo para *Eubateria* ou *Archaea*.

25 Os exemplos de eubactérias incluem *Escherichia coli*, *Thermus thermophilus* e *Bacillus stearothermophilus*. Os exemplos de *Archaea* incluem *Methanococcus jannaschii* (Mj), *Methanosarcina mazei* (Mm), *Methanobacterium*

*thermoautotrophicum* (Mt), *Methanococcus maripaludis*,  
*Methanopyrus kandleri*, *Halobacterium* tais como as espécies  
*Haloferax volcanii* e *Halobacterium* NRC-1, *Archaeoglobus*  
*fulgidus* (Af), *Pyrococcus furiosus* (Pf), *Pyrococcus*  
 5 *horikishii* (Ph), *Pyrobaculum aerophilum*, *Pyrococcus abyssi*,  
*Sulfolobus sofataricus* (Ss), *Sulfolobus tokodaii*, *Aeuiropyrum*  
*pernix* (Ap), *Thermoplasma acidophilum* e *Thermoplasma*  
*volcanium*.

Derivado de: Como aqui utilizado, o termo  
 10 "derivado de" refere-se a um componente que é isolado a  
 partir de ou preparado usando uma molécula ou organismo  
 específico, ou informações a partir da molécula ou organismo  
 especificado. Por exemplo, um polipeptídeo que é derivado de  
 um segundo polipeptídeo pode incluir uma sequência de  
 15 aminoácidos idêntica ou substancialmente similar à sequência  
 de aminoácidos do segundo polipeptídeo. No caso de  
 polipeptídeos, a espécie derivada pode ser obtida, por  
 exemplo, por mutagênese de ocorrência natural, mutagênese  
 artificial direcionada ou mutagênese artificial aleatória. A  
 20 mutagênese usada para derivar polipeptídeos pode ser  
 intencionalmente direcionada ou intencionalmente aleatória,  
 ou uma mistura de cada uma delas. A mutagênese de um  
 polipeptídeo para criar um polipeptídeo diferente derivado  
 do primeiro pode ser um episódio aleatório (por exemplo,  
 25 causado por infidelidade da polimerase), e a identificação  
 do polipeptídeo derivado pode ser feita por métodos de  
 triagem apropriados, como aqui discutido, por exemplo. A  
 mutagênese de um polipeptídeo infere tipicamente a

manipulação do polipeptídeo que codifica o polipeptídeo.

Similarmente, o termo "derivado de" pode se aplicar a polinucleotídeos. Um polipeptídeo derivado de um polinucleotídeo-fonte pode incluir uma seqüência de  
5 nucleotídeos idêntica ou substancialmente similar à seqüência de nucleotídeos-fonte. No caso de polinucleotídeos, as espécies derivadas podem ser obtidas, por exemplo, por mutagênese de ocorrência natural, mutagênese artificial direcionada ou mutagênese artificial  
10 aleatória. A mutagênese usada para derivar polinucleotídeos pode ser intencionalmente direcionada ou intencionalmente aleatória, ou uma mistura de cada uma delas. Em alguns aspectos, um polinucleotídeo derivado é gerado colocando um polinucleotídeo-fonte dentro de um contexto heterólogo, isto  
15 é, dentro de um contexto diferente do seu contexto nativo ou endógeno. Por exemplo, um promotor gênico pode ser derivado de um promotor gênico endógeno removendo esse domínio do promotor endógeno e colocando-o em combinação operável com diferentes seqüências de nucleotídeos com as quais ele não  
20 está normalmente associado.

Seleção positiva ou marcador de triagem: Como aqui utilizado, o termo "seleção positiva ou marcador de triagem" refere-se a um marcador que, quando presente, por exemplo, expressado, ativado, ou similares, resulta na identificação  
25 de uma célula que compreende o traço, por exemplo, uma célula com o marcador de seleção positiva, daquelas sem o traço.

Seleção negativa ou marcador de triagem: Como aqui

utilizado, o termo "seleção negativa ou marcador de triagem" refere-se a um marcador que, quando presente, por exemplo, expressado, ativado, ou similares, permite a identificação de uma célula que não compreende uma propriedade ou traço selecionado (por exemplo, em comparação com uma célula que possui a propriedade ou traço).

Agente de seleção ou triagem: Como aqui utilizado, o termo "agente de seleção ou triagem" refere-se a um agente que, quando presente, permite a seleção/triagem de certos componentes de uma população. Por exemplo, um agente de seleção ou triagem pode ser, porém sem limitações, por exemplo, um nutriente, um antibiótico, um comprimento de onda de luz, um anticorpo, um polinucleotídeo expressado, ou similares. O agente de seleção pode ser variado, por exemplo, por concentração, intensidade, etc.

Repórter: Como aqui utilizado, o termo "repórter", ou termos equivalentes, refere-se em um sentido amplo a qualquer componente que pode ser detectado facilmente em um sistema em estudo, onde a detecção do repórter se correlaciona com a presença ou ausência de alguma outra molécula ou propriedade, ou pode ser usado para identificar, selecionar e/ou triar alvos em um sistema de interesse. A escolha do repórter mais apropriado para usar para uma aplicação específica depende do uso pretendido, e outras variáveis conhecidas pelos versados nessas técnicas. Em alguns aspectos, um repórter é um gene repórter.

Várias moléculas e genes repórteres são conhecidas nessas técnicas. Cada repórter tem um ensaio específico para

a detecção deste repórter. Alguns ensaios de detecção de repórteres podem ser ensaios enzimáticos, enquanto que outros ensaios podem ser de natureza imunológica (por exemplo, análise ELISA ou imunoistoquímica), ou  
 5 colorimétrica, por exemplo. Além disso, um repórter pode incluir uma proteína, por exemplo, uma enzima, que confere resistência ou sensibilidade a antibióticos (por exemplo,  $\beta$ -lactamase, cloranfenicol acetiltransferase (CAT), e similares), um marcador fluorescente (por exemplo, uma  
 10 proteína fluorescente verde, tal como GFP, YFP, EGFP, RFP, etc.), um marcador luminescente (por exemplo, uma proteína luciferase do vaga-lume), um marcador de triagem baseado em afinidade, uma atividade enzimática, tal como *lacZ* ( $\beta$ -galactosidase), ou outros genes marcadores selecionáveis  
 15 positivos ou negativos, tais como ADH (álcool desidrogenase), *his3*, *ura3*, *leu2*, *lys2*, ou similares.

#### BREVE DESCRIÇÃO DAS FIGURAS

A Figura 1 fornece as estruturas e nomes correspondentes de quatro aminoácidos não-naturais que são  
 20 p-benzoil-L-fenilalanina (Bpa), para-acetil-L-fenilalanina (pAcPhe), para-azido-L-fenilalanina (pAzPhe) e para-iodo-L-fenil-alanina (pIPhe).

A Figura 2 fornece um histograma que ilustra as eficiências de supressão de plasmídeos (em relação à  $\beta$ -galactosidase do tipo selvagem) com o promotor *proK* e o  
 25 terminador para o gene *MjtRNA-Tyr(CUA)*, a mutação D286R no gene BpaRS, uma forma mutada do promotor *glnS* para o gene BpaRS, e/ou múltiplas cópias do gene de tRNA. As barras de

erros indicam desvio-padrão e  $n = 3$ .

A Figura 3 fornece uma imagem de quimioluminescência de uma análise *Northern blot* do supressor âmbar *MjtRNA-Tyr(CUA)* expressado a partir dos  
5 plasmídeos de supressão listados.

A Figura 4 fornece uma imagem de quimioluminescência após uma análise *Western blot* de BpaRS expressado sob controle do promotor *glnS* do tipo selvagem e a forma mutada do promotor *glnS*. A mancha usou um conjugado  
10 anticorpo-HRP anti-His(terminal C) (Invitrogen).

A Figura 5 ilustra um mapa plasmídico de pSup-BpaRS-6TRN. Outros genes de sintetase foram subclonados a partir dos seus plasmídeos pBK correspondentes nos sítios NdeI/PstI deste plasmídeo.

15 A Figura 6 ilustra as eficiências de supressão do novo sistema para a incorporação de Bpa, pAcPhe, pAzPhe e pIPhe. As barras de erros indicam desvio-padrão e  $n = 3$ .

A Figura 7 fornece uma imagem de quimioluminescência depois da análise *Western blot* de uma  
20 mioglobina mutante que contém Bpa no lugar de Ser-4 expressado na ausência ou presença de Bpa. A mancha usou um conjugado anticorpo-HRP anti-His(terminal C) (Invitrogen).

A Figura 8 fornece várias seqüências de polinucleotídeos e polipeptídeos que encontram uso com a  
25 invenção.

#### DESCRIÇÃO DETALHADA DA INVENÇÃO

A invenção fornece sistemas aperfeiçoados de vetores de expressão úteis para a expressão bacteriana

eficiente de proteínas mutantes que compreendem um ou mais aminoácidos não-naturais em sítios específicos geneticamente codificados por códons seletores (por exemplo, códons *nonsense* âmbar). Estes sistemas utilizam a tecnologia de  
5 tradução ortogonal conhecida nessas técnicas para a incorporação *in vivo* dos aminoácidos não-naturais. A invenção não está limitada em qualquer aspecto com relação aos componentes ortogonais específicos da tradução, que são usados (isto é, a amino-acil-tRNA sintetase ortogonal  
10 específica ou o tRNA supressor ortogonal específico). Além disso, em algumas modalidades, a invenção fornece composições e métodos aperfeiçoados que encontram amplo uso em sistemas de vetores de expressão bacterianos, que não estão limitados à expressão de amino-acil-tRNA sintetases ou  
15 tRNA's supressores ortogonais.

A invenção fornece características inusitadas de vetores de expressão, que resultam em eficiência significativamente melhorada na incorporação de aminoácidos não-naturais em proteínas em eubactérias (por exemplo, *E.*  
20 *coli*), e resultam na expressão com alto rendimento de proteínas mutantes que contêm os aminoácidos não-naturais em sítios específicos geneticamente designados por códons seletores. A melhor eficiência na incorporação de aminoácidos não-naturais em uma proteína de interesse deve-  
25 se presumivelmente (pelo menos em parte) à melhor expressão da amino-acil-tRNA sintetase e do tRNA supressor ortogonais, embora o entendimento do mecanismo da melhor eficiência não seja necessário para realizar ou usar a invenção.



As inusitadas características dos vetores de expressão da invenção são amplamente compatíveis com uma série de esqueletos de vetores de expressão em *E. coli* e cepas de *E. coli*, e são também facilmente adaptadas para a  
5 expressão de outras proteínas ou tRNA's de interesse, além da expressão de amino-acil-tRNA sintetases ou tRNA's supressores ortogonais.

Em alguns aspectos, as inusitadas características de vetores de expressão, fornecidas pela invenção, são  
10 usadas independentemente de plasmídeos separados. Em outros aspectos, uma única característica inusitada ou combinação de características são usadas em uma pluralidade de plasmídeos. Em ainda outras modalidades, várias destas características são usadas em combinação no mesmo plasmídeo.  
15 A invenção fornece inúmeros aperfeiçoamentos para sistemas de vetores de expressão bacterianos, que podem ser usados para melhorar a expressão de proteínas mutantes que compreendem um ou mais aminoácidos não-naturais, e em alguns casos, podem ser usados mais amplamente para melhorar a  
20 expressão de qualquer polipeptídeo ou tRNA específico de interesse.

A invenção fornece, por exemplo, os seguintes aperfeiçoamentos para sistemas de vetores de expressão bacterianos:

25 (A) A invenção fornece vetores de expressão, onde o gene de amino-acil-tRNA sintetase ortogonal e o gene de tRNA supressor ortogonal são transportados no mesmo plasmídeo. Isto simplifica a expressão destes componentes

ortogonais, onde anteriormente, estes dois componentes eram transportados em plasmídeos separados;

(B) A invenção fornece seqüências promotoras e terminadoras aperfeiçoadas derivadas de operons de tRNA de prolina de *E. coli*, para expressar uma seqüência de tRNA heteróloga, por exemplo, um gene de *MjtRNA-Tyr(CUA)* ortogonal ou qualquer outro gene de tRNA ortogonal. O gene de tRNA de prolina de *E. coli* usado para derivar as seqüências promotoras e terminadoras pode ser os genes de tRNA de *E. coli proK, proL* ou *proM*.

(C) A invenção fornece operons policistrônicos recombinantes aperfeiçoados para a expressão de genes de tRNA, onde quaisquer dois genes de tRNA no operon são separados por uma seqüência ligante heteróloga derivada de um ligante de um operon policistrônico de tRNA de ocorrência natural, por exemplo, o ligante que ocorre naturalmente entre os genes *E. coli valU* e *valX*, ou alternativamente, por exemplo, entre os genes *ileT* e *alaT*.

(D) A invenção fornece uma seqüência promotora inusitada derivada do promotor *E. coli glnS* para a melhor expressão de uma estrutura de leitura aberta, por exemplo, uma estrutura de leitura aberta que codifica uma amino-acil-tRNA sintetase ortogonal.

#### Sistemas de Vetores para Co-expressão de Genes de O-tRNA e O-RS

Em alguns aspectos, a invenção fornece vetores de expressão, onde o gene de amino-acil-tRNA sintetase ortogonal e o gene de tRNA supressor ortogonal são

transportados no mesmo plasmídeo. Esta característica é um aperfeiçoamento sobre as técnicas anteriores, onde anteriormente era necessário co-transformar uma célula hospedeira com dois vetores de expressão separados que transportavam independentemente os genes de O-tRNA e O-RS.

Como descrito nos exemplos, são construídos vários vetores de expressão relacionados que são apropriados para a co-expressão de espécies O-tRNA e O-RS. Estes plasmídeos incluem:

10                   pYR-BpaRS1  
                       pYR-BpaRS5  
                       pYR-BpaRS5 (D286R)  
                       pYR-BpaRS-TRN  
                       pYR-BpaRS-TRN (D286R)  
 15                   pYR-BpaRS-3TRN (D286R)  
                       pYR-BpaRS-6TRN (D286R)  
                       pSup-BpaRS-6TRN (D286R)  
                       pSup-pAcPheRS-6TRN  
                       pSup-pAzPheRS-6TRN  
 20                   pSup-pIPheRS-6TRN

Cada um destes plasmídeos é uma característica da invenção. Entretanto, não se pretende que a invenção esteja limitada a estes plasmídeos, pois os versados nessas técnicas devem reconhecer que a construção de variantes destes plasmídeos estão bem dentro do âmbito da invenção.

Por exemplo, não se pretende que qualquer plasmídeo da invenção esteja limitado à expressão de qualquer espécie de O-tRNA ou O-RS específica para produzir

uma proteína que compreende qualquer aminoácido não-natural específico. Os exemplos aqui fornecidos descrevem o uso exitoso de um *MjtRNA-Tyr(CUA)* (SEQ ID NO: 1) e quatro espécies diferentes de O-RS que têm especificidade de carregamento de tRNA para p-benzoil-L-fenilalanina (Bpa), para-acetil-L-fenilalanina (pAcPhe), para-azido-L-fenilalanina (pAzPhe) e para-iodo-L-fenil-alanina (pIPhe) (vide Figura 1 e Figura 8, SEQ ID NOS: 4, 6, 8 e 10).

Esses exemplos de trabalho acima servem para ilustrar a aplicabilidade mais ampla da invenção a ser usada com outras espécies de O-tRNA e O-RS. Na realidade, a invenção encontra uso na expressão de qualquer O-tRNA ou qualquer O-RS de interesse, e particularmente, componentes ortogonais de tradução que operam otimamente em células eubacterianas. Em alguns aspectos, a invenção encontra uso específico com espécies O-RS derivadas de *Archaea* de aminoacyl-tRNA sintetases de ocorrência natural (por exemplo, *Methanococcus jannaschii*) ou espécies O-tRNA derivadas de *Archaea*. A ampla série de espécies O-tRNA e O-RS que encontram uso com a invenção são conhecidas essas técnicas e estão descritas em inúmeros trabalhos. Vide, por exemplo, publicações internacionais nºs WO 2002/086075, intitulada "Methods and Composition for the Production of Orthogonal tRNA-aminoacyl-tRNA Synthetase Pairs"; WO 2002/085923, intitulada "In Vivo Incorporation of Unnatural Aminoacids"; WO 2004/094593, intitulada "Expandig the Eukariotic Genetic Code"; WO 2005/019415, depositado em 7 de julho de 2004; WO 2005/007870, depositado em 7 de julho de 2004; WO

2005/007624, depositado em 7 de julho de 2004; e WO 2006/110182, depositado em 27 de outubro de 2005, intitulada "Orthogonal Translation Components for the In Vivo Incorporation of Unnatural Aminoacids". Cada um destes

5 pedidos de patente é aqui incorporado como referência em sua totalidade. Para obter uma discussão adicional sobre sistemas de tradução ortogonal que incorporam aminoácidos não-naturais, e métodos para sua produção e uso, vide também Wang e Schultz, "Expanding the Genetic Code", *Angewandte*

10 *Chemie Int. Ed.* 44(1):34-66 (2005); Xie e Schultz, "An Expanding Genetic Code", *Methods* 36(3):227-238 (2005); Xie e Schultz, "Adding Aminoacids to the Genetic Repertoire", *Curr. Opinion in Chemical Biology* 9(6):548-554; e Wang et al., "Expanding the Genetic Code", *Annu. Rev. Biophys. Biomol. Struct.* 13 de janeiro de 2006; cujos teores são aqui

15 incorporados como referência em sua totalidade.

As técnicas anteriores (por exemplo, as técnicas aqui citadas) fornecem também orientação para a construção e uso de inúmeras variantes (por exemplo, variantes

20 conservativas) de fragmentos de espécies O-tRNA e O-RS conhecidas. Estas variantes e fragmentos também encontram uso com os vetores de expressão da invenção. As técnicas fornecem também orientação para a identificação e construção de novas espécies O-tRNA e O-RS, que também encontram uso

25 com a invenção.

#### Construções de Plasmídeos e Células Hospedeiras Eubacterianas

Como descrito e usado no Exemplo 6, os plasmídeos

fornecidos neste relatório descritivo baseiam-se em um esqueleto do vetor pACYC184. Entretanto, não se pretende que os plasmídeos da invenção estejam limitados ao uso daquele vetor com esqueleto específico. Os versados nessas técnicas  
5 devem reconhecer que qualquer um entre uma série de plasmídeos (incluindo outros plasmídeos disponíveis no mercado ou conhecidos) pode ser usado para construir os plasmídeos da invenção. Por exemplo, os plasmídeos pACYC177 e o vetor pRARE2 (Novagen; vide *Innovations*, 12 de junho de  
10 2001) também podem ser usados em conjunto com a invenção. Em alguns aspectos, qualquer plasmídeo portador de uma origem de replicação compatível (por exemplo, a origem de replicação p15A) e pelo menos um marcador de seleção pode ser usado em conjunto com a invenção. Os derivados do  
15 plasmídeo pSC101 também podem ser usados com a invenção.

Também como descrito no Exemplo 6, os plasmídeos fornecidos pela invenção foram usados para transformar *E. coli* eletrocompetente One Shot® TOP-10 (Invitrogen™). Entretanto, não se pretende que a cepa eubacteriana usada  
20 como uma célula hospedeira para produzir proteínas que compreendem um ou mais aminoácidos não-naturais esteja limitada ao uso dessa célula hospedeira específica. Os versados nessas técnicas devem reconhecer que qualquer uma entre uma série de células hospedeiras (incluindo outros  
25 plasmídeos disponíveis no mercado, bem como as cepas produzidas pelo usuário) podem ser facilmente usadas para produzir proteínas que compreendem aminoácidos não-naturais. Por exemplo, outras células hospedeiras tais como a cepa de

*E. coli* DH10B™ (Invitrogen™), Electrocomp™ GeneHogs® (Invitrogen™), BL21 One Shot® (Invitrogen™), e BL21(DE3) One Shot® (Invitrogen™). Na realidade, quaisquer cepas de *E. coli* sem qualquer gene supressor de tRNA endógeno é uma  
 5 célula hospedeira apropriada. Outras espécies de eubactérias além de *E. coli* também encontram uso com invenção. Por exemplo, contempla-se que as cepas de *Bacillus subtilis* também podem ser usadas como células hospedeiras para os vetores da invenção.

10                    Seqüências Promotoras e Terminadoras Aperfeiçoadas para a Expressão de O-tRNA

Para melhorar a eficiência de supressão do sistema de tradução ortogonal, um novo operon de tRNA supressor âmbar com um promotor e terminador de tRNA de *E. coli* de  
 15 ocorrência natural foi construído. Um levantamento de genes de tRNA de *E. coli* revelou que os tRNA's de prolina de *E. coli* têm o mesmo par C1-G72 como tRNA's de *Archaea*; este par de bases é um determinante importante de identidade para o reconhecimento seletivo de *MjtRNA-Tyr(CUA)* por *MjTyrRS* em *E.*  
 20 *coli* (Wang e Schulz, *Chem. Biol.* 8:883-890 (2001)).

Tendo em vista esta observação, um gene de tRNA supressor âmbar sintético foi construído, de tal modo que um gene de O-tRNA heterólogo substituísse o gene *E. coli proK* do mesmo comprimento (77 nucleotídeos) no operon *proK*  
 25 monocistrônico. Um vetor de expressão aperfeiçoado (pYR-BpaRS5) foi gerado substituindo o operon de tRNA supressor original em pYR-BpaRS1 pelo gene *MjtRNA-Tyr(CUA)* sob controle do promotor *proK* (SEQ ID NO: 32) e do terminador

*proK* (SEQ ID NO: 33). Esta construção de expressão apresentou um aumento de 2 vezes na expressão do O-tRNA (vide Figura 3), e resultou em uma eficiência de supressão significativamente melhor (vide Figura 2), ambos em relação  
5 às atividades do vetor pYR-BpaRS1.

Assim sendo, a invenção fornece vetores de expressão aperfeiçoados para a expressão de um tRNA de interesse, onde a expressão de um tRNA policistrônico é acionada pelas seqüências promotora e terminadora derivadas  
10 do gene de tRNA de prolina de *E. coli proK*.

Como descrito no Exemplo 1, o tRNA específico usado para demonstrar este veor de expressão aperfeiçoado foi um tRNA ortogonal (O-tRNA), e mais especificamente, *MjtRNA-Tyr(CUA)*. Entretanto, não se pretende que a melhor  
15 eficiência da expressão de tRNA esteja limitada ao *MjtRNA-Tyr(CUA)*, nem limitada a um O-tRNA. Na realidade, esta característica da invenção pode ser usada para melhorar a expressão de qualquer tRNA de interesse.

Em *E. coli*, três espécies de tRNA são carregadas  
20 com prolina durante a tradução. Além do gene de tRNA *proK*, *E. coli* usa também dois genes *proL*-tRNA adicionais. Eles são *proL* e *proM*. Tendo em vista sua estrutura similar ao locus *proK*, contempla-se que a seqüência promotora de *proL* (SEQ ID NO: 34) e as seqüências terminadoras de *proL* e *proM*  
25 (SEQ ID NOS: 35 e 36, respectivamente) também encontram uso na construção dos vetores de expressão aperfeiçoados da invenção. É também uma característica da invenção que combinações de promotores e terminadores de genes *proL*-



tRNA de *E. coli* também podem ser usadas para conseguir melhor expressão. Por exemplo, a sequência promotora de *proK* (SEQ ID NO: 32) pode ser usada em conjunto com a sequência terminadora de *proL* (SEQ ID NO: 35).

## 5                    Estruturas            Aperfeiçoadas            de            Operons Policistrônicos

A invenção fornece operons policistrônicos recombinantes aperfeiçoados para a expressão de genes de tRNA. Estes operons policistrônicos compreendem múltiplas  
10 cópias (por exemplo, três cópias) de genes de tRNA de interesse, onde as sequências de tRNA são separadas por uma sequência ligante heteróloga derivada de um ligante de um operon policistrônico de tRNA de ocorrência natural, por exemplo, o ligante que ocorre naturalmente entre os genes  
15 *valU* e *ValX* de *E. coli* (SEQ ID NO: 14), ou alternativamente, por exemplo, entre os genes de tRNA *ileT* e *alaT* de *E. coli* (SEQ ID NO: 15).

Assim sendo, a invenção fornece vetores de expressão aperfeiçoados para a expressão de operons de tRNA  
20 policistrônicos, onde o operon compreende pelo menos um ligante de tRNA heterólogo que separa pelo menos duas sequências de tRNA expressadas. Em algumas modalidades, como descrito no Exemplo 3, múltiplos ligantes de tRNA são usados para separar três ou mais sequências de tRNA expressadas no  
25 operon. Neste caso, o ligante de tRNA usado entre cada par de tRNA expressado pode ser diferente (como no Exemplo 3), ou pode ser o mesmo ligante entre cada gene de tRNA.

Não se pretende que a invenção esteja limitada ao

uso do ligante de genes *valU* e *valX* de *E. coli* (SEQ ID NO: 14) ou ao ligante de genes de tRNA *ileT* e *alaT* de *E. coli* (SEQ ID NO: 15). Na realidade, ligantes de tRNA de ocorrência natural adicionais também encontram uso com a invenção. Por exemplo, cada um dos ligantes que seguem, localizados entre os genes de tRNA nativos de *E. coli* listados abaixo, encontram uso com a invenção, onde o ligante usado no sistema recombinante é heterólogo a quaisquer seqüências de tRNA expressadas, está no operon recombinante. Estes ligantes de tRNA úteis incluem:

Ligante de tRNA nativo de <i>E. coli</i>	Seqüência	SEQ ID NO:
<i>valU</i> e <i>valX</i>	ACTACTTTATGTAGTCTCCGCCGTGTAGCAAGAAATTGAGAAGT	14
<i>ileT</i> e <i>alaT</i>	AATTTGCACGGCAAATTTGAAGAGGTTTTAACTACATGTTAT	15
<i>serV</i> e <i>argV</i>	TTT	16
<i>valV</i> e <i>valW</i>	TCCT	17
<i>glyT</i> e <i>thrT</i>	AGATGT	18
<i>metT</i> e <i>leuW</i>	TCTTTTTTTT	19
<i>glnW</i> e <i>metU</i>	TCGAAGAAACAATCT	20
<i>hisR</i> e <i>leuT</i>	TTATTAGAAGTTGTGACAAT	21
<i>glnU</i> e <i>glnW</i>	TCTTCTTCGAGTAAGCGGTTACCCGCCCGGTTAT	22
<i>leuP</i> e <i>leuV</i>	AACGAGGCGATATCAAAAAAGTAAGATGACTGT	23
<i>glnV</i> e <i>glnX</i>	ATTTATTCAAGACGCTTACCTTGTAAGTGCACCCAGT	24
<i>alaW</i> e <i>alaX</i>	AATTTTGCACCCAGCAAACCTGGTACGTAAACGCATCGT	25
<i>ileU</i> e <i>alaU</i>	AATTTGCACGGCAAATTTGAAGAGGTTTTAACTACATGTTAT	26
<i>ileV</i> e <i>alaV</i>	AATTTGCACGGCAAATTTGAAGAGGTTTTAACTACATGTTAT	27
<i>metU</i> e <i>glnV</i>	AATTCCTGAATGTATCGAATATGTTCGGCAAATTCAAAACCAATTTGT	28
<i>glyW</i> e <i>cysT</i>	GTTTAAAAGACATCGGCGTCAAGCGGATGTCTGGCTGAAAGGCCTGAAGAATTT	29
<i>argX</i> e <i>hisR</i>	TTTAGTCCCGCGCTTGAGCTGCGGTGGTAGTAATACCGCGTAACAAGATTTGTAGT	30
<i>argY</i> e <i>argZ</i>	TCTCTTACTTGATATGGCTTTAGTAGCGGTATCAATATCAGCAGTAAAATAAATTTCCCGAT	31

Em algumas modalidades, os ligantes de tRNA preferidos, que encontram uso com a invenção, contêm um dos dois ou ambos nucleotídeos T(-1) e A(77). Estas duas posições de nucleotídeos em ligantes de tRNA demonstraram ser ideais para o processamento eficiente em 5' e 3' de precursores de tRNA quando em seu contexto nativo (isto é, endógeno). Vide, por exemplo, Li e Deutscher, "Maturation pathways for *E. coli* tRNA precursors: A random multienzyme process *in vivo*", *Cell* 86:503-512 (1996) e Zahler et al., "Recognition of the 5' leader of pré-tRNA substrates by the active site of ribonuclease P", *RNA* 9:734-745 (2003). Em outras modalidades, os ligantes de tRNA que encontram uso com a invenção compreendem sítios de restrição (de ocorrência natural ou manipulados).

Em algumas modalidades, a invenção fornece construções que compreendem uma pluralidade do mesmo operon policistrônico, opcionalmente em tandem. Assim sendo, caso um único operon policistrônico compreenda três cópias de uma sequência de nucleotídeos expressável (tal como um gene de tRNA), então dois dos operons resultarão em um total de seis sequências de genes de tRNA sendo expressadas. Este tipo de configuração em *cluster* de genes está demonstrado no Exemplo 3 e na Figura 5.

O operon policistrônico recombinante aperfeiçoado descrito no Exemplo 3 expressa o tRNA ortogonal *MjtRNA-Tyr(CUA)*. Entretanto, não se pretende que a invenção esteja limitada à expressão de *MjtRNA-Tyr(CUA)*; nem a invenção está limitada à expressão de espécies tRNA ortogonais. Na

realidade, os operons policistrônicos aperfeiçoados da invenção podem ser usados para expressar quaisquer espécies tRNA.

Promotor *glnS* de *E. coli* Aperfeiçoado para a

5 Expressão de Polipeptídeos

A invenção fornece uma sequência promotora inusitada derivada do promotor *glnS* de *E. coli* para a expressão aperfeiçoada de um estrutura de leitura aberta. Como descrito no Exemplo 2, o promotor *glnS* mutante (SEQ ID  
10 NO: 12) descrito em Plumbridge e Söll (*Biochemie* 69:539-541 (1987)) foi subclonado em um vetor de expressão da invenção. O seqüenciamento da região do promotor *glnS* subclonado revelou a introdução inadvertida de uma deleção na sequência promotora (além da substituição descrita em Plumbridge e  
15 Söll). Esta outra variante do promotor *glnS* inusitado modificado foi denominada *glnS*-TNR (fornecida em SEQ ID NO: 13). Surpreendentemente, esta mutação resultou em uma melhora na eficiência da tradução do sistema em comparação com a atividade do promotor *glnS* do tipo selvagem, como  
20 determinado por *Western blotting* (vide Figura 4).

Como descrito nos Exemplos 2 e 5, o promotor *glnS*-TNR foi usado para expressar as sintetases ortogonais BpaRS, pAcPheRS, pAzPheRS, e pIPheRS. Entretanto, não se pretende que a invenção esteja limitada à expressão de qualquer  
25 amino-acil-tRNA sintetase ortogonal específica, ou limitada a qualquer amino-acil-tRNA sintetase em geral. Este promotor aperfeiçoado encontra uso na expressão bacteriana de qualquer estrutura de leitura aberta de polipeptídeo

desejada.

## Tecnologia de tRNA/Amino-acil-tRNA Sintetase Ortogonal

Uma compreensão das composições e métodos inusitados da presente invenção é facilitada pela compreensão das atividades associadas com pares de tRNA ortogonal e amino-acil-tRNA sintetase ortogonal. Para adicionar aminoácidos não-naturais adicionais ao código genético, são necessários novos pares ortogonais que compreendem uma amino-acil-tRNA sintetase e um tRNA apropriado, que funcionam eficientemente no mecanismo de tradução no hospedeiro, mas que são "ortogonais" ao sistema de tradução em questão, significando que ele funciona independentemente das sintetases e tRNA's endógenos ao sistema de tradução. As características desejadas do par ortólogo incluem um tRNA que descodifica ou reconhece apenas um códon específico, por exemplo, um códon seletor, que não é descodificado por qualquer tRNA endógeno, e as amino-acil-tRNA sintetases que preferencialmente amino-acilam (ou "carregam") seu tRNA cognato com apenas um aminoácido não-natural específico. O O-tRNA também não é tipicamente amino-acilado por sintetases endógenas. Por exemplo, em *E. coli*, um par ortogonal deve incluir uma amino-acil-tRNA sintetase que não sofre reação cruzada com qualquer um dos tRNA's endógenos, por exemplo, que existem são 40 em *E. coli*, e um tRNA ortogonal que não é amino-acilado por qualquer uma das sintetases endógenas, por exemplo, das quais há 21 em *E. coli*. Até agora, vários aminoácidos não-naturais

estruturalmente diversos foram incorporados em proteínas usando a tecnologia de tradução ortogonal, como é do conhecimento nessas técnicas.

A capacidade de incorporar um aminoácido não-natural de forma específica de sítio em um polipeptídeo pode facilitar o estudo de proteínas, permitindo a modificação altamente seletiva depois da tradução destas proteínas, bem como permitir a engenharia de proteínas com propriedades inusitadas. Por exemplo, a expressão de proteínas que contêm um ou mais aminoácidos não-naturais pode facilitar o estudo de proteínas por marcação específica, alterar a função catalítica de enzimas, melhorar a atividade biológica ou reduzir a reação cruzada com um substrato, reticular uma proteína com outras proteínas, moléculas pequenas ou biomoléculas, reduzir ou eliminar a degradação de proteínas, melhorar a meia-vida de proteínas *in vivo* (por exemplo, por "peguilação" ou outras modificações de sítios reativos introduzidos), etc.

Os sistemas de tradução ortogonais que são apropriados para produzir proteínas que compreendem um ou mais aminoácidos não-naturais são conhecidos nessas técnicas, assim como o são os métodos genéricos para produzir sistemas de tradução ortogonais. Vide, por exemplo, as publicações internacionais n<sup>os</sup> WO 2002/086075, intitulada "Methods and Compositions for the Production of Orthogonal tRNA-Aminoacyl-tRNA Synthetase Pairs"; WO 2002/085923, intitulada "In Vivo Incorporation of Unnatural Amino Acids"; WO 2004/094593, intitulada "Expanding the Eukaryotic Genetic

Code"; WO 2005/019415, depositado em 7 de julho de 2004; WO 205/007870, depositado em 7 de julho de 2004; WO 205/007624, depositado em 7 de julho de 2004; e WO 2006/110182, depositado em 27 de outubro de 2005, intitulado "Orthogonal  
 5 Translation Components for the In Vivo Incorporation of Unnatural Amino Acids". Cada um destes pedidos de patente é aqui incorporado como referência em sua totalidade. Para obter uma discussão sobre sistemas de tradução ortogonal que incorporam aminoácidos não-naturais, e métodos para sua  
 10 preparação e uso, vide também Wang e Schultz, ""Expanding the Genetic Code", *Angewandte Chemie Int. Ed.* 44(1):34-66 (2005); Xie e Schultz, "An Expanding Genetic Code", *Methods* 36(3):227-238 (2005); Xie e Schultz, "Adding Aminoacids to the Genetic Repertoire", *Curr. Opinion in Chemical Biology*  
 15 9(6):548-554; Wang et al., "Expanding the Genetic Code", *Annu. Rev. Biophys. Biomol. Struct.* 13 de janeiro de 2006; e Xie e Schultz, "A chemical toolkit for proteins - an expanded genetic code", *Nat. Rev. Mol. Cell Biol.* 7(10):775-782 (2006; 23 de agosto de 2006), cujos teores são  
 20 incorporados como referência em sua totalidade.

Tais sistemas de tradução compreendem genericamente células (que podem ser células procarióticas tais como *E. coli*, ou células eucarióticas tais como leveduras) que incluem um tRNA ortogonal (O-tRNA), uma  
 25 amino-acil-tRNA sintetase ortogonal (O-RS), e um aminoácido não-natural, onde a O-RS amino-acila o O-tRNA com o aminoácido não-natural. Um par ortogonal da invenção pode incluir um O-tRNA, por exemplo, um tRNA supressor, um tRNA

de deslocamento de estrutura, ou similares, e uma O-RS cognata.

Genericamente, quando um par ortogonal reconhece um códon seletor e carrega um aminoácido em resposta ao  
5 códon seletor, o par ortogonal é dito como "suprimindo" o códon seletor. Isto é, um códon seletor que não é reconhecido pelo mecanismo endógeno do sistema de tradução (por exemplo, da célula) não é normalmente carregado, o que resulta no bloqueio da produção de um polipeptídeo que de  
10 outra forma seria traduzido a partir do ácido nucléico. Em um sistema de par ortogonal, a O-RS amino-acila o O-tRNA com um aminoácido não-natural específico. O O-tRNA carregado reconhece o códon seletor e suprime o bloqueio da tradução causado pelo códon seletor. A célula usa o par O-tRNA/O-RS  
15 para incorporar o aminoácido não-natural em uma cadeia de polipeptídeo em crescimento, por exemplo, por intermédio de um ácido nucléico que compreende um polinucleotídeo que codifica um polipeptídeo de interesse, onde o polinucleotídeo compreende um códon seletor que é  
20 reconhecido pelo O-tRNA. Em certos aspectos desejáveis, a célula pode incluir um par de O-tRNA/O-RS adicional, onde o O-tRNA adicional é carregado pela O-RS adicional com um aminoácido não-natural diferente. Por exemplo, um dos O-tRNA's pode reconhecer um códon de quatro bases e depois o  
25 outro pode reconhecer um códon finalizador. Alternativamente, múltiplos códons finalizadores diferentes ou múltiplos códons com quatro bases, diferentes, podem reconhecer especificamente diferentes códons seletores.



Em certas modalidades, os sistemas compreendem uma célula, tal como uma célula de *E. coli* ou uma célula de levedura, que inclui um tRNA ortogonal (O-tRNA), uma amino-acil-tRNA sintetase ortogonal (O-RS), um aminoácido não-natural e um ácido nucléico que compreende um polinucleotídeo um polipeptídeo de interesse, onde o polinucleotídeo compreende o códon seletor que é reconhecido pelo O-tRNA. O sistema de tradução pode ser também um sistema isento de células, por exemplo, qualquer um entre uma série de sistemas de transcrição/tradução "*in vitro*" disponíveis no mercado em combinação com um par O-tRNA/O-RS e um aminoácido não-natural, como aqui descrito.

Como assinalado, em algumas modalidades, existem múltiplos pares de O-tRNA/O-RS em uma célula ou outro sistema de tradução, o que permite a incorporação de mais do que aminoácido não-natural em um polipeptídeo. Por exemplo, a célula pode incluir ainda um par O-tRNA/O-RS adicional diferente e um segundo aminoácido não-natural, onde este O-tRNA adicional reconhece um segundo códon seletor e esta O-RS adicional preferencialmente amino-acila o O-tRNA com o segundo aminoácido não-natural. Por exemplo, uma célula que inclui um par O-tRNA/O-RS (onde o O-tRNA reconhece, um códon seletor âmbar), pode compreender ainda um segundo par ortogonal, onde o segundo O-tRNA reconhece um códon seletor diferente, por exemplo, um códon opalino, um códon de quatro bases, ou similares. Desejavelmente, os pares ortogonais diferentes são derivados de diferentes origens, que podem facilitar o reconhecimento de diferentes códons seletores.

O O-tRNA e/ou a O-RS podem ser de ocorrência natural ou podem ser, por exemplo, derivados por mutação de um tRNA e/ou RS de ocorrência natural, por exemplo, gerando bibliotecas de tRNA's e/ou bibliotecas de RS's, a partir de qualquer um entre uma série de organismos e/ou usando qualquer uma entre uma série de estratégias de mutação disponíveis. Por exemplo, uma estratégia para produzir um par de tRNA/amino-acil sintetase ortogonal envolve importar um par tRNA/sintetase heterólogo (à célula hospedeira) a partir, por exemplo, de uma fonte que não a célula hospedeira, ou múltiplas fontes, na célula hospedeira. As propriedades da sintetase heteróloga candidata incluem, por exemplo, que ela não carrega qualquer tRNA da célula hospedeira, e as propriedades do tRNA heterólogo candidato incluem, por exemplo, que ele não é amino-acilado por qualquer sintetase da célula hospedeira. Além disso, o tRNA heterólogo é ortogonal a todas sintetases da célula hospedeira.

Uma segunda estratégia para gerar um par ortogonal envolve gerar bibliotecas mutantes a partir das quais se pode triar e/ou selecionar um O-tRNA ou uma O-RS. Estas estratégias podem ser também combinadas.

#### tRNA Ortogonal (O-tRNA)

Um tRNA ortogonal (O-tRNA) desejavelmente medeia a incorporação de um aminoácido não-natural 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*, com uma alta eficiência de supressão. A

eficiência da supressão pode ser determinada por qualquer um entre inúmeros ensaios conhecidos nessas técnicas. Por exemplo, um ensaio de  $\beta$ -galactosidase repórter pode ser usado, por exemplo, um plasmídeo *lacZ* derivado (onde a  
5 construção tem um códon seletor na sequência de ácidos nucleicos de *lacZ*) é introduzido nas células a partir de um organismo apropriado (por exemplo, um organismo no qual os componentes ortogonais podem ser usados) junto com o plasmídeo que compreende um O-tRNA da invenção. Uma  
10 sintetase cognata também pode ser introduzida (como um polipeptídeo ou um polinucleotídeo que codifica a sintetase cognata quando expressado). As células são desenvolvidas no meio até uma densidade desejada, por exemplo, até uma  $DO_{600}$  de cerca de 0,5, e ensaios de  $\beta$ -galactosidase são realizados usando, por exemplo, o Kit de Ensaio de  $\beta$ -Galactosidase  
15 BetaFluor™ (Novagen). A supressão percentual pode ser calculada como a porcentagem de atividade de uma amostra em relação a um controle comparável, por exemplo, o valor observado a partir da construção de *lacZ* derivada, onde a  
20 construção tem um códon *sense* correspondente na posição desejada em vez de um códon seletor.

Os O-tRNA's podem ser derivados também a partir de variações conservativas de O-tRNA's conhecidos. Por exemplo, as variações conservativas de O-tRNA incluem as moléculas  
25 que funcionam como os O-tRNA's específicos, por exemplo, como na listagem de sequências aqui fornecida e que mantêm a estrutura do tRNA com formato em L em virtude de autocomplementaridade apropriada, mas não têm uma sequência

idêntica àquelas, por exemplo, na listagem de seqüências, figuras ou exemplos neste relatório descritivo (e desejavelmente, são diferentes das moléculas de tRNA do tipo selvagem).

5           A composição que compreende um O-tRNA pode incluir ainda uma amino-acil-tRNA sintetase ortogonal (O-RS), onde a O-RS preferencialmente amino-acila o O-tRNA com um aminoácido não-natural. Em certas modalidades, uma composição que inclui um O-tRNA pode incluir ainda um  
10 sistema de tradução (por exemplo, *in vitro* ou *in vivo*). Um ácido nucléico que compreende um polinucleotídeo que codifica um polipeptídeo de interesse, onde o polinucleotídeo compreende um códon seletor que é reconhecido pelo O-tRNA, ou uma combinação de um ou mais  
15 deles, também pode estar presente na célula.

Os métodos para produzir um tRNA ortogonal (O-tRNA) são conhecidos. Em certas modalidades da invenção, os O-tRNA's podem ser produzidos gerando uma biblioteca de mutantes. A biblioteca de tRNA's mutantes pode ser gerada  
20 usando várias técnicas de mutagênese conhecidas nessa área. Por exemplo, os tRNA's mutantes podem ser gerados por mutações específicas de sítio, mutações pontuais aleatórias, recombinação homóloga, embaralhamento de DNA ou outros métodos de mutagênese recursivos, construção quimérica ou  
25 qualquer combinação deles.

Mutações adicionais podem ser introduzidas em uma posição ou posições específicas, por exemplo, e uma posição não-conservada, ou em uma posição conservada, e uma posição

ou posições randomizadas, ou uma combinação de ambas, em uma alça ou região desejada de um tRNA, por exemplo, uma alça anticódon, o tronco acceptor, braço ou alça D, alça variável, braço ou alça TPC, outras regiões da molécula de tRNA, ou  
5 uma combinação delas. Tipicamente, as mutações em um tRNA incluem mutar a alça anticódon de cada membro da biblioteca de tRNA's mutantes, para permitir o reconhecimento de um códon seletor. O método pode incluir ainda adicionar seqüências adicionais ao O-tRNA. Tipicamente, um O-tRNA  
10 possui uma melhora de ortogonalidade para um organismo desejado em comparação com o material de partida, por exemplo, a pluralidade de seqüências de tRNA, e ao mesmo tempo, preservando sua afinidade por uma RS desejada.

Os métodos incluem opcionalmente analisar a  
15 similaridade (e/ou homologia inferida) de seqüências de tRNA's e/ou amino-acil-tRNA sintetases, para determinar candidatos potenciais para um O-tRNA, O-RS e/ou pares deles, que parecem ser ortogonais para um organismo específico. Programas de computador conhecidos nessas técnicas e aqui  
20 descritos podem ser usados para a análise, por exemplo, programas BLAST e de acumulação podem ser usados. Em um exemplo, escolher componentes ortogonais potenciais da tradução para uso em *E. coli*, é escolhida uma sintetase e/ou um tRNA que não apresenta estreita similaridade de seqüência  
25 com organismos eubacterianos.

Tipicamente, um O-tRNA é obtido submetendo, por exemplo, a uma seleção negativa, uma população de células de uma primeira espécie, onde as células compreendem um membro

da pluralidade de O-tRNA's potenciais. A seleção negativa elimina as células que compreendem um membro da biblioteca de O-tRNA's potenciais que é amino-acilado por uma amino-acil-tRNA sintetase (RS) que é endógena para a célula. Isto  
5 produz uma coleção de tRNA's que são ortogonais à célula da primeira espécie.

Em certas modalidades, na seleção negativa, um códon(s) seletor(es) é introduzido em um polinucleotídeo que codifica um marcador de seleção negativa, por exemplo, uma  
10 enzima que confere resistência a antibióticos, por exemplo,  $\beta$ -lactamase, uma enzima que confere um produto detectável, por exemplo, cloranfenicol acetiltransferase (CAT), por exemplo, um produto tóxico, tal como barnase, em uma posição não-essencial (por exemplo, ainda produzindo uma barnase  
15 funcional), etc. A triagem/seleção é feita opcionalmente desenvolvendo a população de células na presença de um agente seletor (por exemplo, um antibiótico, tal como ampicilina). Em uma modalidade, a concentração do agente de seleção é variada.

20 Por exemplo, para medir a atividade de tRNA's supressores, usa-se um sistema de seleção baseado na supressão *in vivo* de códon seletor, por exemplo, mutações *nonsense* (por exemplo, finalizadoras) ou com deslocamento de estrutura, introduzidas em um polipeptídeo que codifica um  
25 marcador de seleção negativa, por exemplo, um gene para  $\beta$ -lactamase (*bla*). Por exemplo, são construídas variantes de polinucleotídeos, por exemplo, variantes *bla*, com um códon seletor em uma certa posição (por exemplo, A184). As

células, por exemplo, bactérias, são transformadas com estes polinucleotídeos. No caso de um tRNA ortogonal, que não pode ser carregado eficientemente por sintetases de *E. coli* endógenas, a resistência a antibióticos, por exemplo, 5 resistência à ampicilina, deve ser próxima ou menor do que aquela para uma bactéria transformada sem qualquer plasmídeo. Caso o tRNA não seja ortogonal, ou caso uma sintetase heteróloga capaz de carregar o tRNA seja co-expressa no sistema, um nível mais alto de resistência a 10 antibióticos, por exemplo, ampicilina, seja observada. São escolhidas células, por exemplo, bactérias, que são incapazes de se desenvolverem sobre placas de ágar LB com concentrações de antibióticos quase iguais às células transformadas sem qualquer plasmídeo.

15           No caso de um produto tóxico (por exemplo, ribonuclease ou barnase), quando um membro de uma pluralidade de tRNA's potenciais é amino-acilado pela hospedeira endógena, por exemplo, sintetases de *Escherichia coli* (isto é, ele não é ortogonal à hospedeira, por exemplo, 20 sintetases de *Escherichia coli*), o códon seletor é suprimido e o produto tóxico de polinucleotídeo leva à morte celular. As células portadoras de tRNA's ortogonais ou tRNA's não-funcionais sobrevivem.

25           Em uma modalidade, a coleção de tRNA's que são ortogonais a um organismo desejado são então submetidos a uma seleção positiva na qual um códon seletor é colocado em um marcador de seleção positiva, por exemplo, codificado por um gene de resistência a fármacos, tal como um gene de  $\beta$ -

lactamase. A seleção positiva é realizada em uma célula que compreende um polinucleotídeo que codifica ou compreende um membro da coleção de tRNA's que são ortogonais à célula, um polinucleotídeo que codifica um marcador de seleção  
5 positiva, e um polinucleotídeo que codifica uma RS cognata. Em certas modalidades, a segunda população de células compreende células que não foram eliminadas pela seleção negativa. Os polinucleotídeos são expressados na célula e a célula é desenvolvida na presença de um agente de seleção,  
10 por exemplo, ampicilina. Os tRNA's são então selecionados quanto à sua capacidade de serem amino-acilados pela sintetase cognata co-expressada, e de inserir um aminoácido em resposta a este códon seletor. Tipicamente, estas células apresentam uma melhora na eficiência da supressão em  
15 comparação com células portadoras de tRNA ou tRNA's não-funcionais, ou tRNA's que não podem ser eficientemente reconhecidos pela sintetase de interesse, são sensíveis ao antibiótico. Portanto, os tRNA's que: (i) não são substratos para a hospedeira endógena, por exemplo, sintetases de  
20 *Escherichia coli*; (ii) podem ser amino-acilados pela sintetase de interesse; e (iii) são funcionais na tradução, sobrevivem a ambas seleções.

Conseqüentemente, o mesmo marcador pode ser um marcador positivo ou negativo, dependendo do contexto no  
25 qual ele é triado. Isto é, o marcador é um marcador positivo caso ele seja triado a favor, mas é um marcador negativo caso seja triado contra.

A severidade da seleção, por exemplo, a seleção



positiva, a seleção negativa, ou a seleção positiva e também a negativa, nos métodos descritos acima, inclui opcionalmente variar a severidade da seleção. Por exemplo, devido ao fato de que a barnase é uma proteína extremamente tóxica, a severidade da seleção negativa pode ser controlada introduzindo diferentes números de códons seletores no gene de barnase, e/ou usando um promotor induzível. Em outro exemplo, a concentração do agente de seleção ou triagem é variada (por exemplo, a concentração de ampicilina). Em alguns aspectos da invenção, a severidade é variada porque a atividade desejada pode ser baixa durante as primeiras rodadas. Assim sendo, são aplicados critérios de seleção menos severos nas primeiras rodadas e critérios mais severos são aplicados em rodas posteriores da seleção. Em certas modalidades, a seleção negativa, a seleção positiva, ou a seleção negativa e também a positiva, podem ser repetidas múltiplas vezes. Múltiplos marcadores diferentes de seleção negativa, marcadores de seleção positiva ou marcadores de seleção negativa e também positiva podem ser usados. Em certas modalidades, o marcador de seleção positiva e negativa pode ser o mesmo.

Outros tipos de seleções/triagens podem ser usados na invenção para produzir componentes ortogonais de tradução, por exemplo, um O-tRNA, uma O-RS, e um par de O-tRNA/O-RS que carrega um aminoácido não-natural em resposta a um códon seletor. Por exemplo, o marcador de seleção negativa, o marcador de seleção positiva, ou o marcador de seleção positiva e também o marcador de seleção negativa

incluem um marcador que fluoresce ou catalisa uma reação luminescente na presença de um reagente apropriado. Em outra modalidade, um produto do marcador é detectado por seleção de células ativadas por fluorescência (FACS) ou por luminescência. Opcionalmente, o marcador inclui um marcador de triagem baseado em afinidade. Vide também Francisco, J.A. et al., "Production and fluorescence-activated cell sorting of *Escherichia coli* expressing a functional antibody fragment on external surface", *Proc. Natl. Acad. Sci. USA* 90:10444-8.

Métodos adicionais para produzir um tRNA ortogonal recombinante podem ser encontrados, por exemplo, nas publicações dos pedidos de patente internacionais nºs WO 2002/086075, intitulado "Methods and Compositions for the Production of Orthogonal tRNA Aminoacyl-tRNA Synthetase Pairs"; WO 2004/094593, intitulado "Expanding the Eukariotic Genetic Code"; e WO 2005/019415, depositado em 7 de julho de 2004. Vide também Forster et al., "Programming peptidomimetic synthetases by translating genetic codes designed de novo", *PNAS* 100(11):6353-6357 (2003); e Feng et al., "Expanding tRNA recognition of a tRNA synthetase by a single amino acid change", *PNAS* 100(10):5676-5681(2003).

#### Amino-acil-tRNA Sintetase Ortogonal (O-RS)

Uma O-RS que encontra uso com a invenção preferencialmente amino-acila um O-tRNA com um aminoácido não-natural, *in vitro* ou *in vivo*. Uma O-RS pode ser disponibilizada para o sistema de tradução, por exemplo, uma célula, por um polipeptídeo que inclui uma O-RS e/ou por um

polinucleotídeo que codifica uma O-RS ou uma parte dela. Por exemplo, uma O-RS compreende uma sequência de aminoácidos conhecida nessas técnicas, ou uma sua variação conservativa. Em outro exemplo, uma O-RS ou uma parte dela é codificada por uma sequência de polinucleotídeo que codifica um aminoácido que compreende a sequência na listagem de sequências ou exemplos neste relatório descritivo, ou uma sua sequência de polinucleotídeo complementar. Vide, por exemplo, Figura 8 para obter as sequências de moléculas de O-RS úteis.

Os métodos para identificar uma amino-acil-tRNA sintetase ortogonal (O-RS), por exemplo, uma O-RS para uso com um O-tRNA são conhecidos. Por exemplo, um método inclui submeter à seleção, por exemplo, seleção positiva, uma população de células de uma primeira espécie, onde as células compreendem individualmente: (1) um membro de uma pluralidade de amino-acil-tRNA sintetases (RS's) (por exemplo, a pluralidade de RS's pode incluir RS's mutantes, RS's derivadas de uma espécie diferente da primeira espécie); (2) o tRNA ortogonal (O-tRNA) (por exemplo, de uma ou mais espécies); e (3) um polinucleotídeo que codifica um marcador de seleção (por exemplo, positiva) e compreende pelo menos um códon seletor. As células são selecionadas ou triadas quanto àquelas que apresentam uma melhora na eficiência da supressão em comparação com células que não têm ou têm uma quantidade reduzida do membro da pluralidade de RS's. A eficiência da supressão pode ser medida por técnicas conhecidas nessa área e como aqui descritas. As

células que têm uma melhora na eficiência da supressão compreendem uma RS ativa que amino-acila o O-tRNA. Um nível de amino-acilação (*in vitro* ou *in vivo*) pela RS ativa de um primeiro conjunto de tRNA's da primeira espécie é comparado  
5 com o nível de amino-acilação (*in vitro* ou *in vivo*) pela RS ativa de um segundo conjunto de tRNA's da segunda espécie. O nível de amino-acilação pode ser determinado por uma substância detectável (por exemplo, um aminoácido não-natural marcado). A RS ativa que amino-acila mais  
10 eficientemente o segundo conjunto de tRNA's em comparação com o primeiro conjunto de tRNA's é tipicamente selecionada, produzindo desta forma uma amino-acil-tRNA sintetase ortogonal eficiente (otimizada) para uso com o O-tRNA. Uma O-RS identificada pelo método também é uma característica da  
15 invenção.

Qualquer um entre inúmeros ensaios pode ser usado para determinar a amino-acilação. Estes ensaios podem ser realizados *in vitro* ou *in vivo*. Por exemplo, os ensaios de amino-acilação *in vitro* estão descritos, por exemplo, em  
20 Hoben e Soll, *Methods Enzymol.* 113:55-59 (1985). A amino-acilação pode ser determinada também usando um repórter junto com componentes ortogonais de tradução e detectar o repórter em uma célula que expressa um polinucleotídeo que compreende pelo menos um códon seletor que codifica uma  
25 proteína. Vide também os documentos n<sup>os</sup> WO 2002/085923, intitulado "In Vivo Incorporation of Unnatural Amino Acids", e WO 2004/094593, intitulado "Expanding the Eukariotic Genetic Code".

A O-RS identificada pode ser manipulada ainda mais para alterar a especificidade da sintetase pelo substrato, de tal modo que apenas um aminoácido não-natural desejado, mas não qualquer um dos 20 aminoácidos comuns, seja  
5 carregado no O-tRNA. Os métodos para gerar uma amino-acil-tRNA sintetase ortogonal com uma especificidade pelo substrato para um aminoácido não-natural incluem mutar a sintetase, por exemplo, no sítio ativo na sintetase, no sítio do mecanismo de alteração na sintetase, em sítios  
10 diferentes, combinando diferentes domínios de sintetases, ou similares, e aplicando um processo de seleção. Usa-se uma estratégia baseada na combinação de uma seleção positiva, e em seguida, uma seleção negativa. Na seleção positiva, a supressão do códon seletor introduzido em uma posição (ou  
15 posições) não-essencial de um marcador positivo permite que as células sobrevivam sob pressão de seleção positiva. Na presença de aminoácidos naturais e também não-naturais, as sobreviventes codificam assim sintetases ativas que carregam o tRNA supressor ortogonal com um aminoácido natural ou não-  
20 natural. Na seleção negativa, a supressão de um códon seletor introduzido em uma posição (ou posições) não-essencial de um marcador negativo remove sintetases com especificidades por aminoácidos naturais. As sobreviventes da seleção negativa e positiva codificam sintetases que amino-  
25 acilam (carregam) o tRNA ortogonal supressor apenas com aminoácidos não-naturais. Estas sintetases podem ser então submetidas a uma mutagênese adicional, por exemplo, embaralhamento de DNA ou outros métodos de mutagênese

recursivos.

Uma biblioteca de O-RS's mutantes pode ser gerada usando várias técnicas de mutagênese conhecidas nessa área. Por exemplo, as RS's mutantes podem ser geradas por mutações  
5 específicas de sítios, mutações pontuais aleatórias, recombinação homóloga, embaralhamento de DNA ou outros métodos de mutagênese recursivos, construção quimérica ou qualquer combinação deles. Por exemplo, uma biblioteca de RS's mutantes pode ser produzida a partir de duas ou mais  
10 outras "subbibliotecas" menos diversas, por exemplo, menores. Bibliotecas quiméricas de RS's também estão incluídas na invenção. Deve-se assinalar que as bibliotecas de tRNA sintetase de vários organismos (por exemplo, microorganismos tais como eubactérias ou archaeobactérias)  
15 tais como bibliotecas que compreendem diversidade natural (vide, por exemplo, patentes nºs US 6.238.884, expedida para Short et al.; 5.756.316, expedida para Schallenberger et al.; 5.783.431, expedida para Petersen et al.; 5.824.485, expedida para Thompson et al.; 5.958.672, expedida para  
20 Short et al.), são opcionalmente construídas e triadas quanto a pares ortogonais.

Depois que as sintetase são submetidas à estratégia de seleção/triagem positiva e negativa, estas sintetases podem ser então submetidas a uma mutagênese  
25 adicional. Por exemplo, um ácido nucléico que codifica a O-RS pode ser isolado; um conjunto de polinucleotídeos que codificam as O-RS's mutadas (por exemplo, por mutagênese aleatória, mutagênese específica de sítio, recombinação ou

qualquer combinação deles) pode ser gerado a partir do ácido nucléico; e estas etapas individuais ou uma combinação destas etapas podem ser repetidas até que seja obtida uma O-RS mutada que preferencialmente amino-acila o O-tRNA com o aminoácido não-natural. Em alguns aspectos da invenção, as etapas são realizadas múltiplas vezes, por exemplo, pelo menos duas vezes.

Níveis adicionais de severidade da seleção/triagem também podem ser usados nos métodos da invenção, para produzir O-tRNA, O-RS ou pares deles. A severidade da seleção ou triagem pode ser variada em uma ou ambas etapas do método, para produzir uma O-RS. Isto incluiria, 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. Selecionar ou triar pode compreender também uma ou mais mudanças na permeabilidade de aminoácidos, uma mudança na eficiência da tradução, uma mudança na fidelidade da tradução, etc. Tipicamente, esta uma ou mais mudanças baseiam-se em uma mutação em um ou mais genes em um organismo no qual um par tRNA-tRNA sintetase ortogonal é usado para produzir a proteína.

Detalhes genéricos adicionais para produzir O-RS, e alterar a especificidade para o substrato da sintetase podem ser encontrados nas publicações internacionais n<sup>os</sup> WO 2002/086075, intitulada "Methods and Compositions for the Production of Orthogonal tRNA Aminoacyl-tRNA Synthetase Pairs", e WO 2004/094593, intitulada "Expanding the

Eukariotic Genetic Code". Vide também Wang e Schultz, "Expanding the Genetic Code", *Angewandte Chemie Int. Ed.* 44(1):34-66 (2005), cujos teores são aqui incorporados como referência em sua totalidade.

## 5                    Origem e Organismos Hospedeiros

Os componentes ortogonais da tradução (O-tRNA e O-RS) que encontram uso com a invenção pode ser derivados de qualquer organismo (ou uma combinação de organismos) para uso em um sistema de tradução hospedeiro a partir de  
10 quaisquer outras espécies, com a advertência de que os componentes O-tRNA/O-RS e os sistema hospedeiro funcionem de uma maneira ortogonal. Não é um requisito que o O-tRNA e a O-RS de um par ortogonal sejam derivados do mesmo organismo. Em alguns aspectos, os componentes ortogonais são derivados  
15 de genes de *Archaea* (isto é, archaeobactérias) para uso em um sistema hospedeiro eubacteriano.

Por exemplo, o O-tRNA ortogonal pode ser derivado de um organismo *Archaea*, por exemplo, uma archaeobactéria, tal como *Methanococcus jannaschii*, *Methanobacterium*  
20 *thermoautotrophicum*, *Halobacterium* tal como *Haloferx volcanii* e espécie *Halobacterium* NRC-1, *Archaeoglobus fulgidus*, *Pyrococcus furiosus*, *Pyrococcus horikoshii*, *Aeuropyrum pernix*, *Methanococcus maripaludis*, *Methanopyrus kandleri*, *Methanosarcina mazei* (Mm), *Pyrobaculum aerophilum*,  
25 *Pyrococcus abyssi*, *Sulfolobus solfataricus* (Ss), *Sulfolobus tokodaii*, *Thermoplasma acidophilum*, *Thermoplasma volcanium*, ou similares, ou uma eubactéria, tal como *Escherichia coli*, *Thermus thermophilus*, *Bacillus stearothermophilus*, ou



similares, enquanto que a O-RS ortogonal pode ser derivada de um organismo ou combinação de organismos, por exemplo, uma archaeobactéria, tal como *Methanococcus jannaschii*, *Methanobacterium thermoautotrophicum*, *Halobacterium* tal como

5 *Haloferx volcanii* e espécie *Halobacterium* NRC-1, *Archaeoglobus fulgidus*, *Pyrococcus furiosus*, *Pyrococcus horikoshii*, *Aeuropyrum pernix*, *Methanococcus maripaludis*, *Methanopyrus kandleri*, *Methanosarcina mazei*, *Pyrobaculum aerophilum*, *Pyrococcus abyssi*, *Sulfolobus solfataricus*,

10 *Sulfolobus tokodaii*, *Thermoplasma acidophilum*, *Thermoplasma volcanium*, ou similares, ou uma eubactéria, tal como *Escherichia coli*, *Thermus thermophilus*, *Bacillus stearothermophilus*, ou similares. Em uma modalidade, origens eucarióticas, por exemplo, plantas, algas, protistas,

15 fungos, leveduras, animais (por exemplo, mamíferos, insetos, artrópodes, etc.), ou similares, também podem ser usados como fontes de O-tRNA's e O-RS's.

Os componentes individuais de um par O-tRNA/O-RS poder ser derivado a partir do mesmo organismo ou organismos

20 diferentes. Em uma modalidade, o par O-tRNA/O-RS é do mesmo organismo. Alternativamente, o tRNA e a O-RS do par O-tRNA/O-RS são de organismos diferentes.

O O-tRNA, O-RS ou o par O-tRNA/O-RS podem ser selecionados ou triados *in vivo* ou *in vitro* e/ou usados em

25 uma célula, por exemplo, uma célula eubacteriana, para produzir um polipeptídeo com um aminoácido não-natural. A célula eubacteriana usada não está limitada, por exemplo, *Escherichia coli*, *Thermus thermophilus*, *Bacillus*

*stearothermophilus*, ou similares. As composições de células eubacterianas, que compreendem componentes da tradução, também são uma característica da invenção.

Vide também o a publicação do pedido de patente internacional nº WO 2004/094593, intitulada "Expanding the Eukariotic Genetic Code", depositado em 16 de abril de 2004, para obter detalhes sobre a triagem de O-tRNA e/ou O-RS em alguma espécie para uso em outra espécie.

Em alguns aspectos, o O-tRNA, O-RS ou o par O-t/RNA/O-RS podem ser selecionados ou triados *in vivo* ou *in vitro* e/ou usados em uma célula, por exemplo, uma célula eucariótica, para produzir um polipeptídeo com um aminoácido não-natural. A célula eucariótica não está limitada; por exemplo, qualquer célula de levedura apropriada, tal como *Saccharomyces cerevisiae* (*S. cerevisiae*) ou similares, pode ser usada. As composições de células eucarióticas, que compreendem componentes da tradução da invenção, também são uma característica da invenção.

Embora os sistemas de tradução ortogonais (por exemplo, que compreendem uma O-RS, um O-tRNA e um aminoácido não-natural) possam utilizar células hospedeiras cultivadas para produzir proteínas que têm aminoácidos não-naturais, não se pretende que um sistema de tradução ortogonal da invenção requeira uma célula hospedeira intacta viável. Por exemplo, um sistema de tradução ortogonal pode utilizar um sistema isento de células, na presença de um extrato celular. Na realidade, o uso de sistemas de transcrição/tradução isentos de células, *in vitro*, para a

produção de proteínas, é uma técnica bem estabelecida. A adaptação destes sistemas *in vitro* para produzir proteínas que têm aminoácidos não-naturais, usando os componentes ortogonais de sistemas de tradução aqui descritos, está dentro do âmbito da invenção.

#### Códons Seletores

Os códons seletores em sistemas de tradução ortogonal expandem a estruturação genética de códons da maquinaria biossintética de proteínas. Por exemplo, um códon seletor inclui, por exemplo, um códon singular com três bases, um códon *nonsense*, tal como um códon finalizador, por exemplo, um códon âmbar (UAG), ou um códon opalino (UGA), um códon não-natural, pelo menos um códon de quatro bases, um códon raro, ou similares. Inúmeros códons seletores podem ser introduzidos em um gene desejado, por exemplo, um ou mais, dois ou mais, mais do que três, etc. Usando diferentes códons seletores, podem ser utilizados múltiplos pares de tRNA/sintetase ortogonais que permitem a incorporação simultânea específica de sítio de múltiplos aminoácidos não-naturais, por exemplo, que incluem pelo menos um aminoácido não-natural, usando estes diferentes códons seletores.

Em uma modalidade, os métodos envolvem o uso de um códon seletor que é um códon finalizador para a incorporação de um aminoácido não-natural *in vivo* em uma célula para dentro de um polipeptídeo. Por exemplo, é produzido um O-tRNA que reconhece o códon finalizador e é amino-acilado por uma O-RS com um aminoácido não-natural. Este O-tRNA não é reconhecido pelas amino-acil-tRNA sintetases de ocorrência

natural da hospedeira. A mutagênese convencional direcionada para sítios pode ser usada para introduzir o códon finalizador no sítio de interesse em um polinucleotídeo que codifica um polipeptídeo de interesse. Vide, por exemplo, 5 Sayers et al., "5',3' Exonuclease in Phosphorothioate-based Oligonucleotide-directed Mutagenesis", *Nucleic Acids Res.* 791-802 (1988). Quando a O-RS, o O-tRNA e o ácido nucléico que codificam um polipeptídeo de interesse são combinados, por exemplo, *in vivo*, o aminoácido não-natural é incorporado 10 em resposta ao códon finalizador, para dar um polipeptídeo que contém o aminoácido não-natural na posição especificada. Em uma modalidade da invenção, o códon finalizador usado como um códon seletor é um códon âmbar, UAG e/ou um códon opalino, UGA. Em um exemplo, um código genético no qual UAG 15 e UGA são ambos usados como um códon seletor pode codificar 22 aminoácidos, e ao mesmo tempo, preservando o códon nonsense ocre, UAA, que é o sinal de terminação mais abundante.

A incorporação de aminoácidos não-naturais *in vivo* 20 pode ser feita sem perturbação significativa da célula hospedeira. Por exemplo, em células não-eucarióticas, tais como *Escherichia coli*, devido ao fato de que a eficiência da supressão para o códon UAG depende da competição entre o O-tRNA, por exemplo, o tRNA supressor âmbar, e o fator de 25 liberação 1 (RF1) (que se liga ao códon UAG e inicia a liberação do peptídeo em crescimento do ribossoma), a eficiência da supressão pode ser modulada, por exemplo, aumentando o nível de expressão de O-tRNA, por exemplo, o

tRNA supressor âmbar, e um fator de liberação eucariótico (por exemplo, eRF) (que se liga a um códon finalizador e inicia a liberação do peptídeo em crescimento do ribossoma), a eficiência da supressão pode ser modulada, por exemplo, aumentando o nível de expressão de O-tRNA, por exemplo, o tRNA supressor. Além disso, compostos adicionais podem estar presentes, por exemplo, agentes redutores, tal como ditionitrositol (DTT).

Os aminoácidos não-naturais podem ser codificados também com códons raros. Por exemplo, quando a concentração de arginina em uma reação de síntese de proteína *in vitro* é reduzida, o códon raro de arginina, AGG, demonstrou ser eficiente para inserção de Ala por um tRNA sintético acilado com alanina. Vide, por exemplo, Ma *et al.*, *Biochemistry* 32:7939 (1993). Neste caso, o tRNA sintético compete com o tRNA<sup>Arg</sup> de ocorrência natural, que existe como uma espécie minoritária em *Escherichia coli*. Além disso, alguns organismos não usam todos os códons triplete. Um códon AGA não designado em *Micrococcus luteus* foi utilizado para a inserção de aminoácidos em um extrato de transcrição/tradução *in vitro*. Vide, por exemplo, Kowal e Oliver, *Nucleic Acids Res.* 25:4685 (1997). Os componentes da invenção podem ser gerados para usar estes códons raros *in vivo*.

Os códons seletores podem compreender também códons estendidos, por exemplo, códons com quatro ou mais bases, tais como códons com quatro, cinco, seis ou mais bases. Os exemplos de códons com quatro bases incluem, por

exemplo, AGGAC, CCCCUC, CCCUC, CUAGA, CUACU, UAGGC, e similares. Os métodos da invenção incluem usar códons estendidos baseado na supressão de deslocamento de estrutura. Códons com quatro ou mais bases podem inserir, 5 por exemplo, um ou múltiplos aminoácidos, na mesma proteína. Em outras modalidades, as alças anticódon pode decodificar, por exemplo, pelo menos um códon com quatro bases, pelo menos um códon com cinco bases, ou pelo menos um códon com seis bases ou mais. Como existem 256 códons com quatro bases 10 possíveis, múltiplos aminoácidos não-naturais podem ser codificados na mesma célula usando um códon com quatro ou mais bases. Vide também Anderson et al., "Exploring the Limits of Codon and Anticodon Size", *Chemistry and Biology* 9:237-244 (2002); e Magliery, "Expanding the Genetic code: 15 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 (2001).

Por exemplo, códons com quatro bases foram usados 20 para incorporar aminoácidos não-naturais em proteínas, usando métodos biossintéticos *in vitro*. Vide, por exemplo, Ma et al., *Biochemistry* 32:7939 (1993); e Hohsaka et al., *J. Am. Chem. Soc.* 121:34 (1999). CGGG e AGGU foram usados para incorporar simultaneamente 2-naftil-alanina e um derivado de 25 NBD de lisina em estreptavidina *in vitro* com dois tRNA's supresores com deslocamento de estrutura, quimicamente acilados. Vide, por exemplo, Hohsaka et al., *J. Am. Chem. Soc.* 121:12194 (1999). Em um estudo *in vivo*, Moore et al.

examinaram a capacidade de derivados de tRNA<sup>Leu</sup> com anticódons NCUA suprimirem códons UAGN (N pode ser U, A, G ou C), e descobriram que o qudruplete UAGA pode ser decodificado por um tRNA<sup>Leu</sup> com um anticódon UCUA com uma  
5 eficiência de 13 a 26% com pouca decodificação na estrutura 0 ou -1. Vide Moore et al., *J. Mol. Biol.* 298:195 (2000). Em uma modalidade, podem ser usados na invenção os códons estendidos baseados em códons raros ou códons *nonsense* que podem reduzir a leitura *missense* e a supressão com  
10 deslocamento de estrutura em outros sítios indesejados. Os códons com quatro bases foram usados como códons seletores em uma série de sistemas ortogonais. Vide, por exemplo, documentos n<sup>os</sup> WO 2005/019415; WO 2005/007870 e WO 2005/07624. Vide também Wang e Schultz, "Expanding the  
15 Genetic Code", *Angewandte Chemie Int. Ed.* 44(1):34-66 (2005), cujo teor é aqui incorporado como referência em sua totalidade. Embora os exemplos abaixo utilizem um códon seletor âmbar, códons com quatro ou mais bases também podem ser usados, modificando os exemplos em questão para incluir  
20 O-tRNA's com quatro bases e sintetases modificadas para incluir mutações similares àquelas descritas anteriormente para várias O-RS's com aminoácidos não-naturais.

Para um dado sistema, um códon seletor pode incluir também um dos códons naturais com três bases, onde o  
25 sistema endógeno não usa (ou usa raramente) o códon com bases naturais. Por exemplo, isto inclui um sistema que não tem um tRNA que reconhece o códon com três bases naturais e/ou um sistema no qual o códon com três bases é um códon

raro.

Os códons seletores incluem opcionalmente pares com bases não-naturais. Estes pares de bases não-naturais expandem ainda mais o alfabeto genético existente. Um par de bases extra aumenta o número de códons triplete de 64 para 125. As propriedades de pares com terceira base incluem pareamento de bases estável e seletivo, incorporação enzimática eficiente no DNA com alta fidelidade por uma polimerase, e extensão eficiente continuada do iniciador depois da síntese do par de bases não-naturais nascente. As descrições de pares de bases não-naturais que podem ser adaptados para os métodos e composições incluem, por exemplo, Hirao et al., "An Unnatural Base Pair for Incorporating Amino Acid Analogues into Protein", *Nature Biotechnology* 20:177-182 (2002). Vide também Wu, Y. et al., *J. Am. Chem. Soc.* 124:14626-14630 (2002). Outras publicações relevantes estão listadas abaixo.

Para uso *in vivo*, o nucleosídeo não-natural é permeável à membrana e é fosforilado para formar o trifosfato correspondente. Além disso, as informações genéticas aumentadas são estáveis e não destruídas por enzimas celulares. Esforços anteriores por Benner e outros se beneficiaram de padrões de ligações hidrogênio que são diferentes daqueles em pares canônicos de Watson-Crick, cujo exemplo mais digno de nota é o par iso-C:iso-G. Vide, por exemplo, Switzer et al., *J. Am. Chem. Soc.* 111:8322 (1989); e Piccirilli et al., *Nature* 343:33 (1990); Kool, *Curr. Opin. Chem. Biol.* 4:602 (2000). Estas bases em geral



desemparelham em algum grau com as bases naturais e não podem ser replicadas enzimaticamente. Kool e colaboradores demonstraram que as interações hidrofóbicas de sobreposição entre bases pode substituir as ligações hidrogênio para  
5 impulsionar a formação de um par de bases. Vide Kool, *Curr.. Opin. Chem. Biol.* 4:602 (2000); e Guckian e Kool, *Angew. Chem. Int. Ed. Engl.* 36:2825 (1998). Em um esforço para desenvolver um par de bases não-naturais que satisfaz a todos os requisitos acima, Schultz, Romesberg e  
10 colaboradores sintetizaram sistematicamente e estudaram uma série de bases hidrofóbicas não-naturais. Descobriu-se que um auto-par PICS:PICS é mais estável do que pares de bases naturais, e pode ser incorporado eficientemente no DNA por um fragmento de Klenow da DNA polimerase I de *Escherichia coli*. Vide, por exemplo, McMinn et al., *J. Am. Chem. Soc.*  
15 121:11586 (1999); e Ogawa et al., *J. Am. Chem. Soc.* 122:3274 (2000). Um auto-par 3MN:3MN pode ser sintetizado por KF com eficiência e seletividade suficientes para função biológica. Vide, por exemplo, Ogawa et al., *J. Am. Chem. Soc.* 122:8803  
20 (2000). Entretanto, ambas bases atuam como um terminador de cadeia para replicação adicional. Recentemente, foi desenvolvida uma DNA polimerase mutante que pode ser usada para replicar o auto-par PICS. Além disso, um auto-par 7<sup>a</sup>1 pode ser replicado. Vide, por exemplo, Tae et al., *J. Am.*  
25 *Chem. Soc.* 123:7439 (2001). Foi desenvolvido também um par de metalobases inusitado, Dipic:Py, que forma um par estável depois de se ligar a Cu(II). Vide Meggers et al., *J. Am. Chem. Soc.* 122:10714 (2000). Devido ao fato de que os códons

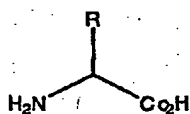
estendidos e os códons não-naturais são intrinsecamente ortogonais aos códons naturais, os métodos da invenção podem se beneficiar desta propriedade para gerar tRNA's ortogonais para eles.

5 Um sistema de desvio da tradução também pode ser usado para incorporar um aminoácido não-natural em um polipeptídeo desejado. Em um sistema de desvio da tradução, uma seqüência grande é inserida em um gene, mas não é traduzida em proteína. A seqüência contém uma estrutura que  
10 serve como sugestão para induzir um ribossoma a saltar a seqüência e retomar a tradução a jusante da inserção.

#### Aminoácidos Não-naturais

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

I



Um aminoácido não-natural é tipicamente qualquer

estrutura que tem a Fórmula I, onde o grupo R é qualquer substituinte que não um usado nos vinte aminoácidos naturais. Vide, por exemplo, "Biochemistry" por L. Stryer, 3ª edição, 1988, Freeman and Company, New York, para obter  
5 as estruturas dos vinte aminoácidos naturais. Deve-se assinalar que os aminoácidos não-naturais da invenção podem ser compostos de ocorrência natural diferentes dos vinte alfa-aminoácidos acima.

Devido ao fato de que os aminoácidos não-naturais  
10 da invenção diferem tipicamente dos aminoácidos naturais na cadeia lateral, os aminoácidos não-naturais formam ligações amida com outros aminoácidos, por exemplo, naturais ou não-naturais, da mesma maneira pela qual são formados em proteínas de ocorrência natural. Entretanto, os aminoácidos  
15 não-naturais têm grupos da cadeia lateral que os distinguem dos aminoácidos naturais.

A Figura 1 fornece as estruturas de aminoácidos não-naturais que são usados nos exemplos de trabalho da presente invenção. Estes aminoácidos não-naturais podem ser  
20 incorporados em proteínas, usando pares de O-RS e O-tRNA apropriados. Por exemplo, p-benzoil-L-fenilalanina (Bpa) pode ser incorporada usando um par de tradução ortogonal que compreende o O-tRNA de SEQ ID NO: 1 e a O-RS cognata de SEQ ID NO: 4. para-acetil-L-fenilalanina (pAcPhe) pode ser  
25 incorporada usando um par de tradução ortogonal que compreende o O-tRNA de SEQ ID NO: 1 e a O-RS cognata de SEQ ID NO: 6. para-azido-L-fenilalanina (pAzPhe) pode ser incorporada usando um par de tradução ortogonal que

compreende o O-tRNA de SEQ ID NO: 1 e a O-RS cognata de SEQ ID NO : 8. para-iodo-L-fenilalanina (pIPhe) pode ser incorporada usando um par de tradução ortogonal que compreende o O-tRNA de SEQ ID NO: 1 e a O-RS cognata de SEQ ID NO: 10.

Entretanto, os aminoácidos não-naturais aqui usados servem apenas para ilustrar a aplicabilidade mais ampla da invenção, e a invenção não está limitada ao uso destes aminoácidos ilustrados na Figura 1.

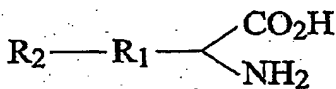
Uma pluralidade de diferentes aminoácidos não-naturais pode ser incorporada simultaneamente em um polipeptídeo de interesse, usando, por exemplo, um segundo par apropriado de O-RS/O-tRNA em conjunto com o primeiro par ortogonal, e onde o primeiro e o segundo pares ortogonais usam códons seletores diferentes.

Em outros aminoácidos não-naturais, por exemplo, R na Fórmula I compreende opcionalmente alquil-, aril-, acil-, hidrazina, ciano-, halo-, hidrazida, alquenila, éter, borato, boronato, fosfo, fosfono, fosfina, enona, imina, éster, hidroxilamina, amina e similares, ou qualquer combinação deles. Outros aminoácidos não-naturais de interesse incluem, porém sem limitações, aminoácidos que compreendem um reticulante fotoativável, aminoácidos marcados com *spin*, aminoácidos fluorescentes, aminoácidos que se ligam a metais, aminoácidos que contêm metais, aminoácidos radioativos, aminoácidos com grupos funcionais inusitados, aminoácidos que interagem de forma covalente ou não-covalente com outras moléculas, aminoácidos *photocaged*

e/ou fotoisomeráveis, aminoácidos que contêm biotina ou análogos de biotina, aminoácidos que contêm ceto, aminoácidos glicosilados, um grupamento sacarídeo anexado à cadeia lateral do aminoácido, aminoácidos que compreendem  
 5 polietilenoglicol ou poliéter, aminoácidos substituídos com átomos pesados, aminoácidos quimicamente cliváveis ou fotocliváveis, aminoácidos com uma cadeia lateral alongada em comparação com aminoácidos naturais (por exemplo, poliéteres ou hidrocarbonetos de cadeia longa, por exemplo,  
 10 maior do que cerca de 5, maior do que cerca de 10 carbonos, etc.), aminoácidos que contêm açúcares ligados a carbonos, aminoácidos que contêm amino-tioácidos, e aminoácidos que contêm um ou mais grupamentos tóxicos.

Em outro aspecto, a invenção fornece aminoácidos  
 15 não-naturais que têm a estrutura genérica ilustrada pela Fórmula IV abaixo:

IV

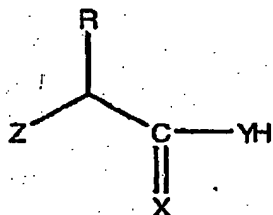


Um aminoácido não-natural que tem esta estrutura é tipicamente qualquer estrutura na qual  $\text{R}_1$  é um substituinte usado em um dos vinte aminoácidos naturais (por exemplo,  
 20 tirosina ou fenilalanina) e  $\text{R}_2$  é um substituinte. Assim sendo, este tipo de aminoácido não-natural pode ser visto como um derivado de aminoácido natural.

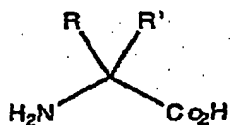
Os aminoácidos não-naturais podem compreender

também opcionalmente estruturas com esqueleto modificado, como ilustrado, por exemplo, pelas estruturas das Fórmulas II e III:

II



III



onde Z compreende tipicamente OH, NH<sub>2</sub>, SH, NH-R',  
 5 ou S-R'; X e Y, que podem ser iguais ou diferentes, compreendem tipicamente S ou O, e R e R', que são opcionalmente iguais ou diferentes, são selecionados tipicamente entre a mesma lista de constituintes para o grupo R descrito acima para os aminoácidos não-naturais que  
 10 têm a Fórmula I, bem como hidrogênio. Por exemplo, os aminoácidos não-naturais da invenção compreendem opcionalmente substituições no grupo amino ou carboxila como ilustrado pelas Fórmulas II e III. Os aminoácidos não-naturais deste tipo incluem, porém sem limitações, α-  
 15 hidróxi-ácidos, α-tioácidos, α-amino-tiocarboxilatos, por exemplo, com cadeias laterais correspondentes aos vinte

aminoácidos naturais comuns ou cadeias laterais não-naturais. Além disso, as substituições no carbono  $\alpha$  incluem opcionalmente L-, D-aminoácidos ou aminoácidos dissustituídos em  $\alpha, \alpha$ , tais como D-glutamato, D-alanina, D-metil-O-tirosina, ácido amino-butírico, e similares. Outras alternativas estruturais incluem aminoácidos cíclicos, tais como análogos de prolina, bem como análogos de prolina com anel de 3, 4, 6, 7, 8 e 9 membros,  $\beta$ - e  $\gamma$ -aminoácidos, tais como  $\beta$ -alanina substituída e ácido  $\gamma$ -amino-butírico.

10 Em alguns aspectos, a invenção utiliza aminoácidos não-naturais na configuração L. Entretanto, não se pretende que a invenção esteja limitada ao uso de aminoácidos não-naturais com configuração L. Contempla-se que os enantiômeros D destes aminoácidos não-naturais também  
15 encontram uso com a invenção.

Os análogos de tirosina incluem tirosinas substituídas em "para", tirosinas substituídas em "orto" e tirosinas substituídas em "meta", onde a tirosina substituída compreende um grupo alquilila, um grupo acetila, um grupo benzoíla, um grupo amino, uma hidrazina, uma hidroxilamina, um grupo tiol, um grupo carboxila, um grupo isopropila, um grupo metila, uma cadeia linear de  $C_6$ - $C_{20}$  ou hidrocarboneto ramificado, um hidrocarboneto saturado ou insaturado, um grupo O-metila, um grupo poliéter, um grupo nitro, ou  
20 similares. Além disso, anéis arila multiplamente substituídos também são contemplados. Os análogos de glutamina da invenção incluem, porém sem limitações, derivados  $\alpha$ -hidróxi, derivados  $\gamma$ -substituídos, derivados  
25

cíclicos, e derivados de glutamina substituídos com amida. Os análogos de fenilalanina exemplificativos incluem, porém sem limitações, fenilalaninas para-substituídas, fenilalaninas orto-substituídas, e fenilalaninas meta-

5 substituídas, onde o substituinte compreende um grupo alquinila, um grupo hidroxila, um grupo metóxi, um grupo metila, um grupo alila, um aldeído, um nitro, um grupo tiol, um grupo ceto, ou similares. Os exemplos específicos de aminoácidos não-naturais incluem, porém sem limitações, p-

10 etil-tiocarbonil-L-fenilalanina, p(3-oxobutanoil)-L-fenilalanina, 1,5-dansil-alanina, ácido 7-amino-cumarínico, ácido 7-hidróxi-cumarínico, nitro-benzil-serina, O-(2-nitro-benzil)-L-tirosina, p-carbóxi-metil-L-fenilalanina, p-ciano-L-fenilalanina, m-ciano-L-fenilalanina, bifenilalanina, 3-

15 amino-L-tirosina, bipyridilalanina, p-(2-amino-1-hidróxi-etil)-L-fenilalanina, p-isopropil-tiocarbonil-L-fenilalanina, 3-nitro-L-tirosina e p-nitro-L-fenilalanina. Além disso, uma p-propargilóxi-fenilalanina, uma 3,4-diidróxi-L-fenilalanina (DHP), uma 3,4,6-triidróxi-L-

20 fenilalanina, uma 3,4,5-triidróxi-L-fenilalanina, 4-nitro-fenilalanina, uma p-acetil-L-fenilalanina, O-metil-L-tirosina, uma L-3-(2-naftil)-alanina, uma 3-metil-fenilalanina, uma O-4-alil-L-tirosina, uma 4-propil-L-tirosina, uma 3-nitro-tirosina, uma 3-tiol-tirosina, uma

25 tri-O-acetil-GlcNAc $\beta$ -serina, uma L-Dopa, um fenilalanina fluorada, uma isopropil-L-fenilalanina, uma p-azido-L-fenilalanina, uma p-acil-L-fenilalanina, uma p-benzoil-L-fenilalanina, uma L-fosfosserina, uma fosfonosserina, uma



fosfonotirosina, uma p-iodo-fenilalanina, uma p-bromo-fenilalanina, uma p-amino-L-fenilalanina, e uma isopropil-L-fenilalanina, e similares. As estruturas de uma série de aminoácidos não-naturais que podem ser incorporados usando sistemas de tradução ortogonal são conhecidas. Vide as referências aqui citadas, sendo cada uma delas aqui incorporada como referência em sua totalidade.

#### Síntese Química de Aminoácidos Não-naturais

Muitos dos aminoácidos não-naturais fornecidos acima estão disponíveis no mercado, por exemplo, na Sigma (EUA) ou Aldrich (Milwaukee, WI, EUA). Aqueles que não estão disponíveis no mercado são opcionalmente sintetizados como descrito em várias publicações ou usando métodos padonizados conhecidos pelos versados nessas técnicas. Para obter detalhes de técnicas de síntese orgânica, vide, por exemplo, "Organic Chemistry", por Fessenden e Fessenden (1982, Segunda Edição, Willard Grant Press, Boston, MA); "Advanced Organic Chemistry" por March (Terceira Edição, 1985, Wiley & Sons, New York); e "Advanced Organic Chemistry" por Carey e Sundberg (Terceira Edição, Partes A e B, 1990, Plenum Press, New York). As publicações adicionais que descrevem a síntese de aminoácidos não-naturais incluem, por exemplo, WO 2002/085923, intitulada "In vivo incorporation of Unnatural Amino Acids"; Matsoukas *et al.*, *J. Med. Chem.* 38:4660-4669 (1995); King e Kidd, "A New Synthesis of Glutamine and of  $\gamma$ -Dipeptides of Glutamic Acid from Phthylated Intermediates", *J. Chem. Soc.* 3315-3319 (1949); Friedman e Chatterrji, "Synthesis of Derivatives of Glutamine as Model Substrates

for Anti-Tumor Agents", *J. Am. Chem. Soc.* 81:3750-3752 (1959); Craig et al., "Absolute Configuration of the Enantiomers of 7-Chloro-4-[[4-(diethylamino)-1-methylbutyl]amino]quinoline (Chloroquine)", *J. Org. Chem.* 53:1167-1170 (1988); Azoulay et al., "Glutamine analogues as Potential Antimalarials", *Eur. J. Med. Chem.* 26:201-5 (1991); Koskinen e Rapoport, "Synthesis of 4-Substituted Prolines as Conformationally Constrained Amino Acid Analogues", *J. Org. Chem.* 54:1859-1866 (1989); Christie e Rapoport, "Synthesis of Optically Pure Pipecolates from L-Asparagine. Application to the Total Synthesis of (+)-Apovincamine through Amino Acid Decarbonylation and Iminium Ion Cyclization", *J. Org. Chem.* 1989:1859-1866 (1985); Barton et al., "Synthesis of Novel  $\alpha$ -Amino-Acids and Derivatives Using Radical Chemistry: Synthesis of L- and D- $\alpha$ -Amino-Adipic Acids, L- $\alpha$ -aminopimelic Acid and Appropriate Unsaturated Derivatives", *Tetrahedron Lett.* 43:4297-4308 (1987); e Subasinghe et al., "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 (1992). Vide também publicação internacional nº WO 2004/058946, intitulada "Protein Arrays", depositada em 22 de dezembro de 2003.

## 25 Absorção de Aminoácidos Não-naturais

A absorção de aminoácidos não-naturais por uma célula é um problema que é tipicamente considerado quando se projeta e seleciona aminoácidos não-naturais, por exemplo,

para incorporação em uma proteína. Por exemplo, a alta densidade de carga de  $\alpha$ -aminoácidos sugere que estes compostos possivelmente não são permeáveis em células. Os aminoácidos naturais são absorvidos por uma célula por intermédio de uma coleção de sistemas de transporte baseados em proteínas, apresentando freqüentemente graus variados de especificidade de aminoácido. Pode ser feita uma rápida triagem que determina quais aminoácidos não-naturais são absorvidos por células. Vide, por exemplo, os ensaios de toxicidade, por exemplo, na publicação internacional nº WO 2004/058946, intitulada "Protein Arrays", depositada em 22 de dezembro de 2003; e Liu e Schultz, "Progress toward the evolution of an organism with an expanded genetic code", *PNAS* 96:4780-4785 (1999). Embora a absorção seja facilmente analisada com vários ensaios, uma alternativa para projetar aminoácidos não-naturais que são suscetíveis às vias de absorção celular é disponibilizar vias biossintéticas para criar aminoácidos *in vivo*.

#### Biossíntese de Aminoácidos Não-naturais

Já existem muitas vias biossintéticas em células para a produção de aminoácidos e outros compostos. Embora um método biossintético para um aminoácido não-natural específico possa não existir na natureza, por exemplo, em uma célula, a invenção fornece esses métodos. Por exemplo, as vias biossintéticas para aminoácidos não-naturais são opcionalmente geradas em uma célula hospedeira, adicionando novas enzimas ou modificando as vias de células hospedeiras existentes. Novas enzimas adicionais são opcionalmente

enzimas de ocorrência natural ou enzimas desenvolvidas artificialmente. Por exemplo, a biossíntese de p-amino-fenilalanina (como apresentada em um exemplo no documento nº WO 2002/085923 supra) baseia-se na adição de uma combinação  
5 de enzimas conhecidas de outros organismos. Os genes para estas enzimas podem ser introduzidos em uma célula transformando a célula com um plasmídeo que compreende os genes. Os genes, quando expressados na célula, fornecem uma via enzimática para sintetizar o composto desejado. Os  
10 exemplos dos tipos de enzimas que são opcionalmente adicionadas estão indicados nos exemplos abaixo. As seqüências de enzimas adicionais são encontradas, por exemplo, no GenBank. As enzimas desenvolvidas artificialmente também são opcionalmente adicionadas a uma  
15 célula da mesma maneira. Desta forma, a maquinaria celular e os recursos de uma célula são manipulados para produzir aminoácidos não-naturais.

Na realidade, qualquer um entre uma série de métodos pode ser usado para produzir enzimas inusitadas para  
20 uso em vias biossintéticas, ou para o desenvolvimento de vias existentes, para a produção de aminoácidos não-naturais, *in vitro* ou *in vivo*. Muitos métodos disponíveis de enzimas em desenvolvimento e outros componentes de vias biossintéticas podem ser aplicados à presente invenção para  
25 produzir aminoácidos não-naturais (ou, na realidade, para desenvolver sintetases para terem novas especificidades de substratos ou outras atividades de interesse). Por exemplo, o embaralhamento de DNA é opcionalmente usado para

desenvolver enzimas inusitadas e/ou vias dessas enzimas para a produção de aminoácidos não-naturais (ou produção de novas sintetases), *in vitro* ou *in vivo*. Vide, por exemplo, Stemmer, "Rapid evolution of a protein *in vitro* by DNA shuffling", *Nature* 370(4):389-391 (1994); e Stemmer, "DNA shuffling by random fragmentation and reassembly: *In vitro* recombination for molecular evolution", *Proc. Natl. Acad. Sci. USA* 91:10747-10751 (1994). Uma abordagem relacionada embaralha famílias de genes relacionados (por exemplo, homólogos) para desenvolver rapidamente as enzimas com características desejadas. Um exemplo desses métodos de "embaralhamento de famílias de genes" é encontrado em Cramer et al., "DNA shuffling of a family of genes from different species accelerates directed evolution", *Nature* 391(6664):288-291 (1998). Novas enzimas (sejam elas componentes de vias biossintéticas ou sintetases) também podem ser geradas usando um procedimento de recombinação de DNA conhecido como "truncamento incremental para a criação de enzimas híbridas" ("ITCHY"), como descrito, por exemplo, em Ostermeier et al., "A combinatorial approach to hybrid enzymes independent of DNA homology", *Nature Biotech.* 17:1205 (1999). Esta abordagem também pode ser usada para gerar uma biblioteca de enzimas ou outras variantes de vias que podem servir como substratos para um ou mais métodos de recombinação *in vitro* ou *in vivo*. Vide também Ostermeier et al., "Combinatorial Protein Engineering by Incremental Truncation", *Proc. Natl. Acad. Sci. USA* 96:3562-67 (1999), e Ostermeier et al., "Incremental Truncation as a Strategy in

the Engineering of Novel Biocatalysts", *Biol. Med. Chem.* 7:2139-44 (1999). Outra abordagem usa mutagênese de conjuntos exponenciais para produzir bibliotecas de enzimas ou outras variantes de vias que são, por exemplo, selecionadas quanto a uma capacidade de catalisar uma reação biossintética relevante para produzir um aminoácido não-natural (ou uma nova sintetase). Nesta abordagem, pequenos grupos de resíduos em uma sequência de interesse são randomizados em paralelo para identificar, em cada posição alterada, aminoácidos que levam a proteínas funcionais. Os exemplos de tais procedimentos, que podem ser adaptados para a presente invenção a fim de produzir novas enzimas para a produção de aminoácidos não-naturais (ou novas sintetases), são encontrados em Delegrave e Youvan, *Biotech. Res.* 11:1548-1552 (1993). Em ainda outra abordagem, uma mutagênese aleatória ou semi-aleatória, usando oligonucleotídeos dopados ou degenerados, para a engenharia de componentes de enzimas e/ou vias, pode ser usada, por exemplo, usando os métodos de mutagênese genéricos, por exemplo, de Arkin e Youvan, "Optimizing nucleotide mixtures to encode specific subsets of amino acids for a semi-random mutagenesis", *Biotechnology* 10:297-300 (1992); ou Reidhaar-Olson et al., "Random mutagenesis of protein sequences using oligonucleotide cassettes", *Methods Enzymol.* 208:564-86 (1991). Ainda outra abordagem, freqüentemente denominada uma mutagênese "não-estocástica", que usa remontagem de polinucleotídeos e mutagênese com saturação em sítio, pode ser usada para produzir enzimas e/ou componentes de vias,

que podem ser triadas quanto a uma capacidade de realizar uma ou mais funções de sintetase ou vias biossintéticas (por exemplo, para a produção de aminoácidos não-naturais *in vivo*). Vide, por exemplo, documento nº WO 2000/046344, intitulado "Non-stochastic Generation of Genetic Vaccines and enzymes", expedido para Short.

Uma alternativa a esses métodos mutacionais envolve recombinar genomas inteiros de organismos e selecionar a progênie resultante quanto a funções de vias específicas (frequentemente referidas como "embaralhamento de genoma inteiro"). Esta abordagem pode ser aplicada à presente invenção, por exemplo, por recombinação genômica e seleção de um organismo (por exemplo, uma célula de *E. coli* ou outra célula) quanto a uma capacidade de produzir um aminoácido não-natural (ou um seu intermediário). Por exemplo, os métodos enunciados nas seguintes publicações podem ser aplicados para projetar vias para a evolução de vias existentes e/ou novas em células para produzir aminoácidos não-naturais *in vivo*: Patnaik et al., "Genome shuffling of lactobacillus for improved acid tolerance", *Nature Biotechnology* 20(7):707-712 (2002); e Zhang et al., "Genome shuffling leads to rapid phenotypic improvement in bacteria", *Nature*, 415(6872):644-646 (7 de fevereiro).

Outras técnicas para a engenharia de organismos e vias metabólicas, por exemplo, para a produção de compostos desejados, também estão disponíveis e também podem ser aplicadas para a produção de aminoácidos não-naturais. Os exemplos de publicações que enunciam abordagens úteis de

engenharia de vias incluem: Nakamura e White, "Metabolic engineering for the microbial production of 1,3-propanediol", *Curr. Opin. Biotechnol.* 14(5):454-9 (2003); Berry et al., "Application of Metabolic Engineering to  
 5 Improve both the Production and Use of Biotech Indigo", *J. Industrial Microbiology and Biotechnology* 28:127-133 (2002); Banta et al., "Optimizing an artificial metabolic pathway: Engineering the cofactor specificity of Corynebacterium 2,5-diketo-D-gluconic acid reductase for use in vitamin C  
 10 biosynthesis", *Biochemistry* 41(20):6226-36 (2002); Selivonova et al., "Rapid Evolution of Novel Traits in Microorganisms", *Applied and Environmental Microbiology* 67:3645 (2001); e muitas outras.

Independente do método usado, tipicamente, o  
 15 aminoácido não-natural, produzido com uma via biossintética de engenharia genética da invenção, é produzido em uma concentração suficiente para a biossíntese eficiente da proteína, por exemplo, uma quantidade celular natural, mas não em um grau que afete significativamente a concentração  
 20 de outros aminoácidos celulares, ou para exaurir recursos celulares. As concentrações típicas produzidas *in vivo* desta maneira são cerca de 10 mM a cerca de 0,05 mM. Depois que a célula é manipulada para produzir enzimas desejadas para uma via específica e um aminoácido não-natural é gerado,  
 25 seleções *in vivo* são opcionalmente usadas para otimizar ainda mais a produção do aminoácido não-natural para a síntese de proteína e crescimento celular.

Componentes Ortogonais que Encontram Uso com a



### Invenção

A incorporação do aminoácido não-natural na proteína é realizada por pares ortogonais que incorporam o aminoácido não-natural em resposta ao sinal genético do códon seletor em *E. coli*, onde os componentes ortogonais não sofrem reação cruzada com componentes endógenos de *E. coli* da maquinaria da tradução da célula hospedeira, mas reconhecem o aminoácido não-natural desejado e o incorporam em proteínas em resposta ao códon seletor (por exemplo, um códon *nonsense* âmbar, TAG). Os componentes ortogonais que encontram uso com a invenção incluem amino-acil-tRNA sintetases ortogonais derivadas de tironil-tRNA-sintetase de *Methanococcus jannaschii*, e o supressor âmbar de tirosil-tRNA<sub>CUA</sub> mutante, que funcionam como um par ortogonal em uma célula hospedeira eubacteriana, tal como *E. coli*. Neste sistema, as amino-acil-tRNA sintetases mutantes amino-acilam o tRNA supressor com seu aminoácido não-natural respectivo e não com qualquer um dos vinte aminoácidos comuns.

Os métodos para produzir componentes ortogonais encontram uso com a invenção, onde estes métodos resultam na incorporação de aminoácidos não-naturais, por exemplo, porém sem limitações, os aminoácidos não-naturais fornecidos na Figura 1, em uma cadeia de polipeptídeo em crescimento, em resposta a um códon seletor, por exemplo, um códon finalizador âmbar, um códon *nonsense*, um códon com quatro ou mais bases, etc., por exemplo, *in vivo*. Por exemplo, os tRNA's ortogonais (O-tRNA's), as amino-acil-tRNA sintetases ortogonais (O-RS's) e seus pares encontram uso com a

invenção.

Em algumas modalidades, esses pares podem ser usados para incorporar um aminoácido não-natural em cadeias de polipeptídeos em crescimento, e subsequente, o polipeptídeo é modificado depois da tradução. Para obter informações adicionais sobre aminoácidos não-naturais que podem ser modificados depois da tradução, vide, por exemplo, os sistemas ortogonais de aminoácidos não-naturais descritos em Chin et al., *Science* 301:964-967 (2003); Zhang et al., *Proc. Natl. Acad. Sci. USA* 101:8882-8887 (2004); Anderson et al., *Proc. Natl. Acad. Sci. USA* 101:7566-7571 (2004); Wang et al., *Science* 292:498-500 (2001); Chin et al., *J. Am. Chem. Soc.* 124:9026-9027; Chun e Schultz, *Chem. Bio. Chem.* 11:1135-1137 (2002); Chin et al., *PNAS USA* 99:11020-11024 (2002); Wang e Schultz, *Chem. Comm.* 1-10 (2002); Wang e Schultz, "Expanding the Genetic Code", *Angewandte Chemie Int. Ed.* 44(1):34-66 (2005); Xie e Schultz, "An Expanding Genetic Code", *Methods* 36:227-238 (2005); e Deiters et al., *Bioorganic & Medicinal Chemistry Letters* 15:1521-1524 (2005), sendo cada um deles incorporado como referência em sua totalidade.

Vide também os sistemas ortogonais de aminoácidos não-naturais descritos nas publicações internacionais n<sup>os</sup> WO 2002/086075, intitulada "Methods and Compositions for the Production of Orthogonal tRNA Aminoacyl-tRNA Synthetase Pairs"; WO 2002/085923, intitulada "In Vivo Incorporation of Unnatural Aminoacids"; WO 2004/094593, intitulada "Expanding the Eukariotic Genetic Code"; WO 2005/019415, depositado em

7 de julho de 2004; WO 2005/007870, depositado em 7 de julho de 2004; WO 2005/007624, depositado em 7 de julho de 2004; WO 2006/034332, depositado em 20 de setembro de 2005; e WO 2006/110182, depositado em 27 de outubro de 2005, intitulada  
5 "Orthogonal Translation Components for the In Vivo Incorporation of Unnatural Aminoacids", em nome de Schultz et al.

Em certas modalidades, a O-RS que encontra uso com a invenção preferencialmente monoacila o O-tRNA em vez de  
10 qualquer tRNA endógeno com um aminoácido não-natural específico, onde a O-RS tem uma inclinação pelo O-tRNA, e onde a razão de O-tRNA carregado com um aminoácido não-natural para o tRNA endógeno carregado com o mesmo aminoácido não-natural é maior do que 1:1, e mais  
15 preferivelmente, onde a O-RS carrega o O-tRNA exclusivamente ou quase que exclusivamente.

A invenção faz uso de tRNA's ortogonais (O-tRNA's), onde o O-tRNA reconhece um códon seletor. Tipicamente, um O-tRNA inclui uma eficiência de supressão de  
20 pelo menos cerca de, por exemplo, 45%, 50%, 60%, 75%, 80% ou 90% na presença de uma sintetase coganata em resposta a um códon seletor, em comparação com a eficiência de supressão de um O-tRNA que compreende ou é codificado por uma sequência de polinucleotídeo enunciada nas listagens de  
25 sequências (por exemplo, SEQ ID NO: 1). Em uma modalidade, a eficiência da supressão da O-RS e do O-tRNA em conjunto é, por exemplo, 5 vezes, 10 vezes, 15 vezes, 20 vezes, 25 vezes ou mais vezes maior do que a eficiência de supressão do O-

tRNA na ausência de uma O-RS. Em alguns aspectos, a eficiência da supressão da O-RS e do O-tRNA em conjunto é pelo menos 45% da eficiência da supressão de um par de tirosil-tRNA sintetase ortogonal derivado de *Methanococcus*  
5 *jannaschii*.

A invenção faz uso de células (por exemplo, *E. coli*) que compreendem um sistema de tradução e seqüências de nucleotídeos que programam a produção de proteínas, onde o sistema de tradução inclui um tRNA orgonal (O-tRNA), uma  
10 amino-acil-tRNA sintetase ortogonal (O-RS), e um aminoácido não-natural. Tipicamente, a O-RS preferencialmente amino-acila o O-tRNA em vez de qualquer tRNA endógeno com o aminoácido não-natural, onde a O-RS tem uma inclinação pelo tRNA, e onde a razão de O-tRNA carregado com o aminoácido  
15 não-natural para o tRNA endógeno carregado com o aminoácido não-natural é maior do que 1:1, e mais preferivelmente, onde a O-RS carrega o O-tRNA exclusivamente ou quase que exclusivamente. O O-tRNA reconhece o primeiro códon seletor, e a O-RS preferencialmente amino-acila o O-tRNA com um  
20 aminoácido não-natural.

Vários polinucleotídeos também encontram uso com a invenção. Estes polinucleotídeos incluem um polinucleotídeo artificial (por exemplo, manfaturado, e de ocorrência não-natural, por exemplo, recombinante) que compreende uma  
25 seqüência de nucleotídeos que codifica uma O-RS. Um polinucleotídeo que encontra uso com a invenção inclui também um ácido nucléico que hibridiza para um polinucleotídeo descrito acima, sob condições altamente

severas, cobrindo substancialmente o comprimento inteiro do ácido nucléico. Os vetores que compreendem polinucleotídeos também encontram uso com a invenção. Por exemplo, um vetor pode incluir um plasmídeo, um cosmídeo, um fago, um viris, um vetor de expressão e/ou similares. Os métodos para produzir componentes de um par O-tRNA/O-RS são conhecidos e encontram uso com a invenção. Vide a presente descrição e a referência aqui citada.

#### Seqüências de Ácidos Nucléicos e Polipeptídeos, e

#### 10 Variantes

Como aqui descrito, as seqüência de polinucleotídeos que codificam, por exemplo, O-tRNA's e O-RS's, encontram uso como a invenção, assim como o fazem as respectivas seqüências de aminoácidos codificadas pelos polinucleotídeos. A descrição fornece exemplos referenciais de seqüências de polinucleotídeos e polipeptídeos que encontram uso com a invenção. Entretanto, deve-se avaliar que o uso da invenção não está limitado às seqüências aqui descritas. Os versados nessas técnicas devem avaliar que a invenção fornece também muitas seqüências relacionadas com as funções aqui descritas, por exemplo, polinucleotídeos e polipeptídeos que codificam variantes conservativas de uma O-RS aqui descrita.

Um polinucleotídeo que encontra uso com a invenção inclui também um polinucleotídeo artificial, isto é, por exemplo, pelo menos 75%, pelo menos 80%, pelo menos 90%, pelo menos 95%, pelo menos 98% ou mais, idêntico àquele de um tRNA de ocorrência natural (mas é diferente de um tRNA de

ocorrência natural). Um polinucleotídeo que encontra uso com a invenção inclui também um polinucleotídeo artificial, isto é, por exemplo, pelo menos 75%, pelo menos 80%, pelo menos 90%, pelo menos 95%, pelo menos 98% ou mais, idêntico (porém  
5 não 100% idêntico) àquele de um tRNA de ocorrência natural.

Em certas modalidades, um vetor que encontra uso com a invenção (por exemplo, um plasmídeo, um cosmídeo, um fago, um vírus, etc.) compreende um polinucleotídeo que encontra uso com a invenção. Em algumas modalidades, o vetor  
10 é um vetor de expressão. Em outras modalidades, o vetor de expressão inclui um promotor ligado operacionalmente a um ou mais dos polinucleotídeos da invenção. Em outras modalidades, uma célula compreende um vetor que inclui um polinucleotídeo que encontra uso com a invenção.

15 Os versados nessas técnicas devem avaliar que muitas variantes das seqüências descritas também encontram uso com a invenção. Por exemplo, as variações conservativas das seqüências descritas que produzem uma seqüência funcionalmente idêntica encontram uso com a invenção. As  
20 variantes das seqüências de polinucleotídeos de ácidos nucléicos, onde as variantes hibridizam para pelo menos uma seqüência descrita, encontram uso com a invenção.

#### Variações Conservativas

Devido à degeneração do código genético,  
25 "substituições silenciosas" (isto é, substituições em uma seqüência de ácido nucléico, que não resultam em uma alteração em um polipeptídeo codificado) são uma característica inferida de qualquer seqüência de ácido

nucléico que codifica uma seqüência de aminoácidos. Similarmente, as "substituições conservativas de aminoácidos", onde um aminoácido ou um número limitado de aminoácidos em uma seqüência de aminoácidos é substituído por aminoácidos diferentes com propriedades altamente similares, também são failmente identificadas como sendo altamente similares a uma construção descrita. Tais variações conservativas de cada seqüência descrita são uma característica da presente invenção.

10           As "variações conservativas" de uma seqüência de ácidos nucléicos específica referem-se aos ácidos nucléicos que codificam seqüências de aminoácidos idênticas ou essencialmente idênticas, ou onde o ácido nucléico não codifica uma seqüência de aminoácidos, a seqüências  
15           essencialmente idênticas. Os versados nessas técnicas devem avaliar que substituições, deleções, ou adições individuais que alteram, adicionam ou deletam um único aminoácido ou uma porcentagem pequena de aminoácidos (tipicamente menos do que 5%, mais tipicamente menos do que 4%, 2% ou 1%) em uma  
20           seqüência codificada são "variações conservativamente modificadas", onde as alterações resultam na deleção de um aminoácido, adição de um aminoácido, ou substituição de um aminoácido por um aminoácido quimicamente similar. Assim sendo, as "variações conservativas" de uma seqüência de  
25           polipeptídeo listada da presente invenção incluem substituições de uma pequena porcentagem, tipicamente menos do que 5%, mais tipicamente menos do que 2% ou 1% dos aminoácidos da seqüência de polipeptídeo, por um aminoácido

do mesmo grupo de substituições conservativas. Finalmente, a  
adição de seqüências que não alteram a atividade codificada  
de uma molécula de ácido nucléico, tal como a adição de uma  
seqüência não-funcional, é uma variação conservativa do  
5 ácido nucléico básico.

As tabelas de substituições conservativas que  
produzem aminoácidos funcionalmente similares são bem  
conhecidas nessas técnicas, onde um resíduo de aminoácido é  
substituído por outro resíduo de aminoácido que tem  
10 propriedades químicas similares (por exemplo, cadeias  
laterais aromáticas ou cadeias laterais positivamente  
carregadas), e portanto, não mudam substancialmente as  
propriedades funcionais da molécula de polipeptídeo. O exto  
que se segue enuncia grupos que contêm aminoácidos naturais  
15 com propriedades químicas semelhantes, onde uma substituição  
dentro de um grupo é uma "substituição conservativa".

TABELA 1

## Substituições Conservativas de Aminoácidos

Cadeias Laterais Apolares e/ou Alifáticas	Cadeias Laterais Polares, Não- carregadas	Cadeias Laterais Aromáticas	Cadeias Laterais Positivamente Carregadas	Cadeias Laterais Negativamente Carregadas
Glicina	Serina			
Alanina	Treonina			
Valina	Cisteína	Fenilalanina	Lisina	Aspartato
Leucina	Metionina	Tirosina	Arginina	Glutamato



Isoleucina	Asparagina	Triptofano	Histidina	
Prolina	Glutamina			

### Hibridização de Ácidos Nucléicos

Uma hibridização comparativa pode ser usada para identificar ácidos nucleicos que encontram uso com a invenção, incluindo as variações conservativas de ácidos nucleicos aqui fornecidas, e este método de hibridização comparativa é um método preferido para distinguir ácidos nucleicos que encontram uso com a invenção. Os ácidos nucleicos-alvo, que hibridizam para ácidos nucleicos aqui fornecidos ou referenciados sob condições de alta, ultra-alta e ultra-ultra-alta severidade também encontram uso com a invenção. Os exemplos de tais ácidos nucleicos incluem aqueles com uma ou poucas substituições conservativas de ácidos nucleicos, em comparação com uma dada seqüência de ácidos nucleicos.

Um ácido nucleico em teste é dito que hibridiza especificamente para um ácido nucleico sonda quando ele hibridiza pelo menos 50% tão bem para como para o alvo complementar perfeitamente emparelhado, isto é, com uma razão de sinal para interferência pelo menos metade tão alta quanto a hibridização da sonda para o alvo sob condições nas quais a sonda perfeitamente emparelhada se liga ao alvo complementar perfeitamente emparelhado com uma razão de sinal para interferência que é pelo menos cerca de 5 vezes a 10 vezes tão alta quanto aquela observada para a hibridização para qualquer um dos ácidos nucleicos-alvo não-emparelhados.

Os ácidos nucléicos "hibridizam" quando eles se associam, tipicamente em solução. Os ácidos nucléicos hibridizam devido a uma série de forças físico-químicas bem caracterizadas, tais como ligação hidrogênio, exclusão por solvente, empilhamento de bases, e similares. Uma orientação extensiva sobre a hibridização de ácidos nucléicos é encontrada 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, New York), bem como em "Current Protocols in Molecular Biology", Ausubel et al., editores, "Current Protocols", uma joint venture entre Greene Publishing Associates, Inc. e John Wiley & Sons, Inc. (suplementado até 2004) ("Ausubel"); Hames e Higgins (1995), "Gene Probes 1", IRL Press at Oxford University Press, Oxford, Inglaterra (Hames e Hughes 1); e Hames e Higgins (1995), "Gene Probes 2", IRL Press at Oxford University Press, Oxford, Inglaterra (Hames e Hughes 2); fornecem detalhes sobre a síntese, marcação, detecção e quantificação de DNA e RNA, incluindo oligonucleotídeos.

Um exemplo de condições severas de hibridização para a hibridização de ácidos nucléicos complementares que têm mais do que 100 resíduos complementares em um filtro, em uma análise *Southern* ou *Northern blot* é 50% formol com 1 mg de heparina a 42 °C, sendo a hibridização conduzida de um dia para o outro. Um exemplos de condições severas de lavagem é uma lavagem com 0,2 x SSC a 65 °C por 15 min (vide

Sambrook supra para obter uma descrição de tampão SSC). Frequentemente, a lavagem de alta severidade é precedida de uma lavagem de baixa severidade para remover o sinal de fundo da sonda. Um exemplo de lavagem de baixa severidade é 2 x SSC a 40 °C por 15 min. Em geral, uma razão de sinal para interferência de 5x (ou mais alta) do que aquela observada para uma sonda não relacionada no ensaio de hibridização específico indica a detecção de uma hibridização específica.

10 "Condições severas de lavagem da hibridização", no contexto de experimentos de hibridização de ácidos nucleicos, tais como hibridizações *Southern* e *Northern*, são dependentes de seqüências, e são diferentes sob parâmetros ambientais diferentes. Uma orientação extensiva sobre a

15 hibridização de ácidos nucleicos é encontrada em Tijssen (1993) supra, e em Hames e Higgins, 1 e 2. As condições severas de hibridização e lavagem podem ser determinadas facilmente de forma empírica para qualquer ácido nucleico em teste. Por exemplo, ao determinar as condições severas de

20 hibridização e lavagem, as condições de hibridização e lavagem são aumentadas gradualmente (por exemplo, aumentando a temperatura, diminuindo a concentração de sais, aumentando a concentração de detergentes e/ou aumentando a concentração de solventes orgânicos tais como formol na hibridização ou

25 lavagem), até que um conjunto selecionado de critérios seja atendido. Por exemplo, em condições de hibridização e lavagem altamente severas, as condições de hibridização e lavagem são aumentadas gradualmente até que uma sonda se

ligue a um alvo complementar perfeitamente emparelhado com uma razão de sinal para interferência que é pelo menos 5x tão alta quanto aquela observada para a hibridização da sonda para um alvo desemparelhado.

5           As condições "muito severas" são selecionadas para serem iguais ao ponto de fusão térmica ( $T_m$ ) para uma sonda específica. A  $T_m$  é a temperatura (sob intensidade iônica e pH definidos) na qual 50% da seqüência em teste hibridiza para uma sonda perfeitamente emparelhada. Para os propósitos  
10 da presente invenção, as condições de hibridização e lavagem "altamente severas" são selecionadas para serem cerca de 5 °C mais baixas do que a  $T_m$  para a seqüência específica em uma intensidade iônica e pH definidos.

          As condições de hibridização e lavagem de  
15 "severidade ultra-alta" são aquelas nas quais a severidade das condições de hibridização e lavagem é aumentada até que a razão de sinal para interferência para a ligação da sonda ao ácido nucléico-alvo complementar perfeitamente emparelhado seja pelo menos 10x tão alta quanto aquela  
20 observada na hibridização para qualquer um dos ácidos nucléicos-alvo não-emparelhados. Um ácido nucléico-alvo que hibridiza para uma sonda sob tais condições, com uma razão de sinal para interferência de pelo menos metade daquela do ácido nucléico-alvo complementar perfeitamente emparelhado,  
25 é dito como se ligando à sonda sob condições de severidade ultra-alta.

          Similarmente, mesmo os níveis mais altos de severidade podem ser determinados aumentando gradualmente as

condições de hibridização e/ou lavagem do ensaio de hibridização relevante. Por exemplo, aqueles nos quais a severidade das condições de hibridização e lavagem são aumentados até que a razão de sinal para interferência para a ligação da sonda ao ácido nucléico-alvo perfeitamente emparelhado seja pelo menos 10x, 20x, 50x, 100x ou 500x ou mais tão alta quanto àquela observada na hibridização para qualquer um dos ácidos nucléicos-alvo não-emparelhados. Um ácido nucléico-alvo que hibridiza para uma sonda sob tais condições, com uma razão de sinal para interferência de pelo menos metade daquela do ácido nucléico-alvo complementar perfeitamente emparelhado é dita como se ligando à sonda sob condições de severidade ultra-alta.

Os ácidos nucléicos que não hibridizam para outro sob condições severas ainda são substancialmente idênticos caso os polipeptídeos que eles codificam sejam substancialmente idênticos. Isto ocorre, por exemplo, quando uma cópia de um ácido nucléico é criada usando a degeneração máxima de códon permitida pelo código genético.

#### 20        Subseqüências Singulares

Em alguns aspectos, a invenção utiliza um ácido nucléico que compreende uma subseqüência singular em um ácido nucléico selecionado entre as seqüências de O-tRNA's e O-RS's aqui descritos ou referenciados. A subseqüência singular é singular em comparação a um ácido nucléico correspondente a qualquer seqüência de ácidos nucléicos de O-tRNA's ou O-RS's conhecidas. O alinhamento pode ser realizado, por exemplo, BLAST ajustada para parâmetros de

omissão. Qualquer subsequência singular é útil, por exemplo, como uma sonda para identificar os ácidos nucleicos da invenção.

5        Similarmente, a invenção utiliza um polipeptídeo que compreende uma subsequência singular em um polipeptídeo selecionado entre as seqüências de O-RS's aqui descritas ou referenciadas. Neste caso, a subsequência singular é singular em comparação a um polipeptídeo correspondente a qualquer uma das seqüências de polipeptídeos conhecidas.

10        A invenção fornece também ácidos nucleicos-alvo que hibridizam sob condições severas para um oligonucleotídeo codificador singular que codifica uma subsequência singular em um polipeptídeo selecionado entre as seqüências de O-RS's, onde a subsequência singular é  
15        singular em comparação com um polipeptídeo correspondete a qualquer um dos polipeptídeos de controle (por exemplo, seqüências originárias a partir das quais as sintetases da invenção foram derivadas, por exemplo, por mutação). As seqüências singulares são determinadas como assinalado  
20        acima.

#### Comparação, Identidade e Homologia de Seqüências

Os termos "idêntica" ou "identidade percentual", no contexto de duas ou mais seqüências de ácidos nucleicos ou polipeptídeos, referem-se a duas ou mais seqüências ou  
25        subsequências que são iguais ou têm uma porcentagem especificada de resíduos de aminoácidos ou nucleotídeos que são iguais, quando comparados e alinhados para correspondência máxima, medida usando um dos algoritmos de

comparação de seqüências descritos abaixo (ou outros algoritmos disponíveis para os versados nessas técnicas) ou por inspeção visual.

O termo "substancialmente idênticos", no contexto  
5 de dois ácidos nucleicos ou polipeptídeos (por exemplo, DNA's que codificam um O-tRNA ou O-RS, ou a seqüência de aminoácidos de uma O-RS), refere-se a duas ou mais seqüências ou subseqüências que têm pelo menos cerca de 60%, cerca de 80%, cerca de 90-95%, cerca de 98%, cerca de 99% ou  
10 mais de identidade de nucleotídeos ou resíduos de aminoácidos, quando comparados e alinhados para correspondência máxima, medida usando um algoritmos de comparação de seqüências ou por inspeção visual. Essas seqüências "substancialmente idênticas" são tipicamente  
15 consideradas como sendo "homólogas", sem referência aos antepassados reais. De preferência, a "identidade substacial" existe em uma região das seqüências, que tem pelo menos cerc de 50 resíduos de comprimento, mais preferivelmente em uma região das seqüências, que tem pelo  
20 menos cerca de 100 resíduos de comprimento, e mais preferivelmente ainda, as seqüências são substancialmente idênticas em pelo menos cerca de 150 resíduos, ou no comprimento inteiro das duas seqüências que estão sendo comparadas.

25 As proteínas e/ou seqüências de proteínas são "homólogas" quando elas são derivadas, naturalmente ou artificialmente, de uma proteína ou seqüência de proteína originária. Similarmente, os ácidos nucleicos e/ou

seqüências de ácidos nucleicos são homólogos quando eles são derivados, naturalmente ou artificialmente, de um ácido nucleico ou seqüência de ácidos nucleicos originária. Por exemplo, qualquer ácido nucleico de ocorrência natural pode ser modificado por qualquer método de mutagênese disponível para incluir um ou mais códons seletores. Quando expressado, este ácido nucleico que sofreu mutagênese codifica um polipeptídeo que compreende um ou mais aminoácidos não-naturais. O processo de mutação, evidentemente, pode alterar adicionalmente um ou mais códons-padrão, mudando também desta forma um ou mais aminoácidos-padrão na proteína mutante resultante. A homologia é genericamente inferida a partir da similaridade de seqüências entre dois ou mais ácidos nucleicos ou proteínas (ou suas seqüências). A porcentagem precisa de similaridade entre seqüências, que é útil para estabelecer a homologia, varia com o ácido nucleico e proteína em questão, mas tão pouco quanto 25% de similaridade de seqüências são usados para estabelecer a homologia. Níveis mais altos de similaridade de seqüências, por exemplo, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95% ou 99% ou mais, também podem ser usados para estabelecer a homologia. Os métodos para determinar as porcentagens de similaridade de seqüências (por exemplo, BLASTP ou BLASTN, usando parâmetros de omissão) estão aqui descritos e estão genericamente disponíveis.

Para comparação de seqüências e determinação da homologia, tipicamente, uma seqüência atua como uma seqüência referencial com a qual as seqüências em teste são



comparadas. Quando se usa um algoritmo de comparação de seqüências, as seqüências em teste e a referencial são alimentadas em um computador, as coordenadas de subsequências são designadas, caso necessário, e os parâmetros do programa de algoritmo de seqüências são designados. O algoritmo de comparação de seqüências, então, calcula a identidade percentual de seqüências para a(s) seqüência(s) em teste em relação à seqüência referencial, baseado nos parâmetros designados do programa.

O alinhamento ótimo de seqüências para comparação pode ser conduzido, por exemplo, pelo algoritmo de homologia local de Smith e Waterman, *Adv. Appl. Math.* 2:482 (1981), pelo algoritmos de alinhamento de homologia de Needeman e Wunsch, *J. Mol. Biol.* 48:443 (1970), pelo método de busca de similaridade de Pearson e Lipman, *Proc. Natl. Acad. Sci. USA* 85:2444 (1988), por implementações computadorizadas destes algoritmos (GAP, BESTIFIT, FASTA e TFASTA no Pacote de Software de Genética de Wisconsin, Genetics Computer Group, 575 Science Dr., Madison, WI, EUA), ou por inspeção visual (vide genericamente "Current Protocols in Molecular Biology", Ausubel et al., editores, "Current Protocols", uma joint venture entre Greene Publishing Associates, Inc. e John Wiley & Sons, Inc., suplementado até 2004).

Um exemplo de um algoritmo apropriado para determinar a identidade percentual de seqüências e similaridade de seqüências é o algoritmo BLAST, que está descrito em Altschul et al. *J. Mol. Biol.* 215:403-410 (1990). O software para realizar análises BLAST está

disponível para o público no website do National Center for Biotechnology Information. Este algoritmo envolve primeiramente identificar pares de seqüências com alta pontuação (HSP's), identificando palavras curtas com comprimento W na seqüência de consulta, emparelham ou satisfazem a alguma pontuação T limite com valor positivo, quando alinhadas com uma palavra do mesmo comprimento em uma seqüência da base de dados. T é referido como o limite de pontuação da palavra vizinha (Altschul et al. supra). Estes acertos de palavras vizinhas iniciais atuam como sementes para iniciar buscas para encontrar HSP's mais longos que as contêm. Estes acertos de palavras são então estendidos em ambas direções ao longo de cada seqüência até uma extensão tão grande quanto o alinhamento cumulativo pode ser aumentado. As pontuações cumulativas são calculadas usando, no caso de seqüências de nucleotídeos, os parâmetros M (pontuação de merecimento para um par de resíduos emparelhados; sempre  $> 0$ ) e N (pontuação de penalização para resíduos não-emparelhados; sempre  $> 0$ ). No caso de seqüências de aminoácidos, uma matriz de pontuação é usada para calcular a pontuação cumulativa. A extensão dos acertos de pontuação em cada direção é detida quando: a pontuação de alinhamentos positivos diminui na quantidade X do seu valor máximo atingido; a pontuação cumulativa cai para zero ou abaixo, devido à acumulação de um ou mais alinhamentos de resíduos com pontuação negativa; ou a extremidade de qualquer seqüência é alcançada. Os parâmetros do algoritmo BLAST, W, T e X, determinam a sensibilidade e a velocidade

do alinhamento. O programa BLAST (no caso de seqüências de nucleotídeos) usa como omissões um comprimento de palavra (W) de 11, uma expectativa (E) de 10, um corte de 100,  $M = 5$ ,  $N = 4$ , e uma comparação de ambos filamentos. No caso de

5 seqüências de aminoácidos, o programa BLAST usa como omissões um comprimento de palavra (W) de 3, uma expectativa (E) de 10, e a matriz de pontuação BLOSUM 62 (vide Henikoff e Henikoff, *Proc. Natl. Acad. Sci. USA* 89:10915 (1989)).

Além de calcular a identidade percentual de

10 seqüências, o algoritmo BLAST realiza também uma análise estatística da similaridade entre duas seqüências (vide, por exemplo, Karlin e Altschul, *Proc. Natl. Acad. Sci. USA* 90:5873-5787 (1993)). Uma mdida da similaridade fornecida pelo algoritmo BLAST é a probabilidade da menor soma ( $P(N)$ ),

15 que fornece uma indicação da probabilidade pela qual um emparelhamento entre duas seqüências de nucleotídeos ou aminoácidos ocorreria por acaso. Por exemplo, um ácido nucléico é considerado similar a uma seqüência referencial caso a probabilidade da menor soma em uma comparação do

20 ácido nucléico em teste com o ácido nucléico referencial seja menor do que cerca de 0,1, mais preferivelmente menor do que cerca de 0,01, e mais preferivelmente ainda, menor do que cerca de 0,001.

#### Mutagênese e Outras Técnicas de Biologia Molecular

25 O polinucleotídeo e os polipeptídeos da invenção e usados na invenção podem ser manipulados usando técnicas de biologia molecular. Os textos genéricos que descrevem as técnicas de biologia molecular incluem Berger e Kimmel,

"Guide to Molecular Cloning Techniques", "Methods in Enzimology", volume 152, Academic Press, Inc., San Diego, CA (Berger); Sambrook et al., "Molecular Cloning - A Laboratory Manual" (3ª edição), volumes 1-3, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, 2001 ("Sambrook") e F.M. Ausubel et al. (editores), "Current Protocols in Molecular Biology", Current Protocols, uma joint venture entre Greene Publishing Associates e John Wiley & Sons, Inc., New York (suplementado até 2004) ("Ausubel")). Estes textos descrevem a mutagênese, o uso de vetores, promotores e muitos outros tópicos relevantes relacionados, por exemplo, à geração de genes que incluem códons seletores para a produção de proteínas que incluem aminoácidos não-naturais, tRNA's ortogonais, sintetases ortogonais, e paredes deles.

Vários tipos de mutagênese podem ser usados em conjunto com a invenção, por exemplo, para mutar moléculas de tRNA, produzir bibliotecas de tRNA's, produzir bibliotecas de sintetases, inserir códons seletores que codificam um aminoácido não-natural em uma proteína ou polipeptídeo de interesse. Eles incluem, porém sem limitações, mutagênese direcionada a sítios, pontual aleatória, recombinação homóloga, embaralhamento de DNA ou outros métodos recursivos de mutagênese, construção quimérica, mutagênese usando modelos que contêm uracila, mutagênese direcionada a oligonucleotídeos, mutagênese de DNA modificada com fosforotioato, mutagênese usando DNA duplex com lacunas, ou similares, ou qualquer combinação

deles. Os métodos adicionais apropriados incluem reparo pontual de desemparelhamentos, mutagênese usando cepas hospedeiras deficientes em reparo, seleção por restrição e purificação por restrição, mutagênese com deleção, 5 mutagênese por síntese total de genes, reparo com rompimento de duplo filamento, e similares. A mutagênese, por exemplo, envolvendo construções quiméricas, também está incluída na presente invenção. Em uma modalidade, a mutagênese pode ser orientada por informações conhecidas da molécula de 10 ocorrência natural ou moléculas de ocorrência natural alteradas ou mutadas, por exemplo, seqüência, comparações de seqüências, propriedades físicas, estrutura cristalina, ou similares.

As células hospedeiras são manipuladas por 15 engenharia genética (por exemplo, transformadas, transduzidas ou transfectadas) com os polinucleotídeos da invenção ou construções que incluem um polinucleotídeo, por exemplo, um vetor que pode ser, por exemplo, um vetor de clonagem ou um vetor de expressão. Por exemplo, as regiões 20 codificadoras para o tRNA ortogonal, a tRNA sintetase ortogonal, e a proteína a serem derivadas estão ligadas operacionalmente a elementos de controle da expressão gênica, que são funcionais na célula hospedeira desejada. Os vetores típicos contêm terminadores de transcrição e 25 tradução, seqüências de iniciação da transcrição e tradução, e promotores úteis para a regulação da expressão do ácido nucléico-alvo específico. Os vetores compreendem opcionalmente cassetes de expressão genéricos que contêm

pelo menos uma seqüência terminadora independente, seqüências que permitem a replicação do cassete em eucariontes, ou procariontes, ou ambos (por exemplo, vetores *shuttle*) e marcadores de seleção para sistemas procarióticos e eucarióticos, ou de preferência, ambos. Vide Gilman e Smith, *Gene* 8:81 (1979); Roberts et al., *Nature* 328:731 (1987); Schneider et al., *Protein Expr. Purif.* 6435:10 (1995); Ausubel, Sambrook, Berger (todos supra). O vetor pode estar, por exemplo, na forma de um plasmídeo, uma bactéria, um vírus, um polinucleotídeo nu, ou um polinucleotídeo conjugado. Os vetores são introduzidos nas células e/ou microrganismos por métodos padronizados, incluindo eletroporação (From et al., *Proc. Natl. Acad. Sci. USA* 82:5824 (1985), infecção por vetores virais, penetração balística em alta velocidade por partículas pequenas com o ácido nucléico dentro da matriz de pérolas ou partículas pequenas, ou sobre a superfície (Klein et al., *Nature* 327:70-73 (1987), e/ou similares.

Um sistema com um único plasmídeo altamente eficiente e versátil foi desenvolvido para a incorporação específica em sítio de aminoácidos não-naturais em proteínas, em resposta ao códon finalizador âmbar (UAG) em *E. coli*. No novo sistema, o par de tRNA<sup>Tyr</sup>(CUA) supressor e tirosil-tRNA sintetase de *M. Jannaschii* é codificado em um único plasmídeo, que é compatível com a maioria dos vetores de expressão em *E. coli*. O operon de tRNA monocistrônico sob controle do promotor *proK* e do terminador foi construído para uma estrutura secundária ótima e processamento do tRNA.

A introdução de uma forma mutada do promotor *glnS* para a sintetase resultou em um aumento significativo na eficiência e fidelidade da supressão. Os aumentos na eficiência da supressão também foram obtidos por múltiplas cópias do gene de tRNA, bem como por uma mutação específica (D286R) na sintetase (Kobayashi et al., "Structural basis for orthogonal tRNA specificities of tyrosyl-tRNA synthetases for genetic code expansion", *Nat. Struct. Biol.* 10(6):425-432 (2003)). A generalidade do sistema otimizado também foi demonstrada por incorporação altamente eficiente e precisa de vários aminoácidos não-naturais diferentes, cujas utilidades singulares ao estudar a função e a estrutura de proteínas foram demonstradas anteriormente.

Um catálogo de Bactérias e Bacteriófagos útil para clonagem é fornecido, por exemplo, pela ATCC, por exemplo, "The ATCC Catalogue of Bacteria and Bacteriophage" (1996), Ghena et al. (editores), publicado pela ATCC. Os procedimentos básicos adicionais para seqüenciamento, clonagem e outros aspectos de biologia molecular e considerações teóricas subjacentes são encontrados também em Sambrook (supra), Ausubel (supra), e em Watson et al. (1992), "Recombinant DNA", Segunda Edição, Scientific American Books, NY. Além disso, essencialmente qualquer ácido nucléico (e virtualmente qualquer ácido nucléico marcado, seja ele padrão ou não-padrão) pode ser solicitado por encomenda ou pela lista de qualquer fornecedor comercial, tal como Midland Certified Reagent Company (midland, TX), The Great American Gene Company (Ramona, CA),

ExpressGen, Inc. (Chicago, IL), Operon Technologies, Inc. (Alameda, CA), e muitos outros.

As células hospedeiras manipuladas geneticamente podem ser cultivadas em meios nutrientes convencionais modificados, conforme apropriado, para atividades tais como, por exemplo, triar etapas, ativar promotores ou selecionar transformantes. Estas células podem ser opcionalmente cultivadas dentro de organismos transgênicos. Outras referências úteis, por exemplo, para isolamento e cultura de células (por exemplo, para subsequente isolamento de ácidos nucleicos) incluem Freshney, "Culture of Animal Cells, a Manual of Basic Technique", terceira edição, Wiley-Liss, New York, e referências lá citadas; Payne et al. (1992), "Plant Cell and Tissue Culture in Liquid Systems", John Wiley & Sons, Inc., New York, NY; Gamborg e Philips (editores) (1995), "Plant Cell, Tissue and Organ Culture: Fundamental Methods", Springer Lab Manual, Springer-Verlag (Berlin Heidelberg New York); e Atlas e Parks (editores), "The Handbook of Microbiological Media (1993), CRC Press, Boca Raton, FL.

#### Proteínas e Polipeptídeos de Interesse

Os métodos para produzir uma proteína que compreende um aminoácido não-natural em uma posição especificada também são uma característica da invenção. Por exemplo, um método pode incluir desenvolver, em um meio apropriado, uma célula (por exemplo, em uma célula de *E. coli*), onde a célula compreende um ácido nucleico que compreende pelo menos um códon seletor e codifica uma



proteína; e fornecer o aminoácido não-natural; onde a célula compreende ainda: um tRNA ortogonal (O-tRNA) que funciona na célula e reconhece o códon seletor; e uma amino-acil-tRNA sintetase ortogonal (O-RS) que preferencialmente amino-acila  
 5 o O-tRNA com o aminoácido não-natural. A proteína assim produzida na *E. coli* compreende um aminoácido não-natural na posição correspondente ao códon seletor. Esta proteína pode ser então opcionalmente reagida sob condições nas quais o aminoácido não-natural sofre modificação covalente,  
 10 produzindo desta forma uma proteína após a tradução.

Em certas modalidades, a O-RS compreende uma inclinação pela amino-acilação do O-tRNA cognato em detrimento de qualquer tRNA endógeno em um sistema de expressão. A razão relativa entre O-tRNA e tRNA endógeno que  
 15 é carregado pela O-RS, quando o O-tRNA e a O-RS estão presentes em concentrações molares iguais, é maior do que 1:1, de preferência pelo menos 2:1, mais preferivelmente 5:1, ainda mais preferivelmente 10:1, ainda mais preferivelmente 20:1, ainda mais preferivelmente 50:1, ainda  
 20 mais preferivelmente 75:1, ainda mais preferivelmente 95:1, 98:1, 99:1 100:1, 500:1, 1.000:1, 5.000:1 o mais alta.

A proteína que compreende um aminoácido não-natural em uma posição especificada, que é modificada depois da tradução, também é uma característica da invenção. A  
 25 proteína é produzida em uma célula, por exemplo, uma célula de *E. coli*. Os pares de O-tRNA/O-RS também residem na célula e utilizam a maquinaria de tradução da célula hospedeira, o que resulta na incorporação *in vivo* de um aminoácido não-

natural dentro de uma proteína de fusão, em resposta a um códon seletor. A capacidade de um sistema de O-tRNA/O-RS para funcionar em uma célula hospedeira para incorporar uma ampla série de aminoácidos não-naturais que podem ser

5 modificadas depois da tradução é conhecida. Vide, por exemplo, Chin et al., *Science* 301:964-967 (2003); Zhang et al., *Proc. Natl. Acad. Sci. USA* 101:8882-8887 (2004); Anderson et al., *Proc. Natl. Acad. Sci. USA* 101:7566-7571 (2004); Wang et al., *Science* 292:498-500 (2001); Chin et

10 al., *J. Am. Chem. Soc.* 124:9026-9027; Chun e Schultz, *Chem. Bio. Chem.* 11:1135-1137 (2002); Chin et al., *PNAS USA* 99:11020-11024 (2002); Wang e Schultz, *Chem. Comm.* 1-10 (2002); Wang e Schultz, "Expanding the Genetic Code", *Angewandte Chemie Int. Ed.* 44(1):34-66 (2005); Xie e

15 Schultz, "An Expanding Genetic Code", *Methods* 36:227-238 (2005); e Deiters et al., *Bioorganic & Medicinal Chemistry Letters* 15:1521-1524 (2005), sendo cada um deles incorporado como referência em sua totalidade.

Vide também os sistemas ortogonais de aminoácidos

20 não-naturais descritos nas publicações internacionais n<sup>os</sup> WO 2002/086075, intitulada "Methods and Compositions for the Production of Orthogonal tRNA Aminoacyl-tRNA Synthetase Pairs"; WO 2002/085923, intitulada "In Vivo Incorporation of Unnatural Aminoacids"; WO 2004/094593, intitulada "Expandig

25 the Eukariotic Genetic Code"; WO 2005/019415, depositado em 7 de julho de 2004; WO 2005/007870, depositado em 7 de julho de 2004; WO 2005/007624, depositado em 7 de julho de 2004; WO 2006/034332, depositado em 20 de setembro de 2005; e WO

2006/110182, depositado em 27 de outubro de 2005, intitulada "Orthogonal Translation Components for the In Vivo Incorporation of Unnatural Aminoacids", em nome de Schultz et al. Cada uma destas referências é incorporada como  
5 referência em sua totalidade.

A incorporação de um aminoácido não-natural pode ser feita, por exemplo, para individualizar mudanças na estrutura e/ou função da proteína, por exemplo, para mudar o tamanho, atividade, caráter nucleofílico, ligação  
10 hidrogênio, caráter hidrofóbico, acessibilidade de marcadores ou grupos reativos, etc. As proteínas que incluem um aminoácido não-natural podem ter uma propriedades catalíticas ou físicas intensificadas ou mesmo inteiramente novas. Por exemplo, as seguintes propriedades são  
15 opcionalmente modificadas pela inclusão de um aminoácido não-natural em uma proteína: toxicidade, biodistribuição, propriedades estruturais, propriedades espectroscópicas, propriedades químicas e/ou fotoquímicas, capacidade enzimática, meia-vida (por exemplo, meia-vida sérica),  
20 capacidade de reagir com outras moléculas, por exemplo, de forma covalente ou não-covalente, e similares. As composições que incluem proteínas que incluem pelo menos um aminoácido não-natural são úteis, por exemplo, para terapias, diagnósticos, enzimas catalíticas, enzimas  
25 industriais, proteínas ligantes (por exemplo, anticorpos) inusitadas, e por exemplo, o para o estudo da estrutura e função de proteínas. Vide, por exemplo, Dougherty, "Unnatural Amino Acids as Probes of Protein Structure and

Function", *Current Opinion in Chemical Biology* 4:645-652 (2000). As proteínas que compreendem um aminoácido não-natural, que podem ser seletivamente modificadas depois da tradução (por exemplo, por uma cicloadição [3+2] ou uma  
5 modificação de Staudinger) podem ser manipuladas para conter qualquer funcionalidade desejada que pode ser acoplada ao parceiro de reação. A natureza do parceiro da reação não é de forma alguma limitada, exceto apenas que ele compreende um grupamento reativo apropriado que resulta em uma anexação  
10 covalente ao resíduo de aminoácido não-natural no polipeptídeo.

Em alguns aspectos, uma composição inclui uma proteína com pelo meno um, por exemplo, pelo menos dois, pelo menos três, pelo menos quatro, pelo menos cinco, pelo  
15 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 iguais ou diferentes, por exemplo, pode haver 1, 2, 3, 4, 5, 6, 7, 8, 9 ou 10 sítios diferentes na proteína, que compreendem 1, 2, 3, 4, 5, 6, 7,  
20 8, 9 ou 10 ou mais aminoácidos não-naturais diferentes. Em outro aspecto, uma composição inclui uma proteína com pelo menos um, mas menos do que todos, de um aminoácido específico presente na proteína é um aminoácido não-natural. Para uma dada proteína com mais do que dois aminoácidos não-  
25 naturais, os aminoácidos não-naturais podem ser iguais, diferentes, ou uma combinação de um múltiplo aminoácido do mesmo tipo com pelo menos um aminoácido não-natural diferente.

Essencialmente qualquer proteína (ou parte dela) que inclui um aminoácido não-natural (e qualquer ácido nucléico codificador correspondente, por exemplo, que inclui um ou mais códons seletores) pode ser produzida usando as composições e métodos da invenção. Nenhuma tentativa é feita para identificar as centenas de milhares de proteínas conhecidas, qualquer uma delas podendo ser modificada para incluir um ou mais aminoácidos não-naturais, por exemplo, individualizando quaisquer métodos de mutação disponíveis para incluir um ou mais códons seletores apropriados em um sistema de tradução relevante. Os repositórios comuns de seqüências para proteínas conhecidas incluem GenBank EMBL, DDBJ e NCBI. Outros repositórios podem ser identificados facilmente buscando 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% ou pelo menos 99% ou mais, idênticas a qualquer proteína disponível (por exemplo, uma proteína terapêutica, uma proteína diagnóstica, uma enzima industrial, ou parte dela, e similares), e elas compreendem um ou mais aminoácidos não-naturais. Os exemplos de proteínas terapêuticas, diagnósticas e outras que podem ser modificadas para compreender um ou mais aminoácidos não-naturais podem ser encontrados, porém sem limitações, nas publicações internacionais nºs WO 2004/094593, depositada em 16 de abril de 2004, intitulada "Expanding the Eukariotic Genetic Code"; e WO 2002/085923, intitulada "In Vivo Incorporation of Unnatural Amino Acids". Os exemplos de

proteínas terapêuticas, diagnósticas e outras que podem ser modificadas para compreender um ou mais aminoácidos não-naturais incluem, porém sem limitações, antitripsina alfa-1, angiostatina, fator anti-hemolítico, anticorpos (outros  
 5 detalhes sobre anticorpos são encontrados abaixo), apolipoproteína, apoproteína, fator natriurético atrial, polipeptídeo natriurético atrial, peptídeos atriais, quimiocinas C-X-C (por exemplo, T39765, NAP-2, ENA-78, Gro-a, Gro-b, Gro-c, IP-10, GCP-2, NAP-4, SDF-1, PF4, MIG),  
 10 calcitonina, quimiocinas CC (por exemplo, proteína-1 quimioatrativa de monócitos, proteína-2 quimioatrativa de monócitos, proteína-3 quimioatrativa de monócitos, proteína-1 alfa inflamatória de monócitos, proteína-1 beta inflamatória de monócitos, RANTES, I309, R83915, R91733,  
 15 HCC1, T58847, D31065, T64262), ligante CD40, Ligante C-kit, colágeno, fator estimulador de colônias (CSF), fator de complemento 5a, inibidor de complemento, receptor 1 de complemento, citocinas (por exemplo, peptídeo-78 ativador de neutrófilos epiteliais, GRO $\alpha$ /MGSA, GRO $\beta$ , GRO $\gamma$ , MIP-1 $\alpha$ , MP-  
 20 1 $\delta$ , MCP-1), fator do crescimento epidérmico (EGF), eritropoietina ("EPO"), toxinas esfoliantes A e B, Fator IX, Fator VII, Fator VIII, Fator X, fator do crescimento de fibroblastos (FGF), fibrinogênio, fibronectina, G-CSF, GM-CSF, glicocerebrosidase, gonadotropina, fatores de  
 25 crescimento, proteínas Hedgehog (por exemplo, *Sonic*, *Indian*, *Desert*), hemoglobina, fator do crescimento de hepatócitos (HGF), hirudina, albumina do soro humano, insulina, fator do crescimento semelhante à insulina (IGF), interferons (por

exemplo, IFN- $\alpha$ , IFN- $\beta$ , IFN- $\gamma$ ), interleucinas (IL-1, IL-2, IL-3, IL-4, IL-5, IL-6, IL-7, IL-8, IL-9, IL-10, IL-11, IL-12, etc.), fator do crescimento de ceratinócitos (KGF), lactoferrina, fator inibitório de leucemia, luciferase, neuriturina, fator inibitório de neutrófilos (NIF), oncostatina M, proteína osteogênica, hormônio paratireóide, PD-ECSF, PDGF, hormônios peptídicos (por exemplo, hormônio do crescimento humano), pleiotropina, proteína A, proteína G, exotoxinas pirogênicas A, B e C, relaxina, renina, SCF, receptor I de complemento solúvel, I-CAM 1 solúvel, receptores de interleucinas solúveis (IL-1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15), receptor de TNF solúvel, somatomedina, somatostatina, somatotropina, estreptocinase, super-antígenos, isto é, enterotoxinas estafilocócicas (SEA, SEB, SEC1, SEC2, SEC3, SED, SEE), superóxido dismutase (SOD), toxina da síndrome do choque tóxico (TSST-1), timosina alfa 1, ativador do plasminogênio tecidual, fator beta de necrose tumoral (TNF beta), fator do crescimento endotelial vascular (VEGF), urocinase, e muitos outros.

Uma classe de proteínas que pode ser fabricada usando as composições e métodos para a incorporação *in vivo* de aminoácidos não-naturais em proteínas, aqui descritos, inclui moduladores transcricionais ou uma parte deles. Os moduladores transcricionais exemplificativos incluem genes e proteínas moduladoras transcricionais que modulam o crescimento, diferenciação, regulação celular, ou similares. Os moduladores transcricionais são encontrados em procariotas, vírus, e eucariotas, incluindo fungos, plantas,

leveduras, insetos, e animais, incluindo mamíferos, fornecendo uma ampla série de alvos terapêuticos. Deve-se avaliar que os ativadores da expressão e transcrpcionais regulam a transcrição por vários mecanismos, por exemplo, ligando-se a receptores, estimulando uma cascata de tradução de sinais, regulando a expressão de fatores de transcrição, ligando-se a promotores e intensificadores, ligando-se a proteínas que se ligam a promotores e intensificadores, desenrolar DNA, dividir pré-RNA, poliadenilar RNA, e degradar RNA.

Uma classe de proteínas da invenção (por exemplo, proteínas com um ou mais aminoácidos não-naturais) incluem proteínas biologicamente ativas tais como citocinas, moléculas inflamatórias, fatores de crescimento, seus receptores, e produtos oncogênicos, por exemplo, interleucinas (por exemplo, IL-1, IL-2, IL-8, etc.), interferons, FGF, IGF-I, IGF-II, FGF, PDGF, TNF, TGF- $\alpha$ , TGF- $\beta$ , EGF, KGF, SCF/c-kit, CD40L/CD40, VLA-4/VCAM-1, ICAM-1/LFA-1, e hialurina/CD44; moléculas de transdução de sinais e produtos oncogênicos correspondentes, por exemplo, Mos, Ras, Raf, e Met; e ativadores e supressores transcrpcionais, por exemplo, p53, Tat, Fos, Myc, Jun, Myb, el, e receptores de hormônios esteróides tais com aqueles para estrogênio, progesterona, testosterona, aldosterona, o ligante de receptores de LDL e corticosterona.

As enzimas (por exemplo, enzimas industriais) ou partes delas com pelo menos um aminoácido não-natural também são fornecidas pela invenção. Os exemplos de enzimas



incluem, porém sem limitações, por exemplo, amidases, aminoácido racemases, acilases, desalogenases, dioxigenases, diarilpropano peroxidases, epimerases, epóxido hidrolases, esterases, isomerases, cinases, glicose isomerases, 5 glicosidasas, glicosil transferases, haloperoxidasas, monooxigenases (por exemplo, p450s), lípases, lignina peroxidases, nitrila hidratases, nitrilases, proteases, fosfatases, subtilisinas, transaminase, e nucleases.

Muitas destas proteínas estão disponíveis 10 comercialmente (vide, por exemplo, catálogo da Sigma BioSciences), e as seqüências e genes de proteínas correspondentes e, tipicamente, muitas variantes delas, são bem conhecidas (vide, por exemplo, GenBank). Qualquer uma delas pode ser modificada pela inserção de um ou mais 15 aminoácidos não-naturais de acordo com a invenção, por exemplo, para alterar a proteína com relação a uma ou mais propriedades terapêuticas, diagnósticas ou enzimáticas de interesse. Os exemplos de propriedades terapeuticamente relevantes incluem meia-vida sérica, meia-vida de 20 prateleira, estabilidade, imunogenicidade, atividade terapêutica, detectabilidade (por exemplo, pela inclusão de grupos-repórter (por exemplo, marcadores ou sítios de ligação a marcadores) nos aminoácidos não-naturais), redução da DL<sub>50</sub> ou outros efeitos colaterais, capacidade de entrar 25 no corpo através do trato gástrico (por exemplo, disponibilidade oral), ou similares. Os exemplos de propriedades diagnósticas incluem meia-vida na prateleira, estabilidade, atividade diagnóstica, detectabilidade, ou

similares. Os exemplos de propriedades enzimáticas relevantes incluem meia-vida na prateleira, estabilidade, atividade enzimática, capacidade de produção, ou similares.

Várias outras proteínas também podem ser modificadas para incluir um ou mais aminoácidos não-naturais, usando as composições e métodos da invenção. Por exemplo, a invenção pode incluir substituir um ou mais aminoácidos naturais em uma ou mais proteínas de vacinas por um aminoácido não-natural, por exemplo, em proteínas de fungos infecciosos, por exemplo, espécies *Aspergillus*, *Candida*; bactérias, particularmente *E. coli*, que serve como um modelo para bactérias patogênicas, bem como bactérias medicamente importantes tais como *Staphilococci* (por exemplo, *aureus*), ou *Streptococci* (por exemplo, *pneumoniae*); protozoários tais como esporozoários (por exemplo, *Plasmodia*), rizópodes (por exemplo, *Entamoeba*), e flagelados (*Trypanosoma*, *Leishmania*, *Trichomonas*, *Giardia*, etc.); vírus tais como (+) RNA vírus (os exemplos incluem poxvírus, por exemplo, *vaccinia*; picornavírus, por exemplo, *pólio*; togavírus, por exemplo, *rubella*; flavivírus, por exemplo, HCV; e coronavírus, (-) RNA vírus (por exemplo, *rabdovírus*, por exemplo, *VSV*; *Paramyxoviruses*, por exemplo, *RSV*; *Orthomyxoviruses*, por exemplo, *influenza*; *Bunyavirus*; e *Arenavirus*), dsDNA vírus (*Reovirus*, por exemplo), vírus de RNA para DNA, isto é, retrovírus, por exemplo, HIV e HTLV, e certos vírus de DNA para RNA, tal como hepatite B.

As proteínas relacionadas à agricultura, tais como as proteínas de resistência a insetos (por exemplo,

proteínas Cry), enzimas da produção de amido e lipídeos, toxinas de plantas e insetos, proteínas de resistência a toxinas, proteínas de desintoxicação por micotoxinas, enzimas de crescimento de plantas (por exemplo, ribulose, 5 1,5-bisfosfato carboxilase/oxigenase, "RUBISCO"), lipoxigenase (LOX), e fosfoenolpiruvato (PEP) carboxilase também são alvos apropriados para modificação com aminoácidos não-naturais.

Em certas modalidades, a proteína de interesse (ou 10 parte dela) é codificada por um ácido nucléico. Tipicamente, o ácido nucléico compreende pelo menos um códon seletor, pelo menos dois códons seletores, pelo menos três códons seletores, pelo menos quatro códons seletores, pelo menos cinco códons seletores, pelo menos seis códons seletores, 15 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 sofrer mutagênese usando métodos bem 20 conhecidos pelos versados nessas técnicas e aqui descritos na seção "Mutagênese e Outras Técnicas de Biologia Molecular", para incluir, por exemplo, um ou mais códons seletores para a incorporação de um aminoácido não-natural. Por exemplo, um ácido nucléico para uma proteína de 25 interesse sofre mutagênese para incluir um ou mais códons seletores, proporcionando a inserção de um ou mais aminoácidos não-naturais. A invenção inclui qualquer uma dessas variantes, por exemplo, mutantes, versões de qualquer

proteína, por exemplo, incluindo pelo menos um aminoácido não-natural. Similarmente, a invenção inclui também ácidos nucleicos correspondentes, isto é, qualquer ácido nucleico com um ou mais códons seletores, que codifica um ou mais aminoácidos não-naturais.

Para produzir uma proteína que inclui um aminoácido não-natural modificado depois da tradução, pode-se usar células hospedeiras e organismos adaptados para a incorporação *in vivo* do aminoácido não-natural por intermédio de pares de tRNA/RS ortogonais. As células hospedeiras são modificadas por engenharia genética (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 derivada. Cada um destes componentes pode estar no mesmo vetor, ou cada um pode estar em um vetor separado, ou 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, uma bactéria, um vírus, um polinucleotídeo nu, ou um polinucleotídeo conjugado.

#### Definição de Polipeptídeos por Imunorreatividade

Devido ao fato de que os polipeptídeos da invenção fornecem uma série de novas seqüências de polipeptídeos (por exemplo, polipeptídeos que compreendem aminoácidos não-naturais no caso de proteínas sintetizadas nos sistemas de tradução aqui fornecidos, ou, por exemplo, no caso de sintetases inusitadas, seqüências inusitadas de aminoácidos-padrão), os polipeptídeos fornecem também novas

características estruturais que podem ser reconhecidas, por exemplo, em ensaios imunológicos. A geração de anti-soros, que se ligam especificamente aos polipeptídeos da invenção, bem como os polipeptídeos que são ligados por tais anti-soros, são uma característica da invenção. O termo "anticorpo", como aqui utilizado, inclui, porém sem limitações, uma polipeptídeo substancialmente codificado por um gene de imunoglobulina ou genes de imunoglobulina, ou fragmentos deles, que se ligam e reconhecem especificamente um analisado (antígeno). Os exemplos incluem anticorpos policlonais, monoclonais, quiméricos, e de cadeia única, e similares. Os fragmentos de imunoglobulinas, incluindo fragmentos Fab e fragmentos produzidos por uma biblioteca de expressão, incluindo apresentação de fagos, também estão incluídos no termo "anticorpo", como aqui utilizado. Vide, por exemplo, Paul, "Fundamental Immunology", 4ª Edição, 1999, Raven Press, New York, para obter a estrutura e terminologia de anticorpos.

Para produzir anti-soros para uso em um imunoensaio, um ou mais dos polipeptídeos imunogênicos é produzido e purificado como aqui descrito. Por exemplo, uma proteína recombinante pode ser produzida em uma célula recombinante. Uma cepa endógama de camundongos (usada neste ensaio porque os resultados são mais reprodutíveis devido à identidade genética virtual dos camundongos) é imunizada com a(s) proteína(s) imunogênica(s) em combinação com um adjuvante-padrão, tal como o adjuvante de freund, e um protocolo de imunização de camundongos padronizado (vide,

por exemplo, Harlow e Lane (1988), "Antibodies: A Laboratory Manual", Cold Spring Harbor Publications, New York, para obter uma descrição-padrão da geração de anticorpos, formatos de imunoensaios e condições que podem ser usadas  
5 para determinar a imunorreatividade específica. Detalhes adicionais sobre proteínas, anticorpos, anti-soros, etc., podem ser encontrados nas publicações internacionais n<sup>os</sup> WO 2004/094593, intitulada "Expanding the Eukariotic Genetic Code"; WO 2002/085923, intitulada "In Vivo Incorporation of  
10 Unnatural Amino Acids"; WO 2004/035605, intitulada "Glycoprotein Synthesis"; e WO 2004/058946, intitulada "Protein Arrays".

*Kits* também são uma característica da invenção. Por exemplo, um *kit* para produzir um polipeptídeo que  
15 compreende pelo menos um aminoácido não-natural é uma característica da invenção, onde o *kit* compreende pelo menos uma construção da invenção. Por exemplo, tais *kits* podem compreender vários componentes selecionados entre: um recipiente para manter os componentes do *kit*, materiais  
20 informativos para produzir o polipeptídeo, um ácido nucléico que compreende uma sequência de polinucleotídeo que codifica um O-tRNA, um ácido nucléico que compreende um polinucleotídeo que codifica uma O-RS, um aminoácido não-natural, reagentes para a modificação do aminoácido não  
25 natural depois da tradução, e uma cepa apropriada de células hospedeiras de *E. coli* para a expressão do O-tRNA/O-RS.

#### EXEMPLOS

Os exemplos que se seguem são fornecidos para

ilustrar, mas não para limitar a invenção reivindicada. Os versados nessas técnicas devem reconhecer uma série de parâmetros não-críticos que podem ser alterados sem fugir do âmbito da invenção reivindicada. Deve-se entender que os  
5 exemplos e modalidades aqui descritas são meramente ilustrativas e que várias mudanças ou modificações, à luz delas, devem ser sugeridas para os versados nessas técnicas e devem estar incluídas dentro do espírito e âmbito deste pedido de patente e âmbito das reivindicações apensadas.

10                    EXEMPLO 1

Construção de um Sistema com um Único Plasmídeo para a Expressão de Polipeptídeos que Compreendem Aminoácidos Não-naturais

O presente exemplo descreve a construção de um  
15 plasmídeo que codifica ambos membros de um par de amino-acil-tRNA ortogonal e amino-acil-tRNA sintetase ortogonal, para a incorporação de p-benzoil-L-fenilalanina.

Foi construído um plasmídeo (denominado pYR-BpaRS1) que contém seqüências de nucleotídeos que codificam  
20 ambos componentes de um par de tradução ortogonal, que funcionam em uma célula hospedeira de *E. coli*. A saber, estes dois componentes são o tRNA ortogonal *Mj*tRNA-Tyr(CUA) e a *Mj*TyrRS sintetase mutante (BpaRS) que especificamente amino-acila o tRNA ortogonal com aminoácido fotorreticulante,  
25 p-benzoil-L-fenilalanina (Bpa, vide Figura 1). Vide Chin et al., *Proc. Natl. Acad. Sci. USA* 99:11020-11024 (2002).

Para determinar a eficiência da supressão deste sistema do plasmídeo pYR-BpaRS1, a atividade de  $\beta$ -

galactosidase foi medida para células de *E. coli* TOP10 (Invitrogen™) co-transformadas com pYR-BpaRS1 e o plasmídeo repórter *lacZ* que codifica  $\beta$ -galactosidase com uma mutação âmbar em um sítio permissivo na sequência líder do gene *lacZ*. Infelizmente, quando as células foram desenvolvidas na presença de Bpa 1 mM, níveis muito baixos de atividade de  $\beta$ -galactosidase foram observados (Figura 2). As tentativas para aumentar a eficiência de supressão modificando as sequências flanqueadoras do gene de tRNA foram malsucedidas.

Para melhorar esta eficiência de supressão, um novo operon de tRNA supressor âmbar com um promotor e terminador de tRNA de *E. coli* de ocorrência natural foi construído. Um levantamento de genes de tRNA de *E. coli* revelou que os prolil-tRNA's de *E. coli* têm o mesmo par C1-G72 que os tRNA's de *Archaea*; este par de bases é um determinante de identidade importante para o reconhecimento seletivo de *MjtRNA-Tyr(CUA)* por *MjTyrRS* em *E. coli* (Wang e Schultz, *Chem. Biol.* 8:883-890 (2001)). Tendo em vista esta observação, um gene de tRNA supressor âmbar sintético foi construído, de tal modo que o gene de *MjtRNA-Tyr(CUA)* substitua o gene *proK* de *E. coli* do mesmo comprimento (77 nucleotídeos) no operon de *proK* monocistrônico. Como o gene *proK* codifica o tRNA que reconhece o códon de prolina mais freqüentemente usado (CCG) em *E. coli* (Nakamura et al., *Nucleic Acids Res.* 28:292 (2000)), esperava-se que o gene de *MjtRNA-Tyr(CUA)* fosse transcrito eficientemente sob o controle do promotor *proK*. Um sítio de ligação de FIS localizado naturalmente a montante do promotor *proK* também



foi incluído na construção gênica sintética para intensificar a transcrição do tRNA (Muskhelishvili et al., *EMBO J.* 16:3655-3665 (1997)). O vetor de expressão final pYR-BpaRS5 foi gerado substituindo o operon do tRNA

5 supressor original em pYR-BPaRS1 pelo o gene de *MjtRNA-Tyr(CUA)* sob o controle do promotor e terminador de *proK*. Quando transformado em *E. coli*, este plasmídeo levou a um aumento de 2 vezes (em relação ao gene de tRNA original sob o controle do promotor *Ipp* e o terminador *rrnC*) na expressão

10 de *MjtRNA-Tyr(CUA)*, como observado por análise *Northern* (vide Figura 3). Este aumento na expressão corresponde a 8% de eficiência de supressão (em relação à expressão de  $\beta$ -galactosidase do tipo selvagem), determinado pelo ensaio da atividade de  $\beta$ -galactosidase (Figura 2).

## 15                    EXEMPLO 2

### Geração de Genes e Promotores de Sintetases Aperfeiçoados

O presente exemplo descreve a construção de sistemas aperfeiçoados para a expressão de componentes

20 ortogonais do sistema de tradução, onde os efeitos de mutações no gene de sintetase e do promotor de sintetase são determinados.

Kobayashi et al. relataram anteriormente que uma única substituição de aminoácido de Asp286 para Arg (D286R)

25 em *MjtRNA-Tyr(CUA)* aumentou significativamente a taxa global de amino-acilação (67 vezes mais alta kcat/Km) de *MjtRNA-Tyr(CUA)* *in vitro*, principalmente devido ao reconhecimento intensificado (57 vezes mais baixo Km) do anti-códon (CUA)

no tRNA supressor âmbar pela sintetase cognata (Kobayashi et al., *Nat. Struct. Biol.* 10:425-432 (2003)). Usando estas informações, esta mesma mutação D286R foi introduzida no gene de BpaRS em pYR-BpaRS5 por mutagênese direcionada para  
5 sítios. Na realidade, o mutante D286R de BpaRS levou a um aumento de 4,5 vezes na atividade de  $\beta$ -galactosidase (vide Figura 2).

Um promotor *glnS* mutante (SEQ ID NO: 12) que tem uma seqüência TATC no lugar de GATC na região 10 demonstrou  
10 anteriormente aumentar a expressão gênica (Plumbridge e Söll, *Biochemie* 69:539-541 (1987)). Sabendo isto, o promotor *glnS* do tipo selvagem em pYR-BpaRS5 foi substituído pela forma mutada do promotor *glnS*, descrita em Plumbridge e Söll em uma tentativa para melhorar a eficiência do sistema.  
15 Entretanto, depois da inserção desta seqüência promotora em pYR-BpaRS5, o seqüenciamento revelou que além da mutação pretendida, mutações por deleção adicionais não pretendidas também foram identificadas. Dentre os mutantes por deleção testados quanto á atividade de  $\beta$ -galactosidase, um mutante  
20 específico, pYR-BpaRS-TRN, que tem uma única deleção de nucleotídeo (resíduo A na posição 15) além da substituição de um nucleotídeo pretendida (G para T na posição 11) no promotor *glnS*, apresentou um aumento de 5 vezes na atividade de  $\beta$ -galactosidase, em comparação com pYR-BpaRS5 (vide  
25 Figura 2). A seqüência de nucleotídeos completa deste novo domínio do promotor *glnS* mutante, denominado *glnS*-TR, identificado deposi do presente seqüenciamento, está fornecida em SEQ ID NO: 13.

Descrição	Seqüência	SEQ ID NO:
Promotor <i>glnS</i> mutante descrito em Plumbridge e Söll	CGATTATCAATTTTAAAAAACTAACAGTTGTCAGCCT GTCCCGCTTATAATATCATACGCC	12
Promotor <i>glnS</i> -TRN	CGATTATCAATTTTAAAAAACTAACAGTTGTCAGCCT GTCCCGCTTTAATATCATACGCC	13

A expressão de BpaRS sob o controle da forma mutada do promotor *glnS* foi aumentada 2 vezes. Determinado por análise *Western blotting* (vide Figura 4). Um aumento de 1,5 vez adicional na atividade de  $\beta$ -galactosidase foi observado por combinação da substituição D286R de BpaRS a o novo promotor *glnS* mutante, correspondendo a uma eficiência global de supressão de 57% para pYR-BpaRS-TRN(D286R).

### EXEMPLO 3

Geração de Sistemas de Expressão de tRNA  
10 Aperfeiçoados

O presente exemplo descreve a construção de sistemas aperfeiçoados para a expressão de componentes ortogonais de sistemas de tradução, onde os efeitos de colocar múltiplas cópias do *MjtRNA-Tyr(CUA)* em um operon  
15 policistrônico são determinados.

O efeito de múltiplas cópias do tRNA supressor âmbar sobre a eficiência da supressão foi observado. Um operon de *MjtRNA-Tyr(CUA)* policistrônico que contém três cópias do gene de tRNA supressor âmbar sob controle de um  
20 único promotor e terminador *proK* foi construído.

As três seqüências de O-tRNA tandem no operon policistrônico foram separadas umas das outras por seqüências ligantes de tRNA derivadas de seqüências ligantes de tRNA de *E. coli* de ocorrência natural. Estas seqüências

ligantes específicas foram escolhidas porque elas contêm os nucleotídeos (T-1) e A(77). Estas duas posições de nucleotídeos em ligantes de tRNA demonstraram ser ótimas para processamento eficiente de 5' e 3' de precursores de tRNA em seu contexto nativo (isto é, endógeno).

O primeiro e segundo genes de *MjtRNA-Tyr(CUA)* no operon policistrônico recombinante são separados por um ligante de tRNA derivado do ligante que ocorre naturalmente entre os genes do tRNA *valU* e *valX* de *E. coli* (SEQ ID NO: 14). O segundo e o terceiro genes de *MjtRNA-Tyr(CUA)* no operon policistrônico recombinante são separados por um ligante de tRNA derivado do ligante que ocorre naturalmente entre os genes de tRNA *ileT* e *alaT* de *E. coli* (SEQ ID NO: 15). O uso destes ligantes tem uma outra vantagem prática pelo fato de que estes polinucleotídeos contêm sítios de restrição convenientes.

Duas cópias idênticas do operon de tRNA policistrônico sintético que contém três cópias do gene de tRNA supressor foram ligadas para gerar *clusters* de genes com seis cópias do gene de *MjtRNA-Tyr(CUA)*. Os *clusters* de genes montados com três e seis cópias foram clonados em pYR-BpaRS-TRN(D286R) para gerar pYR-BpaRS-3TRN(D286R) e pYR-BpaRS-6TRN(D286R), respectivamente. Quando expressados em *E. coli*, estes plasmídeos produziram um aumento de 30% e 50% na expressão de *MjtRNA-Tyr(CUA)*, respectivamente, determinado por análise *Northern* (vide Figura 3). Este aumento na expressão de mensagem produziu um aumento de 30-40% na atividade de  $\beta$ -galactosidase (vide Figura 2).

Devido ao fato de que os tRNA's de códons raros de *E. coli* codificados nesses plasmídeos podem ser desnecessários para a expressão da maioria das proteínas, estes genes de tRNA de *E. coli* foram removidos do plasmídeo pYR-BpaRS-6TRN(D286R) (ilustrado na Figura 5). Como esperado, a eficiência de supressão deste plasmídeo, determinada pelo ensaio de atividade de  $\beta$ -galactosidase *in vitro* permaneceu igual àquela de seu plasmídeo originário (vide Figura 6).

#### EXEMPLO 4

Expressão de uma Proteína-Modelo que Compreende um Aminoácido Não-natural, Usando Sistemas de Expressão Aperfeiçoados

O presente exemplo descreve a expressão de uma proteína-modelo mioglobina do esperma de baleia, que compreende um aminoácido não-natural, usando os sistemas de expressão aperfeiçoados da invenção.

Para examinar ainda mais os aperfeiçoamentos no rendimento e fidelidade da incorporação de aminoácidos não-naturais em proteínas, usando os sistemas da invenção, um mutante de Ser-4 para Bpa da mioglobina do esperma de baleia (descrita em Chin et al., *Proc. Natl. Acad. Sci. USA* 99:11020-11024 (2002)) foi expressada em *E. coli*. As células de *E. coli* TOP10 (Invitrogen™) co-transformadas com pBAD/Myc-His/MB(S4TAG) e pSup-BpaRS-6TRN(D286R) foram desenvolvidas em meio de Luria-Bertani a 37 °C na presença de Bpa 1 mM. Consistentemente com os dados do ensaio de  $\beta$ -galactosidase *in vivo* acima, a mioglobina mutante de

comprimento inteiro foi produzida com um rendimento purificado global de 40 mg/L, enquanto que os sistemas anteriores produziam 2 mg/L de proteína mutante (Chin et al., *Proc. Natl. Acad. Sci. USA* 99:11020-11024 (2002)).

5 Nenhuma proteína mutante foi observada em gel de SDS-PAGE na ausência do aminoácido (Figura 6). A espectrometria de massas MALDI-TOF da mioglobina mutante que contém Bpa no lugar de Ser-4 deu uma massa média de 18.521, o que está de bom acordo com a massa prevista calculada de 18.520. Nenhuma  
10 evidência de qualquer incorporação de aminoácido natural na posição 4 foi detectada no espectro de massas.

#### EXEMPLO 5

Aplicabilidade Ampla dos sistemas de Expressão  
Aperfeiçoados da Invenção

15 O presente exemplo descreve a expressão de  $\beta$ -galactosidase que compreende quatro aminoácidos não-naturais diferentes, onde os sistemas de expressão usam sintetases mtantes diferentes que têm espcificidades de carregamento para aminoácidos não-naturais diferentes.

20 Para testar a generalidade destes sistemas de expressão da inveção, três amino-acil-tRNA sintetases ortogonais adicionais foram testadas no sistema. Estas O-RS's especificamente amino-acilam (isto é, carregam) o O-tRNA MjtRNA-Tyr(CUA) co, alternadamente, p-acetil-L-fenilalanina  
25 (pAcPhe), p-azido-L-fenilalanina (pAzPhe) e p-iodo-L-fenilalanina (pIPhe) (Figura 1). Estes aminoácidos não-naturais são úteis para experimentos de marcação química (Wang et al., *Proc. Natl. Acad. Sci. USA* 100:56-61 (2003)),

fotorreticulação (Chin et al., *J. Am. Chem. Soc.* 124:9026-9027)), e faseamento cristalográfico por raios X (Xie et al., *Nat. Biotechnol.* 22:1297-1301 (2004)). Os vetores de expressão da invenção, que codificam estes genes mutantes

5 *MjTyrRS*, foram construídos subclonando a respectiva O-RS nos sítios NdeI/PstI de pSup-BpaRS-6TRN(D286R), para produzir os vetores pSup-pAcPheRS-6TRN, pSup-pAzPheRS-6TRN e pSup-pIPheRS-6TRN. Como foi o caso com Bpa, quando as células de *E. coli* que carregam estes plasmídeos são desenvolvidas na

10 presença destes aminoácidos (1 mM), o nível da atividade de  $\beta$ -galactosidase é similar àquele da  $\beta$ -galactosidase do tipo selvagem (Figura 7), indicando a produção da  $\beta$ -galactosidase que compreende o respectivo aminoácido não-natural.

#### EXEMPLO 6

15 Construções de Plasmídeos

##### pYR-BpaRS1

pYR-BpaRS1, um replicon de p15A que contém um marcador de resistência a cloranfenicol, *MjtRNA-Tyr(CUA)* sob controle do promotor *lpp* e terminador *rrnC*, e BpaRS sob

20 controle do promotor *GlnS*, foi gerado inserindo os genes de BpaRS e *MjtRNA-Tyr(CUA)* nos sítios SacI e SpeI do plasmídeo pRARE2 (Novagen).

##### Gene de *MjtRNA-Tyr(CUA)* com o Promotor de Terminador *proK*

25 Um operon monocistrônico de *MjtRNA-Tyr(CUA)*, contendo o promotor e terminador *proK*, foi construído por PCR, usando quatro oligonucleotídeos sintéticos em uma estratégia de sobreposição de PCR, para construir o operon

de tRNA inteiro em uma única reação de PCR:

Iniciador	Sequência	SEQ ID NO:
proK P1	GTGCACGGCTAACTAAGCGGCCTGCTGACTTTCTCGCCGATCAAAAGGC	37
proK T1	CTTTCTCGCCGATCAAAAGGCATTTTGCTATTAAGGGATTGACGAGGGCGTATCTGCGCAGTAAGATGCGCCCCGCATTCCGGCGGTAGTTCAGCAGGGC	38
proK T2	CTTTCTCGCCGATCAAAAGGCATTTTGCTATTAAGGGATTGACGAGGGCGTATCTGCGCAGTAAGATGCGCCCCGCATTCCGGCGGTAGTTCAGCAGGGC	39
proK P2	GCATAAGCTTATGCAAAAAAGCCTGCTCGTTGAGCAGGCTTTTCG	40

O amplicon da PCR foi ampliado subsequentemente por PCR usando 2 iniciadores:

Iniciador	Sequência	SEQ ID NO:
proK-F	AGTCTGATCAGTGCACGGCTAACTAAGCGG	41
proK-R	GCATCTCGAGATGCAAAAAAGCCTGCTCGTTG	42

5 O resultado do amplicon foi inserido entre os sítios BcII e XhoI (sublinhados) de YR-BpaRS1, para gerar o plasmídeo pYR-BpaRS5.

#### Promotor *glnS* Mutante

Um promotor *glnS* mutante foi construído por PCR,  
10 usando quatro oligonucleotídeos sintéticos:



Acionador	Sequência	SEQ ID NO:
glnS P1	CCGAGCTCCCGGGTCATC	43
glnS T1	CCGAGCTCCCGGGTCATCAATCATCCCCATAATCCTTGTTAGATTATCAAT TTTAAAAAACTAACAGTTGTCAGCCTGTC	44
glnS T2	GTCCATATGGGATTCCTCAAAGCGTAAACAACGTATAACGGCGTATGATAT TATAAGCGGGACAGGCTGACAACCTGTTAG	45
glnS P2	GTCCATATGGGATTCCTC	46

O produto foi inserido entre os sítios XmaI e NdeI de pYRBpaRS5. Inúmeros clones foram triados com um ensaio de atividade de *lacZ* *in vivo* e um mutante por deleção específico com uma única base e atividade melhorada foi  
5 identificado e sequenciado (denominado pYR-BpaRS-TRN).

#### Cassetes de Genes de tRNA Tandem

Duas sequências ligantes de tRNA, que ocorrem naturalmente entre os genes *valU* e *valX* (SEQ ID NO: 14 e entre os genes *ileT* e *alaT* (SEQ ID NO: 15) no genoma de *E. coli*,  
10 foram usadas como espaçadores entre os genes de *MjtRNA-Tyr(CUA)*. Estas sequências ligantes contêm os sítios de restrição *BsmAI* e *EarI* aos quais os genes de *MjtRNA-Tyr(CUA)* foram ligados. Estas sequências flanqueadoras contêm também resíduos T(-1) e A(77), que são ótimos para o  
15 processamento eficiente de 5' e 3' dos precursores de tRNA. O cassette de genes de tRNA foi ampliado por PCR, usando três conjuntos de iniciadores:

Conjunto 1		
Acionador	Sequência	SEQ ID NO:
Ligante P1	GTGCACGGCTAACTAAGCGGCCTGCTGACTTTCTCGCCGATCAAAA GGC	47
Ligante P2	TACACGGCGGAGACTACATAAAGTAGTTGGTCCGGCGGGCCGGATT TG	48

	Conjunto 2	
Acionador	Seqüência	SEQ ID NO:
Ligante P3	GTAGTCTCCGCCGTGTAGCAAGAAATTGAGAAGTCCGGCGGTAGTT CAGCAG	49
Ligante P4	AAACCTCTTCAAATTTGCCGTGCAAATTTGGTCCGGCGGGCCGGAT TTG	50

	Conjunto 3	
Acionador	Seqüência	SEQ ID NO:
Ligante P5	GCAAATTTGAAGAGGTTTTAACTACATGTTATCCGGCGGTAGTTCA GCAG	51
proK-R	GCATCTCGAGATGCAAAAAAGCCTGCTCGTTG	52

O produto de cada conjunto foi digerido com *BsmAI* (Conjunto 1), *EatI* (Conjunto 2), ou ambos (conjunto 3). A ligação destes três fragmentos de restrição produziu um operon de tRNA policistrônico que contém três cópias do gene de tRNA conectadas por duas seqüências ligantes de ocorrência natural diferentes. O *cluster* de genes resultante foi ampliado por PCR, usando dois conjuntos de iniciadores;

	Set 4	
Acionador	Seqüência	SEQ ID NO:
Tandem P1	ATCAGTGCACGGCTAACTAAGCGG	53
Tandem P2	GCTGGCATGCATGCAAAAAAGCCTGCTCGTTGAGC	54

	Set 5	
Acionador	Seqüência	SEQ ID NO:
Tandem P3	ATCAGCATGCGGCTAACTAAGCGGCCTGCTG	55
Tandem P4	GCTGCTCGAGATGCAAAAAAGCCTGC	56

Os produtos da PCR a partir dos conjuntos 4 e 5 foram digeridos com *SphI* e ligados entre si para gerar uma

montagem de genes de tRNA tandem unidirecionais, que consiste em dois operons de tRNA policistrônicos idênticos, cada um codificando três genes de tRNA sob controle de um único promotor e terminador *proK*. Cada *cluster* de genes de tRNA, que contém uma ou duas cópias idênticas do operon de tRNA policistrônico, foi clonado nos sítios *ApaI* e *XhoI* de pYR-BpaRS-TRN, para produzir pYR-BpaRS-3TRN e pYR-BpaRS-6TRN, respectivamente.

#### Plasmídeos pSup

Cada um dos doze genes do tRNA de *E. coli*, que eram originalmente codificados no plasmídeo pRARE2, foram removidos de pYR-BpaRS-6TRN(D286R) por digestão com *SpeI* e *DrdI*, e em seguida, tratamento com nuclease de feijão-mungo. A religação dos vetores linearizados gerou pSup-BpaRS-6TRN(D286R). Os genes de *MjtRNA-Tyr(CUA)* mutantes para p-acetil-L-fenilalanina, p-azido-L-fenilalanina e p-iodo-L-fenilalanina foram subclonados a partir de seus plasmídeos pBK correspondentes nos sítios *NdeI* e *PstI* de pSup-BpaRS-6TRN(D286R), para gerar pSup-pAcPheRS-6TRN, pSup-pAzPheRS-6TRN e pSup-pIodoPheRS-6TRN, respectivamente.

#### Plasmídeo Repórter *lacZ* e Ensaio da Atividade de $\beta$ -Galactosidase *in vivo*

O códon de fenilalanina (TCC) no resíduo 13 (sublinhado) da sequência líder (MDPLVTAASVLEFGLFET; SEQ ID NO: 57) localizada a montante do gene *lacZ* de pBAD/Myc-His/LacZ (Invitrogen™) foi mutada para um códon âmbar (TAG) por mutagênese direcionada para sítio, para produzir um plasmídeo repórter LacZ pBAD/Myc-His/LacZ(TAG). Este

plasmídeo foi co-transformado com cada plasmídeo supresor em células de *E. coli* TOP10 (Invitrogen™). As células foram incubadas a 37 °C de um dia para o outro em meio de Luria-Bertani, contendo 0,02% de arabinose e aminoácido não-natural 1 mM. A atividade de LacZ ( $\beta$ -galactosidase) foi medida de acordo com o método descrito por Miller (Miller, J.H. "Experiments in Molecular Genetics (cold Spring Harbor Laboratory, New York, (1972)).

#### EXEMPLO 7

##### 10 Metodologias Genéricas

Células de *E. coli* XL1-Blue (Stratagene) foram usadas para clonar e manter plasmídeos. A DNA polimerase PfuUltra™ High-Fidelity (Stratagene®) foi usada para a reação de polimerase em cadeia (PCR). O kit de mutagênese direcionada para sítios QuikChange® (Stratagene®) foi usado para a mutagênese direcionada para sítios. As seqüências de todos plasmídeos construídas foram verificadas por seqüenciamento.

#### Expressão de Proteínas

20 Um gene de mioglobina do esperma de baleia, mutante, marcado com hexa-histidina no terminal C, com um códon âmbar na posição 4 (Ser-4), foi inserido a partir de pBAD-JYAMB-4TAG entre os sítios NcoI e KpnI de pBAD/Myc-His (Invitrogen™), para gerar pBAD/Myc-His/MB(S4TAG). O  
25 plasmídeo foi co-transformado com pSup-BpaRS-6TRN(D286R) em *E. coli* TOP10 (Invitrogen™). As células foram incubadas a 37 °C em LB contendo 100 mg/mL de carbenicilina, 50 mg/mL de cloranfenicol e Bpa 1 mM. A DO<sub>600</sub> = 0,6, as células foram

induzidas pela adição de 0,2% de arabinose e incubadas por 12 h. As células foram colhidas por centrifugação e lisadas com o reagente BugBuster® (Novagen®). A proteína obtida a partir de corpos de inclusão foi purificada com a resina de afinidade por metais TALON® (Contech®) sob condições desnaturantes, de acordo com o protocolo do fabricante. A proteína purificada foi concentrada por ultrafiltração e analisada por espectrometria de massas MALDI-TOF. A concentração da proteína foi medida pelo método de Bradford.

#### Análise Northern

Células de *E. coli* TOP10 (Invitrogen™) transformadas com cada plasmídeo de supressão foram incubadas em LB a 37 °C. Em DO<sub>600</sub> = 0,8, as células foram colhidas. O tRNA total foi isolado por extração com fenol e fracionamento com isopropanol, como descrito anteriormente (Deutscher e Hilderman, "Isolation and partial characterization of *Escherichia coli* mutants with low levels of transfer ribonucleic acid nucleotidyltransferase", *J. Bacteriol.* 118:621-627 (1974)). As amostras de RNA foram separadas em um gel de poliacrilamida desnaturante a 15% e transferidas para a membrana GeneScreen Plus® (PerkinElmer®). A membrana foi hibridizada de um dia para o outro a 55 °C com:

5'-biotina-CCCTGCTGAACTACCGCC-3' (SEQ ID NO: 58).

#### Análise Western da Expressão de Bpa

Um gene de BpaRS marcado com hexa-histidina no terminal C foi construído por PCR e inserido entre os sítios NdeI e PstI de pYR-BpaRS5 e pYR-BpaRS-TRN para gerar pYR-

BpaRS5(c-His) e pYR-BpaRS-TRN(C-His), respectivamente. As células de *E. coli* TOP10 transformadas com cada plasmídeo foram incubadas em LB a 37 °C. As células foram colhidas a  $DO_{600} = 1$  e lisadas com reagente BugBuster. A proteína total  
5 foi separada em gel de poliacrilamida a 10% e transferidas para uma membrana de PVDF (Invitrogen). A membrana foi hibridizada com conjugado de HRP/anticorpo Anti-His (terminal C) (Invitrogen) e detectada por quimioluminescência.

Deve-se entender que os exemplos e modalidades  
10 aqui descritas são meramente com propósitos ilustrativos e que várias modificações ou mudanças, à luz delas, serão sugeridas pelos versados nessas técnicas e estão incluídas dentro do espírito e âmbito deste pedido de patente e do âmbito das reivindicações apensadas.

15 Embora a invenção precedente tenha sido descrita mais detalhadamente com propósitos de clareza e compreensão, deve ficar evidente para os versados nessas técnicas a partir da leitura deste relatório descritivo que várias mudanças na forma e no detalhe podem ser feitas sem fugir do  
20 verdadeiro âmbito da invenção. Por exemplo, todas técnicas e aparelhos descritos acima podem ser usados em várias combinações. Todas publicações, patentes, pedidos de patente, e/ou outros documentos citados neste pedido de patente são incorporados como refer~enci em sua totalidade  
25 com todos os propósitos e até o mesmo grau caso cada publicação, patente, pedido de patente e/ou outro documento fosse indicado individualmente para ser incorporado como referência pra todos propósitos.

## REIVINDICAÇÕES

1. Composição que compreende uma construção de ácido nucléico, **CARACTERIZADA** pelo fato de que compreende:

(a) seqüências de nucleotídeos promotoras e  
5 terminadoras derivadas de um gene de tRNA de prolina de *Escherichia coli* e uma seqüência de nucleotídeos expressável, onde as ditas seqüências promotoras e terminadoras estão ambas ligadas operacionalmente à dita seqüência de nucleotídeos expressável, e onde a dita  
10 seqüência de nucleotídeos expressável é heteróloga às ditas seqüências de nucleotídeos promotoras e terminadoras;

(b) uma seqüência de nucleotídeos promotora correspondente a um promotor *glnS* de *E. coli* modificado que tem uma seqüência de nucleotídeos de SEQ ID NO: 13 e uma  
15 seqüência de nucleotídeos expressável, onde a dita seqüência de nucleotídeo do promotor *glnS* de *E. coli* modificado está ligada operacionalmente à dita seqüência de nucleotídeos expressável;

(c) uma seqüência de nucleotídeos que codifica  
20 que codifica um tRNA ortogonal (O-tRNA) e uma seqüência de nucleotídeos que codifica uma amino-acil-tRNA sintetase ortogonal (O-RS), onde a dita O-RS preferncialmente amino-acila o dito O-tRNA com um aminoácido não-natural; ou

(d) um operon policistrônico que compreende uma  
25 pluralidade de seqüências de nucleotídeos do gene de tRNA, onde pelo menos um gene de tRNA é separado de pelo menos um gene de tRNA adjacente por um ligante de polinucleotídeos heterólogo derivado de um ligante de polinucleotídeo de

ocorrência natural de um operon de tRNA de ocorrência natural.

2. Composição, de acordo com a reivindicação 1, **CARACTERIZADA** pelo fato de que a dita construção compreende  
5 (a) e também (b), onde as ditas seqüências de nucleotídeos expressáveis em (a) e (b) são diferentes.

3. Composição, de acordo com a reivindicação 1, **CARACTERIZADA** pelo fato de que a dita construção compreende  
10 (a), (b) e (c), onde a dita seqüência de nucleotídeos expressável em (a) codifica o dito O-tRNA, e onde a dita seqüência de nucleotídeos expressável em (b) codifica a dita O-RS.

4. Composição, de acordo com a reivindicação 1, **CARACTERIZADA** pelo fato de que o dito gene de tRNA de  
15 prolina de *E. coli* é selecionado entre os genes de tRNA *proK*, *proL* e *proM* de *E. coli*.

5. Composição, de acordo com a reivindicação 1, **CARACTERIZADA** pelo fato de que as ditas seqüências  
20 promotoras e terminadoras de tRNA de prolina de *E. coli* derivadas das seqüências promotoras e terminadoras de *proK* de *E. coli* fornecidas em SEQ ID NOS: 32 (promotora) e 33 (terminadora), respectivamente.

6. Composição, de acordo com a reivindicação 1, **CARACTERIZADA** pelo fato de que a dita seqüência de  
25 nucleotídeos expressável em (a) é um operon policistrônico que compreende uma pluralidade de uma ou mais seqüências de nucleotídeos.

7. Composição, de acordo com a reivindicação 1,



**CARACTERIZADA** pelo fato de que a dita seqüência de nucleotídeos expressável em (a) codifica um tRNA.

8. Composição, de acordo com a reivindicação 7, **CARACTERIZADA** pelo fato de que o dito tRNA é derivado de um  
5 ou mais tRNA's de *Archaea*.

9. Composição, de acordo com a reivindicação 7, **CARACTERIZADA** pelo fato de que o dito tRNA é codificado por uma seqüência de nucleotídeos de SEQ ID NO: 1 (*MjtRNA-Tyr(CUA)*).

10 10. Composição, de acordo com a reivindicação 7, **CARACTERIZADA** pelo fato de que a dita seqüência de nucleotídeos expressável é um operon policistrônico que compreende uma pluralidade da seqüência de nucleotídeos de SEQ ID NO: 1 (*MjtRNA-Tyr(CUA)*).

15 11. Composição, de acordo com a reivindicação 10, **CARACTERIZADO** pelo fato de que a dita seqüência de nucleotídeos expressável é uma pluralidade do dito operon policistrônico.

12. Composição, de acordo com a reivindicação 1,  
20 **CARACTERIZADA** pelo fato de que a dita seqüência de nucleotídeos expressável de (b) codifica um polipeptídeo.

13. Composição, de acordo com a reivindicação 1, **CARACTERIZADA** pelo fato de que a dita seqüência de nucleotídeos expressável de (b) codifica uma amino-acil-tRNA  
25 sintetase.

14. Composição, de acordo com a reivindicação 1, **CARACTERIZADA** pelo fato de que a dita seqüência de nucleotídeos expressável de (b) codifica uma amino-acil-tRNA

sintetase ortogonal (O-RS), onde a dita O-RS preferencialmente amino-acila um O-tRNA correspondente com um aminoácido não-natural.

15 15. Composição, de acordo com a reivindicação 14, **CARACTERIZADA** pelo fato de que a dita O-RS é derivada de uma amino-acil-tRNA sintetase de *Methanococcus jannaschii*.

16. Composição, de acordo com a reivindicação 14, **CARACTERIZADA** pelo fato de que a dita O-RS é derivada de uma tirosil-tRNA sintetase de *Methanococcus jannaschii*.

10 17. Composição, de acordo com a reivindicação 16, **CARACTERIZADA** pelo fato de que a dita O-RS tem uma substituição de ácido aspártico para arginina na posição do aminoácido 286 ou em uma posição análoga à posição 286, em relação à sequência de aminoácidos de tirosil-tRNA sintetase  
15 do tipo selvagem de *Methanococcus jannaschii* fornecida em SEQ ID NO: 2 (MjtRNA-TyrRS do tipo selvagem).

18. Célula hospedeira, **CARACTERIZADA** pelo fato de que compreende a composição de acordo com a reivindicação 1.

20 19. Célula hospedeira, de acordo com a reivindicação 18, **CARACTERIZADA** pelo fato de que a dita célula hospedeira é uma célula hospedeira eubacteriana.

20. Célula hospedeira, de acordo com a reivindicação 18, **CARACTERIZADA** pelo fato de que a dita célula hospedeira é uma célula de *E. coli*.

25 21. Composição, de acordo com a reivindicação 1, **CARACTERIZADA** pelo fato de que o dito operon policistrônico de (d) compreende uma pluralidade de ligantes de polinucleotídeos heterólogos idênticos.

22. Composição, de acordo com a reivindicação 1, **CARACTERIZADA** pelo fato de que o dito operon policistônico de (d) compreende uma pluralidade de ligantes de polinucleotídeos heterólogos, onde pelo menos dois dos  
5 ligantes de polinucleotídeos heterólogos são diferentes.

23. Composição, de acordo com a reivindicação 1, **CARACTERIZADA** pelo fato de que o dito ligante de polinucleotídeo heterólogo de (d) compreende um nucleotídeo timidina no terminal 5', um nucleotídeo adenosina no  
10 terminal 3', ou um nucleotídeo timidina no terminal 5' e também um nucleotídeo adenosina no terminal 3'.

24. Composição, de acordo com a reivindicação 1(d), **CARACTERIZADA** pelo fato de que o dito ligante de polinucleotídeo heterólogo é derivado de um ligante de  
15 polinucleotídeo de ocorrência natural localizado entre os genes de tRNA de *Escherichia coli* endógenos selecionados entre: *valU* e *valX*; *ileT* e *alaT*; *serV* e *argV*; *valV* e *valW*; *glyT* e *thrT*; *metT* e *leuW*; *glnW* e *metU*; *hisR* e *leuT*; *glnU* e *glnW*; *leuP* e *leuV*; *glnV* e *glnX*; *alaW* e *alaX*; *ileU* e *alaU*;  
20 *ileV* e *alaV*; *metU* e *glnV*; *glyW* e *cysT*; *argX* e *hisR*; e *argY* e *argZ*.

25. Composição, de acordo com a reivindicação 1, **CARACTERIZADA** pelo fato de que o dito ligante de polinucleotídeo heterólogo de (d) é derivado da sequência de  
25 nucleotídeos de SEQ ID NO: 14 (ligante *valU/valX*) ou 15 (ligante *ileT/alaT*).

26. Sistema de tradução para a expressão de um polipeptídeo de interesse, compreendendo pelo menos um

aminoácido não-natural em uma posição especificada, **CARACTERIZADO** pelo fato de que o sistema compreende:

(a) um aminoácido não-natural;

(b) uma construção de ácido nucléico, sendo que a dita construção compreende uma seqüência de nucleotídeos que codifica um tRNA ortogonal (O-tRNA) e uma seqüência de nucleotídeos que codifica uma amino-acil-tRNA sintetase ortogonal (O-RS), onde a dita O-RS preferencialmente amino-acila o dito O-tRNA com o dito aminoácido não-natural; e

(c) um polinucleotídeo que codifica o dito polipeptídeo de interesse, sendo que o dito polinucleotídeo compreende pelo menos um códon seletor que é reconhecido pelo dito O-tRNA, onde a posição do códon seletor no polinucleotídeo controla a posição especificada do aminoácido não-natural no polipeptídeo de interesse depois da expressão do polinucleotídeo para produzir o polipeptídeo.

27. Sistema de tradução, de acordo com a reivindicação 26, **CARACTERIZADO** pelo fato de que a dita construção de ácido nucléico compreende pelo menos um entre:

(i) seqüências de nucleotídeos promotora e terminadora derivadas de um gene de tRNA de prolina de *Escherichia coli*, onde as ditas seqüências promotora e terminadora estão ambas ligadas operacionalmente à dita seqüência de nucleotídeos que compreende ou codifica o dito O-tRNA, e onde o dito O-tRNA é heterólogo às ditas seqüências de nucleotídeos promotora e terminadora;

(ii) uma seqüência de nucleotídeos correspondente

a um promotor *glnS* de *E. coli* modificado que tem uma sequência de nucleotídeos de SEQ ID NO: 13, onde a dita sequência de nucleotídeos do *glnS* de *E. coli* modificado está ligada operacionalmente à dita sequência de nucleotídeos que  
5 codifica a dita O-RS; e

(iii) um operon policistrônico que compreende uma pluralidade das ditas seqüências de nucleotídeos do gene de O-tRNA, onde pelo menos um gene de O-tRNA é separado de pelo menos um gene de O-tRNA adjacente por um ligante de  
10 polinucleotídeo heterólogo derivado de um ligante de polinucleotídeo de ocorrência natural de um operon de tRNA de ocorrência natural.

28. Sistema de tradução, de acordo com a reivindicação 27, **CARACTERIZADO** pelo fato de que o dito gene  
15 de tRNA de prolina de *E. coli* é selecionado entre os genes de tRNA *proK*, *proL* e *proM* de *E. coli*.

29. Sistema de tradução, de acordo com a reivindicação 27, **CARACTERIZADO** pelo fato de que as ditas seqüências promotoras e terminadoras de tRNA de prolina de  
20 *E. coli* são derivadas das seqüências promotoras e terminadoras de *proK* de *E. coli* fornecidas em SEQ ID NOS: 32 (promotora) e 33 (terminadora), respectivamente.

30. Sistema de tradução, de acordo com a reivindicação 27, **CARACTERIZADO** pelo fato de que o dito  
25 operon policistrônico compreende uma pluralidade de ligantes de polinucleotídeos heterólogos idênticos.

31. Sistema de tradução, de acordo com a reivindicação 27, **CARACTERIZADO** pelo fato de que o dito

operon policistrônico compreende uma pluralidade de ligantes de polinucleotídeos heterólogos, onde pelo menos dois dos ligantes de polinucleotídeos heterólogos são diferentes.

32. Sistema de tradução, de acordo com a reivindicação 27, **CARACTERIZADO** pelo fato de que dito ligante de polinucleotídeo heterólogo compreende um nucleotídeo timidina no terminal 5', ou um nucleotídeo adenosina no terminal 3', ou um nucleotídeo timidina no terminal 5' e também um nucleotídeo adenosina no terminal 3'.

33. Sistema de tradução, de acordo com a reivindicação 27, **CARACTERIZADO** pelo fato de que dito ligante de polinucleotídeo heterólogo é derivado do ligante de polinucleotídeo de ocorrência natural localizado entre os genes de tRNA de *Escherichhia coli* endógenos selecionados entre: *valU* e *valX*; *ileT* e *alaT*; *serV* e *argV*; *valV* e *valW*; *glyT* e *thrT*; *metT* e *leuW*; *glnW* e *metU*; *hisR* e *leuT*; *glnU* e *glnW*; *leuP* e *leuV*; *glnV* e *glnX*; *alaW* e *alaX*; *ileU* e *alaU*; *ileV* e *alaV*; *metU* e *glnV*; *glyW* e *cystT*; *argX* e *hisR*; e *argY* e *argZ*.

34. Sistema de tradução, de acordo com a reivindicação 27, **CARACTERIZADO** pelo fato de que dito ligante de polinucleotídeo heterólogo é derivado da sequência de nucleotídeos de SEQ ID NO: 14 (ligante *valU/valX*) ou 15 (ligante *ileT/alaT*).

35. Sistema de tradução, de acordo com a reivindicação 26, **CARACTERIZADO** pelo fato de que o dito O-tRNA é derivado de um ou mais tRNA's de *Archaea*.

36. Sistema de tradução, de acordo com a

reivindicação 26, **CARACTERIZADO** pelo fato de que a sequência de nucleotídeos que codifica um O-tRNA é um operon policistrônico que compreende uma pluralidade de sequências de nucleotídeos que codificam um O-tRNA.

5                37. Sistema de tradução, de acordo com a reivindicação 26, **CARACTERIZADO** pelo fato de que a dita sequência de nucleotídeos que codifica um O-tRNA compreende uma sequência de nucleotídeos de SEQ ID NO: 1 (*MjtRNA-Tyr(CUA)*).

10              38. Sistema de tradução, de acordo com a reivindicação 26, **CARACTERIZADO** pelo fato de que a dita sequência de nucleotídeos que codifica um O-tRNA é um operon policistrônico que compreende uma pluralidade da sequência de nucleotídeos de SEQ ID NO: 1 (*MjtRNA-Tyr(CUA)*).

15              39. Sistema de tradução, de acordo com a reivindicação 26, **CARACTERIZADO** pelo fato de que a dita O-RS é derivada de uma amino-acil-tRNA sintetase de *Methanococcus jannaschii*.

20              40. Sistema de tradução, de acordo com a reivindicação 26, **CARACTERIZADO** pelo fato de que a dita O-RS é derivada de uma tirosil-tRNA sintetase de *Methanococcus jannaschii*.

25              41. Sistema de tradução, de acordo com a reivindicação 26, **CARACTERIZADO** pelo fato de que a dita O-RS tem uma substituição de ácido aspártico para arginina na posição do aminoácido 286 ou em uma posição análoga à posição 286, em relação à sequência de aminoácidos de tirosil-tRNA sintetase do tipo selvagem de *Methanococcus*

*jannaschii* fornecida em SEQ ID NO: 2 (*Mjt*tRNA-TyrRS do tipo selvagem).

42. Sistema de tradução, de acordo com a reivindicação 26, **CARACTERIZADO** pelo fato de que compreende  
5 uma célula hospedeira que compreende (a), (b) e (c).

43. Sistema de tradução, de acordo com a reivindicação 42, **CARACTERIZADO** pelo fato de que a dita célula hospedeira é uma célula hospedeira eubacteriana.

44. Sistema de tradução, de acordo com a  
10 reivindicação 42, **CARACTERIZADO** pelo fato de que a dita célula hospedeira é uma célula de *E. coli*.

45. Método para produzir, em uma célula hospedeira, um polipeptídeo de interesse que compreende um aminoácido não-natural em uma posição especificada,  
15 **CARACTERIZADO** pelo fato de que o método compreende:

(a) disponibilizar:

(i) um aminoácido não-natural;

(ii) uma construção de ácido nucléico que compreende uma sequência de nucleotídeos que codifica um  
20 tRNA ortogonal (O-tRNA) e uma sequência de nucleotídeos que codifica uma amino-acil-tRNA sintetase ortogonal (O-RS), onde a dita O-RS preferencialmente amino-acila o dito O-tRNA com o dito aminoácido não-natural; e

(iii) um polinucleotídeo que codifica o dito  
25 polipeptídeo de interesse, sendo que o dito polinucleotídeo compreende pelo menos um códon seletor que é reconhecido pelo dito O-tRNA, onde a posição do códon seletor se correlaciona com a posição especificada do aminoácido não-



natural no polipeptídeo de interesse;

(iv) uma célula hospedeira que compreende (i), (ii) e (iii); e

(b) desenvolver a dita célula hospedeira; e

5 (c) incorporar o dito aminoácido não-natural na posição especificada no dito polipeptídeo durante a tradução do dito polipeptídeo na dita célula hospedeira, produzindo desta forma o dito polipeptídeo de interesse que compreende o dito aminoácido não-natural na posição especificada.

10 46. Método, de acordo com a reivindicação 45, **CARACTERIZADO** pelo fato de que a dita disponibilização de uma construção de ácido nucléico compreende disponibilizar uma sequência de nucleotídeos que codifica um O-tRNA derivado de um ou mais tRNA's de *Archaea*.

15 47. Método, de acordo com a reivindicação 45, **CARACTERIZADO** pelo fato de que a dita disponibilização de uma construção de ácido nucléico compreende disponibilizar um operon policistrônico que compreende uma pluralidade de sequências de nucleotídeos que codificam uma ou mais  
20 espécies de O-tRNA.

48. Método, de acordo com a reivindicação 45, **CARACTERIZADO** pelo fato de que a dita disponibilização de uma construção de ácido nucléico compreende disponibilizar uma sequência de nucleotídeos que codifica um O-tRNA que  
25 compreende uma sequência de nucleotídeos de SEQ ID NO: 1 (MjtRNA-Tyr(CUA)).

49. Método, de acordo com a reivindicação 45, **CARACTERIZADO** pelo fato de que a dita disponibilização de

uma construção de ácido nucléico compreende disponibilizar um operon policistrônico que compreende uma pluralidade da seqüência de nucleotídeos de SEQ ID NO: 1 (*MjtRNA-Tyr(CUA)*).

50. Método, de acordo com a reivindicação 45,  
5 **CARACTERIZADO** pelo fato de que a dita disponibilização de uma construção de ácido nucléico compreende disponibilizar uma seqüência de nucleotídeos que codifica uma O-RS derivada de uma amino-acil-tRNA sintetase de *Methanococcus jannaschii*.

10 51. Método, de acordo com a reivindicação 45, **CARACTERIZADO** pelo fato de que a dita disponibilização de uma construção de ácido nucléico compreende disponibilizar uma seqüência de nucleotídeos que codifica uma O-RS derivada de uma tirosil-tRNA sintetase de *Methanococcus jannaschii*.

15 52. Método, de acordo com a reivindicação 45, **CARACTERIZADO** pelo fato de que a dita disponibilização de uma construção de ácido nucléico compreende disponibilizar uma seqüência de nucleotídeos que codifica uma O-RS que tem uma substituição de ácido aspártico para arginina na posição  
20 do aminoácido 286 ou em uma posição análoga à posição 286, em relação à seqüência de aminoácidos de tirosil-tRNA sintetase do tipo selvagem de *Methanococcus jannaschii* fornecida em SEQ ID NO: 2 (*MjtRNA-TyrRS* do tipo selvagem).

53. Método, de acordo com a reivindicação 45,  
25 **CARACTERIZADO** pelo fato de que a dita disponibilização de uma construção de ácido nucléico compreende disponibilizar pelo menos um entre:

(I) seqüências de nucleotídeos promotora e

terminadora derivadas de um gene de tRNA de prolina de *Escherichia coli*, onde as ditas seqüências promotora e terminadora estão ambas ligadas operacionalmente à dita seqüência de nucleotídeos que codifica o dito O-tRNA, e onde  
 5 a dita seqüência de que codifica o dito O-tRNA é heteróloga às ditas seqüências promotora e terminadora;

(II) uma seqüência de nucleotídeos promotora correspondente a um promotor *glnS* de *E. coli* modificado, onde a dita seqüência de nucleotídeos do *glnS* de *E. coli*  
 10 modificado está ligada operacionalmente à dita seqüência de nucleotídeos que codifica a dita O-RS; e

(III) um operon policistrônico que compreende uma pluralidade das ditas seqüências de nucleotídeos do gene de O-tRNA, onde pelo menos um gene de O-tRNA é separado de  
 15 pelo menos um gene de O-tRNA adjacente por um ligante de polinucleotídeo heterólogo derivado de um ligante de polinucleotídeo de ocorrência natural de um operon de tRNA de ocorrência natural.

54. Método, de acordo com a reivindicação 53,  
 20 **CARACTERIZADO** pelo fato de que a dita disponibilização de uma construção de ácido nucléico de (I) compreende disponibilizar seqüências de nucleotídeos correspondentes às seqüências promotora e terminadora do tRNA de prolina de *Escherichia coli*, sendo o dito tRNA de prolina de *E. coli*  
 25 selecionado entre o tRNA *proK*, *proL* e *proM* de *E. coli*.

55. Método, de acordo com a reivindicação 53,  
**CARACTERIZADO** pelo fato de que a dita disponibilização de uma construção de ácido nucléico de (I) compreende

disponibilizar seqüências de nucleotídeos correspondentes às seqüências promotora e terminadora do tRNA de prolina de *Escherichia coli*, as ditas seqüências promotoras e terminadoras de tRNA de prolina de *E. coli* compreendendo as  
5 seqüências promotoras e terminadoras de *proK* de *E. coli* fornecidas em SEQ ID NOS: 32 (promotora) e 33 (terminadora), respectivamente.

56. Método, de acordo com a reivindicação 53, **CARACTERIZADO** pelo fato de que a dita disponibilização de um  
10 operon policistrônico de (III) compreende disponibilizar uma pluralidade de ligantes de polinucleotídeos heterólogos idênticos.

57. Método, de acordo com a reivindicação 53, **CARACTERIZADO** pelo fato de que a dita disponibilização de um  
15 operon policistrônico de (III) compreende disponibilizar uma pluralidade de ligantes de polinucleotídeos heterólogos, onde pelo menos dois dos ligantes de polinucleotídeos heterólogos são diferentes.

58. Método, de acordo com a reivindicação 53, **CARACTERIZADO** pelo fato de que a dita disponibilização de um  
20 operon policistrônico de (III) compreende disponibilizar um ligante de polinucleotídeo heterólogo que compreende um nucleotídeo timidina no terminal 5', um nucleotídeo adenosina no terminal 3', ou um nucleotídeo timidina no  
25 terminal 5' e também um nucleotídeo adenosina no terminal 3'.

59. Método, de acordo com a reivindicação 53, **CARACTERIZADO** pelo fato de que a dita disponibilização de um operon policistrônico de (III) compreende disponibilizar um

ligante de polinucleotídeo heterólogo derivado do ligante de polinucleotídeo de ocorrência natural localizado entre os genes de tRNA de *Escherichia coli* endógenos selecionados entre: *valU* e *valX*; *ileT* e *alaT*; *serV* e *argV*; *valV* e *valW*;  
 5 *glyT* e *thrT*; *metT* e *leuW*; *glnW* e *metU*; *hisR* e *leuT*; *glnU* e *glnW*; *leuP* e *leuV*; *glnV* e *glnX*; *alaW* e *alaX*; *ileU* e *alaU*; *ileV* e *alaV*; *metU* e *glnV*; *glyW* e *cysT*; *argX* e *hisR*; e *argY* e *argZ*.

60. Método, de acordo com a reivindicação 53,  
 10 **CARACTERIZADO** pelo fato de que a dita disponibilização de um operon policistrônico de (III) compreende disponibilizar um ligante de polinucleotídeo heterólogo derivado da sequência de nucleotídeos de SEQ ID NO: 14 (ligante *valU/valX*) ou 15 (ligante *ileT/alaT*).

15 61. Método, de acordo com a reivindicação 45, **CARACTERIZADO** pelo fato de que a dita disponibilização de uma célula hospedeira compreende disponibilizar uma célula hospedeira eubacteriana.

62. Método, de acordo com a reivindicação 45,  
 20 **CARACTERIZADO** pelo fato de que a dita disponibilização de uma célula hospedeira compreende disponibilizar uma célula hospedeira de *Escherichia coli*.

63. Método para produzir, em uma célula hospedeira, um polipeptídeo de interesse que compreende um  
 25 aminoácido não-natural em uma posição especificada, **CARACTERIZADO** pelo fato de que o método compreende:

(a) disponibilizar:

(i) um aminoácido não-natural;

(ii) uma construção de ácido nucléico que compreende uma seqüência de nucleotídeos que codifica um tRNA ortogonal (O-tRNA);

(iii) uma construção de ácido nucléico que  
5 compreende seqüência de nucleotídeos que codifica uma amino-acil-tRNA sintetase ortogonal (O-RS), onde a dita O-RS preferencialmente amino-acila o dito O-tRNA com o dito aminoácido não-natural;

(iv) um polinucleotídeo que codifica o dito  
10 polipeptídeo de interesse, sendo que o dito polinucleotídeo compreende pelo menos um códon seletor que é reconhecido pelo dito O-tRNA, e onde a posição do códon seletor se correlaciona com a posição especificada do aminoácido não-natural no polipeptídeo de interesse; e

15 (v) uma célula hospedeira que compreende (i), (ii) e (iii) e (iv);

onde as ditas construções de ácidos nucléicos de (ii) e (iii) compreendem coletivamente pelo menos um entre:

(I) seqüências de nucleotídeos promotora e  
20 terminadora derivadas de um gene de tRNA de prolina de *Escherichia coli*, onde as ditas seqüências promotora e terminadora estão ambas ligadas operacionalmente à dita seqüência de nucleotídeos que codifica o dito O-tRNA, e onde a dita seqüência de nucleotídeos que codifica o dito O-tRNA  
25 é heteróloga às ditas seqüências promotora e terminadora;

(II) uma seqüência de nucleotídeos promotora correspondente a um promotor *glnS* de *E. coli* modificado que tem a seqüência de nucleotídeos de SEQ ID NO: 13, onde a

dita seqüência de nucleotídeos do *glnS* de *E. coli* modificado está ligada operacionalmente à dita seqüência de nucleotídeos que codifica a dita O-RS; e

(III) um operon policistrônico que compreende  
 5 uma pluralidade das ditas seqüências de nucleotídeos do gene de O-tRNA, onde pelo menos um gene de O-tRNA é separado de pelo menos um gene de O-tRNA adjacente por um ligante de polinucleotídeo heterólogo derivado de um operon de tRNA de ocorrência natural.

10 (b) desenvolver a dita célula hospedeira; e

(c) incorporar o dito aminoácido não-natural na  
 posição especificada no dito polipeptídeo durante a tradução  
 do dito polipeptídeo na dita célula hospedeira, produzindo  
 desta forma o dito polipeptídeo de interesse que compreende  
 15 o dito aminoácido não-natural na posição especificada.

64. Método, de acordo com a reivindicação 63,  
**CARACTERIZADO** pelo fato de que a dita disponibilização de  
 uma construção de ácido nucléico compreende disponibilizar  
 uma seqüência de nucleotídeos que codifica um O-tRNA  
 20 derivado de um ou mais tRNA's de *Archaea*.

65. Método, de acordo com a reivindicação 63,  
**CARACTERIZADO** pelo fato de que a dita disponibilização de  
 uma construção de ácido nucléico compreende disponibilizar  
 um operon policistrônico que compreende uma pluralidade de  
 25 seqüências de nucleotídeos que codificam uma ou mais  
 espécies de O-tRNA.

66. Método, de acordo com a reivindicação 63,  
**CARACTERIZADO** pelo fato de que a dita disponibilização de

uma construção de ácido nucléico compreende disponibilizar uma seqüência de nucleotídeos que codifica um O-tRNA que compreende uma seqüência de nucleotídeos de SEQ ID NO: 1 (*MjtRNA-Tyr(CUA)*).

5                   67. Método, de acordo com a reivindicação 63, **CARACTERIZADO** pelo fato de que a dita disponibilização de uma construção de ácido nucléico compreende disponibilizar um operon policistrônico que compreende uma pluralidade da seqüência de nucleotídeos de SEQ ID NO: 1 (*MjtRNA-Tyr(CUA)*).

10                   68. Método, de acordo com a reivindicação 63, **CARACTERIZADO** pelo fato de que a dita disponibilização de uma construção de ácido nucléico compreende disponibilizar uma seqüência de nucleotídeos que codifica uma O-RS derivada de uma amino-acil-tRNA sintetase de *Methanococcus*  
15 *jannaschii*.

                  69. Método, de acordo com a reivindicação 63, **CARACTERIZADO** pelo fato de que a dita disponibilização de uma construção de ácido nucléico compreende disponibilizar uma seqüência de nucleotídeos que codifica uma O-RS derivada  
20 de uma tirosil-tRNA sintetase de *Methanococcus jannaschii*.

                  70. Método, de acordo com a reivindicação 63, **CARACTERIZADO** pelo fato de que a dita disponibilização de uma construção de ácido nucléico compreende disponibilizar uma seqüência de nucleotídeos que codifica uma O-RS que tem  
25 uma substituição de ácido aspártico para arginina na posição do aminoácido 286 ou em uma posição análoga à posição 286, em relação à seqüência de aminoácidos de tirosil-tRNA sintetase do tipo selvagem de *Methanococcus jannaschii*



fornecida em SEQ ID NO: 2 (*MjtRNA-TyrRS* do tipo selvagem).

71. Método, de acordo com a reivindicação 63,  
**CARACTERIZADO** pelo fato de que a dita disponibilização de  
uma construção de ácido nucléico de (I) compreende  
5 disponibilizar seqüências de nucleotídeos correspondentes às  
seqüências promotora e terminadora do tRNA de prolina de  
*Escherichia coli*, sendo o dito tRNA de prolina de *E. coli*  
selecionado entre o tRNA *proK*, *proL* e *proM* de *E. coli*.

72. Método, de acordo com a reivindicação 63,  
10 **CARACTERIZADO** pelo fato de que a dita disponibilização de  
uma construção de ácido nucléico de (I) compreende  
disponibilizar seqüências de nucleotídeos correspondentes às  
seqüências promotora e terminadora do tRNA de prolina de  
*Escherichia coli*, as ditas seqüências promotoras e  
15 terminadoras de tRNA de prolina de *E. coli* compreendendo as  
seqüências promotoras e terminadoras de *proK* de *E. coli*  
fornecidas em SEQ ID NOS: 32 (promotora) e 33 (terminadora),  
respectivamente.

73. Método, de acordo com a reivindicação 63,  
20 **CARACTERIZADO** pelo fato de que a dita disponibilização de  
uma construção de ácido nucléico de (III) compreende  
disponibilizar uma pluralidade de ligantes de  
polinucleotídeos heterólogos idênticos.

74. Método, de acordo com a reivindicação 63,  
25 **CARACTERIZADO** pelo fato de que a dita disponibilização de  
uma construção de ácido nucléico de (III) compreende  
disponibilizar uma pluralidade de ligantes de  
polinucleotídeos heterólogos, onde pelo menos dois dos

ligantes de polinucleotídeos heterólogos são diferentes.

75. Método, de acordo com a reivindicação 63, **CARACTERIZADO** pelo fato de que a dita disponibilização de um operon policistrônico de (III) compreende disponibilizar um  
5 ligante de polinucleotídeo heterólogo que compreende um nucleotídeo timidina no terminal 5', um nucleotídeo adenosina no terminal 3', ou um nucleotídeo timidina no terminal 5' e também um nucleotídeo adenosina no terminal 3'.

76. Método, de acordo com a reivindicação 63,  
10 **CARACTERIZADO** pelo fato de que a dita disponibilização de um operon policistrônico de (III) compreende disponibilizar um ligante de polinucleotídeo heterólogo derivado do ligante de polinucleotídeo de ocorrência natural localizado entre os genes de tRNA de *Escherichia coli* endógenos selecionados  
15 entre: *valU* e *valX*; *ileT* e *alaT*; *serV* e *argV*; *valV* e *valW*; *glyT* e *thrT*; *metT* e *leuW*; *glnW* e *metU*; *hisR* e *leuT*; *glnU* e *glnW*; *leuP* e *leuV*; *glnV* e *glnX*; *alaW* e *alaX*; *ileU* e *alaU*; *ileV* e *alaV*; *metU* e *glnV*; *glyW* e *cysT*; *argX* e *hisR*; e *argY* e *argZ*.

20 77. Método, de acordo com a reivindicação 63, **CARACTERIZADO** pelo fato de que a dita disponibilização de um operon policistrônico de (III) compreende disponibilizar um ligante de polinucleotídeo heterólogo derivado da sequência de nucleotídeos de SEQ ID NO: 14 (ligante *valU/valX*) ou 15  
25 (ligante *ileT/alaT*).

78. Método, de acordo com a reivindicação 63, **CARACTERIZADO** pelo fato de que a dita disponibilização de uma célula hospedeira compreende disponibilizar uma célula

hospedeira eubacteriana.

79. Método, de acordo com a reivindicação 63, **CARACTERIZADO** pelo fato de que a dita disponibilização de uma célula hospedeira compreende disponibilizar uma célula  
5 hospedeira de *Escherichia coli*.

80. Método para produzir um sistema de tradução para a expressão de um polipeptídeo de interesse que compreende pelo menos um aminoácido não-natural em uma posição especificada, **CARACTERIZADO** pelo fato de que o  
10 método compreende disponibilizar:

(a) um aminoácido não-natural;

(b) uma construção de ácido nucléico, sendo que a dita construção compreende uma seqüência de nucleotídeos que codifica um tRNA ortogonal (O-tRNA) e uma seqüência de  
15 nucleotídeos que codifica uma amino-acil-tRNA sintetase ortogonal (O-RS), onde a dita O-RS preferencialmente amino-acila o dito O-tRNA com o dito aminoácido não-natural; e

(c) um polinucleotídeo que codifica o dito polipeptídeo de interesse, sendo que o dito polinucleotídeo  
20 compreende pelo menos um códon seletor que é reconhecido pelo dito O-tRNA, onde a posição do códon seletor no polinucleotídeo controla a posição especificada do aminoácido não-natural no polipeptídeo de interesse depois da expressão do polinucleotídeo para produzir o  
25 polipeptídeo.

81. Método, de acordo com a reivindicação 80, **CARACTERIZADO** pelo fato de que compreende uma célula hospedeira que compreende (a), (b) e (c).

82. Método para produzir um polipeptídeo de interesse que compreende um aminoácido não-natural em uma posição especificada, **CARACTERIZADO** pelo fato de que o método compreende:

5           (a) disponibilizar um sistema de tradução, sendo que o dito sistema de tradução compreende:

          (i) um aminoácido não-natural;

          (ii) uma construção de ácido nucléico que compreende uma seqüência de nucleotídeos que codifica um  
10 tRNA ortogonal (O-tRNA) e uma seqüência de nucleotídeos que codifica uma amino-acil-tRNA sintetase ortogonal (O-RS), onde a dita O-RS preferencialmente amino-acila o dito O-tRNA com o dito aminoácido não-natural; e

          (iii) um polinucleotídeo que codifica o dito  
15 polipeptídeo de interesse, sendo que o dito polinucleotídeo compreende pelo menos um códon seletor que é reconhecido pelo dito O-tRNA, onde a posição do códon seletor se correlaciona com a posição especificada do aminoácido não-natural no polipeptídeo de interesse;

20           (iv) uma célula hospedeira que compreende (i), (ii) e (iii); e

          (b) desenvolver a dita célula hospedeira; e

          (c) incorporar o dito aminoácido não-natural na posição especificada no dito polipeptídeo durante a tradução  
25 do dito polipeptídeo na dita célula hospedeira, produzindo desta forma o dito polipeptídeo de interesse que compreende o dito aminoácido não-natural na posição especificada.

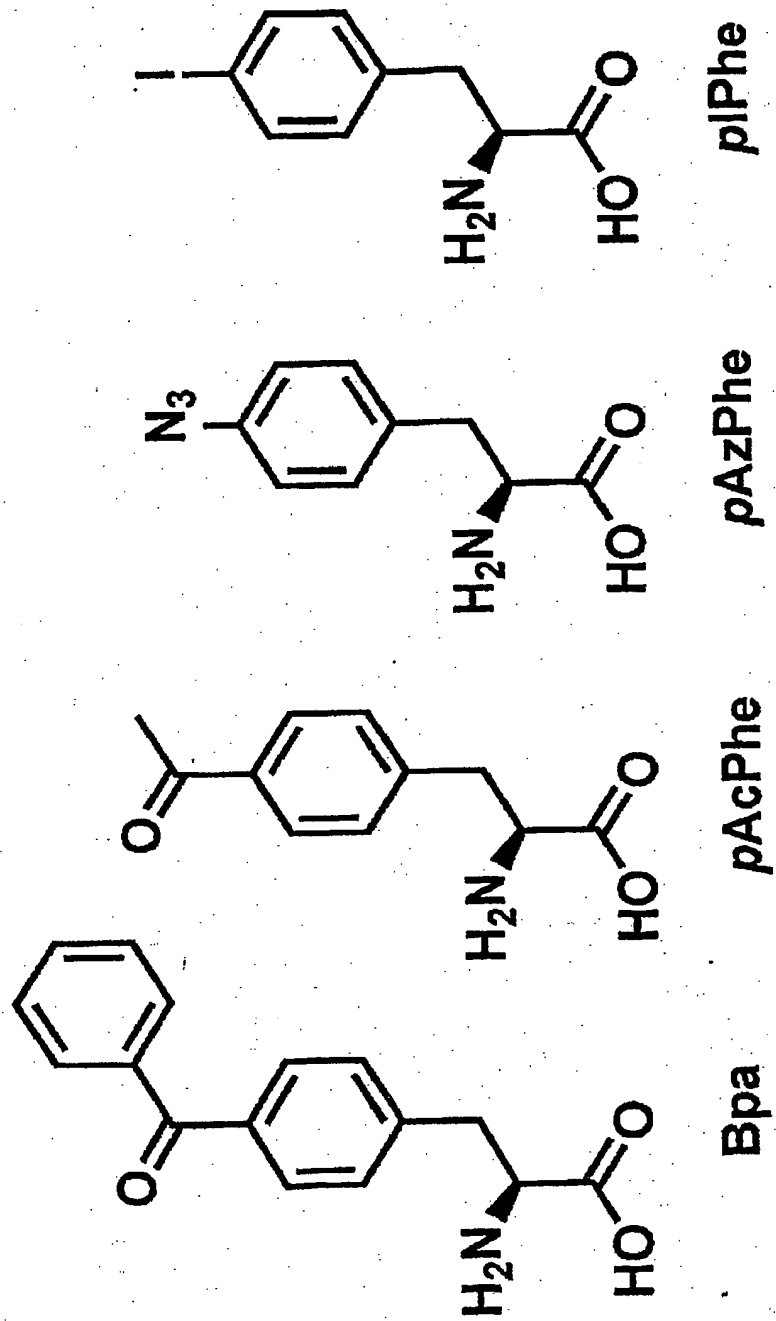


Fig. 1

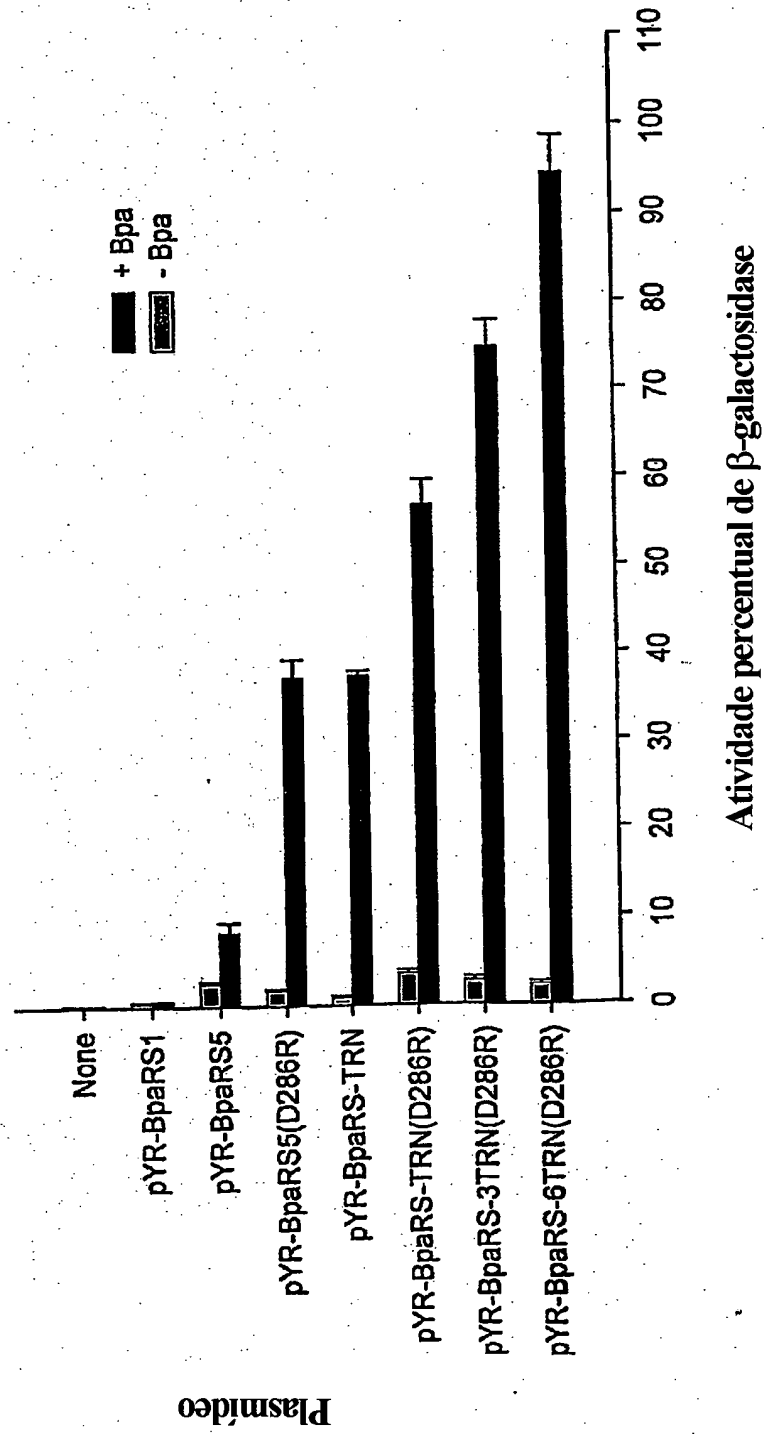


Fig. 2

pYR-BpaRS-6TRN

pYR-BpaRS-3TRN

pYR-BpaRS-TRN

pYR-BpaRS5

pYR-BpaRS1

Nenhum

Plasmídeo

Mj tRNA<sup>Tyr</sup><sub>cua</sub> →

Fig. 3

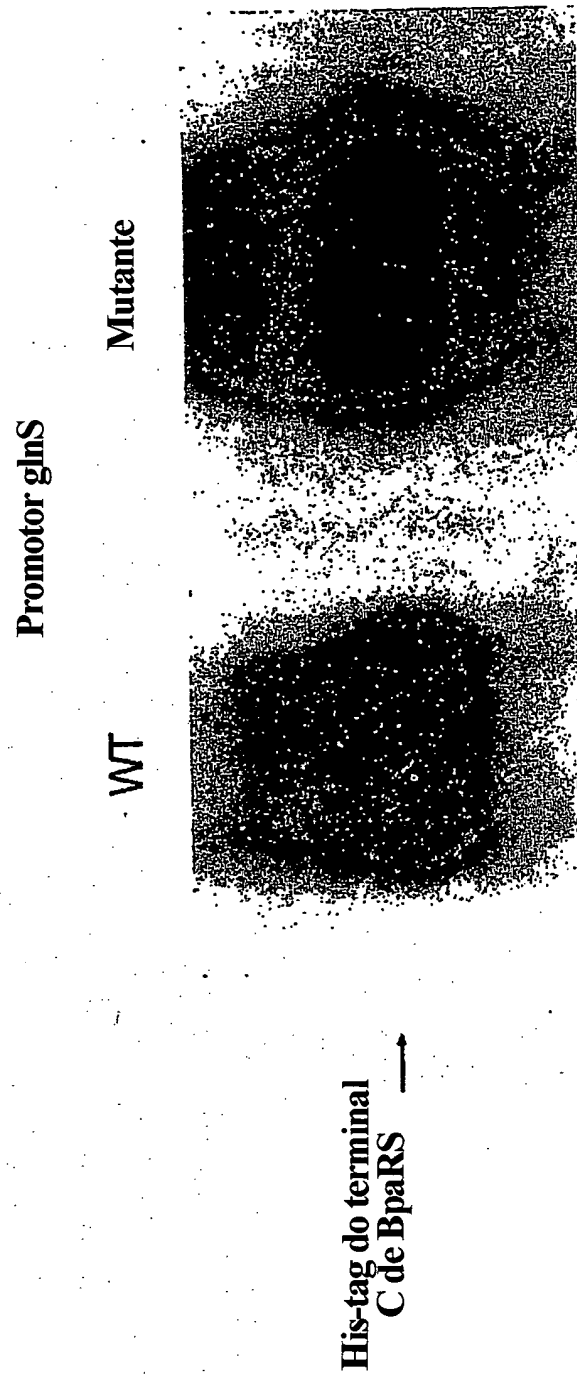


Fig. 4



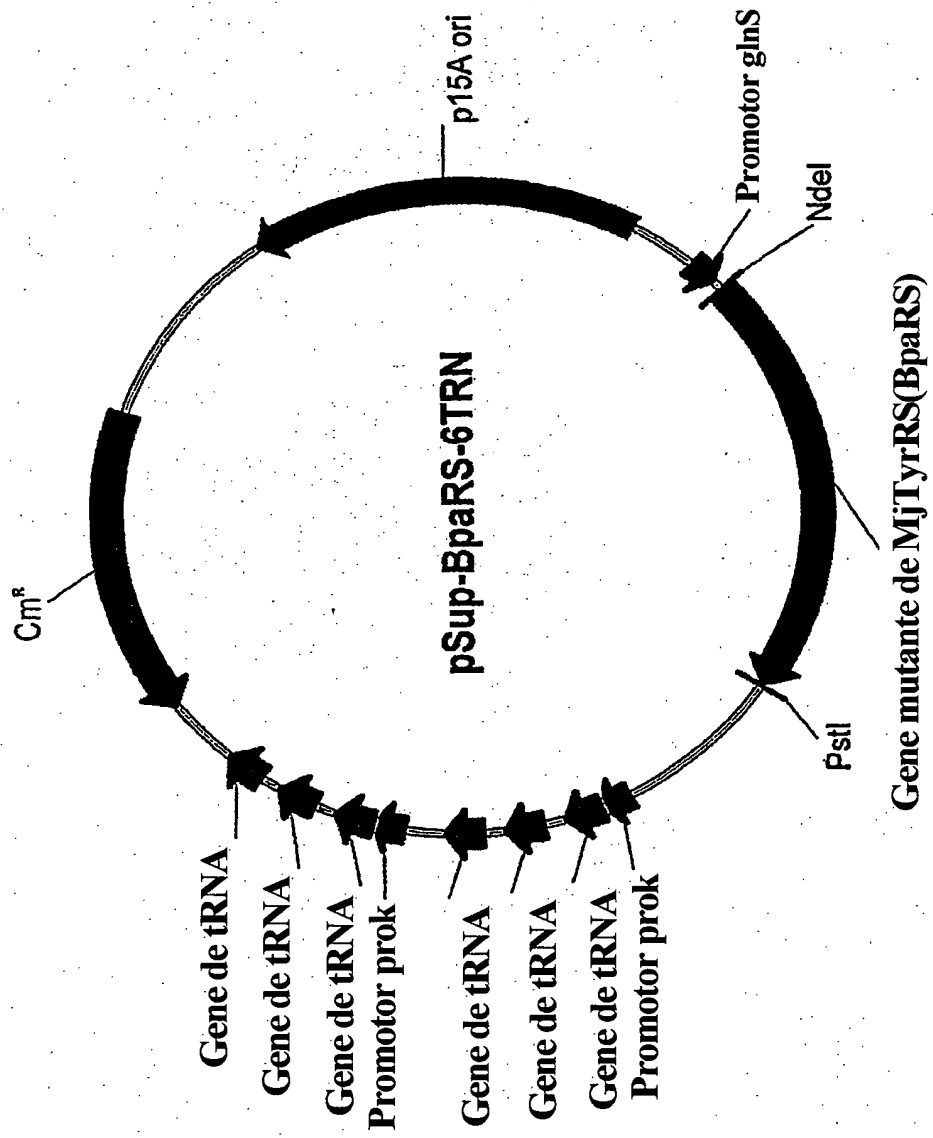


Fig. 5

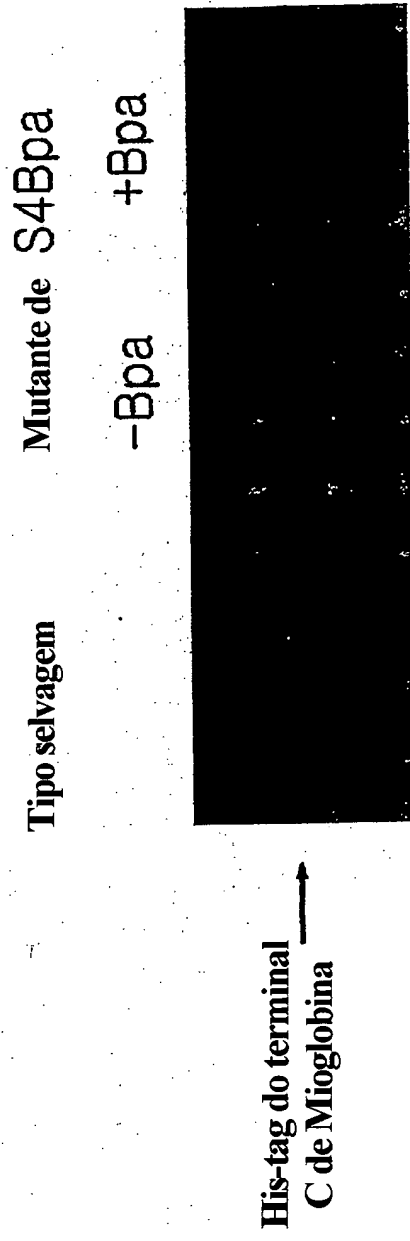


Fig. 6

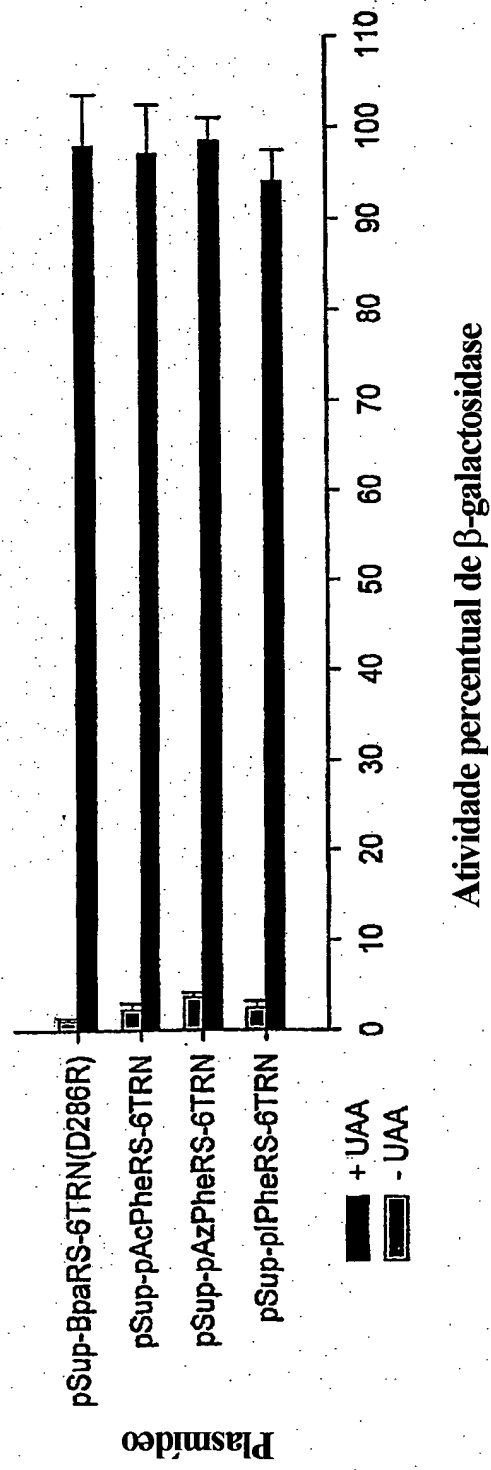


Fig. 7

## Sequências de Nucleotídeos e Aminoácidos

SEQ ID NO:	Descrição	Sequência
1	Supressor tirosil-tRNACUA aka MjtRNA-Tyr(CUA) ou mutRNA <sup>Tyr</sup> CUA de <i>Methanococcus jannaschii</i>	CCGGCGGUAGUUCAGCAGGGCAGAACGGCGGACUCUAAAUCCG CAUGGCGCUGGUUCAAUCCGGCCCCCGGACCA
2	Sequência de aminoácidos de tirosil-tRNA sintetase (MjTyrRS) de <i>Methanococcus jannaschii</i> do tipo selvagem	MDEFEMIKRNTSEIISEELREVLKKDEKSAYIGFEPGSKIHL GHYLQIKKMDLQNAAGFDIIILLADLHAYLNQKGEIDETRKIG DYNKKVFEAMGLKAKYVVGSEFQLDKDYTLNVYRLALKTTLKR ARRSMELIAREDENPKVAEVIYPIMQVNDIHYLGVDVAVGGME QRKIHMLARELLPKKVVCIHNPVLTGLDGEKMSSSKGNFIAY DDSPETIRAKIKKAYCPAGVVEGNPIMEIAKYFLEYPLTIKRP EKFGGDLTVNSYEELSLFKNKELHPMDLKNVAEELIKILEP IRKRL
3	Sequência de nucleotídeos de tirosil-tRNA sintetase (MjTyrRS) de <i>Methanococcus jannaschii</i> do tipo selvagem	ATGACGAATTTGAAATGATAAAGAGAAACACATCTGAAATTATCAGC GAGGAAGAGTTAAGAGAGGTTTAAAAAAGATGAAAAATCTGCTTAC ATAGGTTTGAACCAAGTGGTAAAAATACATTTAGGCCATTATCTCAA ATAAAAAAGATGATTGATTTACAAAATGCTGGATTGATATATTTATA TTGTTGGCTGATTACACGCCCTATTAAACCAGAAAGGAGATTGGAT GAGATTAGAAAAATAGGAGATTATACAAAAAAGTTTGAAGCAATG GGGTTAAAGGCCAAAATATGTTTATGGAAGTGAATTCAGCTTGATAAG GATTACACTGAATGCTATAGATTGGCTTTAAAAACTACCTTAAAA AGAGCAAGAAGGAGTATGGAACCTTAGCAAGAGAGGATGAAATCCA AAGGTTGCTGAAGTTATCTATCCAATAATGCAGGTTAATGATATTCAT TATTTAGGCGTTGATGTTGCAGTTGGAGGGATGGAGCAGAGAAAAATA CACATGTTAGCAAGGGAGCTTTTACCAAAAAAGGTTGTTTGTATTAC AACCTGTCTTAAACGGGTTTGGATGAGAGAGAAAGATGAGTTCTTCA AAAGGGAAATTTATAGCTGTTGATGACTCTCCAGAGAGATTAGGGCT AAGATAAAGAAAGCATACTGCCAGCTGGAGTTGTTGAAGGAATCCA ATAATGGAGATAGCTAAATACTTCTTGAATATCTTTAACCATAAAA AGGCCAGAAAAATTTGGTGGAGATTGACAGTTAATAGCTATGAGGAG TTAGAGAGTTTATTTAAAAATAAGGAATTGCATCAATGGATTAAAA AATGCTGTAGCTGAAGAACTTATAAGATTTTAGAGCCAATTAGAAAG AGATTAA
4	Sequência de aminoácidos de p-benzoil-L-fenilalanina-aminoacil-tRNA sintetase (BpaRS) (derivada de tirosil-tRNA sintetase de <i>Methanococcus jannaschii</i> do tipo selvagem)	MDEFEMIKRNTSEIISEELREVLKKDEKSAGIGFEPGSKIHL GHYLQIKKMDLQNAAGFDIIILLADLHAYLNQKGEIDETRKIG DYNKKVFEAMGLKAKYLYGSPFQLDKDYTLNVYRLALKTTLKR ARRSMELIAREDENPKVAEVIYPIMQVNTSHYLGVDVAVGGME QRKIHMLARELLPKKVVCIHNPVLTGLDGEKMSSSKGNFIAY DDSPETIRAKIKKAYCPAGVVEGNPIMEIAKYFLEYPLTIKRP EKFGGDLTVNSYEELSLFKNKELHPMLKNAVAEELIKILEP IRKRL

Fig. 8

SEQ ID NO:	Descrição	Sequência
5	Sequência de nucleotídeos de p-benzoil-L-fenilalanina-aminoacil-tRNA sintetase	ATGGACGAATTTGAAATGATAAAGAGAAACACATCTGAAATTA TCAGCGAGGAAGAGTTAAGAGAGGTTTTAAAAAAGATGAAAA ATCTGCTGGTATAGGTTTTGAACCAAGTGGTAAAAATACATTTA GGGCATTATCTCCAAATAAAAAAGATGATTGATTTACAAAATG CTGGATTTGATATAATTATATTTGTTGGCTGATTTACACGCCA TTTAAACCAGAAAGGAGAGTTGGATGAGATTAGAAAAATAGGA GATTATAACAAAAAAGTTTTTGAAGCAATGGGGTTAAAGGCCAA AATATCTTTATGGAAGTCCTTTCCAGCTTGATAAGGATTATAC ACTGAATGCTCTATAGATTGGCTTTAAAAACTACCTTAAAAAGA GCAAGAAGGAGTATGGAACCTTATAGCAAGAGAGGATGAAAAATC CAAAGGTTGCTGAAGTTATCTATCCAATAATGCAGGTTAATAC GAGTCATTATTTAGGCGTTGATGTTGCAGTTGGAGGGATGGAG CAGAGAAAAATACACATGTTAGCAAGGGAGCTTTTACCAAAAA AGGTTGTTTGTATTCAACCCCTGTCTTAACGGGTTTGGATGG AGAAGGAAAGATGAGTTCTTCAAAGGGAATTTTATAGCTGTT GATGACTCTCCAGAAGAGATTAGGGCTAAGATAAAGAAAGCAT ACTGCCAGCTGGAGTTGTTGAAGGAAATCCAATAATGGAGAT AGCTAAATACTTCTTGAATATCCTTTAACCATAAAAAGGCCA GAAAAATTTGGTGGAGATTGACAGTTAATAGCTATGAGGAGT TAGAGAGTTTATTTAAAAATAAGGAATTGCATCCAATGGATTT AAAAAATGCTGTAGCTGAAGAACTTATAAAGATTTTAGAGCCA ATTAGAAAGAGATTA
6	Sequência de aminoácidos de p-acetil-L-fenilalanina-aminoacil-tRNA sintetase (pAcPheRS) (derivada de tirosil-tRNA sintetase de <i>Methanococcus jannaschii</i> do tipo selvagem)	MDEFEMIKRNTSEIISEEELREVLKKDEKSALIGFEPGSKIHL GHYLGQIKKMI DLQNA GFDIIILLADLHAYLNQKGELDEIRKIG DYNKVF EAMGLKAKYVYGSEFQLDKDYTLNVYRLALKTTTLR ARRSMELIAREDENPKVAEVIYPIMQVNGCHYRGVDVAVGGM QRKIHMLARELLFKKVVC IHN PVL TGLDGEKMS SSKGNFIAV DDSP EIRAKIKKAYCPAGVVEGNPIMEIAKYFLEYPLTIKRP EKFGGDLTVNSYEELES LFKNKELHPMDLKNVAEELIKILPE IRKRL
7	Sequência de nucleotídeos de p-acetil-L-fenilalanina-aminoacil-tRNA sintetase	ATGGACGAATTTGAAATGATAAAGAGAAACACATCTGAAATTA TCAGCGAGGAAGAGTTAAGAGAGGTTTTAAAAAAGATGAAAA ATCTGCTCTGATAGGTTTTGAACCAAGTGGTAAAAATACATTTA GGGCATTATCTCCAAATAAAAAAGATGATTGATTTACAAAATG CTGGATTTGATATAATTATATTTGTTGGCTGATTTACACGCCA TTTAAACCAGAAAGGAGAGTTGGATGAGATTAGAAAAATAGGA GATTATAACAAAAAAGTTTTTGAAGCAATGGGGTTAAAGGCCAA AATATGTTTATGGAAGTGAATTCCAGCTTGATAAGGATTATAC ACTGAATGCTCTATAGATTGGCTTTAAAAACTACCTTAAAAAGA GCAAGAAGGAGTATGGAACCTTATAGCAAGAGAGGATGAAAAATC CAAAGGTTGCTGAAGTTATCTATCCAATAATGCAGGTTAATGG TTGTCATTATAGGGGCGTTGATGTTGCTGTTGGAGGGATGGAG CAGAGAAAAATACACATGTTAGCAAGGGAGCTTTTACCAAAAA AGGTTGTTTGTATTCAACCCCTGTCTTAACGGGTTTGGATGG AGAAGGAAAGATGAGTTCTTCAAAGGGAATTTTATAGCTGTT GATGACTCTCCAGAAGAGATTAGGGCTAAGATAAAGAAAGCAT ACTGCCAGCTGGAGTTGTTGAAGGAAATCCAATAATGGAGAT AGCTAAATACTTCTTGAATATCCTTTAACCATAAAAAGGCCA GAAAAATTTGGTGGAGATTGACAGTTAATAGCTATGAGGAGT TAGAGAGTTTATTTAAAAATAAGGAATTGCATCCAATGGATTT AAAAAATGCTGTAGCTGAAGAACTTATAAAGATTTTAGAGCCA ATTAGAAAGAGATTA

Fig. 8 Continuação

SEQ ID NO:	Descrição	Sequência
8	Sequência de aminoácidos de p-azido-L-fenilalanina-aminoacil-tRNA sintetase (pAzPheRS) (derivada de <i>Methanococcus jannaschii</i> do tipo selvagem)	MDEFEMIKRNTSEIISEEELREVLKKDEKSALIGFEPGSKIHL GHYLQIKKMLDQAGFDIIILLADLHAYLNQKGELEIRKIG DYNKKVFEAMGLKAKYVYGSPFQLDKDYTLNVYRLALKTTLLKR ARRSMELIAREDENPKVAEVIYPLMQVNQIHSSGVVDVAVGGME QRKIHMLARELLPKKVVCIHNPVLTGLDGEKMSSSKGNFIAV DDSPPEIRAKIKKAYCPAGVVEGNPIMEIAKYFLEYPLTIKRP EKFGGDLTVNSYEELESLEFKNKELHPMDLKNVAEELIKILEP IRKRL
9	Sequência de nucleotídeos de p-azido-L-fenilalanina-aminoacil-tRNA sintetase	ATGGACGAATTGAAATGATAAAGAGAAACACATCTGAAATTA TCAGCGAGGAAGAGTTAAGAGAGGTTTTAAAAAAGATGAAAA ATCTGCTCTGATAGGTTTGAACCAAGTGGTAAATACATTTA GGGCATTATCTCCAAATAAAAAAGATGATTGATTACAAAAATG CTGGATTGATATAATATATTGTTGGCTGATTACACGCCCTA TTTAAACCGAGAAAGGAGAGTTGGATGAGATTAGAAAAATAGGA GATTATAACAAAAAGTTTTTGAAGCAATGGGGTTAAAGGCCAA AATATGTTTATGGAAGTCCGTTCCAGCTTGATAAGGATTATAC ACTGAATGCTCTATAGATTGGCTTTAAAAACTACCTTAAAAAGA GCAAGAAGGAGTATGGAACCTTATAGCAAGAGAGGATGAAATC CAAAGGTTGCTGAAGTTATCTATCCAATAATGCAGGTTAATCA GATTCATTCTAGTGGCGTTGATGTTGCAGTTGGAGGGATGGAG CAGAGAAAAATACACATGTTAGCAAGGGAGCTTTTACCAAAAA AGGTTGTTTGTATTACAAACCCTGTCTTAACGGGTTTGGATGG AGAAGGAAAGATGAGTTCTTCAAAGGGGAATTTTATAGCTGTT GATGACTCTCCAGAAGAGATTAGGGCTAAGATAAAGAAAGCAT ACTGCCAGCTGGAGTTGTTGAAGGAAATCCAATAATGGAGAT AGCTAAATACTTCCTTGAATATCCTTTAACCATAAAAAGGCCA GAAAAATTTGGTGGAGATTTGACAGTTAATAGCTATGAGGAGT TAGAGAGTTTATTTAAAAATAAGGAATTGCATCCAATGGATT AAAAAATGCTGTAGCTGAAGAACTTATAAAGATTTTATAGGCCA ATTAGAAAGAGATTA
10	Sequência de aminoácidos de p-iodo-L-fenilalanina-aminoacil-tRNA sintetase (pIPheRS) (derivada de <i>Methanococcus jannaschii</i> do tipo selvagem)	MDEFEMIKRNTSEIISEEELREVLKKDEKSALIGFEPGSKIHL GHYLQIKKMLDQAGFDIIILLADLHAYLNQKGELEIRKIG DYNKKVFEAMGLKAKYVYGSPFQLDKDYTLNVYRLALKTTLLKR ARRSMELIAREDENPKVAEVIYPLMQVNPLHYEGVDVAVGGME QRKIHMLARELLPKKVVCIHNPVLTGLDGEKMSSSKGNFIAV DDSPPEIRAKIKKAYCPAGVVEGNPIMEIAKYFLEYPLTIKRP EKFGGDLTVNSYEELESLEFKNKELHPMDLKNVAEELIKILEP IRKRL

Fig. 8 cont.

SEQ ID NO:	Descrição	Seqüência
11	Seqüência de nucleotídeos de p-iodo-L-fenilalanina-aminoacil-tRNA sintetase	ATGGACGAATTTGAAATGATAAAGAGAAACACATCTGAAATTA TCAGCGAGGAAGAGTTAAGAGAGGTTTAAAAAAGATGAAAA ATCTGCTCTGATAGGTTTGAACCAAGTGGTAAAATACATTTA GGGCATTATCTCCAAATAAAAAAGATGATTGATTTACAAAATG CTGGATTTGATATAATTATATTGTTGGCTGATTTACACGCCA TTTAAACCAGAAAAGGAGAGTTGGATGAGATTAGAAAAATAGGA GATTATAACAAAAAAGTTTTGAAGCAATGGGGTTAAAGGCAA AATATGTTTATGGAAGTTCGTTCCAGCTTGATAAGGATTATAC ACTGAATGCTATAGATTGGCTTTAAAAACTACCTTAAAAAGA GCAAGAAGGAGTATGAACTTATAGCAAGAGAGGATGAAAAATC CAAAGGTTGCTGAAGTTATCTATCCAATAATGCAGGTTAATCC TCTTCATTATGAGGGCGTTGATGTTGCAGTTGGAGGGATGGAG CAGAGAAAATACACATGTTAGCAAGGGAGCTTTTACCAAAAA AGGTTGTTTGTATTTCACAACCTGTCTTAACGGGTTTGGATGG AGAAGGAAAGATGAGTTCTTCAAAGGGGAATTTTATAGCTGTT GATGACTCTCCAGAAGAGATTAGGGCTAAGATAAAGAAAGCAT ACTGCCAGCTGGAGTTGTTGAAGGAAATCCAATAATGGAGAT AGCTAAATACTTCCCTGAATATCTTTAACCATAAAAAGGCCA GAAAAATTTGGTGGAGATTGACAGTTAATAGCTATGAGGAGT TAGAGAGTTTATTTAAAAATAAGGAATTGCATCCAATGGATT AAAAATGCTGTAGCTGAAGAACTTATAAAGATTTTAGAGCCA ATTAGAAAGAGATTA
12	Promotor de <i>E. coli glnS</i> mutante, que tem uma seqüência TATC no lugar de GATC na região 10 (identificada em Plumbidge e Söll, <i>Biochemie</i> 69:539-541 (1987))	CGATTATCAATTTTAAAAAACTAACAGTTGTCAGCCTGTCCCG CTTATAATATCATACGCC
13	Promotor TRN de <i>E. coli glnS</i> mutante modificada	CGATTATCAATTTTAAAAAACTAACAGTTGTCAGCCTGTCCCG CTTATAATATCATACGCC
14	Seqüência ligante de tRNA entre os genes de <i>E. coli valU</i> e <i>valX</i>	ACTACTTTATGTAGTCTCCGCCGTGTAGCAAGAAATTGAGAAG T
15	Seqüência ligante de tRNA entre os genes de <i>E. coli ileT</i> e <i>alaT</i>	AATTTGCACGGCAAATTTGAAGAGGTTTAACTACATGTTAT
16	Seqüência ligante de tRNA entre os genes de <i>E. coli serV</i> e <i>argV</i>	TTT
17	Seqüência ligante de tRNA entre os genes de <i>E. coli valV</i> e <i>valW</i>	TCCT
18	Seqüência ligante de tRNA entre os genes de <i>E. coli gylT</i> e <i>thrT</i>	AGATGT
19	Seqüência ligante de tRNA entre os genes de <i>E. coli metE</i> e <i>leuW</i>	TCTTTTTTT
20	Seqüência ligante de tRNA entre os genes de <i>E. coli glnW</i> e <i>metU</i>	TCGAAGAAACAATCT
21	Seqüência ligante de tRNA entre os genes de <i>E. coli hisR</i> e <i>leuT</i>	TTATTAGAAGTTGTGACAAAT
22	Seqüência ligante de tRNA entre os genes de <i>E. coli glnU</i> e <i>glnW</i>	TCTTCTTCGAGTAAGCGGTTACCGCCCGGTTAT
23	Seqüência ligante de tRNA entre os genes de <i>E. coli leuP</i> e <i>leuV</i>	AACGAGGCGATATCAAAAAAGTAAGATGACTGT

Fig. 8 cont.

SEQ ID NO:	Descrição	Sequência
24	Sequência ligante de tRNA entre os genes de <i>E. coli glnV</i> e <i>glnX</i>	ATTTATTCAAGACGCTTACCTTGTAAGTGCACCCAGT
25	Sequência ligante de tRNA entre os genes de <i>E. coli alaW</i> e <i>alaX</i>	AATTTTGCACCCAGCAAACCTTGGTACGTAAACGCATCGT
26	Sequência ligante de tRNA entre os genes de <i>E. coli ileU</i> e <i>alaU</i>	AATTTGCACGGCAAATTTGAAGAGGTTTTAACTACATGTTAT
27	Sequência ligante de tRNA entre os genes de <i>E. coli ileV</i> e <i>alaV</i>	AATTTGCACGGCAAATTTGAAGAGGTTTTAACTACATGTTAT
28	Sequência ligante de tRNA entre os genes de <i>E. coli metU</i> e <i>glnV</i>	AATTCTGAATGTATCGAATATGTTCCGGCAAATTCAAAACCAATTTGT
29	Sequência ligante de tRNA entre os genes de <i>E. coli glyW</i> e <i>cysI</i>	GTTTAAAAGACATCGGCGTCAAGCGGATGTCTGGCTGAAAGGCCTGAAGAATTT
30	Sequência ligante de tRNA entre os genes de <i>E. coli argX</i> e <i>hisR</i>	TTTAGTCCCGCGCTTGAGCTGCGGTGGTAGTAATACCGCGTACAAGATTGTAGT
31	Sequência ligante de tRNA entre os genes de <i>E. coli argY</i> e <i>argZ</i>	TCTCTACTTGATATGGCTTTAGTAGCGGTATCAATATCAGCAGTAAATAAATTTCCCGAT
32	Promotor proK de <i>E. coli</i>	AGGCATTTTGCTATTAAAGGGATTGACGAGGGCGTATCTGCGCATGAAGATGCGCCCCGCATT
33	Terminador proK de <i>E. coli</i>	AATTCGAAAAGCCTGCTCAACGAGCAGGCTTTTTT
34	Promotor proL de <i>E. coli</i>	ATCAGTTAGCGAAATATCTTACTTGCAATCGGTGTGGAAAACGGTAGTATTAGCAGCCACGAGT
35	Terminador proL de <i>E. coli</i>	AAAATCCCAAGAAAAAACCAACCTTACGGTTGGTTTTTTTT
36	Terminador proM de <i>E. coli</i>	ATTTTGAACCCCGCTTCGGCGGGGTTTTTTTT
37	Iniciador proK P1	GTGCACGGCTAACTAAGCGGCCTGCTGACTTTCTCGCCGATCAAAAAGGC
38	Iniciador proK T1	CTTTCTCGCCGATCAAAAAGGCATTTTGCTATTAAAGGGATTGACGAGGGCGTATCTGCGCAGTAAGATGCGCCCCGCATTCGGCGGTAGTTTCAGCAGGGC
39	Iniciador proK T2	CTTTCTCGCCGATCAAAAAGGCATTTTGCTATTAAAGGGATTGACGAGGGCGTATCTGCGCAGTAAGATGCGCCCCGCATTCGGCGGTAGTTTCAGCAGGGC
40	Iniciador proK P2	GCATAAGCTTATGCAAAAAGCCTGCTCGTTGAGCAGGCTTTTTCG
41	Iniciador proK-F	AGTCTGATCAGTGACGGCTAACTAAGCGG
42	Iniciador proK-R	GCATCTCGAGATGCAAAAAGCCTGCTCGTTG
43	Iniciador glnS P1	CCGAGCTCCCGGTCATC
44	Iniciador glnS T1	CCGAGCTCCCGGTCATCAATCATCCCCATAATCCTTGTTAGATTATCAATTTTAAAAAACTAACAGTTGTCAGCCTGTC
45	Iniciador glnS T2	GTCCATATGGGATTCTCTAAAGCGTAACAACGTATAACGGCGTATGATATTATAAGCGGGACAGGCTGACAACCTGTTAG
46	Iniciador glnS P2	GTCCATATGGGATTCTCT
47	Ligante do iniciador P1	GTGCACGGCTAACTAAGCGGCCTGCTGACTTTCTCGCCGATCAAAAGGC
48	Ligante do iniciador P2	TACACGGCGGAGACTACATAAAGTAGTTGGTCCGCGGGGCGGATTTG
49	Ligante do iniciador P3	GTAGTCTCCCGGTGTAGCAAGAAATTGAGAAGTCCGGCGGTA GTTCAGCAG
50	Ligante do iniciador P4	AAACCTCTTCAAATTTGCCGTGCAAAATTTGGTCCGCGGGGCGGATTTG

Fig. 8 cont.



SEQ ID NO:	Descrição	Seqüência
51	Ligante do iniciador P5	GCAAATTTGAAGAGGTTTTAACTACATGTTATCCGGCGGTAGT TCAGCAG
52	Iniciador proK-R	GCATCTCGAGATGCAAAAAAGCCTGCTCGTTG
53	Iniciador Tandem P1	ATCAGTGACGGCTAACTAAGCGG
54	Iniciador Tandem P2	GCTGGCATGCATGCAAAAAAGCCTGCTCGTTGAGC
55	Iniciador Tandem P3	ATCAGCATGCGGCTAACTAAGCGGCCTGCTG
56	Iniciador Tandem P4	GCTGCTCGAGATGCAAAAAAGCCTGC
57	Seqüência líder	MDPLVTAASVLEEGLFET
58	Sonda da análise <i>Northern</i>	CCCTGCTGAACTACCGCC-

Fig. 8 cont.

RESUMO

"SISTEMAS PARA A EXPRESSÃO DE COMPONENTES  
ORTOGONAIS DE TRADUÇÃO EM CÉLULAS HOSPEDEIRAS EUBACTERIANAS"

A invenção refere-se a composições e métodos para  
5 a produção *in vivo* de polipeptídeos que compreendem um ou  
mais aminoácidos não-naturais. Especificamente, a invenção  
fornece sistemas plasmídicos para a expressão eubacteriana  
eficiente de polipeptídeos que compreendem um ou mais  
aminoácidos não-naturais em posições geneticamente  
10 programadas.