



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

(21) **PI0709603-8 A2**

(22) Data de Depósito: 13/03/2007
(43) Data da Publicação: 19/07/2011
(RPI 2115)



(51) *Int.Cl.:*
C12P 21/00 2006.01

(54) Título: **EXPRESSÃO DE PROTEÍNAS GENETICAMENTE PROGRAMADA CONTENDO O AMINOÁCIDO NÃO NATURAL FENILSELENOCISTEÍNA**

(30) Prioridade Unionista: 16/03/2006 US 60/783.272,
28/11/2006 US 60/861.456

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

(72) Inventor(es): Jiangyun Wang, Peter G. Schultz

(74) Procurador(es): Nellie Anne Daniel Shores

(86) Pedido Internacional: PCT US2007006382 de 13/03/2007

(87) Publicação Internacional: WO 2007/109035 de 27/09/2007

(57) Resumo: EXPRESSÃO DE PROTEÍNAS GENETICAMENTE PROGRAMADA CONTENDO O AMINOÁCIDO NÃO NATURAL FENILSELENOCISTEÍNA. A invenção diz respeito a pares ortogonais de RNAs e RNAt aminoacil sintetase que podem incorporar o aminoácido não natural fenilselenocisteína em proteínas produzidas em células hospedeiras eubactenanas, tal como E. coli. A invenção fornece, por exemplo, mas sem limitações, RNAt ortogonal aminoacil sintetase inédito, polinucleotídeos que codificam os sistemas de moléculas de sintetase inéditos, fenilselenocisteína e tradução. A invenção fornece adicionalmente métodos para produzir proteínas modificadas (por exemplo, proteínas lipidadas) por meio de modificação programada do resíduo de fenilselenocisteína em uma proteína.

P10709603-8
020080124697/08

"EXPRESSÃO DE PROTEÍNAS GENETICAMENTE PROGRAMADA CONTENDO
O AMINOÁCIDO NÃO NATURAL FENILSELENOCISTEÍNA"

REFERÊNCIA CRUZADA AO PEDIDO RELACIONADO

Este pedido reivindica prioridade e benefício de:

- 5 Pedido de patente provisório dos Estados Unidos 60/783.272, depositado em 16 de março de 2006, e pedido de patente provisório dos Estados Unidos 60/861.456, depositado em 28 de novembro de 2006, cujos conteúdos estão ambos aqui incorporados nas suas íntegras pela referência com todos os propósitos.

10 DECLARAÇÃO DE DIREITOS A INVENÇÕES FEITAS COM PESQUISA E
DESENVOLVIMENTO PATROCINADO PELO GOVERNO FEDERAL

Esta invenção foi feita com suporte do governo do Instituto Nacional de Saúde com concessão No. GM62159. O governo pode ter certos direitos nesta invenção.

CAMPO DA INVENÇÃO

- 15 A invenção é no campo de tradução bioquímica. A invenção diz respeito a composições e métodos para preparar e usar RNAts ortogonais, aminoacil-RNAt sintetase ortogonal, e seus pares, que incorporam aminoácidos não naturais nas proteínas. A invenção também diz respeito a métodos de produzir proteínas nas células usando tais pares e proteínas feitas pelos métodos.

ANTECEDENTES DA INVENÇÃO

- 20 O estudo de estrutura e função da proteína tem historicamente sido baseado nas propriedades e químicas de reações que são disponíveis usando os grupos reativos de aminoácidos de ocorrência natural. Infelizmente, diversos organismos conhecidos, de bactéria até humanos, codificam os mesmos vinte aminoácidos comuns. Estes 20 aminoácidos compreendem um número surpreendentemente limitado de grupos funcionais:
25 bases de nitrogênio, ácidos carboxílicos e amidas, álcoois, e um grupo tiol. Esta seleção limitada de grupos R tem restringido o estudo da estrutura e função da proteína, onde os estudos são confinados pelas propriedades químicas dos aminoácidos de ocorrência natural. Por exemplo, o número limitado de grupos R reativos de ocorrência natural tem limitado a capacidade de fazer as modificações de proteína altamente visadas para a
30 exclusão do outrosaminoácidos em uma proteína.

- Reações de ligação quimiosseletivas envolvendo proteínas são extremamente importantes com uma variedade de propósitos, incluindo, mas sem limitações, estudo da interação proteína-proteína e sinalização celular, e geração de terapêuticas de proteína inéditas. Reações de modificação de proteína mais seletivas atualmente usadas na
35 tecnologia envolvem formação de ligação covalente entre parceiros de reação nucleofílica e eletrofílica que visam resíduos nucleofílicos de ocorrência natural nas cadeias laterais de aminoácidos da proteína, por exemplo, a reação de cadeias laterais de α -halo cetonas com

histidina ou cisteína. A seletividade nestes casos é determinada pelo número e acessibilidade dos resíduos de nucleofílicos na proteína. Infelizmente, proteínas de ocorrência natural freqüentemente contêm sítios de reação ou múltiplos alvos de reação mal posicionados (por exemplo, inacessíveis) (por exemplo, resíduos de lisina, histidina e cisteína), resultando em pouca seletividade nas reações de modificação, tornando a modificação da proteína altamente alvejada pelos reagentes nucleofílicos/eletrofílicos difícil. Além disso, os sítios de modificação são tipicamente limitados às cadeias laterais de lisina, histidina ou cisteína nucleofílicas de ocorrência natural. Modificação em outros sítios é difícil ou impossível.

O que é necessário na tecnologia são novas estratégias para incorporação de aminoácidos não naturais nas proteínas com o propósito de modificar e estudar estrutura e função da proteína, onde os aminoácidos não naturais têm químicas de reação inéditas ou outras propriedades, por exemplo, propriedades biológicas não encontradas nos aminoácidos de ocorrência natural. Existe uma necessidade considerável na tecnologia de criação de novas estratégias para reações de modificação de proteína que modificam as proteínas de uma maneira altamente seletiva e, além disso, modificam proteínas em condições fisiológicas. O que é necessário na tecnologia são novos métodos para produzir modificações de proteína, onde as modificações são altamente específicas, por exemplo, modificações onde nenhum dos aminoácidos de ocorrência natural passa por reações cruzadas ou reações laterais. Químicas inéditas para estratégias da modificação da proteína altamente específicas encontram uma ampla variedade de aplicações no estudo da estrutura e função da proteína e na produção de proteínas terapêuticas.

LIPIDAÇÃO DA PROTEÍNA

A lipidação da proteína é uma modificação pós-translacional chave que está envolvida na localização da proteína, no devido tráfego de proteína intracelular própria e interações proteína-proteína. A lipidação de proteínas é freqüentemente exigida para a devida atividade biológica. Esta característica é crítica para o desenvolvimento de algumas proteínas terapêuticas. Lipidação é também crítica em estudos de interações proteína-proteína e sinalização celular. Infelizmente, a ligação quimiosseletiva *in vitro* para produzir proteínas lipidadas usando a forma não lipidada nativa da proteína é extremamente difícil, e é de um modo geral limitada a modificação de resíduos de cisteína expostos a superfície exclusiva.

As atividades biológicas de muitas proteínas celulares exigem associação com a membrana celular, que depende da modificação pós-tradução de cisteína por resíduos de lipídios tais como frações de farnesila, miristoíla e palmitoíla (Chernomordik and Kozlov (2003), *Annual Review of Chemystres* 72:175-207). Por exemplo, muitos receptores acoplados de proteína G são palmitoilados, proteínas Ras são tanto farnesiladas quanto

palmitoiladas (Chernomordik and Kozlov (2003), *Annual Review of Chemystres* 72:175-207). Embora farnesilação de proteína seja uma modificação estável e irreversível, a palmitoilação é reversível, resultando na regulação dinâmica de função de proteína, e alveijamento específico das membranas celulares (Rocks *et al.* (2005), *Science* 307(5716):1746-1752).

5 Além disso, ácido γ -carboxiglutâmico é uma modificação essencial que é importante para adesão da membrana dependente de cálcio na cascata de coagulação (Davie *et al.* (1991), *Chemystres* 30(43):10363-10370).

SISTEMAS DE TRADUÇÃO ORTOGONAL

Uma estratégia para superar as limitações de um código genético limitado é expandir o código genético e adicionar aminoácidos que têm novas propriedades reativas ao repertório biológico. Uma metodologia geral foi desenvolvida para incorporação sítio-específico *in vivo* de diversos aminoácidos não naturais nas proteínas, tanto em organismos procarióticos quanto eucarióticos. Estes métodos baseiam-se em componentes de tradução de proteína ortogonal que reconhecem um códon seletor adequado para inserir um aminoácido não natural desejado em uma posição definida durante a tradução de polipeptídeo *in vivo*. Estes métodos utilizam um RNAt ortogonal (O-RNAt) que reconhece um códon seletor, e onde um aminoacil-RNAt sintetase ortogonal específico correspondente (um O-RS) carrega o O-RNAt com o aminoácido não natural. Estes componentes não sofrem reação cruzada com nenhum dos aminoácidos ou códons RNAts, RSs, endógenos no organismo hospedeiro (isto é, eles devem ser ortogonais). O uso de tais pares de RNAt-RS ortogonal tornou possível codificar geneticamente um grande número de aminoácidos não naturais estruturalmente diferentes.

10
15
20

A prática de usar sistemas de tradução ortogonal que são adequados para preparar proteínas que compreendem um ou mais aminoácidos não naturais é de um modo geral conhecida na tecnologia, bem como os métodos gerais para produzir sistemas de tradução ortogonal. Por exemplo, ver International Publication Numbers WO 2002/086075, intitulado "METHODS AND COMPOSITION FOR THE PRODUCTION OF ORTOGONAL RNAt-AMINOACYL-RNAt SYNTHETASE PAIRS"; WO 2002/085923, intitulado "IN VIVO INCORPORATION OF UNNATURAL AMINO ACIDS"; WO 2004/094593, intitulado "EXPANDING THE EUKARYOTIC 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, intitulado "ORTOGONAL TRANSLATION COMPONENTS FOR THE VIVO INCORPORATION OF UNNATURAL AMINO ACIDS". Cada um desses pedidos está aqui incorporado pela referência na sua íntegra. Para discussão adicional de sistemas de tradução ortogonal que incorporam aminoácidos não naturais, e métodos para sua produção e uso, Ver também, Wang and Schultz, "Expanding the Genetic Code," *Chem. Commun.*

25
30
35

(*Camb.*) 1:1-11 (2002); Wang and Schultz "Expanding the Genetic Code", *Angewandte Chemie Int. Ed.*, 44(1):34-66 (2005); Xie e Schultz, "A Expanding Genetic Code", *Methods* 36(3):227-238 (2005); Xie and Schultz, "Adding Amino acids to the Repertorie Genetic", *Curr. Opinion in Chemical Biology* 9(6):548-554 (2005); Wang *et al.*, "Expanding the Genetic Code", *Annu. Rev. Biophys. Biomol. Struct.*, 35:225-249 (2006); and Xie and Schultz, "A Chemical Toolkit for Proteins - A Expanded Genetic Code," *Nat. Rev. Mol. Cell Biol.*, 7(10):775-782 (2006).

Existe uma necessidade na tecnologia do desenvolvimento de componentes de tradução ortogonal que incorporam aminoácidos não naturais nas proteínas, onde os aminoácidos não naturais podem ser incorporados em uma posição definida, e onde os aminoácidos não naturais têm propriedades químicas inéditas que permitem que o aminoácido sirva como um alvo para modificação específica (por exemplo, lipidação) para a exclusão de reações cruzadas ou reações laterais com outros sítios nas proteínas. A invenção preenche estas e outras necessidades, que ficarão aparentes mediante a revisão da revelação seguinte.

SUMÁRIO DA INVENÇÃO

A invenção fornece composições e métodos para incorporar o aminoácido não natural fenilselenocisteína em uma cadeia do polipeptídeo em crescimento em resposta a um códon seletor, por exemplo, um códon de parada âmbar, *in vivo* (por exemplo, em uma célula hospedeira). Estas composições incluem pares de RNAts-ortogonais (O-RNAts) e aminoacil-RNAt sintetases ortogonais (O-RSs) que não interagem com o maquinário de tradução da célula hospedeira. Ou seja, o O-RNAt não é carregado (ou não carregado em um nível significativo) com um aminoácido (natural ou não natural) por um aminoacil-RNAt sintetase da célula hospedeira endógena. Similarmente, o O-RSs fornecido pela invenção não carrega nenhum RNAt endógeno com um aminoácido (natural ou não natural) em um nível significativo ou, em alguns casos, detectável. Estas composições inéditas permitem a produção de grandes quantidades de proteínas tendo aminoácidos não naturais translacionalmente incorporados. As propriedades químicas do aminoácido não natural fenilselenocisteína também permitem que a modificação alvejada daquele resíduo produza conjugações desejadas, por exemplo, mas sem limitações, conjugações de lipídio. Tais polipeptídeos especificamente modificados podem encontrar uso como produtos terapêuticos e em pesquisa biomédica.

Em alguns aspectos, a invenção fornece sistemas de tradução. Estes sistemas compreendem um primeiro aminoacil-RNAt sintetase ortogonal (O-RS), um primeiro RNAt ortogonal (O-RNAt) e um primeiro aminoácido não natural que é fenilselenocisteína, onde o primeiro O-RS preferencialmente aminoacila o primeiro O-RNAt com o primeiro aminoácido não natural fenilselenocisteína. Em alguns aspectos, o O-RS preferencialmente aminoacila o

O-RNAt com a dita fenilselenocisteína com uma eficiência que é pelo menos 50 % da eficiência observada para um sistema de tradução compreendendo esse mesmo O-RNAt, a fenilselenocisteína e um aminoacil-RNAt sintetase que compreende a sequência de aminoácidos de SEQ ID NO: 4, 6 ou 8.

5 Os sistemas de tradução podem usar componentes derivados de uma variedade de fontes. Em uma modalidade, o O-RS usado no sistema pode compreender a sequência de aminoácidos de SEQ ID NOS: 4, 6 ou 8, e variantes conservativas desta sequência. Em algumas modalidades, o O-RNAt é um supressor âmbar. Em algumas modalidades, o O-RNAt compreende ou é codificado pela SEQ ID NO: 1.

10 Em alguns aspectos, o sistema de tradução compreende adicionalmente um ácido nucléico que codifica uma proteína de interesse, onde o ácido nucléico tem pelo menos um códon seletor que é reconhecido pelo O-RNAt.

Em alguns aspectos, o sistema de tradução incorpora um segundo par ortogonal (ou seja, um segundo O-RS e um segundo O-RNAt) que utiliza um segundo aminoácido não natural, de maneira que o sistema é agora capaz de incorporar pelo menos dois aminoácidos não naturais diferentes em sítios diferentes selecionados em um polipeptídeo. Neste sistema duplo, o segundo O-RS preferencialmente aminoacila o segundo O-RNAt com um segundo aminoácido não natural que é diferente do primeiro aminoácido não natural, e o segundo O-RNAt reconhece um códon seletor que é diferente do códon seletor reconhecido pelo primeiro O-RNAt.

15

20

Em algumas modalidades, o sistema de tradução reside em uma célula hospedeira (e inclui a célula hospedeira). A célula hospedeira usada particularmente não é limitada, desde que o O-RS e O-RNAt retenham sua ortogonalidade em seu ambiente da célula hospedeira. A célula hospedeira pode ser uma célula eubacteriana, tal como *E. coli*. A célula hospedeira pode compreender um ou mais polinucleotídeos que codificam componentes do sistema de tradução, incluindo o O-RS ou O-RNAt. Em algumas modalidades, o polinucleotídeo que codifica o O-RS compreende uma sequência de nucleotídeos de SEQ ID NO:5,7 ou 9.

25

A invenção também fornece métodos para produzir proteínas com um ou mais aminoácidos não naturais em posições selecionadas. Estes métodos utilizam os sistemas de tradução descritos anteriormente. De um modo geral, estes métodos começam com a etapa de fornecer um sistema de tradução que compreende:

30

- (i) um primeiro aminoácido não natural que é fenilselenocisteína;
- (ii) um primeiro aminoacil-RNAt sintetase ortogonal (O-RS);
- 35 (iii) um primeiro RNAt ortogonal (O-RNAt), em que o O-RS preferencialmente o O-RNAt aminoacila com o aminoácido não natural; e,
- (iv) um ácido nucléico que codifica a proteína, onde o ácido nucléico compreende

pelo menos um códon seletor (opcionalmente um códon âmbar) que é reconhecido pelo primeiro O-RNAt. O método então incorpora o aminoácido não natural na posição selecionada na proteína durante a tradução da proteína em resposta ao códon seletor, produzindo dessa maneira a proteína compreendendo o aminoácido não natural na posição selecionada. Em alguns aspectos destes métodos, o O-RS preferencialmente aminoacila o O-RNAt com a fenilselenocisteína com uma eficiência que é pelo menos 50 % da eficiência observada para um sistema de tradução compreendendo esse mesmo O-RNAt, a fenilselenocisteína, e um aminoacil-RNAt sintetase compreendendo a sequência de aminoácidos de SEQ ID NO: 4, 6 ou 8. Em alguns aspectos, o sistema de tradução inclui um ácido nucléico que codifica o O-RS.

Estes métodos podem ser amplamente aplicados usando uma variedade de reagentes. Em algumas modalidades, um polinucleotídeo que codifica o O-RS é fornecido. Em algumas modalidades, o O-RS compreende uma sequência de aminoácidos de SEQ ID NO: 4, 6 ou 8, ou variantes conservativas destes. Opcionalmente, o sistema de tradução usado nos métodos inclui um ácido nucléico que codifica o O-RS, por exemplo, um ácido nucléico de SEQ ID NO: 5, 7 ou 9.

Em algumas modalidades destes métodos, a provisão de uma etapa do sistema de tradução compreende mutar uma bolsa de ligação de aminoácido de um aminoacil RNAt sintetase tipo selvagem por mutagênese de sítio dirigido, e selecionar um O-RS resultante que preferencialmente aminoacila o O-RNAt com o aminoácido não natural. A etapa de seleção pode compreender selecionar positiva ou negativamente o O-RS de um agrupamento de moléculas de aminoacil-RNAt sintetase resultantes depois da mutagênese de sítio dirigido. Em algumas modalidades, a etapa de prover fornece um polinucleotídeo que codifica o O-RNAt, por exemplo, um O-RNAt que é um supressor âmbar de RNAt, ou um O-RNAt que compreende ou é codificado por um polinucleotídeo de SEQ ID NO: 1. Nestes métodos, a etapa de fornecimento pode também fornecer um ácido nucléico compreendendo um códon seletor de âmbar que é utilizado pelo sistema de tradução.

Estes métodos podem também ser modificados para incorporar mais que um aminoácido não natural em uma proteína. Nestes métodos, é empregado um segundo sistema de ORTOGONAL TRANSLATION em conjunto com o primeiro sistema de tradução, onde o segundo sistema tem diferentes especificidades de aminoácido e códon seletor. Por exemplo, a etapa de fornecimento pode incluir fornecer um segundo O-RS e um segundo O-RNAt, onde o segundo O-RS preferencialmente aminoacila o segundo O-RNAt com um segundo aminoácido não natural que é diferente do primeiro aminoácido não natural, e onde o segundo O-RNAt reconhece um códon seletor no ácido nucléico que é diferente do códon seletor reconhecido pelo primeiro O-RNAt.

Os métodos para produzir uma proteína com um aminoácido não natural podem

também ser conduzidos no contexto de uma célula hospedeira. Nestes casos, é fornecida uma célula hospedeira, onde a célula hospedeira compreende o aminoácido não natural, o O-RS, o O-RNAt e o ácido nucléico com pelo menos um códon seletor que codifica a proteína, e onde cultivar a célula hospedeira resulta na incorporação do aminoácido não natural. Em algumas modalidades, a etapa de fornecimento compreende fornecer uma

5 célula eubacteriana hospedeira (por exemplo, *E. coli*). Em algumas modalidades, a etapa de fornecimento inclui fornecer uma célula hospedeira contendo um polinucleotídeo que codifica o O-RS. Por exemplo, o polinucleotídeo que codifica o O-RS pode compreender uma seqüência de nucleotídeos de SEQ ID NO: 5, 7 ou 9.

10 Em algumas variações destes métodos, os procedimentos incluem adicionalmente modificação da fenilselenocisteína após sua incorporação em um polipeptídeo. Por exemplo, a fenilselenocisteína pode reagir em certas condições que resultam na conversão para deidroalanina na posição selecionada. Esta reação pode ser por eliminação oxidativa. A reação pode ser realizada expondo a fenilselenocisteína ao peróxido de hidrogênio.

15 A invenção também fornece métodos para produzir proteínas lipidadas, onde o lipídio é conjugado em uma posição designada selecionada. Estes métodos utilizam os sistemas de tradução descritos anteriormente. De um modo geral, estes métodos começam com a etapa de fornecer um sistema de tradução que compreende:

(i) um aminoácido não natural fenilselenocisteína;

20 (ii) um aminoacil-RNAt sintetase ortogonal (O-RS);

(iii) um RNAt ortogonal (O-RNAt), onde o O-RS preferencialmente aminoacila o O-RNAt com a fenilselenocisteína; e,

(iv) um ácido nucléico que codifica a proteína, onde o ácido nucléico compreende pelo menos um códon seletor (opcionalmente um códon âmbar) que é reconhecido pelo O-RNAt. O método então incorpora a fenilselenocisteína na posição selecionada na proteína

25 durante a tradução da proteína em resposta ao códon seletor. Essa fenilselenocisteína reage em seguida para produzir deidroalanina na posição selecionada, que, por sua vez, reage com um lipídio para produzir uma fração de aminoácido lipidado, produzindo dessa maneira uma proteína com um lipídio na posição selecionada na proteína.

30 Em alguns aspectos destes métodos, a reação da fenilselenocisteína é por eliminação oxidativa ou exposição a peróxido de hidrogênio. Em alguns aspectos, o lipídio conjugado que reagiu com a deidroalanina pode ser ácido tiopalmítico, farnesilmercaptano ou 1-hexadecanetiol. Os aminoácidos lipidados assim formados são palmitoilcisteína, farnesilcisteína e S-hexadecilcisteína. A modificação da deidroalanina é tipicamente por uma

35 reação de adição de Michael.

A invenção também fornece métodos para produzir uma proteína tendo um resíduo de deidroalanina em uma posição selecionada. Estes métodos utilizam os sistemas de

tradução descritos anteriormente. De um modo geral, estes métodos começam com a etapa de fornecer um sistema de tradução que compreende:

(i) um aminoácido não natural fenilselenocisteína

(ii) um aminoacil-RNAt sintetase ortogonal (O-RS);

5 (iii) um RNAt ortogonal (O-RNAt), onde o O-RS preferencialmente aminoacila o O-RNAt com a fenilselenocisteína; e,

(iv) um ácido nucléico que codifica a proteína, onde o ácido nucléico compreende pelo menos um códon seletor (opcionalmente um códon âmbar) que é reconhecido pelo O-RNAt. O método então incorpora a fenilselenocisteína na posição selecionada na proteína
10 durante a tradução da proteína em resposta ao códon seletor. Essa fenilselenocisteína reage em seguida para produzir deidroalanina na posição selecionada. Em alguns aspectos destes métodos, a reação da fenilselenocisteína é por eliminação oxidativa ou exposição ao peróxido de hidrogênio.

A invenção também fornece uma variedade de composições, incluindo ácidos
15 nucléicos e proteínas. A natureza da composição não é particularmente limitada, desde que a composição compreenda o ácido nucléico ou proteína especificados. As composições da invenção podem compreender qualquer número de componentes adicionais de qualquer natureza.

Por exemplo, a invenção fornece composições que compreendem polipeptídeos,
20 onde os polipeptídeos compreendem a sequência de aminoácidos de SEQ ID NO: 4, 6 ou 8, ou uma variante conservativa deste. Em alguns aspectos, o polipeptídeo da variante conservativa aminoacila um RNAt ortogonal cognato (O-RNAt) com um aminoácido não natural com uma eficiência que é pelo menos 50 % da eficiência observada para um sistema de tradução que compreende o O-RNAt, o aminoácido não natural, e uma sintetase
25 aminoacil-RNAt compreendendo a sequência de aminoácidos de SEQ ID NO: 4, 6, 8 ou 10. A invenção também fornece polinucleotídeos que codificam qualquer desses polipeptídeos anteriores. Em algumas modalidades, estes polinucleotídeos podem compreender uma sequência de nucleotídeos de SEQ ID NO: 5, 7 ou 9. Em algumas modalidades, os polipeptídeos estão em uma célula.

30 A invenção também fornece composições de polinucleotídeo que compreendem uma sequência de nucleotídeos de SEQ ID NO: 5, 7 ou 9. Em algumas modalidades, a invenção fornece vetores que compreendem os polinucleotídeos, por exemplo, vetores de expressão. Em algumas modalidades, a invenção fornece células que compreendem um vetor descrito anteriormente.

35 DEFINIÇÕES

Antes de descrever a invenção em detalhes, deve-se entender que esta invenção não está limitada a sistemas biológicos particulares, que podem, é claro, variar. Deve-se

também entender que a terminologia aqui usada tem o propósito de descrever apenas modalidades particulares, e não deve ser limitante. Conforme usado nesta especificação e nas reivindicações anexas, as formas singulares "um", "uma", "o" e "a" incluem referentes plurais, a menos que o conteúdo determine claramente de outra forma. Assim, por exemplo, referência a "uma célula" inclui combinações de duas ou mais células; referência a "um polinucleotídeo" inclui, como uma questão prática, muitas cópias de que polinucleotídeo.

A menos que aqui definido e a seguir no resto da especificação, todos termos técnicos e científicos aqui usados têm o mesmo significado normalmente entendido pelos versados na tecnologia a qual a invenção diz respeito.

Ortogonal: Da maneira aqui usada, o termo "ortogonal" refere-se a uma molécula (por exemplo, um RNAt ortogonal (O-RNAt) e/ou um aminoacil-RNAt sintetase ortogonal (O-RS)) que funciona com componentes endógenos de uma célula com baixa eficiência, comparada a uma molécula correspondente que é endógena para a célula ou sistema de tradução, ou que não pode funcionar com componentes endógenos da célula. No contexto de RNAts e aminoacil-RNAt sintetase, ortogonal refere-se a uma incapacidade ou baixa eficiência, por exemplo, menos que 20 % de eficiência, menos que 10 % de eficiência, menos que 5 % de eficiência, ou menos que 1% de eficiência, de um RNAt ortogonal funcionar com um RNAt sintetase endógeno, comparado a de um RNAt endógeno funcionar com o RNAt sintetase endógeno, ou de um aminoacil-RNAt sintetase ortogonal funcionar com um RNAt endógeno, comparado a de um RNAt sintetase endógeno de funcionar com o endógeno RNAt. A molécula falta uma molécula complementar ortogonal endógena funcionalmente normal na célula. Por exemplo, um RNAt ortogonal em uma célula é aminoacilado por qualquer RS endógeno da célula com eficiência baixa, ou mesmo zero, quando comparado a aminoacilação de um RNAt endógeno pelo RS endógeno. Em um outro exemplo, um RS ortogonal aminoacila qualquer RNAt endógeno de uma célula de interesse com eficiência baixa, ou mesmo zero, comparado a aminoacilação do RNAt endógeno por um RS endógeno. A segunda molécula ortogonal pode ser introduzida na célula que funciona com a primeira molécula ortogonal. Por exemplo, um par de RNAt ortogonal/RS inclui componentes complementares introduzidos que funcionam juntos na célula com uma eficiência (por exemplo, 45 % de eficiência, 50 % de eficiência, 60 % de eficiência, 70 % de eficiência, 75 % de eficiência, 80 % de eficiência, 90 % de eficiência, 95 % de eficiência, ou 99 % ou mais eficiência), comparado aquela de um controle, por exemplo, um par de RNAt/RS endógeno correspondente, ou um par ortogonal ativo (por exemplo, um par de tirosil RNAt ortogonal/RS).

Tirosil-RNAt Ortogonal: Da maneira aqui usada, um tirosil-RNAt ortogonal (tirosil-O-RNAt) é um RNAt que é ortogonal para um sistema de tradução de interesse, onde o RNAt é:

(1) idêntico ou substancialmente similar a um tirosil-RNAt de ocorrência natural,
(2) derivado de um tirosil-RNAt de ocorrência natural por mutagênese natural ou artificial,

(3) derivado por qualquer processo que leva uma seqüência de um tipo selvagem ou seqüência de tirosil-RNAt mutante de (1) ou (2) em consideração, (4) homóloga a um tirosil-RNAt tipo selvagem ou mutante; (5) homóloga a qualquer exemplo RNAt que é designado como um substrato para um tirosil-RNAt sintetase na figura 2, ou (6) uma variante conservativa de qualquer exemplo RNAt que é designado como um substrato para um tirosil-RNAt sintetase na figura 2. O tirosil-RNAt pode existir carregado com um aminoácido ou em um estado descarregado. Deve-se também entender que um "tirosil-O-RNAt" opcionalmente é carregado (aminoacilado) por uma sintetase cognata com um aminoácido a não ser tirosina, respectivamente, por exemplo, com um aminoácido não natural. Realmente, percebe-se que um tirosil-O-RNAt da invenção é vantajosamente usado para inserir essencialmente qualquer aminoácido, quer natural ou não natural, em um polipeptídeo em crescimento, durante a tradução, em resposta a um códon seletor.

Aminoácido tirosil sintetase ortogonal: Da maneira aqui usada, um aminoácido tirosil sintetase ortogonal (tirosil-O-RS) é uma enzima que preferencialmente aminoacila o tirosil-O-RNAt com um aminoácido em um sistema de tradução de interesse. O aminoácido que o tirosil-O-RS carrega no tirosil-O-RNAt pode ser qualquer aminoácido, quer natural, não natural ou artificial, e não está aqui limitado. A sintetase é opcionalmente a mesma ou homóloga a uma sintetase de aminoácido tirosil de ocorrência natural, ou a mesma ou homóloga a uma sintetase designada como um O-RS na figura 2. Por exemplo, o O-RS pode ser uma variante conservativa de um tirosil-O-RS da figura 2, e/ou pode ser pelo menos 50 %, 60 %, 70 %, 80 %, 90 %, 95 %, 98 %, 99 % ou mais idêntica em seqüência a um O-RS da figura 2.

Cognato: O termo "Cognato" refere-se a componentes que funcionam juntos, ou têm algum aspecto de especificidade um com o outro, por exemplo, um RNAt ortogonal e um aminoacil-RNAt ortogonal sintetase. Os componentes podem também ser referidos como complementares.

Preferencialmente aminoacila: Da maneira aqui usada em referência aos sistemas de tradução ortogonal, um O-RS "preferencialmente aminoacila" um O-RNAt cognato quando o O-RS carrega o O-RNAt com um aminoácido mais eficientemente do que ele carrega qualquer RNAt endógeno em um sistema de expressão. Ou seja, quando o O-RNAt e qualquer dado RNAt endógeno estão presentes em um sistema de tradução em razões molares aproximadamente iguais, o O-RS carregará o O-RNAt mais freqüentemente do que ele carregará o RNAt endógeno. Preferivelmente, a razão relativa do O-RNAt carregado pelo O-RS com o RNAt endógeno carregado pelo O-RS é alta, preferivelmente resultando no O-

RS carregando o O-RNAt exclusivamente, ou quase exclusivamente, quando o O-RNAt e RNAt endógeno estão presentes em concentrações molares iguais no sistema de tradução. A razão relativa entre O-RNAt e RNAt endógeno que é carregado pelo O-RS, quando o O-RNAt e O-RS estão presentes em concentrações molares iguais, é maior que 1:1, preferivelmente pelo menos cerca de 2:1, mais preferivelmente 5:1, ainda mais preferivelmente 10:1, acima de tudo mais preferivelmente 20:1, ainda mais preferivelmente 50:1, acima de tudo mais preferivelmente 75:1, ainda mais preferivelmente 95:1, 98:1, 99:1, 100:1, 500:1, 1.000:1, 5.000:1 ou mais.

O O-RS "preferencialmente aminoacila um O-RNAt com um aminoácido não natural" quando (a) o O-RS preferencialmente aminoacila o O-RNAt comparado a um RNAt endógeno, e (b) onde essa aminoacilação é específica para o aminoácido não natural, comparado a aminoacilação do O-RNAt pelo O-RS com qualquer aminoácido natural. Ou seja, quando os aminoácidos não naturais e naturais estão presentes em quantidades molares iguais em um sistema de tradução compreendendo o O-RS e O-RNAt, o O-RS carregará o O-RNAt com o aminoácido não natural mais freqüentemente do que com o ácido natural. Preferivelmente, a razão relativa do O-RNAt carregado com o aminoácido não natural para O-RNAt carregado com o aminoácido natural é alta. Mais preferivelmente, O-RS carrega o O-RNAt exclusivamente, ou quase exclusivamente, com o aminoácido não natural. A razão relativa entre o carregamento do O-RNAt com o aminoácido não natural e carregamento do O-RNAt com o aminoácido natural, quando ambos os aminoácidos naturais e não naturais estão presentes no sistema de tradução em concentrações molares iguais, é maior que 1:1, preferivelmente pelo menos cerca de 2:1, mais preferivelmente 5:1, ainda mais preferivelmente 10:1, acima de tudo mais preferivelmente 20:1, ainda mais preferivelmente 50:1, acima de tudo mais preferivelmente 75:1, ainda 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 a códons reconhecidos pelo O-RNAt no processo de tradução e não reconhecidos por um RNAt endógeno. O laço anticódon de O-RNAt reconhece o códon seletor no mRNA e incorpora seu aminoácido, por exemplo, um aminoácido não natural, neste sítio no polipeptídeo. Códon seletores podem incluir, por exemplo, códons não sentido, tais como, códons de parada, por exemplo, códons âmbar, ocre e opala; códons de quatro ou mais bases; códons raros; códons derivados de pares de base natural ou não natural e/ou similares.

RNAt supressor: Um RNAt supressor é um RNAt que altera a leitura de um RNA mensageiro (RNAm) em um dado sistema de tradução, tipicamente permitindo a incorporação de um aminoácido em resposta a um códon de parada (isto é, "leitura total") durante a tradução de um polipeptídeo. Em alguns aspectos, um códon seletor da invenção é um códon supressor, por exemplo, um códon de parada (por exemplo, um códon âmbar,

ocre ou opala), um códon de quatro bases, um códon raro, etc.

Atividade de supressão: Da maneira aqui usada, o termo "atividade de supressão" refere-se em geral à capacidade de um RNAt (por exemplo, um supressor RNAt) permitir leitura de um códon translacional (por exemplo, um códon seletor que é um códon âmbar ou um códon de 4 bases ou mais) que possa resultar de outra forma na terminação de tradução ou não tradução (por exemplo, deslocamento de quadro). Atividade de supressão de um supressor RNAt pode ser expressa como uma porcentagem de atividade de leitura translacional observada, comparada a um segundo supressor RNAt, ou comparada a um sistema de controle, por exemplo, um sistema de controle faltando um O-RS.

A presente invenção fornece vários métodos pelos quais a atividade de supressão pode ser quantificada. A supressão percentual de um O-RNAt e O-RS particular em função de um códon seletor (por exemplo, um códon âmbar) de interesse refere-se à porcentagem de atividade de um dado marcador de teste expresso (por exemplo, LacZ), que inclui um códon seletor, em um ácido nucléico que codifica o marcador de teste expresso, em um sistema de tradução de interesse, onde o sistema de tradução de interesse inclui um O-RS e um O-RNAt, comparado a uma construção de controle positivo, onde o controle positivo falta o O-RNAt, o O-RS e o códon seletor. Assim, por exemplo, se uma construção do marcador de controle positivo ativo que falta um códon seletor tem uma atividade observada de X em um dado sistema de tradução, em unidades relevantes ao ensaio do marcador em questão, então a supressão da porcentagem de uma construção de teste compreendendo o códon seletor é a porcentagem de X que o marcador de construção de teste apresenta nas mesmas condições essencialmente ambientais que o marcador de controle positivo foi expresso, exceto que o marcador de construção de teste é expresso em um sistema de tradução que também inclui o O-RNAt e o O-RS. Tipicamente, o sistema de tradução que expressa o marcador de teste também inclui um aminoácido que é reconhecido pelo O-RS e O-RNAt. Opcionalmente, as medições de supressão percentual podem ser refinadas por comparação do marcador de teste com uma construção de marcador de controle "de fundo" ou "negativa", que inclui o mesmo códon seletor como o marcador de teste, mas em um sistema que não inclui o O-RNAt, O-RS e/ou aminoácido relevante reconhecido pelo O-RNAt e/ou O-RS. Este controle negativo é usado na normalização das medições de supressão percentual para levar em conta os efeitos de sinal de fundo do marcador no sistema de tradução de interesse.

A eficiência de supressão pode ser determinada por qualquer dos diversos ensaios conhecidos na tecnologia. Por exemplo, um ensaio reportador de β -galactosidase pode ser usado, por exemplo, um plasmídeo lacZ derivatizado (onde a construção tem um códon seletor na sequência de ácido nucléico lacZ) é introduzido nas células a partir de um organismo apropriado (por exemplo, um organismo onde os componentes ortogonais podem

ser usados) junto com o plasmídeo compreendendo um O-RNAt da invenção. Uma sintetase cognata pode também ser introduzida (tanto como um polipeptídeo quanto como um polinucleotídeo que codifica uma sintetase cognata quando expressa). As células crescem em meio a uma densidade desejada, por exemplo, com um OD₆₀₀ de cerca de 0,5, e ensaios β -galactosidase são realizados, por exemplo, usando o conjunto de ensaio β -Galactosidase Betaflúor™ (Novagen). A supressão percentual pode ser calculada como a porcentagem de atividade para uma amostra com relação a um controle comparável, por exemplo, o valor observado da construção lacZ derivatizada, onde a construção tem um códon sentido correspondente na posição desejada em vez de um códon seletor.

10 Sistema de tradução: O termo "sistema de tradução" refere-se aos componentes que incorporaram um aminoácido em uma cadeia do polipeptídeo em crescimento (proteína). Componentes de um sistema de tradução podem incluir, por exemplo, ribossomos, RNAs, sintetase, mRNA e similares. O O-RNAt e/ou o O-RSSs da invenção podem ser adicionados ou ser parte de um sistema de tradução *in vitro* ou *in vivo*, por exemplo, em uma célula não eucariótica, por exemplo, uma bactéria (tal como *E. coli*), ou em uma célula eucariótica, por exemplo, uma célula de levedura, uma célula de mamífero, uma célula de planta, uma célula de alga, uma célula de fungo, uma célula de inseto e/ou similares.

20 Aminoácido não natural: Da maneira aqui usada, o termo "aminoácido não natural" refere-se a qualquer aminoácido, aminoácido modificado e/ou análogo de aminoácido, que não é um dos 20 aminoácidos de ocorrência natural comuns. Por exemplo, o aminoácido não natural fenilselenocisteína (Ver figura 1, estrutura 1) encontra uso com a invenção.

25 Derivados de: Da maneira aqui usada, o termo "derivados de" refere-se a um componente que é isolado feito usando uma molécula ou organismo especificado, ou informação da molécula ou organismo especificado. Por exemplo, um polipeptídeo que é derivado de um segundo polipeptídeo pode incluir uma sequência de aminoácidos que é idêntica ou substancialmente similar à sequência de aminoácidos do segundo polipeptídeo. No caso de polipeptídeos, as espécies derivadas podem ser obtidas, por exemplo, por mutagênese de ocorrência natural, mutagênese artificial dirigida ou mutagênese artificial aleatória. A mutagênese usada para derivar polipeptídeos pode ser intencionalmente dirigida ou intencionalmente aleatória, ou uma mistura de cada. A mutagênese de um polipeptídeo para criar um polipeptídeo diferente derivado do primeiro pode ser um evento aleatório (por exemplo, causado por infidelidade de polimerase) e a identificação do polipeptídeo derivado pode ser feita por métodos de classificação apropriados, por exemplo, da maneira aqui discutida. Mutagênese de um polipeptídeo tipicamente implica na manipulação do polinucleotídeo que codifica o polipeptídeo.

35 Marcador de seleção ou classificação positiva: Da maneira aqui usada, o termo

"marcador de seleção ou classificação positiva" refere-se a um marcador que, quando presente, por exemplo, expresso, ativado ou similares, resulta na identificação de uma célula, que compreende o trato, por exemplo, uma célula com o marcador de seleção positiva, daquelas sem o traço.

5 Marcador de seleção ou classificação negativa: Da maneira aqui usada, o termo "marcador de seleção ou classificação negativa" refere-se a um marcador que, quando presente, por exemplo, expresso, ativado, ou similares, permite a identificação de uma célula que não compreende uma propriedade ou traço selecionados (por exemplo, comparado a uma célula que não possui a propriedade ou trato).

10 Reportador: Da maneira aqui usada, o termo "reportador" refere-se a um componente que pode ser usado para identificar e/ou selecionar componentes alvos de um sistema de interesse. Por exemplo, um reportador pode incluir uma proteína, por exemplo, uma enzima, que confere resistência ou sensibilidade a antibiótico (por exemplo, D-lactamase, cloranfenicol acetiltransferase (CAT), e similares), um marcador de classificação
15 fluorescente (por exemplo, proteína fluorescente verde (por exemplo, (GFP), YFP, EGFP, RFP, etc.), um marcador luminescente (por exemplo, uma proteína da luciferase de vagalume), uma afinidade baseada no marcador de classificação, ou genes marcadores selecionáveis positivos ou negativos tais como lacZ, β -gal/lacZ (β -galactosidase), ADH (álcool deidrogenase), his3, ura3, leu2, lis2, ou similares.

20 Eucarioto: Da maneira aqui usada, o termo "eucarioto" refere-se a organismos pertencentes ao reino Eucarya. Eucariotos são de um modo geral distinguíveis de procariotos por sua organização tipicamente multicelular (mas não exclusivamente multicelular, por exemplo, levedura), a presença de um núcleo ligado na membrana e outros organelos ligados na membrana, material genético linear (isto é, cromossomos lineares), a
25 ausência de operons, a presença de introns, capeamento da mensagem e poli-A mRNA, e outras características bioquímicas, tais como uma estrutura ribossomal distinguível. Organismos eucarióticos incluem, por exemplo, animais (por exemplo, mamíferos, insetos, répteis, pássaros, etc.), ciliatos, plantas (por exemplo, monocots, dicots, algas, etc.), fungo, levedura, flagelados, microsporidia, protista, etc.

30 Procarioto: Da maneira aqui usada, o termo "procarioto" refere-se a organismos pertencentes ao reino Monera (também denominado Procarya). Organismos procarióticos são de um modo geral distinguíveis dos eucariotos por sua organização unicelular, reprodução assexuada por brotamento ou fissão, a falta de um núcleo ligado a membrana ou outros organelos ligados na membrana, um cromossomo circular, a presença de operons, a
35 ausência de introns, capeamento da mensagem e poli-A mRNA, e outras características bioquímicas, tal como uma estrutura ribossomal distinguível. Os Procarias incluem sub-reinos Eubactéria e Archaea (algumas vezes denominadas "Archaeobacteria"). Cianobactéria

(as algas azuis esverdeadas) e micoplasma são algumas vezes dados classificações separadas no Reino Monera.

Bactéria: Da maneira aqui usada, os termos "bactéria" e "eubactéria" referem-se a organismos procarióticos que são distinguíveis da Archaea. Similarmente, Archaea refere-se a procariotos que são distinguíveis da eubactéria. Eubactéria e Archaea podem ser distinguidas por um número de critério morfológico e bioquímico. Por exemplo, diferenças nas seqüências de RNA ribossomais, estrutura polimerase RNA, a presença ou ausência de introns, sensibilidade a antibiótico, a presença ou ausência da parede celular peptidoglicanos e outros componentes da parede celular, as estruturas ramificadas versus não ramificadas de lipídios da membrana, e a presença/ausência de histonas e proteínas tipo histona são usadas para projetar um organismo para Eubactéria ou Archaea.

Exemplos de Eubactérias incluem *Escherichia coli*, *Thermus thermophilus*, *Bacillus subtilis* e *Bacillus stearothermophilus*. Exemplo de Archaea incluem *Methanococcus jannaschii* (Mj), *Methanosarcina mazei* (Mm), *Methanobacterium thermoautotrophicum* (Mt), *Methanococcus maripaludis*, *Methanopyrus kandleri*, *Halobacterium* tais como *Haloferax volcanii* e *Halobacterium species NRC-1*, *Archaeoglobus fulgidus* (AD), *Pyrococcus furiosus* (PD), *Pyrococcus horikoshii* (Ph), *Pyrobaculum aerophilum*, *Pyrococcus abyssi*, *Sulfolobus solfataricus* (Ss), *Sulfolobus tokodaii*, *Aeuropyrum pernix* (Ap), *Thermoplasma acidophilum* e *Thermoplasma volcanium*.

Variante conservativa: Da maneira aqui usada, o termo "variante conservativa", no contexto de um componente de tradução, refere-se a um componente de tradução, por exemplo, uma variante conservativa O-RNAt ou uma variante conservativa O-RS, que desempenha funcionalidade similar a um componente de base que a variante conservativa é similar, por exemplo, a um O-RNAt ou O-RS, tendo variações na seqüência comparada a uma referência O-RNAt ou O-RS. Por exemplo, um O-RS, ou uma variante conservativa de O-RS, aminoacilará um O-RNAt cognato com um aminoácido não natural fenilselenocisteína. Neste exemplo, o O-RS e a variante conservativa O-RS não tem 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 seqüência, desde que a variante conservativa seja ainda complementar ao (por exemplo, função) cognato correspondente O-RNAt ou O-RS.

Em algumas modalidades, uma variante conservativa O-RS compreende um ou mais substituições de aminoácido conservativas comparada ao O-RS do qual ele foi derivado. Em algumas modalidades, uma variante conservativa O-RS compreende uma ou mais substituições de aminoácido conservativas comparada ao O-RS da qual ele foi derivado e, além disso, retém atividade biológica de O-RS; por exemplo, uma variante conservativa O-RS que retém pelo menos 10 % da atividade biológica da molécula O-RS pai

da qual ela foi derivada ou, alternativamente, pelo menos 20 %, pelo menos 30 %, ou pelo menos 40 %. Em algumas modalidades preferidas, a variante conservativa O-RS retém pelo menos 50 % da atividade biológica da molécula O-RS pai da qual ela foi derivada. As substituições de aminoácido conservativas de uma variante conservativa O-RS pode ocorrer em qualquer domínio do O-RS, incluindo a bolsa de ligação de aminoácido.

Agente de seleção ou classificação: Da maneira aqui usada, o termo "agente de seleção ou classificação" refere-se a um agente que, quando presente, permite seleção/classificação de certos componentes de uma população. Por exemplo, um agente de seleção ou classificação pode ser, mas sem limitações, por exemplo, um nutriente, um antibiótico, um comprimento de onda da luz, um anticorpo, um polinucleotídeo expresso, ou similares. O agente de seleção pode ser variado, por exemplo, por concentração, intensidade, etc.

Em resposta a: Da maneira aqui usada, o termo "em resposta a" refere-se ao processo em que um O-RNAt da invenção reconhece um códon seletor e media a incorporação do aminoácido não natural, que é acoplado ao RNAt, na cadeia do polipeptídeo em crescimento.

Codificar: Da maneira aqui usada, o termo "codificar" refere-se a qualquer processo pelo qual a informação em uma macromolécula polimérica ou cadeia de seqüência é usada para direcionar a produção de uma segunda molécula ou cadeia de seqüência que é diferente da primeira molécula ou cadeia de seqüência. Da maneira aqui usada, o termo é usado de forma geral, e pode ter uma variedade de aplicações. Em alguns aspectos, o termo "codificar" descreve o processo de replicação de DNA semiconservativo, onde uma fita de uma molécula de DNA de fita dupla é usada como um molde para codificar uma fita irmã complementar sintetizada por um DNA polimerase DNA-dependente.

Em um outro aspecto, o termo "codificar" refere-se a qualquer processo pelo qual a informação em uma molécula é usada 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, incorporar pelo processo de transcrição uma enzima de RNA polimerase DNA-dependente). Também, uma molécula de RNA pode codificar um polipeptídeo, como no processo de tradução. Quando usado para descrever o processo de tradução, o termo "codificar" também estende-se ao 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 um DNA polimerase RNA-dependente. Em um outro aspecto, uma molécula de DNA pode codificar um polipeptídeo, onde entende-se que "codifica" na forma usada nesse caso incorpora tanto os processos de transcrição quanto de tradução.

DESCRIÇÃO RESUMIDA DAS FIGURAS

A figura 1 fornece as estruturas químicas do aminoácido não natural fenilselenocisteína (estrutura 1) bem como estruturas que podem ser derivadas de fenilselenocisteína. Estas estruturas incluem deidroalanina (2), palmitoilcisteína (3), famesilcisteína (4), S-hexadecilcisteína (5) e ácido gama-carboxiglutâmico (6).

5 A figura 2 fornece vários nucleotídeo e seqüências de aminoácidos que encontram uso com a invenção.

DESCRIÇÃO DETALHADA DA INVENÇÃO

A invenção fornece soluções para as limitações inerentes ao uso de um sistema de tradução confinado pelos 20 aminoácidos de ocorrência natural. As soluções incluem 10 composições e métodos relacionados a incorporação bio-sintética de sítio-específico programada e o aminoácido não natural fenilselenocisteína (figura 1, estrutura 1) nas proteínas usando sistemas de tradução ortogonal. A incorporação de fenilselenocisteína na proteína pode ser programada para ocorrer em qualquer posição desejada pela engenharia do polinucleotídeo que codifica a proteína de interesse para conter um códon seletor que 15 sinaliza a incorporação do aminoácido não natural na cadeia do polipeptídeo em crescimento.

A invenção fornece composições inéditas incluindo novos aminoacil-RNAt sintetase (O-RS) que têm a capacidade de carregar um Cognato supressor O-RNAt adequado (por exemplo, o O-RNAt de SEQ ID NO: 1) com fenilselenocisteína. Estes O-RS são novos 20 mutantes do *Methanococcus jannaschii* tirosil-RNAt sintetase que carregam seletivamente o O-RNAt com o aminoácido não natural fenilselenocisteína em células hospedeiras bacterianas. Mais preferivelmente, os componentes ortogonais não apresentam reação cruzada com componentes hospedeiros endógenos do maquinário translacional da célula hospedeira (por exemplo, uma célula *E. coli*).

25 O O-RS da invenção pode incluir o O-RS de SEQ ID NOS: 4, 6 ou 8. A invenção também fornece polinucleotídeos que codificam estes polipeptídeos. O presente revelação também descreve o metodologia para evolução dos pares de O-RNAt/O-RS inéditos que funcionam na eubactéria para incorporar especificamente no sítio um aminoácido não natural fenilselenocisteína em resposta a códons seletores.

30 A invenção também fornece métodos para a incorporação genética e sítio-específico altamente eficiente de fenilselenocisteína (figura 1, estrutura 1) nas proteínas (preferivelmente *in vivo*) em resposta a um códon seletor (por exemplo, o códon não sentido âmbar, TAG). Estes métodos e composições inéditos podem ser usados, por exemplo, em um sistema hospedeiro bacteriano.

35 Em alguns casos, o aminoácido não natural fenilselenocisteína pode em seguida ser de modo específico e regio-seletivamente modificados após sua incorporação em um polipeptídeo, da maneira descrita na presente revelação. Em virtude da reação química do

grupo fenilseleno, proteínas nas quais o aminoácido não natural é incorporado podem ser modificadas com seletividade extremamente alta. Em alguns casos, grupo reativo de aminoácido não natural fenilselenocisteína tem a vantagem de ser completamente alienado aos sistemas *in vivo*, melhorando dessa maneira a seletividade da reação. Em alguns aspectos, as reações de modificação podem ser conduzidas usando condições de reação relativamente brandas que permitem reações de conjugação tanto *in vitro* quanto *in vivo* envolvendo proteínas, e preservando disponibilidade da célula hospedeira e/ou atividade biológica da proteína.

Em alguns aspectos, a fração de fenilselenocisteína incorporada é modificada, cujo produto modificado é em seguida um a um novamente modificado, por exemplo, por uma reação de conjugação.

A natureza do material que é ultimamente conjugado na posição selecionada na proteína (correspondente ao códon seletor no quadro de leitura que codifica a proteína) não é particularmente limitada, e pode ser qualquer entidade desejada, por exemplo, lipídios, corantes, fluoróforos, agentes de reticulação, derivados de sacarídeo, polímeros (por exemplo, derivados de polietileno glicol), fotoreticuladores, compostos citotóxicos, marcadores de afinidade, derivados de biotina, resinas, microesferas, uma segunda proteína ou polipeptídeo (ou mais), polinucleotídeo(s) (por exemplo, DNA, RNA, etc.), quelantes de metais, cofatores, ácidos graxos, carboidratos, e similares.

Em alguns aspectos, para demonstrar (mas não limitar) a presente invenção, a revelação descreve aqui a fração de aminoácido não natural fenilselenocisteína incorporada em uma proteína modelo, por exemplo, mioglobina, hormônio do crescimento humano e GFP. Não se pretende que a incorporação do aminoácido não natural fenilselenocisteína seja limitada a qualquer particular proteína. Deve ficar claro que a incorporação de aminoácido não natural fenilselenocisteína em qualquer proteína de interesse desejada pode ser realizada usando as diretrizes da presente revelação. A geração de proteínas que compreendem um ou mais aminoácidos não naturais fenilselenocisteína (tanto sozinhas quanto em combinação com outros aminoácidos não naturais diferentes) é vantajosa com uma variedade de propósitos, por exemplo, para o uso em proteínas terapêuticas e com propósitos de pesquisa.

MODIFICAÇÃO DE FENILSELENOCISTEÍNA

Em alguns aspectos, a invenção fornece métodos para a modificação do resíduo de aminoácido fenilselenocisteína após sua incorporação em um polipeptídeo. Uma modificação como essa, por exemplo, clivagem oxidativa, pode converter benéficamente o resíduo de aminoácido fenilselenocisteína no aminoácido de deidroalanina α,β -insaturado (Ver figura 1, estrutura 2). Este aminoácido de deidroalanina não natural é também reativo e pode ser subsequentelemente modificado.

Não pretende-se que a invenção seja limitada a nenhum mecanismo particular (por exemplo, eliminação oxidativa) ou condições de reação específicas (por exemplo, exposição ao peróxido de hidrogênio) para a conversão de fenilselenocisteína em deidroalanina. Versados na tecnologia reconhecerão uma variedade de mecanismos e condições de reação alternativos adequados que encontram igual uso com a invenção para a conversão de um resíduo de fenilselenocisteína em deidroalanina.

MODIFICAÇÃO DE DEIDROALANINA

Em alguns aspectos, a invenção fornece métodos para modificação adicional de um resíduo de aminoácido de deidroalanina não natural em um polipeptídeo. O resíduo de deidroalanina é reativo e pode ser alvejado em reações de modificação altamente específicas. Estes métodos são especialmente usados na formação de lipídio conjugado para produzir proteínas lipidadas.

Da maneira aqui descrita, reações de adição de Michael de deidroalanina do aminoácido não natural resultam em proteínas com modificações pós-translacionais sítio-específico programadas. Por exemplo, reação de deidroalanina com um tio-lipídio pode gerar uma proteína lipidada. Por exemplo, reação com ácido tiopalmítico resulta em palmitoilcisteína (Ver figura 1, estrutura 3), reação com farnesilmercaptano produz farnesilcisteína (Ver figura 1, estrutura 4), e reação com malonato produz ácido γ -carboxiglutâmico (Ver figura 1, estrutura 6). Além disso, reação com 1-hexadecanetiol resulta em S-hexadecilcisteína (Ver figura 1, estrutura 5).

Embora a presente especificação forneça estes exemplos, não se pretende que a invenção seja limitada a qualquer mecanismo particular (por exemplo, um caminho adição de Michael) ou reagentes específicos (tio-lipídios) na modificação (conjugação) de deidroalanina. Versados na tecnologia percebem uma variedade de mecanismos alternativos, condições de reação e reagentes adequados que encontram igual uso com a invenção para a modificação do resíduo de deidroalanina. Certamente, modificação do resíduo de deidroalanina não é limitada a conjugação de lipídio, e este mecanismo de conjugação pode ser usado para conjugar qualquer fração desejada com o polipeptídeo no sítio de deidroalanina.

TECNOLOGIA DE RNAt/AMINOACIL-RNAt SINTETASE ORTOGONAL

Um entendimento das composições inéditas e métodos da presente invenção exige um entendimento das atividades associadas com pares de RNAt ortogonal e aminoacil-RNAt sintetase ortogonal. A fim de adicionar mais aminoácidos não naturais ao código genético, são necessários novos pares ortogonais compreendendo um aminoacil-RNAt sintetase e um RNAt adequado que possam funcionar eficientemente no maquinário translacional hospedeiro, mas que sejam "ortogonais" ao sistema de tradução em questão, significando que eles funcionam independentemente da sintetase e RNAts endógenos ao sistema de

tradução. Características desejadas do par ortogonal incluem RNAt que codifica ou reconhece apenas um códon específico, por exemplo, um códon seletor, que não é decodificado por nenhum RNAt endógeno, e aminoacil-RNAt sintetase que preferencialmente aminoacila (ou "carrega") seu cognato RNAt com apenas um aminoácido não natural específico. O O-RNAt também não é tipicamente aminoacilado (ou é fracamente aminoacilado, isto é, carregado) por sintetase endógena. Por exemplo, em um sistema hospedeiro *E. coli*, um par ortogonal incluirá um aminoacil-RNAt sintetase que não reage de forma cruzada com nenhum um RNAt endógeno, por exemplo, que existem 40 em *E. coli*, e um RNAt ortogonal que não é aminoacilado por nenhuma das sintetases endógenas, por exemplo, das quais existem 21 em *E. coli*.

Os princípios gerais de sistemas de tradução ortogonal que são adequados para preparar proteínas que compreendem um ou mais aminoácidos não naturais são conhecidos na tecnologia, tais como: os métodos gerais para produzir sistemas de tradução ortogonal. Por exemplo, ver International Publication Numbers WO 2002/086075, intitulado "METHODS AND COMPOSITION FOR THE PRODUCTION OF ORTOGONAL AMINOACIL-RNAt RNAt-SYNTHETASE PAIRS", WO 2002/085923, intitulado "IN VIVO INCORPORATION OF UNNATURAL AMINO ACIDS", WO 2004/094593, intitulado "EXPANDING THE EUKARYOTIC 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, intitulado "COMPONENTS OF TRANSLATION ORTOGONAL FOR THE IN VIVO INCORPORATION OF UNNATURAL AMINO ACIDS". Cada uma dessas publicações está aqui incorporada pela referência na sua íntegra. Para discussão de sistemas de tradução ortogonal que incorporam aminoácidos não naturais, e métodos para sua produção e uso, Ver também, Wang and Schultz "Expanding the Genetic Code," *Angewandte Chemie Int. Ed.*, 44(1):34-66 (2005), Xie and Schultz, "A Expanding Genetic Code", *Methods* 36(3):227-238 (2005); Xie and Schultz, "Adding Amino acids to the Genetic Repertoire", *Curr. Opinion in Chemical Biology* 9(6):548-554 (2005); Wang *et al.*, "Expanding the Genetic Code", *Annu. Rev. Biophys. Biomol. Struct.*, 35:225-249 (2006); and Ryu and Schultz, Efficient Incorporation of Unnatural Amino Acids in "Proteins *Escherichia coli*", *Nat. Methods* (4):263-265 (2006), cujos conteúdos estão cada qual incorporados pela referência na sua íntegra.

SISTEMAS DE TRADUÇÃO ORTOGONAL

Sistemas de tradução ortogonal de um modo geral compreendem células (que podem ser células procarióticas tal como *E. coli*) que incluem um RNAt ortogonal (O-RNAt), um aminoacil RNAt sintetase ortogonal (O-RS), e um aminoácido não natural, onde o O-RS aminoacila o O-RNAt com o aminoácido não natural. Um par ortogonal da invenção pode incluir um O-RNAt, por exemplo, um supressor RNAt, um deslocamento de quadro RNAt, ou

similares, e um Cognato O-RS. Os sistemas ortogonais da invenção podem tipicamente compreender pares O-RNAt/O-RS, tanto no contexto de uma célula hospedeira quanto sem a célula hospedeira. Além dos sistemas multicomponentes, a invenção também fornece componentes individuais inédito, por exemplo, polipeptídeos aminoacil-RNAt sintetase
 5 ortogonal inéditos (por exemplo, SEQ ID NO: 4, 6 ou 8), e os polinucleotídeos que codificam aqueles polipeptídeos (por exemplo, SEQ ID NO: 5, 7 ou 9).

Em geral, quando um par ortogonal reconhece um códon seletor e carrega um aminoácido em resposta ao códon seletor, diz-se que o par ortogonal "suprime" o códon seletor. Ou seja, um códon seletor que não é reconhecido pelo sistema de tradução (por
 10 exemplo, as células) maquinário endógeno não é ordinariamente carregado, que resulta em bloqueio da produção de um polipeptídeo que de outra forma será traduzido do ácido nucléico. Em um par ortogonal sistema, o O-RS aminoacila o O-RNAt com um aminoácido não natural específico. O O-RNAt carregado reconhece o códon seletor e suprime o bloqueio translacional causado pelo códon seletor.

Em alguns aspectos, um O-RNAt da invenção reconhece um códon seletor e inclui pelo menos cerca de, por exemplo, 45 %, a 50 %, a 60 %, a 75 %, a 80 %, ou a 90 % ou
 15 mais eficiência de supressão na presença de uma sintetase cognata em resposta a um códon seletor, comparada à eficiência de supressão de um O-RNAt que compreende ou codifica uma seqüência de polinucleotídeo apresentada nas listagens de seqüência aqui.

Em algumas modalidades, a eficiência de supressão do O-RS e o O-RNAt juntos é cerca de, por exemplo, 5 vezes, 10 vezes, 15 vezes, 20 vezes, ou 25 vezes ou mais que a
 20 eficiência de supressão do O-RNAt que falta no O-RS. Em alguns aspectos, a eficiência de supressão do O-RS e o O-RNAt juntos é pelo menos cerca de, por exemplo, 35 %, 40 %, 45 %, 50 %, 60 %, 75 %, 80 %, ou 90 % ou mais da eficiência de supressão de um par
 25 sintetase ortogonal como apresentado nas listagens de seqüência aqui.

A célula hospedeira usa o par O-RNAt/O-RS para incorporar o aminoácido não natural em uma cadeia do polipeptídeo em crescimento, por exemplo, por meio de 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-RNAt. Em
 30 certos aspectos preferidos, a célula pode incluir um ou mais pares de O-RNAt/ O-RS adicionais, onde o O-RNAt adicional é carregado pelo O-RS adicional com um aminoácido não natural diferente. Por exemplo, um dos O-RNAts podem reconhecer um códon de quatro bases e o outro O-RNAt pode reconhecer um códon de parada. Alternativamente, múltiplos códons de parada diferentes ou múltiplos códons de quatro bases diferentes
 35 podem ser usados no mesmo ácido nucléico de codificação.

Conforme notado, em algumas modalidades, existem múltiplos pares de O-A/O-RS em uma célula ou outro sistema de tradução que permitem a incorporação de mais que um

aminoácido não natural em um polipeptídeo. Por exemplo, a célula pode incluir adicionalmente um par de O-RNAt/O-R diferente adicional e um segundo aminoácido não natural, onde este O-RNAt adicional reconhece um segundo códon seletor e este O-RS adicional preferencialmente aminoacila o O-RNAt com o segundo aminoácido não natural.

- 5 Por exemplo, uma célula que inclui um par de O-RNAt/O-R (onde o O-RNAt reconhece, por exemplo, um códon seletor frâmbar), pode compreender adicionalmente um segundo par ortogonal, onde o segundo O-RNAt reconhece um códon seletor diferente, por exemplo, um códon opala, um códon de quatro bases, ou similares. Desejavelmente, os pares ortogonais diferentes são derivados de diferentes fontes, que podem facilitar reconhecimento de
- 10 diferentes códons seletores.

Em certas modalidades, os sistemas compreendem uma célula tal como uma célula *E. coli* que inclui um RNAt ortogonal (O-RNAt), um aminoacil- RNAt sintetase ortogonal (O-RS), um aminoácido não natural e um ácido nucléico que compreende um polinucleotídeo que codifica um polipeptídeo de interesse, onde o polinucleotídeo

15 compreende o códon seletor que é reconhecido pelo O-RNAt. O sistema de tradução pode também ser um sistema sem célula, por exemplo, qualquer um de uma variedade de sistemas de transcrição/tradução "*in vitro*" comercialmente disponíveis em combinação com um par de O-RNAt/O-R e um aminoácido não natural da maneira aqui descrita.

O O-RNAt e/ou o O-RS podem ser de ocorrência natural ou podem ser, por

20 exemplo, derivados por mutação de um RNAt e/ou RS de ocorrência natural, por exemplo, gerando bibliotecas de RNAts e/ou bibliotecas de RSs, a partir de qualquer uma de uma variedade de organismos e/ou usando qualquer uma de uma variedade de estratégias de mutação disponível. Por exemplo, uma estratégia para produzir um par de RNAt ortogonal/aminoacil-RNAt sintetase envolve importar um par de RNAt/ sintetase heterólogo

25 (para a célula hospedeira), por exemplo, de uma fonte a não ser a célula hospedeira, ou múltiplas fontes, na célula hospedeira. As propriedades do candidato a sintetase heteróloga incluem, por exemplo, que ele não carrega nenhuma célula hospedeira RNAt, e as propriedades do candidato a RNAt heterólogo incluem, por exemplo, que eles não são aminoacilado por nenhuma célula hospedeira sintetase. Além disso, o RNAt heterólogo é

30 ortogonal a todas células hospedeiras sintetase. A segunda estratégia para gerar um par ortogonal envolve gerar bibliotecas mutantes das quais se classifica e/ou seleciona um O-RNAt ou O-RS. Estas estratégias podem também ser combinadas.

RNA_t ORTOGONAL (O-RNAt)

Um RNAt ortogonal (O-RNAt) da invenção desejavelmente media a incorporação de

35 um aminoácido não natural em uma proteína que é codificada por um polinucleotídeo que compreende um códon seletor que é reconhecido pelo O-RNAt, por exemplo, *in vivo* ou *in vitro*. Em certas modalidades, um O-RNAt da invenção inclui pelo menos cerca de, por

exemplo, um 45 %, a 50 %, a 60 %, a 75 %, a 80 %, ou a 90 % ou mais eficiência de supressão na presença de uma sintetase cognata em resposta a um códon seletor comparado a um O-RNAt compreendendo ou codificado por uma seqüência de polinucleotídeos tal como apresentada nas seqüências O-RNAt nas listagens de seqüência aqui.

A eficiência de supressão pode ser determinada por qualquer dos diversos ensaios conhecidos na tecnologia. Por exemplo, um ensaio reportador β -galactosidase pode ser usado, por exemplo, um plasmídeo lacZ derivatizado (onde a construção tem um códon seletor na seqüência de ácido nucléico lacZ) é introduzida nas células de um organismo apropriado (por exemplo, um organismo onde os componentes ortogonais podem ser usados) junto com plasmídeo que compreende um O-RNAt da invenção. Uma sintetase cognata pode também ser introduzida (tanto como um polipeptídeo quanto como um polinucleotídeo que codifica uma sintetase cognata quando expressa). As células crescem em meio até uma densidade desejada, por exemplo, até um OD₆₀₀ de cerca de 0,5, e ensaios de β -galactosidase são realizados, por exemplo, usando o estojo de ensaio Beta Fluor™ β -Galactosidase (Novagen). Supressão percentual pode ser calculada como a porcentagem de atividade para uma amostra com relação a um controle comparável, por exemplo, o valor observado da construção lacZ derivatizada, onde a construção tem um códon sentido correspondente na posição desejada em vez de um códon seletor.

Exemplos de O-RNAts da invenção são apresentados nas listagens de seqüência aqui, por exemplo, ver figura 2 e SEQ ID NO: 1. A revelação também fornece aqui orientação para o projeto de espécies O-RNAt equivalentes adicionais. Em uma molécula de RNA, tais como uma molécula O-RS mRNA, ou O-t RNA, timina (T) é substituída com uracila (U) com relação a uma dada seqüência (ou vice-versa para um DNA codificado), ou complemento destes. Modificações adicionais para as bases podem também estar presentes para gerar moléculas equivalentes funcionalmente em grande medida.

A invenção também engloba variações conservativas de O-RNAts correspondentes aos O-RNAts particulares aqui. Por exemplo, variações conservativas de O-RNAt incluem aquelas moléculas que funcionam como os O-RNAts particulares, por exemplo, como nas listagens de seqüência aqui e que mantêm a estrutura RNAt em forma de L em virtude da auto-complementaridade apropriada, mas que não tem uma seqüência idêntica a esta, por exemplo, na seqüência listada ou figura 2 e, desejavelmente, são moléculas sem ser as moléculas de RNAt tipo selvagem.

A composição compreendendo um O-RNAt pode incluir adicionalmente um aminoacil-RNAt sintetase ortogonal (O-RS), onde o O-RS preferencialmente aminoacila o O-RNAt com um aminoácido não natural. Em certas modalidades, uma composição incluindo um O-RNAt pode inclui adicionalmente um 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-RNAt, ou uma combinação de um ou mais desses pode também estar presente na célula.

Métodos de produzir um RNAt ortogonal (O-RNAt) são também um recurso da invenção. Um O-RNAt produzido pelo método é também um recurso da invenção. Em certas modalidades da invenção, os O-RNAts podem ser produzidos gerando uma biblioteca de mutantes. A biblioteca de RNAts mutantes pode ser gerada usando várias técnicas de mutagênese conhecidas na tecnologia. Por exemplo, os RNAts mutantes podem ser gerados pelas mutações de sítio-específico, mutações de ponto aleatório, recombinação homóloga, embaralhamento de DNA ou outros métodos de mutagênese recursivos, construções quiméricas ou qualquer combinação destes, por exemplo, do O-RNAt de SEQ ID NO: 1.

Mutações adicionais podem ser introduzidas em posição(s) específica(s), por exemplo, em posição(s) não conservativa(s), ou em uma posição conservativa, em uma(s) posição(s) aleatória(s), ou uma combinação de ambos em um laço ou região desejada de um RNAt, por exemplo, um alça anticódon, o braço acceptor, braço ou alça D, alça variável, braço ou alça TPC, outras regiões da molécula de tRNA, ou uma combinação destes. Tipicamente, mutações em um RNAt incluem mutação do laço anticódon de cada membro da biblioteca de mutantes RNAts para permitir o reconhecimento de um códon seletor. O método pode incluir adicionalmente adicionar mais seqüências ao O-RNAt. Tipicamente, um O-RNAt possui uma melhoria de ortogonalidade para um organismo desejado comparado ao material de partida, por exemplo, a pluralidade de seqüências de RNAt, preservando ainda sua afinidade no sentido de um RS desejado.

Os métodos opcionalmente incluem analisar a similaridade (e/ou homologia inferida) de seqüências de RNAts e/ou aminoacil-RNAt sintetase para determinar candidatos potenciais para um O-RNAt, O-RS e/ou seus pares, que parecem ser ortogonais a um organismo específico. Programas computadorizados conhecidos na tecnologia e aqui descritos podem ser usados para a análise, por exemplo, programas BLAST e de empilhamento podem ser usados. Em um exemplo, para escolher componentes translacionais ortogonais potenciais para o uso em *E. coli*, é escolhida uma sintetase e/ou um RNAt que não apresenta similaridade de seqüência próxima com os organismos eubacterianos.

Tipicamente, um O-RNAt é obtido submetendo, por exemplo, a seleção negativa uma população de células de uma primeira espécie, onde as células compreendem um membro da pluralidade de O-RNAts potenciais. A seleção negativa elimina células que compreendem um membro de bibliotecas de O-RNAts potenciais que é aminoacilado por um aminoacil-RNAt sintetase (RS) que é endógeno à célula. Isto fornece um agrupamento de

RNAs que são ortogonais à célula da primeira espécie.

Em certas modalidades, na seleção negativa, um(s) códon(s) seletor(s) é(são) introduzido(s) em um polinucleotídeo que codifica um marcador de seleção negativa, por exemplo, uma enzima que confere resistência a antibiótico, por exemplo, β -lactamase, uma
 5 enzima que confere um produto detectável, por exemplo, β -galactosidase, cloranfenicol acetiltransferase (CAT), por exemplo, um produto tóxico, tal como barnase, em uma posição não essencial (por exemplo, produzindo ainda um barnase funcional), etc. Classificação/seleção é opcionalmente feita crescendo a população de células na presença de um agente seletivo (por exemplo, um antibiótico, tal como ampicilina). Em uma
 10 modalidade, a concentração do agente de seleção é variada.

Por exemplo, para medir a atividade de RNAs supressores, é usado um sistema de seleção que é baseado na supressão *in vivo* de códon seletor, por exemplo, mutações não sentido (por exemplo, parada) ou deslocamento de quadro introduzidas em um polinucleotídeo que codifica um marcador de seleção negativa, por exemplo, um gene para
 15 β -lactamase (*bla*). Por exemplo, variantes de polinucleotídeo, por exemplo, variantes *bla*, com um códon seletor em uma certa posição (por exemplo, A184), são construídos. Células, por exemplo, bactérias, são transformadas com estes polinucleotídeos. No caso de um RNA ortogonal, que não pode ser eficientemente carregado por sintetase de *E. coli* endógena, a resistência a antibiótico, por exemplo, resistência a ampicilina, deve ser
 20 aproximadamente igual ou menor que aquela para uma bactérias transformadas sem plasmídeo. Se o RNA não é ortogonal, ou se um sintetase heteróloga capaz de carregar o RNA é co-expresso no sistema, um maior nível de antibiótico, por exemplo, resistência a ampicilina, é observado. Células, por exemplo, bactérias, são escolhidas que são incapazes de crescer em placas ágar LB com concentrações de antibiótico aproximadamente iguais às
 25 das células transformadas sem plasmídeos.

No caso de um produto tóxico (por exemplo, ribonuclease ou barnase), quando um membro da pluralidade de RNAs potenciais é aminoacilado pelo hospedeiro endógeno, por exemplo, sintetase de *Escherichia coli* (isto é, ele não é ortogonal ao hospedeiro, por exemplo, *Escherichia coli* sintetase), o códon seletor é suprido e o produto polinucleotídeo
 30 tóxico produzido leva à morte celular. células que abrigam RNA ortogonais ou RNAs não funcionais sobrevivem.

Em uma modalidade, o agrupamento de RNAs que são ortogonais a um organismo desejado é em seguida submetido a uma seleção positiva na qual um códon seletor é colocado em um marcador de seleção positiva, por exemplo, codificado por um gene de
 35 resistência a medicamento, um gene β -lactamase como esse. A seleção positiva é realizada em uma célula compreendendo um polinucleotídeo que codifica ou que compreende um membro de agrupamentos de RNAs que são ortogonais a célula, um polinucleotídeo que

codifica um marcador de seleção positiva, e um polinucleotídeo que codifica um RS cognato. Em certas modalidades, a segunda população de células compreende células que não foram eliminadas pela seleção negativa. Os polinucleotídeos são expressos na célula e a célula cresce na presença de um agente de seleção, por exemplo, ampicilina. RNAts são em seguida selecionados por sua capacidade de ser aminoacilado pela sintetase cognata co-expressa e de inserir um aminoácido em resposta a este códon seletor. Tipicamente, estas células mostram uma melhoria na eficiência de supressão, comparadas às células que abrigam RNAt(s), ou RNAts não funcionais que podem não ser eficientemente reconhecidos pela sintetase de interesse. Uma célula que abrigam o RNAts ou RNAts não funcionais que não são eficientemente reconhecidos pela sintetase de interesse, são sensíveis ao antibiótico. Portanto, RNAts que: (i) não são substratos para hospedeiro endógeno, por exemplo, *Escherichia coli*, sintetase; (ii) podem ser aminoacilados pela sintetase de interesse; e (iii) são funcionais em tradução, sobrevivem a ambas seleções.

Conseqüentemente, o mesmo marcador pode ser tanto um marcador positivo quanto negativo, dependendo do contexto no qual ele é classificado. Isto é, o marcador é um marcador positivo se ele for classificado, mas um marcador negativo se desclassificado.

O rigor da seleção, por exemplo, a seleção positiva, a seleção negativa, ou tanto a seleção positiva quanto negativa, nos métodos supradescritos opcionalmente inclui variar o rigor da seleção. Por exemplo, em virtude da barnase ser uma proteína extremamente tóxica, o rigor da seleção negativa pode ser controlado introduzindo diferentes números de códons seletores no gene barnase e/ou usando um promotor indutível. Em um outro exemplo, a concentração do agente de seleção ou classificação é variada (por exemplo, concentração de ampicilina). Em alguns aspectos da invenção, o rigor varia em virtude de a atividade desejada poder ser baixa durante as etapas iniciais. Assim, critérios de seleção menos rigorosos são aplicados em etapas iniciais e critérios mais rigorosos são aplicados em etapas finais de seleção. Em certas modalidades, a seleção negativa, a seleção positiva, ou tanto a seleção negativa quanto a positiva podem ser repetidas múltiplas vezes. Múltiplos diferentes marcadores de seleção negativa, marcadores de seleção positiva, ou tanto seleção negativa quanto marcadores positivos podem ser usados. Em certas modalidades, o marcador de seleção positiva e negativa pode ser o mesmo.

Outros tipos de seleções/classificação podem ser usados na invenção para produzir componentes translacionais ortogonais, por exemplo, um O-RNAt, um O-RS, e um par de O-RNAt/O-R 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 tanto os marcadores de positiva quanto negativa podem incluir um marcador que fluoresce ou catalisa uma reação luminescente na presença de um reagente adequado. Em uma outra modalidade, um produto do marcador é detectado por avaliação da emissão de

fluorescência das células (FACS) ou por luminescência. Opcionalmente, o marcador inclui uma afinidade baseada no marcador de classificação. Ver também, Francisco, J. A., *et al.*, (1993) *Production and fluorescence-activated cell sorting of Escherichia coli expressing a functional antibody fragment on the external surface*. Proc Natl Acad Sci U S A. 90:10444-8.

5 Métodos adicionais para produzir um RNAt ortogonal recombinante podem ser encontrados, por exemplo, nas publicações de pedido de patente internacionais WO 2002/086075, intitulado "METHODS AND COMPOSITIONS FOR THE PRODUTUION OF ORTOGONAL RNAt SYNTHETASE AMINOACYL-RNAt PAIRS", WO 2004/094593, intitulado "EXPANDING THE EUKARYOTIC GENETIC CODE GENETIC" e WO
10 2005/019415, depositado em 7 de julho de 2004. Ver também Forster *et al.*, (2003) *Programming peptidomimetic sintetase by translating genetic codes designed de novo* PNAS 100(11):6353-6357, e, Feng *et al.*, (2003), *Expanding RNAt recognition of a RNAt sintetase for the single amino acid change*, PNAS 100(10): 5676-5681.

AMINOACIL-RNAt SINTETASE ORTOGONAL (O-RS)

15 Um O-RS da invenção preferencialmente aminoacila um O-RNAt com um aminoácido não natural, *in vitro* ou *in vivo*. Um O-RS da invenção pode ser fornecido ao sistema de tradução, por exemplo, uma célula, por um polipeptídeo que inclui um O-RS e/ou por um polinucleotídeo que codifica um O-RS ou uma porção destes. Por exemplo, um exemplo O-RS compreende uma seqüência de aminoácidos apresentada em SEQ ID NO: 4,
20 6 ou 8, ou um variação conservativa destes. Em um outro exemplo, um O-RS, ou uma porção deste, é codificado por uma seqüência de polinucleotídeos que codifica um aminoácido que compreende a seqüência na lista ou exemplos de seqüência aqui, ou uma seqüência de polinucleotídeos complementar desta. Ver, por exemplo, o polinucleotídeo de SEQ ID NO: 5, 7 ou 9.

25 Métodos para identificar um aminoacil-RNAt sintetase ortogonal (O-RS), por exemplo, um O-RS, para uso com um O-RNAt, são também um recurso da invenção. Por exemplo, um método inclui submeter à seleção, por exemplo, seleção positiva, uma população de células de uma primeira espécie, onde as células individualmente compreendem: 1) um membro de pluralidade de aminoacil-RNAt sintetase (RSs), (por
30 exemplo, a pluralidade de RSs pode incluir RSs mutante e RSs derivados de uma espécie a não ser a primeira espécie, ou tanto RSs mutantes quanto RSs derivados de uma espécie a não ser a primeira espécie); 2) o RNAt ortogonal (O-RNAt) (por exemplo, de uma ou mais espécies); e 3) um polinucleotídeo que codifica um marcador de seleção (por exemplo, positivo) e compreende pelo menos um códon seletor. Células são selecionadas ou
35 classificadas em relação àquelas que apresentam uma melhoria na eficiência de supressão, comparadas com células não têm, ou que têm uma quantidade reduzida do membro da pluralidade de RSs. A eficiência de supressão pode ser medida por técnicas conhecidas na

tecnologia e da maneira aqui descrita. Células tendo uma melhoria na eficiência de supressão compreendem um RS ativo que aminoacila o O-RNAt. Um nível de aminoacilação (*in vitro* ou *in vivo*) pelo RS ativo de um primeiro conjunto de RNAts da primeira espécie é comparado com o nível de aminoacilação (*in vitro* ou *in vivo*) pelo RS ativo de um segundo conjunto de RNAts da segunda espécie. O nível de aminoacilação pode ser determinado por uma substância detectável (por exemplo, um aminoácido marcado não natural). O RS ativo que aminoacila mais eficientemente o segundo conjunto de RNAts comparado ao primeiro conjunto de RNAts é tipicamente selecionado, fornecendo dessa maneira um aminoacil-RNAt sintetase ortogonal eficiente (otimizada) para uso com o O-RNAt. Um O-RS, identificado pelo método, é também uma recurso da invenção.

Qualquer dos diversos ensaios pode ser usado para determinar aminoacilação. Estes ensaios podem ser realizados *in vitro* ou *in vivo*. Por exemplo, ensaios de aminoacilação *in vitro* são descritos, por exemplo, em Hoben and Soil (1985) Methods Enzymol. 113:55-59. Aminoacilação pode também ser determinada usando um reportador junto com componentes de tradução ortogonal e detectando o reportador *em* uma célula que expressa um polinucleotídeo compreendendo pelo menos um códon seletor que codifica uma proteína. Ver também, WO 2002/085923, intitulado "IN VIVO INCORPORATION OF UNNATURAL AMINO ACIDS", e WO 2004/094593, intitulado "EXPANDING THE EUKARYOTIC GENETIC CODE".

O O-RS identificado pode ser manipulado adicionalmente para alterar a especificidade do substrato da sintetase, de maneira que apenas um aminoácido não natural desejado, mas não qualquer dos 20 aminoácidos comuns, seja carregado no O-RNAt. Métodos para gerar um aminoacil-RNAt sintetase ortogonal com uma especificidade de substrato para um aminoácido não natural incluem na mutação da sintetase, por exemplo, no sítio ativo na sintetase, no sítio de mecanismo de edição na sintetase, nos diferentes sítios combinando diferentes domínios de sintetase, ou similares, e na aplicação de um processo de seleção. Uma estratégia é usada, que é baseada na combinação de uma seleção positiva seguida por uma seleção negativa. Na seleção positiva, supressão do códon seletor introduzido em uma posição não essencial(s) de um marcador positivo permite que a células sobreviva sob pressão de seleção positiva. Na presença tanto de aminoácidos naturais quanto não naturais, sobreviventes assim codificam sintetase ativa carregando o supressor RNAt ortogonal tanto com um aminoácido natural quanto não natural. Na seleção negativa, supressão de um códon seletor introduzida em uma posição não essencial(s) de um marcador negativo remove especificidades de sintetase com aminoácido natural. Sobreviventes da seleção negativa e positiva codificam sintetase que aminoacila (carrega) o ortogonal supressor RNAt seleção apenas com aminoácidos não naturais. Esta sintetase pode em seguida ser submetida a mutagênese adicional, por exemplo, embaralhamento de

DNA ou outros métodos de mutagênese recursivos.

Uma biblioteca de O-RSs mutantes pode ser gerada usando várias técnicas de mutagênese conhecidas na tecnologia. Por exemplo, os RSs mutantes podem ser gerados por mutações de sítio-específico, mutações de ponto aleatório, recombinação homóloga, embaralhamento de DNA ou outros métodos de mutagênese recursiva, construções quiméricas ou qualquer combinação destes. Por exemplo, uma biblioteca de RSs mutantes pode ser produzida, por exemplo, de duas ou mais outras "sub-bibliotecas" menores, menos diversas. Bibliotecas quiméricas de RSs estão também incluídas na invenção. Deve-se notar que bibliotecas de RNAt sintetase de vários organismos (por exemplo, microorganismos tais como eubactéria ou archaeobacteria) tal como bibliotecas que compreendem diversidade natural (Ver, por exemplo, patente U.S. No. 6.238.884 de Short et al, patente U.S. No. 5.756.316 de Schallenberger et al; patente U.S. No. 5.783.431 de Petersen et al, patente U.S. No. 5.824.485 de Thompson et al, patente U.S. No. 5.958.672 de Short et al), são opcionalmente construídas e classificadas pelos pares ortogonais.

Uma vez que a sintetase é submetida a estratégia de seleção/classificação positiva e negativa, esta sintetase pode em seguida ser submetida a mutagênese adicional. Por exemplo, um ácido nucléico que codifica o O-RS pode ser isolado; um conjunto de polinucleotídeos que codifica O-RSs mutado (por exemplo, por mutagênese aleatória, mutagênese de sítio-específico, recombinação ou qualquer combinação destes) pode ser gerado do ácido nucléico; e estas etapas individuais ou uma combinação destas etapas, podem ser repetidas até que um O-RS mutado seja obtido, que preferencialmente aminoacila o O-RNAt 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 rigor de seleção/classificação podem também ser usados nos métodos da invenção, para produzir O-RNAt, O-RS, ou seus pares. O rigor de seleção ou classificação pode variar em uma ou ambas etapas do método para produzir um O-RS. Isto pode incluir, por exemplo, variar a quantidade de agente de seleção/classificação que é usada, etc. Etapas adicionais de seleções positivas e/ou negativas podem também ser realizadas. Seleção ou classificação pode também compreender uma ou mais de uma mudança na permeabilidade do aminoácido, uma mudança na eficiência de tradução, uma mudança na fidelidade translacional, etc. Tipicamente, uma ou mais mudança é baseada em uma mutação em um ou mais gene em um organismo no qual um par de RNAt ortogonal-RNAt sintetase é usado para produzir proteína.

Detalhes gerais adicionais para produzir O-RS e alterar a especificidade do substrato da sintetase podem ser encontrados em Publicação internacional WO 2002/086075, intitulado "METHODS AND COMPOSITIONS FOR THE PRODUCTION OF ORTOGONAL RNAt SINTETASE AMINOACIL-RNAt PARES", e WO 2004/094593, intitulado

"EXPANDING THE EUKARYOTIC GENETIC CODE". Ver também, Wang and Schultz "Expanding the genetic Code", *Angewandte Chemie Int. Ed.*, 44(1):34-66 (2005), cujos conteúdos estão aqui incorporados pela referência nas suas íntegras.

ORGANISMOS FONTES E HOSPEDEIROS

Os componentes translacionais ortogonais (O-RNAt e O-RS) da invenção podem ser derivados de qualquer organismo (ou uma combinação de organismos) para uso em um sistema hospedeiro de tradução de qualquer outra espécie, com a condição de que os componentes O-RNAt/O-RS e o sistema hospedeiro funcionem de uma maneira ortogonal. Não é uma exigência que o O-RNAt e o O-RS de um par ortogonal sejam derivados do mesmo organismo. Em alguns aspectos, os componentes ortogonais são derivados de genes Archaea (isto é, archaeobacteria) para uso em um sistema hospedeiro eubacteriano.

Por exemplo, o O-RNAt ortogonal pode ser derivado de um organismo Archaea, por exemplo, um archaeobacterium, tais como *Methanococcus jannaschii*, *Methanobacterium thermoautotrophicum*, *Halobacterium* tais como *Haloferax volcanii* e *Halobacterium* espécie NRC-1, *Archaeoglobus fulgidus*, *Pyrococcus furiosus*, *Pyrococcus horikoshii*, *Aeuropyrum pernix*, *Methanococcus maripaludis*, *Methanopyrus kandleri*, *Methanosarcina mazei* (Mm), *Pyrobaculum aerophilum*, *Pyrococcus abyssi*, *Sulfolobus solfataricus* (Ss), *Sulfolobus tokodaii*, *Thermoplasma acidophilum*, *Thermoplasma volcanium*, ou similares, ou um eubactéria, tais como *Escherichia coli*, *Thermus thermophilus*, *Bacillus subtilis*, *Bacillus stearothermophilus*, ou similares. Enquanto o O-RS ortogonal pode ser derivado de um organismo ou combinação de organismos, por exemplo, um archaeobacterium, tais como *Methanococcus jannaschii*, *Methanobacterium thermoautotrophicum*, *Halobacterium* tais como *Haloferax volcanii* e *Halobacterium* espécie NRC-1, *Archaeoglobus fulgidus*, *Pyrococcus furiosus*, *Pyrococcus horikoshii*, *Aeuropyrum pernix*, *Methanococcus maripaludis*, *Methanopyrus kandleri*, *Methanosarcina mazei*, *Pyrobaculum aerophilum*, *Pyrococcus abyssi*, *Sulfolobus solfataricus*, *Sulfolobus tokodaii*, *Thermoplasma acidophilum*, *Thermoplasma volcanium*, ou similares, ou uma eubactéria, tais como *Escherichia coli*, *Thermus thermophilus*, *Bacillus subtilis*, *Bacillus stearothermophilus*, ou similares. Em uma modalidade, fontes eucarióticas, por exemplo, plantas, algas, protista, fungo, levedura, animais (por exemplo, mamíferos, insetos, arthropods, etc.), ou similares, podem também ser usadas como fontes de O-RNAts e O-RSs.

Os componentes individuais de um par de O-RNAt/O-R podem ser derivados do mesmo organismo ou de organismos diferentes. Em uma modalidade, o par de O-RNAt/O-R é do mesmo organismo. Alternativamente, o O-RNAt e o O-RS do par de O-RNAt/O-R são de organismos diferentes.

O O-RNAt, O-RS ou o par O-RNAt/O-R podem ser selecionados ou classificados *in vivo* ou *in vitro* e/ou usados em uma célula, por exemplo, uma célula eubacteriana, para

produzir um polipeptídeo com um aminoácido não natural. Uma célula eubacteriana usada não é limitada, por exemplo, *Escherichia coli*, *Thermus thermophilus*, *Bacillus subtilis*, *Bacillus stearothermophilus*, ou similares. Composições de células eubacterianas que compreende componentes translacionais da invenção são também um recurso da invenção.

5 Ver também, publicação de pedido de patente internacional 2004/094593, intitulado "EXPANDING THE EUKARYOTIC GENETIC CODE", depositado em 16 de abril de 2004, para classificar O-RNAt e/ou O-RS em uma espécie para uso em uma outra espécie.

Embora sistemas de tradução ortogonal (por exemplo, compreendendo um O-RS, um O-RNAt e um aminoácido não natural) possam utilizar células hospedeiras cultivadas para produzir proteínas tendo aminoácidos não naturais, não pretende-se que um sistema de ORTOGONAL TRANSLATION da invenção exija uma célula hospedeira intacta viável. Por exemplo, um sistema de ORTOGONAL TRANSLATION pode utilizar um sistema sem célula na presença de um extrato da célula. Realmente, o uso de transcrição/sistemas sem célula *in vitro*, de tradução para produção de proteína é uma técnica bem estabelecida. A adaptação destes sistemas *in vitro* para produzir proteínas tendo aminoácidos não naturais usando sistema de componente de translações ortogonais aqui descritos está bem no escopo da invenção.

CÓDONS SELETORES

Códons seletores da invenção expandem a estrutura do códon genético do maquinário bio-sintético da proteína. Por exemplo, um códon seletor inclui, por exemplo, um códon de três bases exclusive, um códon não sentido, tal como um códon de parada, por exemplo, um códon âmbar (UAG), ou um códon opala (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 que três, etc. Usando códons seletores diferentes, múltiplos pares RNAt ortogonal/sintetase podem ser usado que permitem a incorporação sítio-específico simultânea de múltiplos aminoácidos não naturais por exemplo, incluindo 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 de parada para a incorporação de um aminoácido não natural *in vivo* em uma célula. Por exemplo, um O-RNAt é produzido que reconhece o códon de parada e é aminoacilado por um O-RS com um aminoácido não natural. Este O-RNAt não é reconhecido pela aminoacil-RNAt sintetase do hospedeiro de ocorrência natural. Mutagênese de sítio dirigido convencional pode ser usado para introduzir o códon de parada no sítio de interesse em um polinucleotídeo que codifica um polipeptídeo de interesse. Ver, por exemplo, Sayers, J.R., *et al.* (1988), *5',3' Exonuclease in phosphorotioate-based oligonucleotide-directed mutagenesis*. Nucleic Acids Res. 791-802. Quando o O-RS, O-RNAt e o ácido nucléico que

codifica um polipeptídeo de interesse são combinados, por exemplo, *in vivo*, o aminoácido não natural é incorporado em resposta ao códon de parada para dar um polipeptídeo contendo o aminoácido não natural na posição especificada. Em uma modalidade da invenção, o códon de parada usado como um códon seletor é um códon âmbar, UAG, e/ou
 5 um códon opala, UGA. Em um exemplo, um código genético no qual UAG e UGA são ambos usados como um códon seletor pode codificar 22 aminoácidos preservando ainda o códon não sentido ocre, UAA, que é o sinal de terminação mais abundante.

A incorporação de aminoácidos não naturais *in vivo* pode ser feita sem perturbação significativa da célula hospedeira. Por exemplo, em células não eucarióticas, tal como
 10 *Escherichia coli*, em virtude da eficiência de supressão para o códon UAG depender da competição entre o O-RNAt, por exemplo, o supressor âmbar de RNAt, e do fator 1 de liberação (RF1) (que se liga ao códon UAG e inicia a liberação do peptídeo de crescimento do ribossomo), a eficiência de supressão pode ser modulada, por exemplo, tanto aumentando o nível de expressão de O-RNAt, por exemplo, o supressor RNAt, quanto
 15 usando uma cepa deficiente em RF1. Em células eucarióticas, em virtude da eficiência de supressão com o códon UAG depender da competição entre o O-RNAt, por exemplo, o supressor âmbar de RNAt, e um fator de liberação eucariótica (por exemplo, eRF) (que liga a um códon de parada e inicia a liberação do peptídeo de crescimento do ribossomo), a eficiência de supressão pode ser modulada, por exemplo, aumentando a nível de expressão
 20 de O-RNAt, por exemplo, o supressor RNAt. Além disso, compostos adicionais podem também estar presentes, por exemplo, agentes de redução tal como ditionetol (DTT).

Aminoácidos não naturais podem também ser codificados com códons raros. Por exemplo, quando a concentração de arginina em uma reação de síntese de proteína *in vitro* é reduzido, o códon de arginina raro, AGG, demonstrou ser eficiente para inserção de Ala
 25 por um RNAt sintético acilado com alanina. Ver, por exemplo, Ma *et al.*, Biochemistry, 32:7939 (1993). Neste caso, o RNAt sintético compete com o RNAt^{Arg} de ocorrência natural, que existe como uma espécie secundária em *Escherichia coli*. Além disso, alguns organismos não usam todos os códons triplete. Um códon AGA não projetado em *Micrococcus luteus* foi utilizado para inserção de aminoácidos em um extrato transcrição/tradução *in vitro*. Ver, por exemplo, Kowal e Oliver, Nucl. Ácido. Res., 25:4685
 30 (1997). Componentes da invenção podem ser gerados para usar esses códons raros *in vivo*.

Códons seletores podem também compreender códons estendidos, por exemplo, códons de quatro ou mais bases, tais como, códons de quatro, cinco, seis ou mais bases. Exemplos de códons de quatro bases incluem, por exemplo, AGGA, CUAG, UAGA, CCCU, e similares. Exemplos de códons de cinco bases incluem, por exemplo, AGGAC, CCCCUC,
 35 CCCUC, CUAGA, CUACU, UAGGC e similares. Métodos da invenção incluem usar códons estendidos baseado na supressão de deslocamento de quadro. Códons de quatro ou mais

bases podem inserir, por exemplo, um ou múltiplos aminoácidos não naturais, na mesma proteína. Em outras modalidades, as alças anticódon podem decodificar, por exemplo, pelo menos um códon de quatro bases, pelo menos um códon de cinco bases, ou pelo menos um códon de seis bases ou mais. Subseqüentemente, existem 256 códons de quatro bases possíveis, múltiplos aminoácidos não naturais podem ser codificados na mesma célula usando um códon de quatro ou mais bases. Ver também, Anderson *et al.*, (2002) *Exploring the Limits of Codon and Anticodon Size*, Chemistry and Biology, 9:237-244, e Magliery, (2001) *Expanding the genetic Code: Selection of Efficient Suppressors of Four-base Codons and Identification of "Shifty" Four-base Codons whiht a Library Approach in Escherichia coli*, J. Mol. Biol. 307: 755-769.

Por exemplo, códons de quatro bases têm sido usados para incorporar aminoácidos não naturais nas proteínas usando métodos bio-sintéticos *in vitro*. Ver, por exemplo, Ma *et al.*, (1993) Biochemistry, 32:7939; e Hohsaka *et al.*, (1999) J. Am. Chem. Soc., 121:34. CGGG e AGGU foram usados para incorporar simultaneamente 2-naftilalanina e um NBD derivado de lisina em estreptavidina *in vitro* com dois supressores RNAts de deslocamento de quadro quimicamente acilados. Ver, por exemplo, Hohsaka *et al.*, (1999) J. Am. Chem. Soc., 121:12194. Em um estudo *in vivo*, Moore *et al.* examinou a capacidade de RNAt^{Leu} derivados com anticódons UCUA suprir códons UAGN (N pode ser U, A, G, ou C), e descobriu que o UAGA quadruplete pode ser decodificado por um RNAt^{Leu} com um anticódon UCUA com uma eficiência de 13 a 26 % com pequena decodificação no quadro 0 ou -1. Ver Moore *et al.*, (2000) J. Mol. Biol., 298:195. Em uma modalidade, códons estendidos baseados nos códons raros ou códons não sentido podem ser usados na invenção, que podem reduzir leitura de sentido incorreto e supressão de deslocamento de quadro em outros sítios indesejados. Códons de quatro base têm sido usados como códons seletores em uma variedade de sistemas ortogonais. Ver, por exemplo, WO 2005/019415; WO 2005/007870 and WO 2005/07624. Ver também, Wang and Schultz "Expanding the genetic Code," *Angewandte Chemie Int. Ed.*, 44(1):34-66 (2005), cujos conteúdos estão aqui incorporados nas suas íntegras pela referência. Embora os exemplos a seguir utilizem um códon seletor fr âmbar, códons de quatro ou mais bases podem ser igualmente usados, modificando os exemplos aqui para incluir O-RNAts de quatro bases e sintetase modificados para incluir mutações similares àquelas descritas anteriormente para vários O-RSs de aminoácido não natural:

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

Códons seletores opcionalmente incluem pares de base não naturais. Estes pares

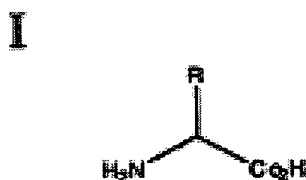
de base não naturais expandem adicionalmente o alfabeto genético existente. Um par de base extra aumenta o número de códons triplete de 64 para 125. Propriedades dos terceiros pares de base incluem pareamento de base estável e seletivo, incorporação enzimática eficiente no DNA com alta fidelidade por um polimerase, e a extensão do iniciador continuou eficiente depois da síntese do par de base não natural nascente. Descrições de pares de base não naturais que podem ser adaptadas para métodos e composições incluem, por exemplo, Hirao, *et al.*, (2002) *Na unnatural base pair for incorporating amino acid analogues into protein*, Nature Biotechnology, 20:177-182. Ver também Wu, Y., *et al.*, (2002) J. Am. Chem. Soc. 124:14626-14630. Outras publicações relevantes são listadas a seguir.

Para uso *in vivo*, o nucleosídeo não natural é permeável a membrana e é fosforilado para formar o trifosfato correspondente. Além disso, a maior informação genética é estável e não destruída pelas enzimas celulares. Esforços anteriores de Benner e outros tiram vantagem de padrões de ligação de hidrogênio que são diferentes daqueles em pares canônicos de Watson-Crick, o mais notável exemplo de que é o par iso-C:iso-G. Ver, por exemplo, Switzer *et al.*, (1989) J. Am. Chem. Soc., 111:8322; and Piccirilli *et al.*, (1990) Nature, 343:33; Kool, (2000) Curr. Opin. Chem. Biol., 4:602. Essas bases em geral faltam até um certo ponto de emparelhamento com bases naturais e não podem ser enzimaticamente replicadas. Kool e colaboradores demonstram que interações e empacotamento hidrofóbico entre bases podem substituir ligação de hidrogênio para conduzir a formação de par de base. Ver Kool, (2000) Curr. Opin. Chem. Biol., 4:602; and Guckian and Kool, (1998) Angew. Chem. Int. Ed. Engl., 36, 2825. Em um esforço para desenvolver um par de base não natural satisfazendo todas as exigências anteriores, Schultz, Romesberg e colaboradores têm sistematicamente sintetizado e estudado uma série de bases hidrofóbicas não naturais. Um auto-par PICS:PICS deve ser mais estável do que pares de base naturais, e podem ser eficientemente incorporados no DNA por fragmento Klenow de *Escherichia coli* DNA polimerase I (KF). Ver, por exemplo, McMinn *et al.*, (1999) J. Am. Chem. Soc., 121:11586 e Ogawa *et al.*, (2000) J. Am. Chem. Soc., 122:3274. Um auto-par 3MN:3MN pode ser sintetizado por KF com eficiência e seletividade suficiente para função biológica. Ver, por exemplo, Ogawa *et al.*, (2000) J. Am. Chem. Soc., 122:8803. Portanto, ambas bases agem como um terminador de cadeia para replicação adicional. Foi recentemente desenvolvido um DNA polimerase mutante que pode ser usado para replicar o auto-par PICS. Além disso, um auto-par 7AI pode ser replicado. Ver, por exemplo, Tae *et al.*, (2001) J. Am. Chem. Soc., 123:7439. Tem também sido desenvolvido um par de metalobase inédito, Dipic:Py, que forma um par estável mediante ligação Cu(II). Ver Meggers *et al.*, (2000) J. Am. Chem. Soc., 122:10714. Em virtude de códons estendidos e códons não naturais serem intrinsecamente ortogonais aos códons naturais, os métodos da invenção podem tirar vantagem desta propriedade para gerar RNAt ortogonais para eles.

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

AMINOÁCIDOS NÃO NATURAIS

Da maneira aqui usada, um aminoácido não natural refere-se a qualquer aminoácido, aminoácido modificado, ou aminoácido análogo a não ser 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, fenilalanina, prolina, serina, treonina, triptofano, tirosina, valina. A estrutura genérica de um alfa-aminoácido é ilustrada pela Fórmula I:



Um aminoácido não natural é tipicamente qualquer estrutura tendo a Fórmula I em que o grupo R é qualquer substituinte a não ser um usado nos vinte aminoácidos naturais. Ver, por exemplo, Biochemistry por L. Stryer, P ed. 1988, Freeman and Company, New York, para as estruturas dos vinte aminoácidos naturais. Note que os aminoácidos não naturais da invenção podem ser compostos de ocorrência natural sem ser os vinte alfa-aminoácidos anteriores.

Em virtude de os aminoácidos não naturais da invenção tipicamente diferirem dos aminoácidos naturais na cadeia lateral, os aminoácidos não naturais formam amida ligada com outros aminoácidos, por exemplo, natural ou não natural, da mesma maneira da qual eles são formados em proteínas de ocorrência natural. Portanto, os aminoácidos não naturais têm grupos de cadeia lateral que os distingue dos aminoácidos naturais.

O aminoácido não natural fenilselenocisteína é de particular interesse aqui (Ver figura 1, estrutura 1). Além do aminoácido não natural fenilselenocisteína, outros aminoácidos não naturais podem ser simultaneamente incorporados em um polipeptídeo de interesse, por exemplo, usando um segundo par O-RS/O-RNAt apropriado em conjunto com um par ortogonal fornecido pela presente invenção. Muitos desses aminoácidos não naturais adicionais e pares ortogonais adequados são conhecidos. Ver a presente revelação e as referências aqui citadas. Por exemplo, ver Wang and Schultz "Expanding o Genetic Code", *Angewandte Chemie Int. Ed.*, 44(1):34-66 (2005), Xie and Schultz, "A Expanding Genetic

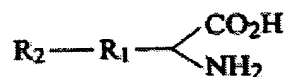
Code", *Methods* 36(3):227-238 (2005), Xie e Schultz, "Adding Amino acids to the Genetic Repertoire", *Curr. Opinion in Chemical Biology* 9(6):548-554 (2005) e Wang *et al.*, "Expanding the Genetic Code", *Annu. Rev. Biophys. Biomol. Struct.*, 35:225-249 (2006); cujos conteúdos estão cada qual incorporados pela referência nas suas íntegras.

Embora o aminoácido não natural fenilselenocisteína mostrado na figura 1, estrutura 1, seja de interesse primário nos exemplos aqui descritos, não pretende-se que a invenção seja estritamente limitada àquela estrutura. Realmente, uma variedade de análogos estruturalmente relacionados facilmente derivados pode ser imediatamente produzida que retêm a característica principal da fenilselenocisteína mostrada na figura 1, e também são especificamente reconhecidos pelo aminoacil-RNAt sintetase da invenção (por exemplo, o O-RS de SEQ ID NOS: 4, 6 e 8). Pretende-se que estes aminoácidos análogos relacionados estejam no escopo da invenção.

Em outros aminoácidos não naturais, por exemplo, R na Fórmula I opcionalmente compreende um alquil-, aril-, acil-, hidrazina, ciano-, halo-, hidrazida, alquenila, éter, borato, boronato, fosfo, fosfono, fosfina, enona, imina, éster, hidroxilamina, amina, e similares, ou qualquer combinação destes. Outros aminoácidos não naturais de interesse incluem, mas sem limitações, aminoácidos que compreendem um reticulador fotoativável, aminoácido com marcador de spin, aminoácidos fluorescentes, aminoácidos de ligação de metal, aminoácidos contendo metal, aminoácidos radioativos, aminoácidos com grupos funcionais inéditos, aminoácidos que interagem covalente ou não covalentemente com outras moléculas, aminoácidos fotoativados e/ou fotoisomerizáveis, biotina ou biotina-análoga contendo aminoácidos, ceto contendo aminoácidos, aminoácidos glicosilados, uma fração de sacarídeo anexada à cadeia lateral de aminoácido, aminoácidos compreendendo polietileno glicol ou poliéter, aminoácidos substituídos com átomo pesado, aminoácidos quimicamente cliváveis ou fotocliváveis, aminoácidos com uma cadeia lateral alongada comparada com os ácidos naturais (por exemplo, poliéteres ou hidrocarbonetos de cadeia longa, por exemplo, maior que cerca de 5, maior que cerca de 10 carbonos, etc.), aminoácidos contendo açúcar ligados a carbono, amino tioácido contendo aminoácidos, e aminoácidos contendo uma ou mais frações tóxicas.

Em um outro aspecto, a invenção fornece aminoácidos não naturais tendo a estrutura geral ilustrada pela Fórmula IV a seguir:

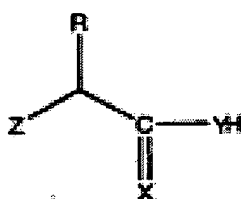
IV



Um aminoácido não natural tendo esta estrutura é tipicamente qualquer estrutura onde R_1 é um substituinte usado em um dos vinte ácidos naturais (por exemplo, tirosina ou fenilalanina) e R_2 é um substituinte. Assim, este tipo de aminoácido não natural pode ser visto como um aminoácido natural derivado.

- 5 Além dos aminoácidos não naturais que contêm a estrutura fenilselenocisteína mostrado na figura 1, estrutura 1, aminoácidos não naturais podem também compreender opcionalmente estruturas da espinha dorsal modificadas, por exemplo, da maneira ilustrada pelas estruturas das Fórmulas II e III:

II



III



- em que Z tipicamente compreende OH, NH₂, SH, NH-R', ou S-R'; X e Y, que podem ser os mesmos ou diferentes, tipicamente compreendem S ou O, e R e R', que são opcionalmente os mesmos ou diferentes, são tipicamente selecionados da mesma lista de constituintes para o grupo R descrito anteriormente para os aminoácidos não naturais tendo Fórmula I bem como hidrogênio. Por exemplo, aminoácidos não naturais da invenção, opcionalmente compreendem substituições no grupo amino ou carboxila da maneira
- 10 ilustrada pelas Fórmulas II e III. Aminoácidos não naturais deste tipo incluem, mas sem limitações, ácidos α -hidróxi, α -tioácido α -aminotiocarboxilatos, por exemplo, com cadeias laterais correspondentes às cadeias laterais de vinte aminoácidos naturais ou não naturais comuns. Além disso, substituições no α -carbono opcionalmente incluem L, D, ou α - α -aminoácidos di-substituídos tais como D-glutamato, D-alanina, D-metil-O-tirosina, ácido
- 15 aminobutírico, e similares. Outras alternativas estruturais incluem aminoácidos cíclicos, tal como prolina análoga bem como prolina análoga de anel de 3, 4, 6, 7, 8 e 9 membros, aminoácidos β e γ tais como β -alanina e ácido γ -amino butírico substituídos.

Em alguns aspectos, a invenção utiliza aminoácidos não naturais na configuração

L. Portanto, não pretende-se que a invenção seja limitada ao uso de aminoácidos não naturais de configuração L. Observou-se que o enantiômeros D destes aminoácidos não naturais também encontram uso na invenção.

Os aminoácidos não naturais que encontram uso na invenção não estão estritamente limitados ao aminoácido não natural fenilselenocisteína mostrado na figura 1, estrutura 1. Versados na tecnologia reconhecerão que uma ampla variedade de análogos não naturais de aminoácidos de ocorrência natural é facilmente derivada. Por exemplo, mas sem limitações, derivados não naturais de tirosina são facilmente produzidos. Análogos de tirosina incluem, por exemplo, tirosinas para-substituídas, tirosinas orto-substituídas, e tirosinas meta-substituídas, em que a tirosina substituída compreende um grupo alquil, um grupo acetila, um grupo benzoila, um grupo amino, uma hidrazina, uma hidroxiamina, um grupo tiol, um grupo carboxi, um grupo isopropila, um grupo metila, uma cadeia ramificada $C_6 - C_{20}$ ou hidrocarboneto ramificado, um hidrocarboneto saturado ou insaturado, um grupo O-metila, um grupo poliéter, um grupo nitro, ou similares. Além disso, anéis arila

15 multiplamente substituídos são também observados. Análogos glutamina da invenção incluem, mas sem limitações, derivados de α -hidróxi, derivados de γ -substituído, derivados cíclicos, e derivados de glutamina substituído por amida. Exemplos de análogos de fenilalanina incluem, mas sem limitações, fenilalaninas para-substituídas, fenilalaninas orto-substituído, e fenilalaninas meta-substituído, em que o substituinte compreende um grupo

20 alquil, um grupo hidróxi, um grupo metóxi, um grupo metila, um grupo alila, um aldeído, um nitro, um grupo tiol, ou grupo ceto, ou similares. Exemplos específicos de aminoácidos não naturais incluem, mas sem limitações, fenilselenocisteína, sulfotirosina, p-etiltiocarbonil-1-fenilalanina, p-(3-oxobutanoil)-1-fenilalanina, 1,5-dansil-alanina, aminoácido de 7-amino-cumarina, aminoácido de 7-hidróxi-cumarina, nitrobenzil-serina, O-(2-nitrobenzil)-1-tirosina, p-carboximetil-1-fenilalanina, p-ciano-1-fenilalanina, m-ciano-1-fenilalanina, bifenilalanina, 3-amino-1-tirosina, biperidil alanina, p-(2-amino-1-hidroxietil)-1-fenilalanina, p-isopropiltiocarbonil-1-fenilalanina, 3-nitro-1-tirosina e p-nitro-1-fenilalanina. Também, um p-propargiloxifenilalanina, um 3,4-dihidróxi-1-fenilalanina (DHP), um 3, 4, 6-triidróxi-1-fenilalanina, um 3,4,5-triidróxi-1-fenilalanina, 4-nitro-fenilalanina, ap-acetil-1-fenilalanina, O-

30 metil-1-tirosina, um L-3-(2-naftil)alanina, um 3-metil-fenilalanina, um O-4-alil-1-tirosina, um 4-propil-1-tirosina, um 3-nitro-tirosina, um 3-tiol-tirosina, um tri-O-acetil-GlcNAc β -serina, um L-Dopa, um fenilalanina fluorado, um isopropil-1-fenilalanina, um p-azido-1-fenilalanina, um p-acil-1-fenilalanina, um p-benzoil-1-fenilalanina, um L-fosfoserina, um fosfonoserina, um fosfontirosina, um p-iodo-fenilalanina, um p-bromofenilalanina, um p-amino-1-fenilalanina, e um isopropil-1-fenilalanina, e similares. As estruturas de uma variedade de aminoácidos não naturais são reveladas nas referências aqui citadas. Ver também, pedidos de patente internacionais publicados WO 2004/094593, intitulado "EXPANDING THE EUKARYOTICA

GENETIC CODE," e WO 2006/110182, intitulado " ORTOGONAL TRANSLATION COMPONENTS FOR THE *IN VIVO* INCORPORATION OF UNNATURAL AMINO ACIDS", depositado em 27 de outubro de 2005.

SÍNTESE QUÍMICA DE AMINOÁCIDOS NÃO NATURAIS

Muitos dos aminoácidos não naturais fornecidos anteriormente estão comercialmente disponíveis, por exemplo, pela Sigma (USA) ou Aldrich (Milwaukee, WI, USA). Aqueles que não estão comercialmente disponíveis são opcionalmente sintetizados, fornecidos em várias publicações ou usando métodos padrões conhecidos pelos versados na tecnologia. Para técnicas de síntese orgânica, Ver, por exemplo, Organic Chemistry de Fessenden e Fessenden, (1982, Second Edition, Willard Grant Press, Boston Mass.); Advanced Organic Chemistry by March (Third Edition, 1985, Wiley and Sons, New York); e Advanced Organic Chemistry by Carey and Sundberg (Third Edition, Parts A e B, 1990, Plenum Press, New York). Publicações adicionais descrevendo a síntese de aminoácidos não naturais incluem, por exemplo, WO 2002/085923 intitulado "In vivo incorporation of Unnatural Amino Acids", Matsoukas *et al.*, (1995) J. Med. Chem., 38, 4660-4669; King and Kidd (1949) *A New Synthesis of Glutamine and the γ -Dipeptides of Glutamic Acid from Phthylated Intermediates.* J. Chem. Soc., 3315-3319; Friedman and Chatterji (1959) *Synthesis of Derivatives of Glutamine as Model Substrates for Anti-Tumor Agents.* J. Am. Chem. Soc. 81, 3750-3752; Craig *et al.* (1988) *Absolute Configuration of the Enantiomers of 7-Chloro-4 [[4-(diethylamino)-1-methylbutyl]amino]quinoline (Cloroquine).* J. Org. Chem. 53, 1167-1170; Azoulay *et al.* (1991) *Analogues of glutamine as Potential Antimalarials.* Eur. J. Med. Chem. 26, 201-5; Koskinen, e Rapoport (1989) *Synthesis of 4-Substituted Prolines as Conformationally Constrained Amino Acid Analogs.* J. Org. Chem. 54, 1859-1866; Christie e Rapoport (1985) *Synthesis of Optically Pure Pivalates from L-Asparagine. Application to the Total Synthesis of (+)-Apovincamine through Amino Acid Decarbonylation and Iminium Ion Cyclization.* J. Org. Chem. 1989:1859-1866; Barton *et al.*, (1987) *Synthesis of Novel α -Amino-Acids and Derivatives Using Radical Chemistry: Synthesis of L- and D- α -Amino-Adipic Acids, L-aminopimelic Acid and Appropriate Unsaturated Derivatives.* Tetrahedron Lett. 43:4297-4308; and Subasinghe *et al.*, (1992) *Quisqualic Acid analogs: synthesis of beta-heterocyclic 2-aminopropanoic acid derivatives and their activity at a novel quisqualate-sensitized site.* J. Med. Chem. 35:4602-7. Ver também, International Publication WO 2004/058946, intitulado "PROTEIN ARRAYS", depositado em 22 de dezembro de 2003.

Captação celular de aminoácidos não naturais

Captação de aminoácido não natural por uma célula é uma questão que é tipicamente considerado quando se projeta e seleciona aminoácidos não naturais, por exemplo, para incorporação em uma proteína. Por exemplo, a densidade de carga alta de α -aminoácidos sugere que estes compostos são improvavelmente permeáveis para a célula.

Ácido naturais são captados na célula por meio de uma coleção de sistemas de transporte baseados em proteína exibindo freqüentemente variação de graus de especificidade de aminoácido. Pode ser feita uma classificação rápida, que avalia quais aminoácidos não naturais, se houver, são captados pelas células. Ver, por exemplo, os ensaios de toxicidade, por exemplo, em International Publication WO 2004/058946, intitulado "PROTEIN ARRAYS", depositado em 22 de dezembro de 2003, e Liu and Schultz (1999) *Progress toward the evolution of the organism with expanded Genetic Code*. PNAS 96:4780-4785. Embora a captação seja facilmente analisada com vários ensaios, uma alternativa para projetar aminoácidos não naturais que são amenos a caminhos de captação celular, é fornecer caminhos bio-sintéticos para criar aminoácidos *in vivo*.

Bio-síntese de Aminoácidos não naturais

Muitos caminhos bio-sintéticos já existem em células para a produção de aminoácidos e outros compostos. Embora um método bio-sintético para um aminoácido não natural particular não possa existir na natureza, por exemplo, em uma célula, a invenção fornece tais métodos. Por exemplo, caminhos bio-sintéticos para aminoácidos não naturais são opcionalmente gerados em célula hospedeira adicionando novas enzimas ou modificando caminhos de célula hospedeira existentes. Novas enzimas adicionais são opcionalmente enzimas de ocorrência natural ou artificialmente enzimas desenvolvidas. Por exemplo, a bio-síntese de p-aminofenilalanina (apresentado em um exemplo em WO 2002/085923) conta com a adição de uma combinação de enzimas conhecidas de outros organismos. Os genes para estas enzimas podem ser introduzidos em uma célula transformando a célula com um plasmídeo compreendendo os genes. Os genes, quando expressos na célula, fornecem um caminho enzimático para sintetizar o composto desejado. Exemplos dos tipos de enzimas que são opcionalmente adicionados são fornecidos nos exemplos a seguir. Seqüências de enzimas adicionais são encontradas, por exemplo, em Genbank. Enzimas desenvolvidas artificialmente são também opcionalmente adicionadas em uma célula da mesma maneira. Desta maneira, o maquinário celular e recursos de uma célula são manipulados para produzir aminoácidos não naturais.

Realmente, qualquer um de uma variedade de métodos pode ser usado para produzir enzimas inéditas para uso em caminhos bio-sintéticos, ou para evolução de caminhos existentes, para a produção de aminoácidos não naturais, *in vitro* ou *in vivo*. Muitos métodos disponíveis de desenvolver enzimas e outros componentes de caminho bio-sintético podem ser aplicados à presente invenção para produzir aminoácidos não naturais (ou, realmente, para desenvolver sintetase para ter novas especificidades de substrato ou outras atividades de interesse). Por exemplo, embaralhamento de DNA é opcionalmente usado para desenvolver enzimas inéditas e/ou caminhos de tais enzimas para a produção de aminoácidos não naturais *in vitro* ou *in vivo* (ou produção de nova sintetase). Ver, por

exemplo, Stemmer (1994), *Rapid evolution of a proteína in vitro by DNA shuffling*, Nature 370(4):389-391; and Stemmer, (1994), *DNA shuffling by random fragmentation and reassembly: In vitro recombination for molecular evolution*, Proc. Natl. Acad. Sci. USA., 91:10747-10751. Uma abordagem relacionada embaralha famílias de genes relacionados

5 (por exemplo, homóloga) que desenvolvem facilmente enzimas com características desejadas. Um exemplo de tais métodos "embaralhamento de gene de família" é encontrado em Cramer *et al.* (1998) "DNA shuffling of a family of genes from diverse species accelerates directed evolution" Nature, 391(6664): 288-291. Novas enzimas (quer componentes por caminho bio-sintético quer sintetase) podem também ser geradas usando

10 um procedimento de recombinação de DNA conhecido como "truncação incremental para a criação de enzimas híbridas" ("ITCHY"), por exemplo, da maneira descrita em Ostermeier *et al.* (1999) "A combinatorial approach to hybrid enzymes independent of DNA homology" Nature Biotech 17:1205. Esta abordagem pode também ser usada para gerar uma biblioteca de enzima ou outras variantes de caminhos que podem servir como substratos para um ou

15 mais métodos de recombinação *in vitro* ou *in vivo*. Ver, também, Ostermeier *et al.* (1999) "Combinatorial Protein Engineering by Incremental Truncation", Proc. Natl. Acad. Sci. USA, 96: 3562-67, and Ostermeier *et al.* (1999), "Incremental Truncation as a Strategy in the Engineering of Novel Biocatalysts", Biological and Medicinal Chemistry, 7: 2139-44. Uma outra abordagem usa mutagênese de montagem exponencial para produzir bibliotecas de

20 enzima ou outras variantes de caminho que são, por exemplo, selecionadas pela capacidade de catalisar uma reação bio-sintética relevante para produzir um aminoácido não natural (ou uma sintetase inédita). Nesta abordagem, pequenos grupos de resíduos em uma sequência de interesse são randomizados em paralelo para identificar, em cada posição alterada, aminoácidos que levam às proteínas funcionais. Exemplos de tais

25 procedimentos, que podem ser adaptados para a presente invenção para produzir novas enzimas para a produção de aminoácidos não naturais (ou sintetase inédita) são encontrados em Delegrave and Youvan (1993) Biotechnology Research 11:1548-1552. Ainda em uma outra abordagem, mutagênese aleatória ou semi-aleatória usando oligonucleotídeos dopados ou degenerados para enzima e/ou engenharia de componente

30 de caminho pode ser usado, por exemplo, usando os métodos gerais de mutagênese por exemplo, de Arkin and Youvan (1992) "Optimizing nucleotide mixtures to encode specific subsets of amino acids for semi-random mutagenesis" Biotechnology 10:297-300; ou Reidhaar-Olson *et al.* (1991) "Random mutagenesis of protein sequences using oligonucleotide cassettes" Methods Enzymol. 208:564-86. Ainda em uma outra abordagem,

35 freqüentemente denominada mutagênese "não estocásticas", que usa mutagênese de ré-montagens de polinucleotídeo e saturação do sítio pode ser usada para produzir enzimas e/ou componentes de caminho, que pode em seguida ser classificado pela capacidade de

realizar uma ou mais funções de sintetase ou de caminho bio-sintético (por exemplo, para a produção de aminoácidos não naturais *in vivo*). Ver, por exemplo, Short "NON-STOCHASTIC GENERATION OF GENETIC VACCINES AND ENZYMS" WO 00/46344.

Uma alternativa para tais métodos mutacionais envolve recombinar todos os genomas de organismos e selecionar progênie resultante para funções de caminho particular (frequentemente referido como "embaralhamento de genoma completo"). Esta abordagem pode ser aplicada à presente invenção, por exemplo, por recombinação e seleção genômica de um organismo (por exemplo, um *E. coli* ou outra célula) pela capacidade de produzir um aminoácido não natural (ou intermediário deste). Por exemplo, métodos preceituados nas publicações seguintes podem ser aplicados para projetar caminho para a evolução de caminhos existentes e/ou novos em células para produzir aminoácidos não naturais *in vivo*: Patnaik *et al.* (2002) "Genome shuffling of lactobacillus for improved acid tolerance" *Nature Biotechnology* 20(7):707-712, e Zhang *et al.* (2002) "Genome shuffling Leads to Rapid Phenotypic Improvement in Bacteria" *Nature* 415(6872):644-646.

Outras técnicas para engenharia de organismo e caminho metabólico, por exemplo, para a produção de compostos desejados estão também disponíveis e podem também ser aplicados à produção de aminoácidos não naturais. Exemplos de publicações que preceituam abordagens de engenharia de caminho usadas incluem: Nakamura and White (2003) "Metabolic engineering for the microbial production of 1,3 propanediol" Curr. Opin. Biotechnol. 14(5):454-9; Berry *et al.* (2002) "Application of Metabolic Engineering to improve both the production and use of Biotech Indigo" J. Industrial Microbiology and Biotechnology 28:127-133; Banta *et al.* (2002) "Optimizing a artificial metabolic pathway: Engineering o cofactor especificity of Corynebacterium 2,5-diketo-D-gluconic acid reductase for use in vitamin C biosynthesis" Biochemistry, 41(20), 6226-36; Selivonova *et al.* (2001) "Rapid Evolution of Novel Traits in Microorganisms" Applied and Environmental Microbiology, 67:3645, e muitos outros.

Independente do método usado, tipicamente, o aminoácido não natural produzido com um caminho bio-sintético de engenharia da invenção é produzido em uma concentração suficiente para bio-síntese de proteína eficiente, por exemplo, uma quantidade celular natural, mas não ao ponto de afetar significativamente a concentração de outros aminoácidos celulares ou exaurir os recursos celulares. Concentrações típicas produzidos *in vivo* desta maneira são cerca de 10 mM a cerca de 0,05 mM. Uma vez que uma célula é projetada por engenharia para produzir enzimas desejadas para um caminho específico e um aminoácido não natural é gerado, seleções *in vivo* são opcionalmente usadas para otimizar ainda mais a produção do aminoácido não natural tanto para síntese de proteína ribossomal quanto crescimento celular.

Componentes ortogonais para Incorporar Aminoácidos não naturais

Uma invenção fornece composições e métodos para produzir componentes ortogonais para incorporar o aminoácido não natural fenilselenocisteína (Ver figura 1, estrutura 1) em uma cadeia do polipeptídeo em crescimento em resposta a um códon seletor, por exemplo, um códon de parada âmbar, um códon não sentido, um códon de quatro ou mais bases, etc., por exemplo, *in vivo*. Por exemplo, a invenção fornece ortogonal-RNAs (O-RNAs), aminoacil-RNA sintetase ortogonal (O-RSs) e seus pares. Estes pares podem ser usados para incorporar um aminoácido não natural nas cadeias do polipeptídeo em crescimento.

Uma composição da invenção inclui um aminoacil-RNA sintetase ortogonal (O-RS), onde o O-RS preferencialmente aminoacila um O-RNA com fenilselenocisteína. Em certas modalidades, o O-RS compreende uma seqüência de aminoácidos compreendendo SEQ ID NO: 4, 6 ou 8, e variações conservativas destas. Em certas modalidades da invenção, o O-RS preferencialmente aminoacila o O-RNA sobre qualquer RNA endógeno com um aminoácido não natural particular, onde o O-RS tem uma predisposição para o O-RNA, e onde a razão de O-RNA carregado com um aminoácido não natural para o RNA endógeno carregado com o mesmo aminoácido não natural é maior que 1:1, e mais preferivelmente onde o O-RS carrega o O-RNA exclusivamente ou quase exclusivamente.

Uma composição que inclui um O-RS por opcionalmente incluir adicionalmente um RNA ortogonal (O-RNA), onde o O-RNA reconhece um códon seletor. Tipicamente, um O-RNA da invenção inclui eficiência de supressão de pelo menos cerca de, por exemplo, 45 %, 50 %, 60 %, 75 %, 80 % ou 90 % ou mais na presença de uma sintetase cognata em resposta a um códon seletor comparado à eficiência de supressão de um O-RNA que compreende ou é codificado por uma seqüência de polinucleotídeo apresentada na seqüência listada e exemplos aqui (por exemplo, SEQ ID NO: 1). Em uma modalidade, a eficiência de supressão do O-RS e do O-RNA juntos, por exemplo, é 5 vezes, 10 vezes, 15 vezes, 20 vezes, 25 vezes ou mais, maior do que a eficiência de supressão do O-RNA na ausência de um O-RS. Em alguns aspectos, a eficiência de supressão do O-RS e do O-RNA juntos é pelo menos 45 % da eficiência de supressão de um par de tirosil-RNA ortogonal sintetase derivado de *Methanococcus jannaschii*.

Uma composição que inclui um O-RNA pode opcionalmente incluir uma célula (por exemplo, uma célula eubacteriana, tal como uma célula *E. coli* e similares, ou uma célula eucariótica tal como uma célula de levedura), e/ou um sistema de tradução.

Uma célula (por exemplo, uma célula eubacteriana ou uma célula de levedura) que compreende um sistema de tradução é também fornecida pela invenção, onde o sistema de tradução inclui um RNA (O-RNA) ortogonal; um aminoacil-RNA sintetase ortogonal (O-RS); e, um aminoácido não natural fenilselenocisteína. Tipicamente, o O-RS

preferencialmente aminoacila o O-RNAt sobre qualquer RNAt endógeno com o aminoácido não natural, onde o O-RS tem uma predisposição for o O-RNAt, e onde a razão de O-RNAt carregado com o aminoácido não natural para o RNAt endógeno carregado com o aminoácido não natural é maior que 1:1, e mais preferivelmente onde o O-RS carrega o O-RNAt exclusivamente ou quase exclusivamente. O O-RNAt reconhece o primeiro códon seletor, e o O-RS preferencialmente aminoacila o O-RNAt com um aminoácido não natural. Em uma modalidade, o O-RNAt compreende ou é codificado por uma seqüência de polinucleotídeo apresentada em SEQ ID NO: 1, ou uma seqüência de polinucleotídeo complementar destas. Em uma modalidade, o O-RS compreende uma seqüência de aminoácidos apresentada em SEQ ID NO: 4, 6, 8 ou 10, e variações conservativas destas.

Uma célula da invenção pode compreender opcionalmente mais um par de O-RNAt/O-R diferente adicional e um segundo aminoácido não natural, por exemplo, onde este O-RNAt reconhece um segundo códon seletor e este O-RS preferencialmente aminoacila o O-RNAt correspondente com o segundo aminoácido não natural, onde o segundoamino ácido é diferente from o primeiro aminoácido não natural. Opcionalmente, uma célula da invenção inclui 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-RNAt.

Em certas modalidades, uma célula da invenção é uma célula eubacteriana (tais como *E. coli*), que inclui um -RNAt ortogonal (O-RNAt), um aminoacil-RNAt sintetase ortogonal (O-RS), um aminoácido não natural, e um ácido nucléico que compreende um polinucleotídeo que codifica um polipeptídeo de interesse, onde o polinucleotídeo compreende o códon seletor que é reconhecido pelo O-RNAt. Em certas modalidades da invenção, o O-RS preferencialmente aminoacila o O-RNAt com o aminoácido não natural com uma eficiência que é maior que a eficiência com a qual o O-RS aminoacila qualquer RNAt endógeno.

Em certas modalidades da invenção, um O-RNAt da invenção compreende ou é codificado por uma seqüência de polinucleotídeo apresentada na listas de seqüência ou exemplos aqui (por exemplo, SEQ ID NO: 1), ou uma seqüência de polinucleotídeo complementar desta. Em certas modalidades da invenção, um O-RS compreende uma seqüência de aminoácidos apresentada nas listas de seqüência, ou uma variação conservativa destas. Em uma modalidade, o O-RS ou uma porção deste é codificado por uma seqüência de polinucleotídeo que codifica um aminoácido apresentado nas listas de seqüência ou exemplos aqui, ou uma seqüência de polinucleotídeo complementar destas.

O O-RNAt e/ou o O-RS da invenção podem ser derivados de qualquer uma de uma variedade de organismos (por exemplo, organismos eucariótico e/ou não eucariótico).

Polinucleotídeos são também um recurso da invenção. Um polinucleotídeo da

invenção (por exemplo, SEQ ID NO: 5, 7 ou 9) inclui um polinucleotídeo artificial (por exemplo, feito pelo homem, e ocorrência não natural) que compreende uma seqüência de nucleotídeos que codifica um polipeptídeo apresentado na listagens de seqüência aqui, e/ou é complementar a seqüência de polinucleotídeo. Um polinucleotídeo da invenção pode também incluir um ácido nucléico que hibridiza com um polinucleotídeo descrito anteriormente, em condições altamente rigorosas substancialmente por todo o comprimento do ácido nucléico. Um polinucleotídeo da invenção também inclui um polinucleotídeo que é, por exemplo, pelo menos 75 %, pelo menos 80 %, pelo menos 90 %, pelo menos 95 %, pelo menos 98 % ou mais idêntico àquele de um RNAt de ocorrência natural ou ácido nucléico de codificação correspondente (mas um polinucleotídeo da invenção é sem ser um RNAt de ocorrência natural ou ácido nucléico de codificação correspondente), onde o RNAt reconhece um códon seletor, por exemplo, um códon de quatro bases. Polinucleotídeos artificiais que são, por exemplo, pelo menos 80 %, pelo menos 90 %, pelo menos 95 %, pelo menos 98 % ou mais idêntico a qualquer dos anteriores e/ou um polinucleotídeo que compreende uma variação conservativa de qualquer da anterior, estão também incluídas em polinucleotídeos da invenção.

Vetores que compreendem um polinucleotídeo da invenção são também uma recurso da invenção. Por exemplo, um vetor da invenção pode incluir um plasmídeo, um cosmídeo, um fago, um vírus, um vetor de expressão, e/ou similares. Uma célula que compreende um vetor da invenção é também um recurso da invenção.

Métodos de produzir componentes de um par de O-RNAt/O-R são também recursos da invenção. Componentes produzidos por estes métodos são também um recurso da invenção. Por exemplo, métodos de produzir pelo menos um RNAt que é ortogonal a uma célula (O-RNAt) incluem gerar uma biblioteca de RNAts mutantes; mutar um laço anticódon de cada membro da biblioteca de RNAts mutantes para permitir o reconhecimento de um códon seletor, fornecendo dessa maneira uma biblioteca de O-RNAts potencial, e submeter a seleção negativa uma primeira população de células de uma primeira espécie, onde as células compreendem um membro de biblioteca de O-RNAts potencial. A seleção negativa elimina as células que compreendem um membro de biblioteca de O-RNAts potencial que é aminoacilado por um aminoacil-RNAt sintetase (RS) que é endógena à célula. Isto fornece um agrupamento de RNAts que são ortogonais às célula da primeira espécie, fornecendo dessa maneira pelo menos um O-RNAt. É também fornecido um O-RNAt produzido pelos métodos da invenção.

Em certas modalidades, os métodos compreendem adicionalmente submeter a seleção positiva uma segunda população de células da primeira espécie, onde as células compreendem um membro de agrupamento de RNAts que são ortogonais às célula da primeira espécie, um aminoacil-RNAt sintetase cognato, e um marcador de seleção positiva.

Usando a seleção positiva, células são selecionadas ou classificadas por aquelas células que compreendem um membro de agrupamento de RNAts que é aminoacilado pelo aminoacil-RNAt sintetase cognato e que mostra uma resposta desejada na presença do marcador de seleção positiva, fornecendo dessa maneira um O-RNAt. Em certas

5 modalidades, a segunda população de células compreende células que não foram eliminadas pela seleção negativa.

Métodos para identificar um aminoacil-RNAt sintetase ortogonal que carrega um O-RNAt com um aminoácido não natural são também fornecidos. Por exemplo, métodos incluem submeter uma população de células de uma primeira espécie a uma seleção, onde

10 cada uma das células compreende: 1) um membro da pluralidade de aminoacil-RNAt sintetase (RSs), (por exemplo, a pluralidade de RSs pode incluir RSs mutante e, RSs derivados de uma espécie sem ser uma primeira espécie ou tanto RSs mutante quanto RSs derivados de uma espécie sem ser uma primeira espécie); 2) o RNAt ortogonal (O-RNAt) (por exemplo, de uma ou mais espécie); e 3) um polinucleotídeo que codifica um marcador

15 de seleção positiva e compreende pelo menos um códon seletor.

Células (por exemplo, uma célula hospedeira) são selecionadas ou classificadas por aquelas que apresentam uma melhoria na eficiência de supressão comparadas às células que não tem, ou têm uma quantidade reduzida de membros da pluralidade de RSs. Estas células selecionadas/classificadas compreendem um RS ativo que aminoacila o O-

20 RNAt. Um aminoacil-RNAt sintetase ortogonal identificado pelo método é também um recurso da invenção.

Métodos de produzir uma proteína em uma célula (por exemplo, em uma célula eubacteriana tal como um célula *E. coli* ou similares, ou em uma célula de levedura) tendo o aminoácido não natural em uma posição selecionada são também um recurso da invenção.

25 Por exemplo, um método inclui crescer uma célula em um meio apropriado onde a célula compreende um ácido nucléico que compreende pelo menos um códon seletor e codifica uma proteína, fornecer o aminoácido não natural, e incorporar o aminoácido não natural na posição especificada na proteína durante a tradução do ácido nucléico com o pelo menos um códon seletor, produzindo dessa maneira a proteína. Uma célula compreende

30 adicionalmente: um RNAt ortogonal (O-RNAt) que funciona na célula e reconhece o códon seletor; e, um aminoacil-RNAt sintetase ortogonal (O-RS) que preferencialmente aminoacila o O-RNAt com o aminoácido não natural. Uma proteína produzida por este método é também um recurso da invenção.

A invenção também fornece composições que incluem proteínas, onde as proteínas

35 compreendem fenilselenocisteína. Em certas modalidades, a proteína compreende uma seqüência de aminoácidos que é pelo menos 75 % idêntico àquela de uma proteína conhecida, por exemplo, hormônio do crescimento humano, uma proteína terapêutica, uma

proteína de diagnóstico, uma enzima industrial, ou porção destes. Opcionalmente, a composição compreende um veículo farmaceuticamente aceitável.

ÁCIDO NUCLÉICO E SEQÜÊNCIAS DE POLIPEPTÍDEO E VARIANTES

Da maneira aqui descrita, a invenção fornece seqüências de polinucleotídeo que codificam, por exemplo, O-RNAs e O-RSs, e seqüências de aminoácido de polipeptídeo, por exemplo, O-RSs, e, por exemplo, composições, sistemas e métodos que compreendem o dito polinucleotídeo ou seqüências de polipeptídeo. Exemplos das ditas seqüências, por exemplo, O-RNA e aminoácido de O-RS e seqüência de nucleotídeos são revelados aqui (Ver figura 2). Portanto, versados na tecnologia perceberão que a invenção não está limitada àquelas seqüências especificamente aqui recitadas, por exemplo, nos Exemplos e listagem de seqüência. Versados na tecnologia perceberão que a invenção também fornece muitas seqüências relacionadas com as funções aqui descritas, por exemplo, polinucleotídeos e polipeptídeos que codificam variantes conservativas de um O-RS aqui revelado.

As construções e análise de espécie de sintetase ortogonal (O-RS) que são capazes de aminoacilar um O-RNA cognato com fenilselenocisteína são descritos no Exemplo 1. Este exemplo descreve as construções e análise da espécie O-RS que são capazes de incorporar o aminoácido não natural fenilselenocisteína.

A invenção fornece polipeptídeos (O-RSs) e polinucleotídeos, por exemplo, polinucleotídeos O-RNA, que codificam O-RSs ou porções destes, oligonucleotídeos usados para isolar clones de aminoacil-RNA sintetase, etc. Polinucleotídeos da invenção incluem aqueles que codificam proteínas ou polipeptídeos de interesse da invenção com um ou mais códons seletores. Além disso, polinucleotídeos da invenção incluem, por exemplo, um polinucleotídeo compreendendo uma seqüência de nucleotídeos apresentada em SEQ ID NO: 5, 7 ou 9, e um polinucleotídeo que é complementar ou que codifica uma seqüência de polinucleotídeo destes. Um polinucleotídeo da invenção também inclui qualquer polinucleotídeo que codifica uma seqüência de aminoácidos O-RS compreendendo SEQ ID NO: 4, 6 ou 8. Similarmente, um ácido nucléico artificial que hibridiza para um polinucleotídeo indicado anteriormente em condições altamente rigorosas substancialmente por todo o comprimento do ácido nucléico (e é um sem ser um polinucleotídeo de ocorrência natural) é um polinucleotídeo da invenção. Em uma modalidade, uma composição inclui um polipeptídeo da invenção e um excipiente (por exemplo, tampão, água, excipiente farmaceuticamente aceitável, etc.). A invenção também fornece um anticorpo ou anti-soro especificamente imuno-reativo com um polipeptídeo da invenção. Um polinucleotídeo artificial é um polinucleotídeo que é feito pelo homem, e não é de ocorrência natural.

Um polinucleotídeo da invenção também inclui um polinucleotídeo artificial que é, por exemplo, pelo menos 75 %, pelo menos 80 %, pelo menos 90 %, pelo menos 95 %, pelo

menos 98 % ou mais idêntico àquele de um RNAt de ocorrência natural (mas é um RNAt sem ser de ocorrência natural). Um polinucleotídeo também inclui um polinucleotídeo artificial que é, por exemplo, pelo menos 75 %, pelo menos 80 %, pelo menos 90 %, pelo menos 95 %, pelo menos 98 % ou mais idêntico (mas não 100 % idêntico) àquele de um

5 RNAt de ocorrência natural.

Em certas modalidades, um vetor (por exemplo, um plasmídeo, um cosmídeo, um fago, um vírus, etc.) compreende um polinucleotídeo da invenção. Em uma modalidade, o vetor é um vetor de expressão. Em uma outra modalidade, o vetor de expressão inclui um promotor operável ligado a um ou mais dos polinucleotídeos da invenção. Em uma outra

10 modalidade, uma célula compreende um vetor que inclui um polinucleotídeo da invenção.

Versados na tecnologia também percebem que muitas variantes das seqüências reveladas são incluídas na invenção. Por exemplo, variações conservativas das seqüências reveladas que rendem uma seqüência funcionalmente idêntica são incluídas na invenção. Variantes da seqüência de ácido nucléico de polinucleotídeos, em que as variantes

15 hibridizam com pelo menos uma seqüência revelada, são consideradas incluídas na invenção. Subseqüências exclusivas das seqüências aqui reveladas, determinadas, por exemplo, por técnicas de comparação de seqüência padrão, estão também incluídas na invenção.

Variações Conservativas

Por causa da degeneração do código genético, "substituições silenciosas" (isto é, substituições em uma seqüência de ácido nucléico que não resultam em uma alteração em um polipeptídeo codificado) são um recurso implícito de cada seqüência de ácido nucléico que codifica uma seqüência de aminoácido. Similarmente, "substituições de aminoácido conservativas", onde um ou um número limitado de aminoácidos em uma seqüência de

20 aminoácidos é substituído com aminoácidos diferentes com propriedades altamente similares, é também imediatamente identificado como altamente similar a uma construção revelada. Tais variações conservativas de cada seqüência revelada são um recurso da presente invenção.

"Variações conservativas" de uma seqüência de ácido nucléico particular referem-se àqueles ácidos nucléicos que codificam seqüências de aminoácido idênticas ou essencialmente idênticas, ou, onde o ácido nucléico não codifica uma seqüência de aminoácido, com seqüências essencialmente idênticas. Versados na tecnologia reconhecerão que substituições individuais, deleções ou adições que alteram, adicionam ou deletam um aminoácido simples ou uma pequena porcentagem de aminoácidos (tipicamente

30 menos que 5 %, mais tipicamente menos que 4 %, 2 % ou 1 %) em uma seqüência codificada são "variações conservativamente modificadas" onde a alterações resultam na deleção de um aminoácido, adição de um aminoácido, ou substituição de um aminoácido

com um aminoácido quimicamente similar. Assim, "variações conservativas" de uma seqüência de polipeptídeo listada da presente invenção incluem substituições de uma pequena porcentagem, tipicamente menos que 5 %, mais tipicamente menos que 2 % ou 1 %, dos aminoácidos da seqüência de polipeptídeo, com um aminoácido do mesmo grupo de substituição conservativa. 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 ácido nucléico básico.

Tabelas de substituição conservativa fornecendo aminoácidos funcionalmente similares são bem conhecidas na tecnologia, onde um resíduo de aminoácido é substituído por um outro resíduo de aminoácido tendo propriedades químicas similares (por exemplo, cadeias laterais aromáticas ou cadeias laterais positivamente carregadas), e portanto não muda substancialmente as propriedades funcionais da molécula de polipeptídeo. Os seguintes grupos de exemplo apresentados que contêm aminoácidos naturais de propriedades químicas similares, onde substituições dentro de um grupo é uma "substituição conservativa".

Substituições de aminoácido conservativas

| Cadeias laterais alifáticas polar e/ou | Cadeias laterais Não Descarregadas Polar, | Cadeias laterais aromáticas | Cadeias laterais Carregadas Positivamente | Cadeias laterais Carregadas negativamente |
|--|--|--|--|--|
| Glicina Alanina Valina Leucina Isoleucina Prolina | Serina Treonina Cisteína Metionina Asparagina Glutamina | Fenilalanina Tirosina Tryptofano | Lisina Arginina Histidina | Aspartato Glutamato |

HIBRIDIZAÇÃO DE ÁCIDO NUCLÉICO

Hibridização comparativa pode ser usada para identificar ácidos nucléicos da invenção, incluindo variações conservativas de ácidos nucléicos da invenção, e este método de hibridização comparativa é um método preferido de ácidos nucléicos distinguíveis da invenção. Além disso, ácidos nucléicos alvos que hibridizam com um ácido nucléico representado pela SEQ ID NO: 5, 7, 9 ou 11, em condições de alto, ultra alto e ultra ultra alto rigor são um recurso da invenção. Exemplos de tais ácidos nucléicos incluem aqueles com uma ou umas poucas substituições de ácido nucléico silenciosas ou conservativas,

comparadas a uma dada sequência de ácido nucléico.

Acredita-se que um ácido nucléico teste hibridiza especificamente com um ácido nucléico sonda quando ele hibridiza pelo menos 50 %, bem como a sonda para o alvo complementar perfeitamente pareado, isto é, com uma razão sinal para ruído alta de pelo menos a metade da hibridização da sonda para o alvo em condições nas quais a sonda perfeitamente pareada se liga ao alvo complementar perfeitamente pareado com uma razão sinal para ruído que é pelo menos cerca de 5x - 10x tão alta quanto aquela observada para hibridização com qualquer dos ácidos nucléicos alvos não pareados.

Ácidos nucléicos "hibridizam" quando eles associam, tipicamente em solução. Ácidos nucléicos hibridizam em virtude de uma variedade de forças físico-químicas bem caracterizadas, tais como ligação de hidrogênio, exclusão de solvente, empilhamento de base e similares. Um guia abrangente para a hibridização de ácidos nucléicos é encontrado em Tijssen (1993) *Laboratory Techniques in Biochemistry and Molecular Biology—Hibridization with Acid nucleic Probes*, Part I, Chapter 2, "Overview of Principles of Hibridization and the Estrategy of Acid nucleic Probes Assays," (Elsevier, New York), bem como em *Current Protocols in Molecular Biology*, Ausubel *et al.*, eds., *Current Protocols*, a joint venture between Greene Publishing Associates, Inc. and John Wiley & Sons, Inc., (suplementado em 2004); and Hames and Higgins (1995), *Gene Probes 1* and *Gene Probes 2*, ambos de IRL Press at Oxford University Press, Oxford, England, fornecem detalhes nas sínteses, marcação, detecção e quantificação de DNA e RNA, incluindo oligonucleotídeos.

Um exemplo de condições de hibridização rigorosa para hibridização de ácidos nucléicos complementares que têm mais do que 100 resíduos complementares em um filtro em um Southern ou northern blot é 50 % formalina com 1 mg de heparina a 42 °C, com a hibridização sendo realizada por toda a noite. Um exemplo de condições de lavagem rigorosa é uma lavagem 0,2 x SSCat 65 °C por 15 minutos (Ver, Sambrook *et al.* para uma descrição de tampão SSC; Sambrook *et al.*, *Molecular Cloning - A Laboratory Manual* (3rd Ed.), Vol. 1-3, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, 2001). Frequentemente a lavagem de alta rigorosidade é precedida por uma lavagem de baixa rigorosidade para remover sinal de sonda de fundo. Um exemplo de lavagem de baixa rigorosidade é 2 x SSC a 40 °C por 15 minutos. Em geral, uma razão sinal para ruído de 5 x (ou maior) do que aquela observada por uma sonda não relacionada na ensaio de hibridização particular indica detecção de uma hibridização específica.

"Condições de lavagem de hibridização rigorosa" no contexto de hibridização de Ácido nucléico experimentam tais como hibridizações Southern e northern são dependentes de sequência, e são diferente em parâmetros ambientais diferentes. Um guia abrangente para a hibridização de ácidos nucléicos é encontrado em Tijssen (1993) *Laboratory Techniques in Biochemistry and Molecular Biology—Hibridization with Acid nucleic Probes*,

Part I, Chapter 2, "Overview of Principles of Hybridization and the Strategy of Acid nucleic Probe Assay," (Elsevier, New York); e em Hames and Higgins (1995), *Gene Probes 1* e *Gene Probes 2*, ambos de IRL Press at Oxford University Press, Oxford, England.

Condições de hibridização e lavagem rigorosa podem facilmente ser determinadas empiricamente por qualquer ácido nucléico teste. Por exemplo, na determinação de condições de hibridização e lavagem rigorosa, as condições de hibridização e lavagem são gradualmente aumentadas (por exemplo, aumentando a temperatura, diminuindo a concentração de sal, aumentando concentração de detergente e/ou aumentando a concentração de solventes orgânicos tal como formalina na hibridização ou lavagem), até que um conjunto selecionado de critérios seja alcançado. Por exemplo, em condições de hibridização e lavagem altamente rigorosas, as condições de hibridização e lavagem são gradualmente aumentadas até que uma sonda se ligue a um alvo perfeitamente complementar pareado com uma razão sinal para ruído que é pelo menos 5x maior que a observada para hibridização da sonda com um alvo não pareado.

Condições "muito rigorosas" são selecionadas iguais ao ponto de fusão (T_m) por uma sonda particular. A T_m é a temperatura (em concentração iônica e pH definidos) na qual 50 % da seqüência teste hibridiza com uma sonda perfeitamente pareada. Com os propósitos da presente invenção, de um modo geral, condições de hibridização e lavagem "altamente rigorosas" são selecionadas em cerca de 5 °C menor que a T_m para a seqüência específica a uma concentração iônica e pH definidos.

"Condições de ultra alta rigorosidade" de hibridização e lavagem são aquelas nas quais a rigorosidade das condições de hibridização e lavagem são aumentadas até que a razão sinal para ruído para ligação da sonda para o ácido nucléico alvo perfeitamente complementar pareado é pelo menos 10 x maior do a observada para hibridização com qualquer dos ácidos nucléicos alvos não pareados. Um ácido nucléico alvo que hibridiza com uma sonda em tais condições, com uma razão sinal para ruído de pelo menos $\frac{1}{2}$ que do ácido nucléico alvo complementar perfeitamente pareado se liga à sonda em condições de ultra-alta rigorosidade.

Similarmente, níveis de rigorosidade ainda maiores podem ser determinados aumentando gradualmente as condições de hibridização e/ou lavagem do ensaio de hibridização relevante. Por exemplo, aqueles nos quais a rigorosidade das condições de hibridização e lavagem são aumentadas até que a razão sinal para ruído para ligação da sonda com o ácido nucléico alvo complementar perfeitamente pareado é pelo menos 10 x, 20 X, 50 X, 100 X, ou 500 X ou mais maior que a observada para hibridização com qualquer dos ácidos nucléicos alvos não pareados. Um ácido nucléico alvo que hibridiza com uma sonda em tais condições, com uma razão sinal para ruído de pelo menos $\frac{1}{2}$ que do ácido nucléico alvo complementar perfeitamente pareado se liga a sonda em condições de ultra-

ultra-alta rigurosidade.

Ácidos nucleicos que não hibridizam uns com os outros em condições rigorosas são ainda substancialmente idênticas se os polipeptídeos com o qual eles codificam forem substancialmente idênticos. Isto ocorre, por exemplo, quando uma cópia de um ácido nucleico é criada usando a degeneração de códon máxima permitida pelo código genético.

Subseqüências Exclusivas

Em alguns aspectos, a invenção fornece um ácido nucleico que compreende uma subseqüência exclusiva em um ácido nucleico selecionado das seqüências de O-RNAs e O-RSs aqui reveladas. A subseqüência exclusiva é exclusiva comparada a um ácido nucleico correspondente a qualquer seqüência de ácido nucleico O-RNAt ou O-RS conhecida. O alinhamento pode ser realizado usando, por exemplo, o BLAST ajustado nos parâmetros padrões. Qualquer subseqüência exclusiva é usada, por exemplo, como uma sonda para identificar os ácidos nucleicos da invenção ou ácidos nucleicos relacionados.

Similarmente, a invenção inclui um polipeptídeo que compreende uma subseqüência exclusiva em um polipeptídeo selecionado das seqüências de O-RSs aqui reveladas. Aqui, a subseqüência exclusiva é exclusiva comparada a um polipeptídeo correspondente a qualquer seqüência de polipeptídeo conhecida.

A invenção também fornece ácidos nucleicos alvos que hibridizam em condições rigorosas com um oligonucleotídeo de codificação exclusiva que codifica uma subseqüência exclusiva em um polipeptídeo selecionado das seqüências do O-RSs nas quais a subseqüência exclusiva é exclusiva, comparada a um polipeptídeo correspondente a qualquer dos polipeptídeos controle (por exemplo, seqüências parentais das quais sintetase da invenção foi derivada, por exemplo, por mutação). Seqüências exclusivas são determinadas conforme notado anteriormente.

Comparação de seqüência, identidade e homologia

Os termos "idêntico" ou "porcentagem de identidade" no contexto de duas ou mais seqüências de ácidos nucleicos ou polipeptídeos refere-se a duas ou mais seqüências ou subseqüências que são as mesmas ou têm uma porcentagem especificada de resíduo de aminoácidos ou nucleotídeos que são os mesmos, quando comparados e alinhados para correspondência máxima, medidos usando um dos algoritmos de comparação de seqüência descritos a seguir (ou outros algoritmos disponíveis pelos versados na tecnologia), ou por inspeção visual.

A frase "substancialmente idêntico", no contexto de dois ácidos nucleicos ou polipeptídeos (por exemplo, DNAs que codificam um O-RNAt ou O-RS, ou a seqüência de aminoácidos de um 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 mais identidade de nucleotídeo ou resíduo de aminoácido, quando comparadas e

alinhadas para correspondência máxima, medidas usando um algoritmo de comparação de seqüência ou por inspeção visual. Tais seqüências "substancialmente idênticas" são tipicamente consideradas "homólogas," sem referência a origem real. Preferivelmente, a "identidade substancial" existe em uma região das seqüências que é pelo menos cerca de 50 resíduos em comprimento, mais preferivelmente em uma região de pelo menos cerca de 100 resíduos, e mais preferivelmente, as seqüências são substancialmente idênticas em pelo menos cerca de 150 resíduos, ou no comprimento total das duas seqüências comparadas.

Seqüências de proteínas e/ou proteína são "homólogas" quando elas são derivadas, naturalmente ou artificialmente, de uma seqüência de proteína ou proteína ancestral comum. Similarmente, ácidos nucleicos e/ou seqüência de ácidos nucleicos são homólogas quando elas são derivadas, naturalmente ou artificialmente, de um ácido nucleico ou seqüência de ácido nucleico ancestral comum. Por exemplo, qualquer ácido nucleico de ocorrência natural pode ser modificado por qualquer método de mutagenese disponível para incluir um ou mais códon seletor. Quando expresso, este ácido nucleico mutagenizado codifica um polipeptídeo que compreende um ou mais aminoácidos não naturais. O processo de mutação pode, é claro, adicionalmente alterar um ou mais códons padrões, carregando igualmente dessa maneira um ou mais aminoácidos padrões na proteína mutante resultante. Homologia é de um modo geral inferido da similaridade de seqüência entre dois ou mais ácidos nucleicos ou proteínas (ou seqüências destes). A porcentagem precisa de similaridade entre seqüências que é usada no estabelecimento da homologia varia com o ácido nucleico e proteína em questão, mas apenas 25 % de similaridade de seqüência são rotineiramente usados para estabelecer a homologia. Níveis de similaridade de seqüência maiores, por exemplo, 30 %, 40 %, 50 %, 60 %, 70 %, 80 %, 90 %, 95 %, ou 99 % ou mais, podem também ser usados para estabelecer a homologia. Métodos para determinar as porcentagens de similaridade de seqüência (por exemplo, BLASTP e BLASTN usando parâmetros padrões) são aqui descritos e estão de um modo geral disponíveis.

Para comparação de seqüência e determinação da homologia, uma seqüência age tipicamente como uma seqüência de referência na qual seqüências teste são comparadas. Durante o uso de um algoritmo de comparação de seqüência, teste e seqüência de referências são alimentados em um computador, subseqüências coordenadas são designadas, se necessário, e parâmetros do programa de algoritmo de seqüência são designados. O algoritmo de comparação de seqüência em seguida calcula a porcentagem de identidade de seqüência para a(s) seqüência(s) teste(s) com relação a seqüência de referência, baseado nos parâmetros de programa designados.

Alinhamento de seqüências ideal para comparação pode ser conduzido, por

exemplo, pelo algoritmo de homologia local de Smith and Waterman, Adv. Appl. Math. 2:482 (1981), pelo algoritmo de alinhamento de homologia de Needleman and Wunsch, J. Mol. Biol. 48:443 (1970), pela pesquisa para método de similaridade de Pearson e Lipman, Proc. Nat'l. Acad. Sci. USA 85:2444 (1988), pelas implementações computadorizadas destes

5 algoritmos (GAP, BESTFIT, FASTA, e TFASTA em Wisconsin Genetics Software Package, Genetics Computer Group, 575 Science Dr., Madison, WI), ou por inspeção visual (Ver de um modo geral *Current Protocols In Molecular Biology*, Ausubel *et al.*, eds., *Current Protocols*, a joint venture da Greene Publishing Associates, Inc. and John Wiley & Sons, Inc., (suplementado em 2004).

10 Um exemplo de um algoritmo que é adequado para determinar a porcentagem de identidade de seqüência e similaridade de seqüência é o algoritmo BLAST, que é descrito em Altschul *et al.*, *J. Mol. Biol.*, 215:403-410 (1990). O software para realizar análise BLAST é publicamente disponível pela National Center for Biotechnology Information (Ver o endereço de rede NCBI). Este algoritmo envolve primeiro identificar pares de seqüência de

15 alta pontuação (HSPs) identificando palavras pequenas de comprimento W na seqüência consulta, que tanto pareia quanto satisfaz alguns pontuações limites de valor positivo T quando alinhados com um palavra do mesmo comprimento em uma seqüência da base de dados. T é referido como o limite de pontuações de palavra vizinha (Altschul *et al.*, (1990) *J. Mol. Biol.*, 215:403-410). Estas colisões de palavra vizinha inicial agem como sementes para

20 iniciar pesquisas para encontrar conteúdos de HSPs maiores contendo-as. As colisões de palavra são em seguida estendidas em ambas direções ao longo de cada seqüência até que as pontuações de alinhamento cumulativo possam ser aumentadas. As pontuações cumulativas são calculadas usando, para seqüência de nucleotídeos, os parâmetros M (pontuações para trás para um par de resíduos pareados; sempre > 0) e N (pontuações de

25 penalidade para resíduos não pareados; sempre < 0). Para seqüências de aminoácidos, uma matriz de pontuação é usada para calcular as pontuações cumulativas. A extensão das colisões de palavra em cada direção pára quando: as pontuações de alinhamento cumulativas caem em uma quantidade X de seu valor máximo obtido; a pontuação cumulativa vai de zero ou menos, em virtude do acúmulo de um ou mais alinhamentos de

30 resíduo de pontuação negativa; ou a extremidade de qualquer seqüência é atingida. Os parâmetros e algoritmo BLAST W, T, e X determinam a sensibilidade e velocidade do alinhamento. O programa BLASTN (para seqüência de nucleotídeos) usa como padrões um comprimento de palavra (W) de 11, uma expectativa (E) de 10, um corte de 100, M=5, N=-4, e uma comparação de ambas fitas. Para seqüências de aminoácido, o programa BLASTP

35 usa como padrões um comprimento de palavra (W) de 3, um expectativa (E) de 10, e a matriz de pontuação BLOSUM62 (Ver Henikoff and Henikoff (1989) Proc. Natl. Acad. Sci. USA 89:10915).

Além de calcular a porcentagem de identidade de seqüência, o algoritmo BLAST também realiza uma análise estatística da similaridade entre duas seqüências (Ver, por exemplo, Karlin and Altschul, Proc. Nat'l. Acad. Sci. USA 90:5873-5787 (1993)). Uma medida de similaridade fornecida pelo algoritmo BLAST é a probabilidade da soma mínima (P(N)), que fornece uma indicação da probabilidade pela qual um pareamento entre duas seqüências de nucleotídeo ou aminoácido ocorreria por chance. Por exemplo, um ácido nucléico é considerado similar a uma seqüência de referência se a probabilidade da soma mínima em uma comparação do ácido nucléico teste com a ácido nucléico de referência for menos que cerca de 0,1, mais preferivelmente menos que cerca de 0,01, e mais preferivelmente menos que cerca de 0,001.

Mutagênese e Outras Técnicas de Biologia Molecular

Polinucleotídeo e polipeptídeos da invenção usados na invenção podem ser manipulados usando técnicas biológicas moleculares. Textos gerais que descrevem técnicas biológicas moleculares incluem Berger and Kimmel, *Guide to Molecular Cloning Techniques, Methods in Enzymology*, volume 152 (1987), Academic Press, Inc., San Diego, CA; Sambrook *et al.*, *Molecular Cloning - A Laboratory Manual* (3rd Ed.), Vol. 1-3, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, 2001 and *Current Protocols In Molecular Biology*, Ausubel *et al.*, eds., *Current Protocols*, uma joint venture da Greene Publishing Associates, Inc. and John Wiley & Sons, Inc., (suplementado em 2004)). Estes textos descrevem mutagênese, o uso de vetores, promotores e muitos outros tópicos relevantes relacionados por exemplo a geração de genes que incluem códons seletores para produção de proteínas que incluem aminoácidos não naturais, RNAs ortogonais, sintetase ortogonal, e seus pares.

Vários tipos de mutagênese são usados na invenção, por exemplo, para mutar moléculas de RNAs, para produzir bibliotecas de RNAs, para produzir bibliotecas de sintetase, para inserir códons seletores que codificam aminoácidos não naturais em uma proteína ou polipeptídeo de interesse. Eles incluem, mas sem limitações, sítio-dirigido, mutagênese de ponto aleatório, recombinação homóloga, embaralhamento de DNA ou outros métodos de mutagênese recursivos, construções quiméricas, mutagênese usando uracila contendo gabaritos, mutagênese direcionado a oligonucleotídeo, mutagênese de DNA modificado por fosforotioato, mutagênese usando DNA duplo aberto ou similares, ou qualquer combinação destes. Métodos adequados adicionais incluem reparo não pareado do ponto, mutagênese usando cepas hospedeiras deficientes de reparo, restrição-seleção e restrição-purificação, mutagênese de deleção, mutagênese por síntese de gene total, reparo de ruptura da fita dupla, e similares. Mutagênese, por exemplo, envolvendo construções quiméricas, é também incluída na presente invenção. Em uma modalidade, mutagênese pode ser guiada por informação conhecida da molécula de ocorrência natural ou molécula

de ocorrência natural alterada ou mutada, por exemplo, seqüência, comparações de seqüências, propriedades físicas, estrutura cristalina ou similares.

Células hospedeiras são geneticamente projetadas por engenharia (por exemplo, transformadas, convertidas ou transfectadas) com os polinucleotídeos da invenção ou
 5 construções que incluem um polinucleotídeo da invenção, por exemplo, um vetor da invenção, que pode ser, por exemplo, um vetor de clonagem ou um vetor de expressão. Por exemplo, as regiões que codificam o RNAt ortogonal, o RNAt sintetase ortogonal, e a proteína a ser derivatizada são operáveis ligados aos elementos de controle de expressão de gene que são funcionais na célula hospedeira desejada. Vetores típicos contêm
 10 terminadores de transcrição e tradução, seqüências de iniciação de transcrição e tradução, e promotores usados para regulação da expressão do ácido nucléico alvo particular. Os vetores opcionalmente compreendem cassetes de expressão genérica contendo pelo menos uma seqüência terminadora independente, replicação permitindo seqüências do cassete em eucariotos, ou procariotos, ou ambos (por exemplo, vetores de transferência) e marcador de
 15 seleções tanto para sistemas procarióticos quanto eucarióticos. Vetores são adequados para replicação e/ou integração em procariotos, eucariotos, ou preferivelmente em ambos. Ver Gillman and Smith, Gene 8:81 (1979); Roberts *et al.*, Nature, 328:731 (1987); Schneider *et al.*, Protein Expr. Puff. 6435:10 (1995); *Current Protocols In Molecular Biology*, Ausubel *et al.*, Eds., *Current Protocols*, a joint venture between Greene Publishing Associates, Inc. and
 20 John Wiley & Sons, Inc., (suplementado em 2004); Sambrook *et al.*, *Molecular Cloning - A Laboratory Manual* (3rd Ed.), Vol. 1-3, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, 2001; e Berger and Kimmel, *Guide to Molecular Cloning Technics, Methods in Enzymology*, volume 152 (1987), Academic Press, Inc., San Diego, CA. O vetor pode ser, por exemplo, na forma de um plasmídeo, uma bactéria, um vírus, um polinucleotídeo descoberto, ou um polinucleotídeo conjugado. Os vetores são introduzidos nas células e/ou
 25 microorganismos por métodos padrões incluindo eletroporação (Fromm *et al.*, Proc. Natl. Acad. Sci. USA 82, 5824 (1985), infection by viral vectors, high velocity ballistic penetration by small particles with the acid nucleic either within the matrix of small beads or particles, or on the surface (Klein *et al.*, Nature 327, 70-73 (1987)), e/ou similares.

30 Um sistema de plasmídeo simples altamente eficiente e versátil foi desenvolvido para incorporação de sítio-específico de aminoácidos não naturais nas proteínas em resposta ao códon de parada âmbar (UAG) em *E. coli*. No sistema inédito, o par de supressor *M. jannaschii* RNAtttyr(CUA) e tirosil-RNAt sintetase são codificados em um plasmídeo simples, que é compatível com a maioria dos vetores de expressão *E. coli*.
 35 Operon RNAt monocistrônico sob o controle de promotor e terminador proK foi construído por processamento de estrutura secundária e RNAt ideal. Introdução de uma forma mutada de promotor glnS para a sintetase resultou em um aumento significativo tanto na eficiência

de supressão quanto fidelidade. Aumento na eficiência de supressão foi também obtido por múltiplas cópias de gene RNAt, bem como por uma mutação específica (D286R) em uma sintetase (Kobayashi *et al.*, "Structural basis for RNAt ortogonal especificities of tirosil-,RNA synthetases for genetic code expansion," Nat. Struct. Biol., 10(6):425-432 [2003]). A

5 generalidade do sistema otimizado foi também demonstrado por incorporação altamente eficiente e precisa de diversos aminoácidos não naturais diferentes, cujas utilidades exclusivas no estudo da estrutura e função de proteína foram previamente demonstradas.

Um catálogo de Bactérias e Bacteriófagos usados para clonagem é fornecido, por exemplo, pelo ATCC, por exemplo, The ATCC Catalogue of Bacteria and Bacteriophage

10 (1996) Gherna *et al.* (eds) publicado pelo ATCC. Procedimentos básicos adicionais para seqüenciamento, clonagem e outros aspectos de biologia molecular e considerações teóricas fundamentais são também encontrados em Sambrook *et al.*, *Molecular Cloning - A Laboratory Manual (3rd Ed.)*, Vol. 1-3, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, 2001; *Current Protocols In Molecular Biology*, Ausubel *et al.*, eds., *Current*

15 *Protocols*, uma joint venture da Greene Publishing Associates, Inc. and John Wiley & Sons, Inc., (suplementado em 2004); e em Watson *et al.* (1992) Recombinant DNA Second Edition Scientific American Books, NY. Além disso, essencialmente qualquer ácido nucléico (e qualquer ácido nucléico virtualmente marcado, quer padrão ou não ou padrão) pode ser ordenado de forma customizada ou padrão a partir de qualquer uma de uma variedade de

20 fontes comerciais, tal como o 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 projetadas por engenharia podem ser cultivadas em meio nutriente convencional modificado da maneira apropriada para tais atividades, por exemplo,

25 como classificar etapas, ativar promotores ou selecionar transformantes. Estas células podem opcionalmente ser cultivadas em organismos transgênicos. Outras referências usadas, por exemplo, para isolamento e cultura de célula (por exemplo, para isolamento de ácido nucléico subsequente) incluem Freshney (1994) Culture of Animal Cells, a Manual of Basic Technique, third edition, Wiley- Liss, New York e as referências aqui citadas; Payne *et al.* (1992) Cell of plant and Tissue Culture in Liquid Systems, John Wiley & Sons, Inc. New

30 York, NY; Gamborg and Phillips (eds) (1995) Cell of plant, Tissue and Organ Culture: Fundamental Methods Springer Lab Manual, Springer-Verlag (Berlin Heidelberg New York) e Atlas and Parks (eds) The Handbook of Microbiological Media (1993) CRC Press, Boca Raton, FL.

35 PROTEÍNAS E POLIPEPTÍDEOS DE INTERESSE

Métodos de produzir uma proteína em uma célula com um aminoácido não natural em um posição especificada são também um recurso da invenção. Por exemplo, um método

inclui crescer, em um meio apropriado, a célula, onde a célula compreende um ácido nucléico que compreende pelo menos um códon seletor e codifica uma proteína; e fornecer o aminoácido não natural, onde a célula compreende adicionalmente: um RNAt ortogonal (O-RNAt) que funciona na célula e reconhece o códon seletor; e, um aminoacil-RNAt
 5 sintetase ortogonal (O-RS) que preferencialmente aminoacila o O-RNAt com o aminoácido não natural. Uma proteína produzida por este método é também um recurso da invenção.

Em certas modalidades, o O-RS compreende uma predisposição para a aminoacilação do O-RNAt cognato em qualquer RNAt endógeno em um sistema de expressão. Uma razão relativa entre O-RNAt e RNAt endógeno que é carregado pelo O-RS,
 10 quando o O-RNAt e O-RS estão presentes em concentrações molares iguais, é maior que 1:1, preferivelmente pelo menos cerca de 2:1, mais preferivelmente 5:1, ainda mais preferivelmente 10:1, acima de tudo mais preferivelmente 20:1, ainda mais preferivelmente 50:1, acima de tudo mais preferivelmente 75:1, ainda mais preferivelmente 95:1, 98:1, 99:1, 100:1, 500:1, 1,000:1, 5,000:1 ou maior.

A invenção também fornece composições que incluem proteínas, onde as proteínas compreendem um aminoácido não natural. Em certas modalidades, a proteína compreende uma seqüência de aminoácidos que é pelo menos 75 % idêntica àquela de uma proteína terapêutica, uma proteína de diagnóstico, uma enzima industrial, ou porção destas.

As composições da invenção e composições feitas pelos métodos da invenção
 20 opcionalmente são em uma célula. Os pares de O-RNAt/O-RS ou componentes individuais da invenção podem em seguida ser usados em um maquinário de tradução do sistema hospedeiro, que resulta em um aminoácido não natural sendo incorporado em uma proteína. Publicação Internacional 2004/094593, depositado em 16 de abril de 2004, intitulado "EXPANDING THE EUCARYOTIC GENETIC CODE", e WO 2002/085923, intitulado "IN
 25 VIVO INCORPORATION OF UNNNATURAL AMINO ACIDS," descrevem este processo, e estão incorporados aqui pela referência. Por exemplo, quando um par de O-RNAt/O-R é introduzido em um hospedeiro, por exemplo, uma célula *Escherichia coli*, o par leva a incorporação *in vivo* de um aminoácido não natural tal como fenilselenocisteína em uma proteína em resposta a um códon seletor. O aminoácido não natural que é adicionado ao
 30 sistema pode ser um aminoácido sintético, tal como um derivado de um fenilalanina ou tirosina, que pode ser exogenamente adicionado ao meio e crescimento. Opcionalmente, as composições da presente invenção podem ser em um sistema de tradução *in vitro*, ou em um sistema(s) *in vivo*.

Uma célula da invenção fornece a capacidade de sintetizar proteínas que
 35 compreendem aminoácidos não naturais usados em grandes quantidades. Em alguns aspectos, a composição opcionalmente inclui, por exemplo, pelo menos 10 microgramas, pelo menos 50 microgramas, pelo menos 75 microgramas, pelo menos 100 microgramas,

pelo menos 200 microgramas, pelo menos 250 microgramas, pelo menos 500 microgramas, pelo menos 1 miligrama, pelo menos 10 miligramas ou mais da proteína que compreende um aminoácido não natural, ou uma quantidade que pode ser obtida com métodos de produção de proteína *in vivo* (detalhes na produção e purificação da proteína recombinante são fornecidos aqui). Em um outro aspecto, a proteína está opcionalmente presente na composição a uma concentração de, por exemplo, pelo menos 10 microgramas de proteína por litro, pelo menos 50 microgramas de proteína por litro, pelo menos 75 microgramas de proteína por litro, pelo menos 100 microgramas de proteína por litro, pelo menos 200 microgramas de proteína por litro, pelo menos 250 microgramas de proteína por litro, pelo menos 500 microgramas de proteína por litro, pelo menos 1 miligrama de proteína por litro, ou pelo menos 10 miligramas de proteína por litro ou mais, por exemplo, em um lisado de celular, um tampão, um tampão farmacêutico, ou outras suspensões líquidas (por exemplo, em um volume, por exemplo, de qualquer valor a partir de cerca de 1 nL a cerca de 100 L). A produção de grandes quantidades (por exemplo, maior que aquela tipicamente possível com outros métodos, por exemplo, tradução *in vitro*) de uma proteína em uma célula incluindo pelo menos um aminoácido não natural é um recurso da invenção.

A incorporação de um aminoácido não natural pode ser feita, por exemplo, para adequar mudanças em estrutura e/ou função da proteína, por exemplo, mudar o tamanho, acidez, nucleofilicidade, ligação de hidrogênio, hidrofobicidade, acessibilidade de sítios alvos de protease, alvejar uma fração (por exemplo, para um arranjo de proteína), incorporação de marcadores ou grupos reativos, etc. Proteínas que incluem um aminoácido não natural podem propriedades físicas ou catalíticas melhores ou ainda totalmente novas. Por exemplo, as propriedades seguintes são opcionalmente modificadas pela inclusão de um aminoácido não natural em uma proteína: toxicidade, biodistribuição, propriedades estruturais, propriedades espectroscópicas, propriedades químicas e/ou fotoquímicas, capacidade catalítica, meia-vida (por exemplo, meia-vida do soro), capacidade de reagir com outras moléculas, por exemplo, covalente ou não covalentemente, e similares. As composições incluindo proteínas que incluem pelo menos um aminoácido não natural são usadas, por exemplo, para novas terapêuticas, diagnósticos, enzimas catalíticas, enzimas industriais, proteínas de ligação (por exemplo, anticorpos), e por exemplo, o estudo de estrutura e função da proteína. Ver, por exemplo, Dougherty, (2000) *Unnnaturals Amino Acids as Probes of Protein Estruture and Funtion*, Current Opinion in Chemical Biology, 4:645-652.

Em alguns aspectos da invenção, uma composição inclui pelo menos uma proteína com pelo menos um, por exemplo, pelo menos dois, pelo menos três, pelo menos quatro, pelo menos cinco, pelo menos seis, pelo menos sete, pelo menos oito, pelo menos nove, ou pelo menos dez ou mais aminoácidos não naturais. Os aminoácidos não naturais podem ser

os mesmos ou diferente, por exemplo, pode haver 1, 2, 3, 4, 5, 6, 7, 8, 9 ou 10 ou mais diferentes sítios na proteína que compreendem 1, 2, 3, 4, 5, 6, 7, 8, 9 ou 10 ou mais aminoácidos não naturais diferentes. Em um outro aspecto, uma composição inclui uma proteína com pelo menos um, mas não todos, de um aminoácido particular presente na

5 proteína é um aminoácido não natural. Para uma dada proteína com mais que um aminoácidos não naturais, os aminoácidos não naturais podem ser idênticos ou diferentes (por exemplo, a proteína pode incluir dois ou mais diferente tipos de aminoácidos não naturais, ou pode incluir dois do mesmo aminoácido não natural). Para uma dada proteína com mais que dois aminoácidos não naturais, os aminoácidos não naturais podem ser os

10 mesmos, diferentes, ou um combinação de um aminoácido não natural múltiplo do mesmo tipo com pelo menos um aminoácido não natural diferente.

Essencialmente qualquer proteína (ou porção desta) que inclui um aminoácido não natural (e qualquer ácido nucléico de codificação correspondente, por exemplo, que inclui um ou mais códons seletores) pode ser produzida usando as composições e métodos aqui.

15 Nenhuma tentativa é feita para identificar as centenas de milhares de proteínas conhecidas, qualquer das quais pode ser modificada para incluir um ou mais aminoácidos não naturais, por exemplo, qualquer método de mutação adequando disponível para incluir um ou mais códons seletores apropriados em um sistema de tradução relevante. Repositórios de sequência comuns para proteínas conhecidas incluem GenBank EMBL, DDBJ e o NCBI.

20 Outros repositórios podem facilmente ser identificados pesquisando na Internet.

Tipicamente, a 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 de diagnóstico, uma enzima industrial, ou porção destas, e similares), e elas

25 compreendem um ou mais aminoácidos não naturais. Exemplos de proteínas terapêuticas, de diagnóstico e outras proteínas que podem ser modificadas para compreender um ou mais aminoácidos não naturais podem ser encontrados, mas sem limitações, em Publicações Internacionais 2004/094593, depositado em 16 de abril de 2004, intitulado "EXPANDING THE EUKARYOTIC GENETIC CODE", e WO 2002/085923, intitulado "IN

30 VIVO INCORPORATION OF UNNATURAL AMINO ACIDS". Exemplos de proteínas terapêuticas, de diagnóstico, e outras proteínas que podem ser modificados para compreender um ou mais aminoácidos não naturais incluem, mas sem limitações, por exemplo, hirudina, hormônio do crescimento humano, RAS, alfa-1 antitripsina, Angiostatina, fator Antiemolítico, anticorpos (mais detalhes nos anticorpos são encontrados a seguir),

35 Apolipoproteína, Apoproteína, fator Atrial natriurético, polipeptídeo natriurético Atrial, peptídeos Atriais, C-X-C quimiocinas (por exemplo, T39765, NAP-2, ENA-78, Gro-a, Gro-b, Gro-c, IP-10, GCP-2, NAP-4, SDF-1, PF4, MIG), Calcitonina, quimiocinas CC (por exemplo,

proteína quimioatrativa do monócito-1, proteína quimioatrativa do monócito-2, Proteína quimioatrativa do monócito-3, Proteína inflamatória do monócito-1 alfa, proteína inflamatória do monócito-1 beta, RANTES, I309, R83915, R91733, HCC1, T58847, D31065, T64262), ligante CD40, Ligante C-kit, Colágeno, fator estimulante de Colônia (CSF), fator do
 5 Complemento 5a, inibidor do Complemento, receptor do Complemento 1, citocinas, (por exemplo, Peptídeo-78 de ativação do Neutrófilo epitelial, GRO α /MGSA, GRO β , GRO γ , MIP-1a, MIP-1 δ , MCP-1), Fator de crescimento epidérmico (EGF), Eritropoietina ("EPO"), Toxinas esfoliantes A e B, Fator IX, Fator VII, Fator VIII, Fator X, Fator do crescimento Fibroblasto (FGF), Fibrinogênio, Fibronectina, G-CSF, GM-CSF, Glucocerebrosidase,
 10 Gonadotropina, fatores do crescimento, proteínas Hedgehog (por exemplo, Sonic, Indian, Desert), Hemoglobina, Fator do crescimento Hepatócito (HGF), hirudina, Albumina sérica humana, Insulina, Fator do crescimento tipo Insulina (IGF), interferons (por exemplo, IFN- α , IFN- β , IFN- γ), interleucinas (por exemplo, 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 Queratinócito (KGF), Lactoferrina, fator inibitório
 15 de leucemia, Luciferase, Neurturina, fator inibitório de Neutrófilo (NIP), oncostatina M, proteína Osteogênica, Hormônio do paratireóide, PD-ECSF, PDGF, hormônios do peptídeo (por exemplo, Hormônio do crescimento humano), Pleiotropino, Proteína A, Proteína G, exotoxinas Pirogênicas A, B, e C, Relaxina, Renina, SCF, receptor I do complemento Solúvel, I-CAM 1 Solúvel, receptores de interleucina Solúvel (IL-1, 2, 3, 4, 5, 6, 7, 9, 10, 11,
 20 12, 13, 14, 15), receptor de TNF Solúvel, Somatomedina, Somatostatina, Somatotropina, Estreptoquinase, Superantígenos, isto é, enterotoxinas Estafilococais (SEA, SEB, SEC1, SEC2, SEC3, SED, VER), superóxido de dismutase (SOD), toxina síndrome do choque tóxico (TSST-1), timosina alfa 1, ativador plasminogênio do Tecido, fator beta de necrose tumoral (TNF beta), receptor do fator de necrose tumoral (TNFR), fator-alfa de necrose
 25 tumoral (TNF alfa), Fator do crescimento endotelial vascular (VEGEF), uroquinase e muitos outros.

Uma classe de proteínas que pode ser feita usando as composições e métodos para incorporação de aminoácidos não naturais *in vivo* aqui descritos inclui moduladores transcrpcionais ou uma porção destes. Exemplo de moduladores transcrpcionais inclui genes
 30 e proteínas moduladoras transcrpcionais que modulam o crescimento, diferenciação e regulação celular, ou similares. Moduladores transcrpcionais são encontrados em procariotos, vírus e eucariotos, incluindo fungo, plantas, levedura, insetos e animais, incluindo mamíferos, fornecendo uma ampla faixa de alvos terapêuticos. Percebe-se que ativadores de expressão e transcrpcionais regulam transcrição por muitos mecanismos, por
 35 exemplo, ligando a receptores, estimulando uma cascata de transdução de sinal, regulando expressão de fatores de transcrição, ligando a promotores e melhoradores, ligando a proteínas que se ligam a promotores e melhoradores, desdobrando DNA, unindo pre-mRNA,

poliadenilando RNA, e degradando RNA.

Uma classe de proteínas da invenção (por exemplo, proteínas com um ou mais aminoácidos não naturais) inclui proteínas biologicamente ativas tais como citocinas, moléculas inflamatórias, fatores do crescimento, seus receptores, e produtos do oncogene, por exemplo, interleucinas (por exemplo, IL-1, IL-2, IL-8, etc.), interferons, FGF, IGF-I, IGF-II, FGF, PDGF, TNF, TGF- α , TGF- β , EGF, KGF, SCF/c-Kit, CD40L/CD40, VLA-4/VCAM-1, ICAM-1/LFA-1, e hialurina/CD44; moléculas de transdução de sinal e produtos do oncogene correspondentes, por exemplo, Mos, Ras, Raf, e Met; e ativadores e supressores transcricionais, por exemplo, p53, Tat, Fos, Myc, Jun, Myb, Rel, e receptores do hormônio esteróide tal como aquele para estrogênio, progesterona, testosterona, aldosterona, o ligante do receptor de LDL e corticosterona.

Enzimas (por exemplo, enzimas industriais) ou porções destas com pelo menos um aminoácido não natural são também fornecidas pela invenção. Exemplos de enzimas incluem, mas sem limitações, por exemplo, amidases, racemases de aminoácido, acilases, dealogenases, dioxigenases, diarilpropano peroxidases, epimerases, epóxido hidrolases, esterases, isomerases, quinases, glicose isomerases, glicosidases, glicosil transferases, haloperoxidases, monooxigenases (por exemplo, p450s), lipases, lignina peroxidases, nitrila hidratases, nitrilases, proteases, fosfatases, subtilisinas, transaminase, e nucleases.

Muitas destas proteínas estão comercialmente disponíveis (Ver, por exemplo, Sigma BioSciences), e as seqüências de proteína e genes correspondentes, e tipicamente, muitos variantes destes, são bem conhecidas (Ver, por exemplo, Genbank). Qualquer delas pode ser modificada pela inserção de um ou mais aminoácidos não naturais de acordo com a invenção, por exemplo, para alterar a proteína com relação a uma ou mais propriedades terapêutica, de diagnóstico ou enzimática de interesse. Exemplos de propriedades terapeuticamente relevantes incluem meia-vida sérica, meia-vida em prateleira, estabilidade, imunogenicidade, atividade terapêutica, detectabilidade (por exemplo, pela inclusão de grupos reportadores (por exemplo, marcadores ou sítios de ligação de marcador) nos aminoácidos não naturais), redução de LD₅₀ ou outros efeitos colaterais, capacidade de entrar no corpo através do trato gástrico (por exemplo, disponibilidade oral), ou similares. Exemplos de propriedades de diagnóstico incluem meia-vida em prateleira, estabilidade, atividade de diagnóstico, detectabilidade ou similares. Exemplos de propriedades enzimáticas relevantes incluem meia-vida em prateleira, estabilidade, atividade enzimática, capacidade de produção, ou similares.

Uma variedade de outras proteínas pode também ser modificada para incluir um ou mais aminoácidos não naturais usando composições e métodos da invenção. Por exemplo, a invenção pode incluir substituir um ou mais aminoácidos naturais em uma ou mais proteínas de vacina com um aminoácido não natural, por exemplo, em proteínas de fungos

infecciosos, por exemplo, *Aspergillus*, espécie de *Candida*; bactéria, particularmente *E. coli*, que serve de modelo para bactéria patogênica, bem como bactéria medicamente importantes tais como *Staphylococci* (por exemplo, *aureus*), ou *Streptococci* (por exemplo, *pneumoniae*); protozoário tal como esporozoário (por exemplo, *Plasmodia*), *rhizopods* (por exemplo, *Entamoeba*) e flagelados (*Trypanosome*, *Leishmania*, *Trichomonas*, *Giardia*, etc.); vírus tais como vírus RNA (+) (exemplos incluem Poxvírus por exemplo, *vaccinia*; Picornavírus, por exemplo, *polio*; Togavírus, por exemplo, *rubella*; Flavivírus, por exemplo, HCV; e Coronavírus), vírus RNA (-) (por exemplo, Rabdovírus, por exemplo, VSV; Paramixovírus, por exemplo, RSV; Ortomixovírus, por exemplo, influenza; Buniavírus; e Arenavírus), vírus dsDNA (Reovírus, por exemplo), vírus RNA a DNA, isto é, Retrovírus, por exemplo, HIV e HTLV, e certos vírus DNA a RNA tais como Hepatite B.

Proteínas agriculturalmente relacionados tais como proteínas de resistência a inseto (por exemplo, o Cri proteínas), enzimas de produção de amido e lipídio, toxinas de planta e inseto, proteínas de resistência a toxina, proteínas de desintoxicação de Micotoxina, enzimas de crescimento de planta (por exemplo, Ribulose 1,5-bisfosfato Carboxilase/Oxygenase, "RUBISCO"), lipoxigenase (LOX), e Fosfoenolpiruvato (PEP) carboxilase são também alvos adequados para modificação de aminoácido não natural.

Em certas modalidades, a proteína ou polipeptídeo de interesse (ou porção destes) nos métodos e/ou composições da invenção 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, pelo menos sete códons seletores, pelo menos oito códons seletores, pelo menos nove códons seletores, dez ou mais códons seletores.

Genes que codificam proteínas ou polipeptídeos de interesse podem ser mutagenizados usando métodos bem conhecidos pelos versados na tecnologia e aqui descritos em "Mutagenesis and Other Molecular Biology Techniques" 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 interesse é mutagenizado para incluir um ou mais códons seletores, fornecendo a inserção de um ou mais aminoácidos não naturais. A invenção inclui qualquer variante deste, por exemplo, mutante, versões de qualquer proteína, por exemplo, incluindo pelo menos um aminoácido não natural. Similarmente, a invenção também inclui ácidos nucléicos correspondentes, isto é, qualquer ácido nucléico com um ou mais códons seletores que codificam um ou mais aminoácidos não naturais.

Para preparar uma proteína que inclui um aminoácido não natural, pode-se usar células hospedeiras e organismos que são adaptados para a incorporação *in vivo* do

aminoácido não natural por meio de pares RNA^t/RS ortogonais. Células hospedeiras são geneticamente projetadas por engenharia (por exemplo, transformadas, convertidas ou transfectadas) com um ou mais vetores que expressam o RNA^t ortogonal, o RNA^t ortogonal sintetase, e um vetor que codifica a proteína a ser derivatizada. Cada um desses

5 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 ser, por exemplo, na forma de um plasmídeo, uma bactéria, um vírus, um polinucleotídeo nu, ou um polinucleotídeo conjugado.

Definição de Polipeptídeos por Imunoreatividade

10 Em virtude de os polipeptídeos da invenção fornecerem uma variedade de novas seqüências de polipeptídeo (por exemplo, polipeptídeos que compreendem aminoácidos não naturais no caso de proteínas sintetizadas nos sistemas de tradução aqui, ou, por exemplo, no caso das sintetases inéditas, seqüências inéditas de aminoácidos padrões), os polipeptídeos também fornecem novos recursos estruturais que podem ser reconhecidos,

15 por exemplo, em ensaios imunológicos. A geração de anti-soro, que liga especificamente os polipeptídeos da invenção, bem como os polipeptídeos que são ligados por anti-soro como esse, são um recurso da invenção. O termo "anticorpo", da maneira aqui usada, inclui, mas sem limitações, um polipeptídeo substancialmente codificado por um gene de imunoglobulina ou genes de imunoglobulina, ou fragmentos destes que ligam

20 especificamente e reconhecem um analito (antígeno). Exemplos incluem anticorpos de cadeia policlonal, monoclonal, quimérica e simples, e similares. Fragmentos de imunoglobulinas, incluindo fragmentos Fab e fragmentos produzidos por uma biblioteca de expressão, incluindo exibição de fago, são também incluídos no termo "anticorpo" da maneira aqui usada. Ver, por exemplo, Paul, Fundamental Immunology, 4th Ed., 1999,

25 Raven Press, New York, para estrutura e terminologia de anticorpo.

A fim de produzir anti-soro para uso em um imunoensaio, um ou mais dos polipeptídeos imunogênicos são produzidos e purificados da maneira aqui descrita. Por exemplo, proteína recombinante pode ser produzida em uma célula recombinante. Uma cepa natural de camundongos (usada neste ensaio em virtude de os resultados serem mais

30 reprodutíveis em virtude da identidade genética virtual dos camundongos) é imunizada com a(s) proteína(s) imunogênica(s) em combinação com um adjuvante padrão, tal como adjuvante de Freund, e um protocolo de imunização de camundongo padrão (Ver, por exemplo, Harlow and Lane (1988) Antibodies, A Laboratory Manual, Cold Spring Harbor Publications, New York, para uma descrição padrão de geração de anticorpo, formatos e

35 condições de imunoensaio que podem ser usados para determinar imunoreatividade específica. Detalhes adicionais em proteínas, anticorpos, anti-soro, etc. podem ser encontrados em International Publication Numbers WO 2004/094593, intitulado

"EXPANDING THE EUCARYOTIC GENETIC CODE;" WO 2002/085923, intitulado "IN VIVO INCORPORATION OF UNNATURAL AMINO ACIDS", WO 2004/035605, intitulado "GLYCOPROTEIN SYNTHESIS", e WO 2004/058946, intitulado "PROTEIN ARRAYS."

USE DE PARES O-RNA E O-RS E O-RNAt/O-RS

5 As composições da invenção e composições feitas pelos métodos da invenção estão opcionalmente em uma célula. Os pares de O-RNAt/O-RS ou componentes individuais da invenção podem em seguida ser usados em um maquinário de tradução do sistema hospedeiro, o que faz com que um aminoácido não natural seja incorporado em uma proteína. International Publication Number WO 2002/085923 by Schultz, *et al.*, intitulado "IN

10 VIVO INCORPORATION OF UNNATURAL AMINO ACIDS", descreve este processo e está incorporado aqui pela referência. Por exemplo, quando um par de O-RNAt/O-RS é introduzido em um hospedeiro, por exemplo, *Escherichia coli* ou levedura, o par leva à incorporação *in vivo* de um aminoácido não natural, que pode ser exogenamente adicionado ao meio de crescimento, em uma proteína, por exemplo, uma proteína teste de mioglobina

15 ou uma proteína terapêutica, em resposta a um códon seletor, por exemplo, um códon não sentido âmbar. Opcionalmente, as composições da invenção podem ser em um sistema de tradução *in vitro*, ou em um(s) sistema(s) celular(s) *in vivo*. Proteínas com o aminoácido não natural podem ser usadas em qualquer de uma ampla faixa de aplicação. Por exemplo, a fração não natural incorporada em uma proteína pode servir como um alvo para qualquer de

20 uma ampla faixa de modificações, por exemplo, reticulando com outras proteínas, com pequenas moléculas tais como marcadores ou corantes e/ou biomoléculas. Com estas modificações, a incorporação do aminoácido não natural pode resultar em melhores proteínas terapêuticas e pode ser usada para alterar ou melhorar a função catalítica de enzimas. Em alguns aspectos, a incorporação e modificação subsequente de um

25 aminoácido não natural em uma proteína pode facilitar estudos na estrutura da proteína, interações com outras proteínas, e similares.

ESTOJOS

Estojos são também um recurso da invenção. Por exemplo, é fornecido um estojo para produzir uma proteína que compreende pelo menos um aminoácido não natural em

30 uma célula, onde o estojo inclui pelo menos um recipiente contendo uma sequência de polinucleotídeos que codifica um O-RNAt, e/ou um O-RNAt, e/ou uma sequência de polinucleotídeos que codifica um O-RS, e/ou um O-RS. Em uma modalidade, o estojo inclui adicionalmente o aminoácido não natural fenilselenocisteína. Em uma outra modalidade, o estojo compreende adicionalmente materiais instrucionais para produzir a proteína e/ou uma

35 célula hospedeira.

EXEMPLOS

Os exemplos seguintes são oferecidos para ilustrar, mas não para limitar, a

invenção reivindicada. Versados na tecnologia reconhecerão uma variedade de parâmetros não críticos que podem ser alterados sem fugir do escopo da invenção reivindicada. Entende-se que os exemplos e modalidades aqui descritos têm propósitos apenas ilustrativos e que várias modificações ou mudanças à luz desta serão sugeridos pelos versados na tecnologia e que deverão estar incluído no espírito e propósito deste pedido e escopo das reivindicações anexas.

EXEMPLO

Seleção genética de Sintetases Mutantes Específicas para Fenilselenocisteína

QUÍMICA REATIVA DE FENILSELENOCISTEÍNA

Selênio é um elemento essencial tanto em síntese orgânica quanto biológica. Demonstrou-se que uma fração de fenilseleno pode ser oxidada em condições brandas, e a eliminação espontânea de syn do selenóxido resulta na formação de olefina. Esta estratégia tem sido usada extensivamente em síntese total.

A incorporação de um aminoácido não natural contendo selênio em proteínas criará um alvo atrativo para modificação pós-translacional altamente seletiva de proteínas, por exemplo, para produzir proteínas seletivamente lipidadas. Um aminoácido seleno não natural como esse é fenilselenocisteína (CAS Registry Number 71128-82-0), sinonimamente denominada fenilselenida ou Se-Fenil-L-selenocisteína. Esta estrutura é fornecida na figura 1, estrutura 1. Este aminoácido não natural pode ser obtido, por exemplo, da Sigma®-Aldrich® Co., Inc., número de catálogo 50827.

Clivagem oxidativa de um resíduo de aminoácido fenilselenocisteína resulta no aminoácido deidroalanina α,β -insaturado (Ver figura 1, estrutura 2). Deidroalanina é encontrada em inúmeros peptídeos de ocorrência natural (Chatterjee *et al.* (2005), *Chemical Reviews* 105(2):633-683). Deidropeptídeos contendo resíduos de cisteína não protegidos sofrem adição de conjugado estereosseletivo intramolecular para produzir lantioninas cíclicas (Chatterjee *et al.* (2005), *Chemical Reviews* 105(2):633-683). Lantioninas são estruturas que são encontradas em lantibióticos, uma classe de antibióticos modificados pós-translacionalmente (Chatterjee *et al.* (2005), *Chemical Reviews* 105(2):633-683).

Reações de adição de Michael do aminoácido não natural deidroalanina (Ver figura 1, estrutura 2) podem resultar em proteínas com modificações pós-translacionais, por exemplo, por reação com um tio-lípido para gerar uma proteína lipidada. Mais especificamente, reação com ácido tiopalmítico resulta em palmitoilcisteína (Ver figura 1, estrutura 3), reação com farnesilmercaptano disponibiliza farnesilcisteína (Ver figura 1, estrutura 4), e reação com malonato produz ácido γ -carboxiglutâmico (Ver figura 1, estrutura 6). Além disso, reação com 1-hexadecanetiol resulta em S-hexadecilcisteína (Ver figura 1, estrutura 5). Embora S-hexadecilcisteína não seja uma modificação pós-translacional nativa, a presença deste resíduo em proteína pode ter propriedades desejadas; por exemplo, ela

pode resultar em ligação de albumina sérica humana, e maior estabilidade *in vivo* da proteína.

A lipidação de polipeptídeos por modificação alvejada de resíduos de fenilselenocisteína pode ter propriedades desejadas, tais como localização de membrana, melhor estabilidade *in vivo* (isto é, maior meia-vida) e solubilidade de lipídio. Além disso, as formas lipidadas de algumas proteínas são as formas biologicamente ativas.

PRODUÇÃO DE PROTEÍNAS QUE COMPREENDE FENILSELENOCISTEÍNA

Metodologias que permitem a adição sistemática de aminoácidos não naturais aos códigos genéticos de *E. coli*, levedura e célula mamíferas foram descritas anteriormente.

Tais métodos são baseados na evolução de um supressor RNAt não sentido um par aminoacil-RNAt sintetase (RS) que tem a propriedade de ortogonalidade, definida como a capacidade de incorporar seletivamente um dado aminoácido em resposta a um códon exclusivo sem passar por reação cruzada com RNAts do hospedeiro endógeno, aminoacil-RNAt sintetase, ou aminoácidos.

Uma presente invenção fornece composições e métodos para a geração de polipeptídeos *in vivo* contendo o aminoácido não natural fenilselenocisteína pela criação de reagentes ortogonais que permitem a incorporação translacional geneticamente programada deste aminoácido não natural diretamente em uma cadeia do polipeptídeos em crescimento. Além disso, o uso destes sistemas ortogonais permite a produção de grandes quantidades de polipeptídeos contendo o aminoácido não natural (isto é, quantidades muito grandes dos polipeptídeos que seria possível usando outros métodos de síntese). Polipeptídeos que compreendem fenilselenocisteína podem ser usados em reações de modificação alveadas, por exemplo, mas sem limitações, a geração de polipeptídeos artificialmente lipidados de acordo com as instruções da presente especificação. Tais proteínas modificadas encontram uma variedade de usos, incluindo, mas sem limitações, melhores agentes terapêuticos.

Para gerar um par de RNAt ortogonal/aaRS que insere exclusivamente fenilselenocisteína (figura 1, estrutura 1), uma biblioteca de mutantes de sítio ativo do *Methanococcus jannaschii* tirosil-RNAt sintetase (M)TyrRS), que especificamente carrega um supressor não sentido *M. jannaschii* projetado por engenharia (MjRNAt^{Tyr_{CUA}}) não reconhecido pela *E. coli* sintetase, foi usado. Uma biblioteca de pBK-lib5 sintetase, da maneira descrita em Wang *et al.* (2006), *Annual Review of Biophysics and Biomolecular Structure* 35:225-249, foi usada na classificação e submetida a uma série de seleções positiva e negativa. A sobrevivência na seleção positiva foi contingente mediante a supressão de uma mutação âmbar no gene cloranfenicol acetiltransferase (CAT) na presença de fenilselenocisteína; a sobrevivência na seleção negativa foi contingente mediante supressão inadequada de mutações âmbar em um gene que codifica a proteína barnase tóxica na ausência de fenilselenocisteína. Clones sobrevivem tanto através de

etapas positiva quanto negativa de seleção apenas se eles incorporarem exclusivamente fenilselenocisteína em resposta ao códon âmbar.

Após estas seleções, foram identificados candidatos a clone que permitiram que as células abrigassem o gene CAT com um códon seletor de mutação âmbar em um sítio permissivo para sobreviver na presença de cloranfenicol apenas na presença de fenilselenocisteína. Na ausência de fenilselenocisteína, as mesmas células não crescem na presença de cloranfenicol, consistente com incorporação de fenilselenocisteína eficiente com pouca ou nenhuma sobrevivência de fundo da incorporação de aminoácidos endógenos.

Seqüenciamento dos clones sintetase mutante candidatos revelaram três diferente sintetase isoladas, cada uma das quais é capaz de funcionar em um sistema de ORTOGONAL TRANSLATION

Clone 1 (PhSeRS-SD): Y32W, L65E, H70G, D158Q, L162S

Clone 2 (PhSeRS-K4): Y32W, L65H, H70G, F108N, Q109S, D158S, L162E

Clone 3 (PhSeRS-K5): Y32W, L65H, A67G, H70G, F108K, Q109S, D158E, L162E

Estes clones são sumarizados mais detalhadamente na tabela a seguir.

Seqüências de aminoácidos tirosil-RNAt sintetase de *Methanococcus jannaschii*

Tipo selvagem e Mutantes

| | 32 | 65 | 67 | 70 | 108 | 109 | 158 | 162 | SEQ ID NO: |
|-----------------------------|-----|-----|-----|-----|-----|-----|-----|-----|------------|
| Tipo selvagem | r | Leu | Ala | His | Phe | Gln | Asp | Leu | 2 |
| Clone 1 Mutante (SD) | Trp | Glu | Ala | Gly | Phe | Gln | Gln | Ser | 4 |
| Clone 2 Mutante (K4) | Trp | His | Ala | Gly | Asn | Ser | Ser | Glu | 6 |
| Clone 3 Mutante (KS) | Trp | His | Gly | Gly | Lys | Ser | Glu | Glu | 8 |

Das três sintetases mutantes isoladas, a sintetase do clone 2 (PhSeRSK4) teve a mais alta atividade e especificidade. O nucleotídeo e seqüências de aminoácidos completos de cada um desses clones e a espécie tipo selvagem correspondente são fornecidos na

figura 2.

VALIDAÇÃO DA SINTETASE ORTOGONAL

Usando a sintetase mutante aqui descrita, uma proteína modelo mioglobina contendo um códon seletor fr âmbar na posição 4 (TAG4) foi expressa com 1,5-2 mg/L de rendimento sem nenhuma expressão de fundo na ausência de fenilselenocisteína. Esta proteína modelo contendo fenilselenocisteína foi adicionalmente submetida a modificação pós-translacional para produzir ácido γ -carboxiglutâmico (figura 1 estrutura 6), que foi confirmado usando espectrometria de massa de alta resolução.

Duas proteínas modelos adicionais compreendendo fenilselenocisteína foram também expressas com 3-5 mg/L de rendimento usando o sistema de produção de sintetase mutante e ortogonal *in vivo* aqui descrito. Hormônio do crescimento humano (hGH) foi produzido contendo fenilselenocisteína. Após a produção da forma de fenilselenocisteína, a proteína foi submetida a modificação para produzir hGH que contém S-hexadecilcisteína. As propriedades de ligação da albumina sérica humana e a meia-vida *in vivo* do hGH modificado em camundongos estão atualmente sendo testadas.

Os reagentes ortogonais aqui descritos foram também usados para produzir proteína fluorescente verde (GFP) contendo fenilselenocisteína. Essa forma da proteína foi também subsequente modificada para formar, alternativamente, S-hexadecilcisteína, farnesilcisteína e palmitoilcisteína de GFP. As propriedades de alvejando da membrana dessas formas lipidadas de GFP estão atualmente sendo testadas.

Assim, a presente invenção fornece composições e métodos para a produção de proteínas geneticamente programadas contendo o aminoácido não natural fenilselenocisteína (Ver figura 1, estrutura 1). A subsequente eliminação oxidativa por peróxido de hidrogênio produz deidroalanina (Ver figura 1, estrutura 2) na posição desejada. Esta fração de deidroalanina pode em seguida ser alvejada em várias reações secundárias de modificação.

Embora a invenção apresentada tenha sido descrita com algum detalhe com propósitos de clareza e entendimento, fica claro aos versados na tecnologia, a partir da leitura dessa revelação, que várias mudanças nesta revelação na forma e detalhes podem ser feitas sem fugir do espírito e escopo da invenção. Todas as publicações, patentes, pedidos de patente, e/ou outros documentos citados neste pedido estão incorporados pela referência na sua íntegra com todos os propósitos, como se cada publicação, patente, pedido de patente individual e/ou outro documento estivessem individualmente indicados para ser incorporados pela referência com todos os propósitos.

LISTAGEM DE SEQUÊNCIA

<110> The Scripps Research Institute
WANG, JIANGYUN
SCHULTZ, PETER G

<120> "EXPRESSÃO DE PROTEÍNAS GENETICAMENTE PROGRAMADA
CONTENDO O AMINOÁCIDO NÃO NATURAL FENILSELENOCISTEÍNA"

<130> 54-001820US

<140> US 11/724,039

<141> 2007-03-13

<160> 9

<170> PatentIn version 3.4

<210> 1
<211> 77
<212> RNA
<213> Artificial

<220>
<223> nonsense suppressor tyrosyl-tRNA CUA derived from Methanococcus
jannaschii

<400> 1
ccggcgguag uucagcaggg cagaacggcg gacucuaaaau ccgcauggcg cugguucaaa 60
uccggccccgc cggacca 77

<210> 2
<211> 306
<212> PRT
<213> Methanococcus jannaschii

<400> 2

Met Asp Glu Phe Glu Met Ile Lys Arg Asn Thr Ser Glu Ile Ile Ser
1 5 10 15

Glu Glu Glu Leu Arg Glu Val Leu Lys Lys Asp Glu Lys Ser Ala Tyr
20 25 30

Ile Gly Phe Glu Pro Ser Gly Lys Ile His Leu Gly His Tyr Leu Gln
35 40 45

Ile Lys Lys Met Ile Asp Leu Gln Asn Ala Gly Phe Asp Ile Ile Ile
50 55 60

Leu Leu Ala Asp Leu His Ala Tyr Leu Asn Gln Lys Gly Glu Leu Asp
65 70 75 80

Glu Ile Arg Lys Ile Gly Asp Tyr Asn Lys Lys Val Phe Glu Ala Met
85 90 95

Gly Leu Lys Ala Lys Tyr Val Tyr Gly Ser Glu Phe Gln Leu Asp Lys
100 105 110

Asp Tyr Thr Leu Asn Val Tyr Arg Leu Ala Leu Lys Thr Thr Leu Lys
 115 120 125

Arg Ala Arg Arg Ser Met Glu Leu Ile Ala Arg Glu Asp Glu Asn Pro
 130 135 140

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

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

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

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

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

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

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

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

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

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

Arg Leu
 305

<210> 3
 <211> 918
 <212> DNA
 <213> Methanococcus jannaschii

<400> 3
 atggacgaat ttgaaatgat aaagagaaac acatctgaaa ttatcagcga ggaagagtta 60
 agagaggttt taaaaaaga tgaaaaatct gcttacatag gttttgaacc aagtggtaaa 120
 atacatttag ggcattatct ccaataaaa aagatgattg atttacaaaa tgctggattt 180
 gatataatta tattgttggc tgatttacac gcctatttaa accagaaagg agagttggat 240
 gagattagaa aaataggaga ttataacaaa aaagtttttg aagcaatggg gttaaaggca 300


```

aaatatgttt atggaagtga attccagctt gataaggatt atacactgaa tgtctataga 360
ttggctttta aaactacctt aaaaagagca agaaggagta tggaacttat agcaagagag 420
gatgaaaatc caaaggttgc tgaagttatc tatccaataa tgcagggttaa tgatattcat 480
tatttaggcg ttgatgttgc agttggaggg atggagcaga gaaaaatata catgttagca 540
agggagcttt taccaaaaaa ggttgtttgt attcacaacc ctgtcttaac gggtttggat 600
ggagaaggaa agatgagttc ttcaaaaggg aattttatag ctgttgatga ctctccagaa 660
gagattaggg ctaagataaa gaaagcatat tgcccagctg gagttgttga aggaaatcca 720
ataatggaga tagctaaata cttccttgaa tctcctttaa ccataaaaag gccagaaaaa 780
tttggaggag atttgacagt taatagctat gaggagttag agagtttatt taaaaataag 840
gaattgcatc caatggattt aaaaaatgct gtagctgaag aacttataaa gattttagag 900
ccaattagaa agagatta 918

```

```

<210> 4
<211> 306
<212> PRT
<213> Artificial

```

```

<220>
<223> Phenylselenocysteine aminoacyl tRNA synthetase

```

```

<400> 4

```

```

Met Asp Glu Phe Glu Met Ile Lys Arg Asn Thr Ser Glu Ile Ile Ser
1          5          10          15

```

```

Glu Glu Glu Leu Arg Glu Val Leu Lys Lys Asp Glu Lys Ser Ala Trp
20          25          30

```

```

Ile Gly Phe Glu Pro Ser Gly Lys Ile His Leu Gly His Tyr Leu Gln
35          40          45

```

```

Ile Lys Lys Met Ile Asp Leu Gln Asn Ala Gly Phe Asp Ile Ile Ile
50          55          60

```

```

Glu Leu Ala Asp Leu Gly Ala Tyr Leu Asn Gln Lys Gly Glu Leu Asp
65          70          75          80

```

```

Glu Ile Arg Lys Ile Gly Asp Tyr Asn Lys Lys Val Phe Glu Ala Met
85          90          95

```

```

Gly Leu Lys Ala Lys Tyr Val Tyr Gly Ser Glu Phe Gln Leu Asp Lys
100          105          110

```

```

Asp Tyr Thr Leu Asn Val Tyr Arg Leu Ala Leu Lys Thr Thr Leu Lys
115          120          125

```

```

Arg Ala Arg Arg Ser Met Glu Leu Ile Ala Arg Glu Asp Glu Asn Pro
130          135          140

```

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

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

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

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

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

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

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

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

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

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

Arg Leu
305

<210> 5
<211> 921
<212> DNA
<213> Artificial

<220>
<223> Phenylselenocysteine aminoacyl tRNA synthetase

<400> 5
atggacgaat ttgaaatgat aaagagaaac acatctgaaa ttatcagcga ggaagagtta 60
agagaggttt taaaaaaga tgaaaaatct gcttggatag gttttgaacc aagtggtaaa 120
atacatttag ggcattatct ccaataaaa aagatgattg atttcaaaa tgctggattt 180
gatataatta tagagttggc tgatttaggg gcctatttaa accagaaagg agagttggat 240
gagattagaa aaataggaga ttataacaaa aaagtttttg aagcaatggg gttaaaggca 300
aaatatgttt atggaagtga attccagctt gataaggatt atacactgaa tgtctataga 360
ttggctttta aaactacctt aaaaagagca agaaggagta tggaacttat agcaagagag 420

gatgaaaatc caaaggttgc tgaagttatc tatccaataa tgcagggttaa tcaaattcat 480
 tatagtggcg ttgatgttgc agttggaggg atggagcaga gaaaaataca catgttagca 540
 agggagcttt taccaaaaaa ggttggttgc attcacaacc ctgtcttaac gggtttggat 600
 ggagaaggaa agatgagttc ttcaaaaggg aattttatag ctgttgatga ctctccagaa 660
 gagattaggg ctaagataaa gaaagcatac tgcccagctg gagttgttga aggaaatcca 720
 ataatggaga tagctaaata cttccttgaa tatcctttaa ccataaaaag gccagaaaaa 780
 tttggtggag atttgacagt taatagctat gaggagttag agagtttatt taaaaataag 840
 gaattgcata caatggattt aaaaaatgct gtagctgaag aacttataaa gattttagag 900
 ccaattagaa agagattata a 921

<210> 6
 <211> 306
 <212> PRT
 <213> Artificial

<220>
 <223> Phenylselenocysteine aminoacyl tRNA synthetase

<400> 6

Met Asp Glu Phe Glu Met Ile Lys Arg Asn Thr Ser Glu Ile Ile Ser
1 5 10 15

Glu Glu Glu Leu Arg Glu Val Leu Lys Lys Asp Glu Lys Ser Ala Trp
20 25 30

Ile Gly Phe Glu Pro Ser Gly Lys Ile His Leu Gly His Tyr Leu Gln
35 40 45

Ile Lys Lys Met Ile Asp Leu Gln Asn Ala Gly Phe Asp Ile Ile Ile
50 55 60

His Leu Ala Asp Leu Gly Ala Tyr Leu Asn Gln Lys Gly Glu Leu Asp
65 70 75 80

Glu Ile Arg Lys Ile Gly Asp Tyr Asn Lys Lys Val Phe Glu Ala Met
85 90 95

Gly Leu Lys Ala Lys Tyr Val Tyr Gly Ser Glu Asn Ser Leu Asp Lys
100 105 110

Asp Tyr Thr Leu Asn Val Tyr Arg Leu Ala Leu Lys Thr Thr Leu Lys
115 120 125

Arg Ala Arg Arg Ser Met Glu Leu Ile Ala Arg Glu Asp Glu Asn Pro
130 135 140

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

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

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

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

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

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

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

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

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

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

Arg Leu
 305

<210> 7
 <211> 921
 <212> DNA
 <213> Artificial

<220>
 <223> Phenylselenocysteine aminoacyl tRNA synthetase

<400> 7
 atggacgaat ttgaaatgat aaagagaaac acatctgaaa ttatcagcga ggaagagtta 60
 agagaggttt taaaaaaga tgaaaaatct gcttgatag gttttgaacc aagtggtaaa 120
 atacatttag gccattatct ccaaataaaa aagatgattg atttcaaaa tgctggattt 180
 gatataatta tacatttggc tgatttaggc gcctatttaa accagaaagg agagttggat 240
 gagattagaa aaataggaga ttataacaaa aaagtttttg aagcaatggg gttaaaggca 300
 aaatatgttt atggaagtga aaattctctt gataaggatt atacactgaa tgtctataga 360
 ttggctttta aaactacctt aaaaagagca agaaggagta tggaacttat agcaagagag 420
 gatgaaaatc caaaggttgc tgaagttatc tatccaataa tgcagggttaa ttctattcat 480
 tatgagggcg ttgatgttgc agttggaggg atggagcaga gaaaaataca catgttagca 540

agggagcttt taccaaaaaa ggttgttgt attcacaacc ctgtcttaac gggtttggat 600
 ggagaaggaa agatgagttc ttcaaaaggg aattttatag ctgttgatga ctctccagaa 660
 gagattaggg ctaagataaa gaaagcatac tgcccagctg gagttgttga aggaaatcca 720
 ataatggaga tagctaaata cttccttgaa tatcctttaa ccataaaaag gccagaaaaa 780
 tttggtggag atttgacagt taatagctat gaggagttag agagtttatt taaaaataag 840
 gaattgcatc caatggattt aaaaaatgct gtagctgaag aacttataaa gatttttagag 900
 ccaattagaa agagattata a 921

<210> 8
 <211> 306
 <212> PRT
 <213> Artificial

<220>
 <223> Phenylselenocysteine aminoacyl tRNA synthetase

<400> 8

Met Asp Glu Phe Glu Met Ile Lys Arg Asn Thr Ser Glu Ile Ile Ser
1 5 10 15

Glu Glu Glu Leu Arg Glu Val Leu Lys Lys Asp Glu Lys Ser Ala Trp
20 25 30

Ile Gly Phe Glu Pro Ser Gly Lys Ile His Leu Gly His Tyr Leu Gln
35 40 45

Ile Lys Lys Met Ile Asp Leu Gln Asn Ala Gly Phe Asp Ile Ile Ile
50 55 60

His Leu Gly Asp Leu Gly Ala Tyr Leu Asn Gln Lys Gly Glu Leu Asp
65 70 75 80

Glu Ile Arg Lys Ile Gly Asp Tyr Asn Lys Lys Val Phe Glu Ala Met
85 90 95

Gly Leu Lys Ala Lys Tyr Val Tyr Gly Ser Glu Lys Ser Leu Asp Lys
100 105 110

Asp Tyr Thr Leu Asn Val Tyr Arg Leu Ala Leu Lys Thr Thr Leu Lys
115 120 125

Arg Ala Arg Arg Ser Met Glu Leu Ile Ala Arg Glu Asp Glu Asn Pro
130 135 140

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

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

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

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

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

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

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

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

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

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

Arg Leu
 305

<210> 9
 <211> 921
 <212> DNA
 <213> Artificial

<220>
 <223> Phenylselenocysteine aminoacyl tRNA synthetase

<400> 9
 atggacgaat ttgaaatgat aaagagaaac acatctgaaa ttatcagcga ggaagagtta 60
 agagagggttt taaaaaaga tgaaaaatct gcttggatag gttttgaacc aagtggtaaa 120
 atacatttag ggcattatct ccaaataaaa aagatgattg atttcaaaa tgctggattt 180
 gatataatta tacatttggg tgatttaggc gcctatttaa accagaaagg agagttggat 240
 gagattagaa aaataggaga ttataacaaa aaagtttttg aagcaatggg gttaaaggca 300
 aaatatgttt atggaagtga aaagtctctt gataaggatt atacactgaa tgtctataga 360
 ttggcttttaa aaactacctt aaaaagagca agaaggagta tggaacttat agcaagagag 420
 gatgaaaatc caaagggttc tgaagttatc tatccaataa tgcagggttaa tgagattcat 480
 tatgagggcg ttgatgttc agttggaggg atggagcaga gaaaaataca catgttagca 540
 agggagcttt taccaaaaaa gggtgtttgt attcacaacc ctgtcttaac gggtttggat 600
 ggagaaggaa agatgagttc ttcaaaaggg aattttatag ctgttgatga ctctccagaa 660

| | |
|--|-----|
| gagattaggg ctaagataaa gaaagcatac tgcccagctg gagttgttga aggaaatcca | 720 |
| ataatggaga tagctaaata cttccttgaa tatcctttta ccataaaaag gccagaaaaa | 780 |
| tttgggtggag atttgacagt taatagctat gaggagttag agagtttatt taaaaataag | 840 |
| gaattgcatc caatggattt aaaaaatgct gtagctgaag aacttataaa gatttttagag | 900 |
| ccaattagaa agagattata a | 921 |

REIVINDICAÇÕES

1. Sistema de tradução, **CARACTERIZADO** pelo fato de que compreende:

um primeiro aminoácido não natural que é fenilselenocisteína;

um primeiro RNAt ortogonal aminoacil sintetase (O-RS); e

um primeiro RNAt ortogonal (O-RNAt);

em que o dito primeiro O-RS preferencialmente aminoacila o dito primeiro O-RNAt com a dita fenilselenocisteína.

2. Sistema de tradução, de acordo com a reivindicação 1, **CARACTERIZADO** pelo fato de que o dito primeiro O-RS preferencialmente aminoacila o dito primeiro O-RNAt com a dita fenilselenocisteína com uma eficiência que é pelo menos 50 % da eficiência observada para um sistema de tradução que compreende o dito O-RNAt, a dita fenilselenocisteína, e uma RNAt aminoacil sintetase que compreende a sequência de aminoácido de SEQ ID NO: 4, 6 ou 8.

3. Sistema de tradução, de acordo com a reivindicação 1, **CARACTERIZADO** pelo fato de que o dito primeiro O-RS compreende uma sequência de aminoácido selecionada do grupo que consiste em sequências de aminoácido apresentadas em SEQ ID NOS: 4, 6, 8, e variantes conservadoras destes.

4. Sistema de tradução, de acordo com a reivindicação 1, **CARACTERIZADO** pelo fato de que o dito primeiro O-RNAt é um RNAt supressor âmbar.

5. Sistema de tradução, de acordo com a reivindicação 1, **CARACTERIZADO** pelo fato de que o dito primeiro O-RNAt compreende ou é codificado por uma sequência de polinucleotídeo apresentada em SEQ ID NO: 1.

6. Sistema de tradução, de acordo com a reivindicação 1, **CARACTERIZADO** pelo fato de que compreende adicionalmente um ácido nucléico que codifica uma proteína de interesse, o dito ácido nucléico que compreende pelo menos um códon seletor, em que o dito códon seletor é reconhecido pelo dito primeiro O-RNAt.

7. Sistema de tradução, de acordo com a reivindicação 6, **CARACTERIZADO** pelo fato de que compreende adicionalmente um segundo O-RS e um segundo O-RNAt, em que o segundo O-RS preferencialmente aminoacila o segundo O-RNAt com um segundo aminoácido não natural que é diferente do primeiro aminoácido não natural, e em que o segundo O-RNAt reconhece um códon seletor que é diferente do códon seletor reconhecido pelo primeiro O-RNAt.

8. Sistema de tradução, de acordo com a reivindicação 1, **CARACTERIZADO** pelo fato de que o dito sistema compreende uma célula hospedeira que compreende o dito primeiro aminoácido não natural, o dito primeiro O-RS e o dito primeiro O-RNAt.

9. Sistema de tradução, de acordo com a reivindicação 8, **CARACTERIZADO** pelo fato de que a dita célula hospedeira é uma célula eubacteriana.

10. Sistema de tradução, de acordo com a reivindicação 9, **CHARACTERIZADO** pelo fato de que a dita célula eubacteriana é uma célula *E. coli*.

11. Sistema de tradução, de acordo com a reivindicação 8, **CHARACTERIZADO** pelo fato de que a dita célula hospedeira compreende um polinucleotídeo que codifica o dito primeiro O-RS.

12. Sistema de tradução, de acordo com a reivindicação 11, **CHARACTERIZADO** pelo fato de que o dito polinucleotídeo compreende uma sequência de nucleotídeo apresentada em SEQ ID NO: 5, 7 ou 9.

13. Sistema de tradução, de acordo com a reivindicação 8, **CHARACTERIZADO** pelo fato de que a dita célula hospedeira compreende um polinucleotídeo que codifica o dito primeiro O-RNAt.

14. Método para produzir em um sistema de tradução uma proteína que compreende um aminoácido não natural em uma posição selecionada, **CHARACTERIZADO** pelo fato de que o método compreende:

(a) fornecer um sistema de tradução que compreende:
um primeiro aminoácido não natural que é fenilselenocisteína;
uma primeira RNAt ortogonal aminoacil sintetase (O-RS);
um primeiro RNAt ortogonal (O-RNAt), em que o dito primeiro O-RS preferencialmente aminoacila o dito primeiro O-RNAt com a dita fenilselenocisteína; e,

um ácido nucléico que codifica a dita proteína, em que o dito ácido nucléico compreende pelo menos um códon seletor que é reconhecido pelo dito primeiro O-RNAt; e,

(b) incorporar o dito aminoácido não natural na dita posição selecionada na dita proteína durante a tradução da dita proteína em resposta ao dito códon seletor, produzindo dessa maneira a dita proteína que compreende o dito aminoácido não natural na posição selecionada.

15. Método, de acordo com a reivindicação 14, **CHARACTERIZADO** pelo fato de que o dito primeiro O-RS preferencialmente aminoacila o dito primeiro O-RNAt com a dita fenilselenocisteína com uma eficiência que é pelo menos 50 % da eficiência observada para um sistema de tradução que compreende o dito O-RNAt, a dita fenilselenocisteína, e um RNAt aminoacil sintetase que compreende a sequência de aminoácido de SEQ ID NO: 4, 6 ou 8

16. Método, de acordo com a reivindicação 14, **CHARACTERIZADO** pelo fato de que o dito sistema de tradução compreende um polinucleotídeo que codifica o dito O-RS.

17. Método, de acordo com a reivindicação 14, **CHARACTERIZADO** pelo fato de que o dito fornecimento de um sistema de tradução compreende fornecer um O-RS que compreende uma sequência de aminoácido selecionada do grupo que consiste nas seqüências de aminoácido apresentadas em SEQ ID NOS: 4, 6, 8, e variantes conservadoras

destes.

18. Método, de acordo com a reivindicação 14, **CARACTERIZADO** pelo fato de que o dito fornecimento de um primeiro RNAt ortogonal aminoacil sintetase compreende mutar um aminoácido ligando a cavidade de um RNAt aminoacil sintetase tipo selvagem por mutagênese direcionada ao sítio, e selecionar um O-RS resultante que preferencialmente aminoacila o dito O-RNAt com o dito aminoácido não natural, fornecendo dessa maneira o dito primeiro RNAt ortogonal aminoacil sintetase.

19. Método, de acordo com a reivindicação 18, **CARACTERIZADO** pelo fato de que a dita etapa de seleção compreende selecionar positivamente e selecionar negativamente para o dito O-RS de um pool que compreende uma pluralidade de moléculas de RNAt aminoacil sintetase mutantes produzidas depois do dito said mutagênese direcionada ao sítio.

20. Método, de acordo com a reivindicação 14, **CARACTERIZADO** pelo fato de que o dito fornecimento de um sistema de tradução compreende fornecer um polinucleotídeo que codifica o dito O-RNAt.

21. Método, de acordo com a reivindicação 14, **CARACTERIZADO** pelo fato de que o dito fornecimento de um sistema de tradução compreende fornecer um O-RNAt que é um RNAt supressor âmbar.

22. Método, de acordo com a reivindicação 14, **CARACTERIZADO** pelo fato de que o dito fornecimento de um sistema de tradução compreende fornecer um ácido nucléico que codifica ou compreende o dito O-RNAt, em que o dito ácido nucléico compreende uma seqüência de polinucleotídeo apresentada em SEQ ID NO: 1.

23. Método, de acordo com a reivindicação 14, **CARACTERIZADO** pelo fato de que o dito fornecimento de um sistema de tradução compreende fornecer um ácido nucléico que codifica o dito O-RS, em que o dito ácido nucléico compreende uma seqüência de polinucleotídeo apresentada em SEQ ID NO: 5, 7 ou 9.

24. Método, de acordo com a reivindicação 14, **CARACTERIZADO** pelo fato de que o dito fornecimento de um sistema de tradução compreende fornecer um ácido nucléico que compreende um códon seletor de âmbar.

25. Método, de acordo com a reivindicação 14, **CARACTERIZADO** pelo fato de que a dita proteína compreende adicionalmente um segundo aminoácido não natural que é diferente do dito primeiro aminoácido não natural, e em que o dito sistema de tradução compreende adicionalmente um segundo O-RS e um segundo O-RNAt, em que o segundo O-RS preferencialmente aminoacila o segundo O-RNAt com um segundo aminoácido não natural que é diferente do primeiro aminoácido não natural, e em que o segundo O-RNAt reconhece um códon seletor no ácido nucléico que é diferente do códon seletor reconhecido pelo primeiro O-RNAt.

26. Método, de acordo com a reivindicação 14, **CARACTERIZADO** pelo fato de que o dito fornecimento de um sistema de tradução compreende fornecer uma célula hospedeira, em que a dita célula hospedeira compreende o dito primeiro aminoácido não natural, o dito primeiro O-RS, o dito primeiro O-RNAt e o dito ácido nucléico, e em que a dita etapa de incorporação compreende cultivar a dita célula hospedeira.

27. Método, de acordo com a reivindicação 26, **CARACTERIZADO** pelo fato de que a dita célula hospedeira é uma célula eubacteriana hospedeira.

28. Método, de acordo com a reivindicação 27, **CARACTERIZADO** pelo fato de que a dita célula eubacteriana hospedeira é uma célula hospedeira *E. coli*.

29. Método, de acordo com a reivindicação 26, **CARACTERIZADO** pelo fato de que a dita célula hospedeira compreende um ácido nucléico que codifica o dito O-RS.

30. Método, de acordo com a reivindicação 29, **CARACTERIZADO** pelo fato de que o dito ácido nucléico que codifica o dito O-RS compreende uma sequência de nucleotídeo apresentada em SEQ ID NO: 5, 7 ou 9.

31. Método, de acordo com a reivindicação 14, **CARACTERIZADO** pelo fato de que o dito método compreende adicionalmente etapa (c) reagir a dita fenilselenocisteína na dita posição selecionada na dita proteína para produzir uma proteína que compreende deidroalanina na dita posição selecionada.

32. Método, de acordo com a reivindicação 31, **CARACTERIZADO** pelo fato de que a dita reação é por eliminação oxidativa.

33. Método, de acordo com a reivindicação 31, **CARACTERIZADO** pelo fato de que a dita reação compreende expor a dita fenilselenocisteína ao peróxido de hidrogênio.

34. Método, para produzir em um sistema de tradução uma proteína que compreende uma fração de lipídio em uma posição selecionada, **CARACTERIZADO** pelo fato de que o método compreende:

(a) fornecer um sistema de tradução que compreende:

um aminoácido não natural de fenilselenocisteína;

um RNAt ortogonal aminoacil sintetase (O-RS);

um RNAt ortogonal (O-RNAt), em que o dito O-RS preferencialmente aminoacila o dito O-RNAt com a dita fenilselenocisteína; e,

um ácido nucléico que codifica a dita proteína, em que o dito ácido nucléico compreende um códon seletor na dita posição selecionada que é reconhecido pelo dito O-RNAt;

(b) incorporar fenilselenocisteína na dita posição selecionada na dita proteína durante tradução da dita proteína em resposta ao dito códon seletor;

(c) reagir a dita fenilselenocisteína na posição selecionada na dita proteína para produzir uma proteína que compreende deidroalanina na posição selecionada; e,

(d) reagir a dita deidroalanina na posição selecionada na dita proteína com uma fração de lipídio para produzir uma proteína que compreende um lipídeo na fração do ácido na posição selecionada na proteína.

35. Método, de acordo com a reivindicação 34, **CARACTERIZADO** pelo fato de que a dita reação da dita fenilselenocisteína na posição selecionada na dita proteína é por eliminação oxidativa.

36. Método, de acordo com a reivindicação 34, **CARACTERIZADO** pelo fato de que a dita reação da dita fenilselenocisteína na posição selecionada na dita proteína compreende expor a dita fenilselenocisteína ao peróxido de hidrogênio.

37. Método, de acordo com a reivindicação 34, **CARACTERIZADO** pelo fato de que a dita fração de lipídio é selecionada do grupo que consiste em ácido tiopalmítico, famesilmercaptano e 1-hexadecanetiol.

38. Método, de acordo com a reivindicação 34, **CARACTERIZADO** pelo fato de que a dita fração de aminoácido lipidado é selecionada do grupo que consiste em palmitoilcisteína, farnesilcisteína e S-hexadecilcisteína.

39. Método, de acordo com a reivindicação 34, **CARACTERIZADO** pelo fato de que a dita reação da dita deidroalanina na posição selecionada na dita proteína é por uma reação de adição de Michael.

40. Método para produzir em um sistema de tradução uma proteína que compreende um aminoácido de deidroalanina não natural em uma posição selecionada, **CARACTERIZADO** pelo fato de que o método compreende:

(a) fornecer um sistema de tradução que compreende:

um aminoácido não natural de fenilselenocisteína;

um RNAt ortogonal aminoacil sintetase (O-RS);

um RNAt ortogonal (O-RNAt), em que o dito O-RS preferencialmente aminoacila o dito O-RNAt com a dita fenilselenocisteína; e,

um ácido nucléico que codifica a dita proteína, em que o dito ácido nucléico compreende um códon seletor na dita posição selecionada que é reconhecido pelo dito O-

(b) incorporar o dito aminoácido não natural de fenilselenocisteína na dita posição selecionada na dita proteína durante a tradução da dita proteína em resposta ao dito códon seletor,

(c) reagir a dita fenilselenocisteína na posição selecionada na dita proteína, produzindo dessa maneira a dita proteína que compreende o dito aminoácido de deidroalanina não natural na posição selecionada.

41. Método, de acordo com a reivindicação 40, **CARACTERIZADO** pelo fato de que a dita reação da dita fenilselenocisteína na posição selecionada na dita proteína é por eliminação oxidativa.

42. Método, de acordo com a reivindicação 40, **CHARACTERIZADO** pelo fato de que a dita reação da dita fenilselenocisteína na posição selecionada na dita proteína compreende expor a dita fenilselenocisteína ao peróxido de hidrogênio.

5 43. Composição, **CHARACTERIZADA** pelo fato de que compreende um polipeptídeo que compreende uma sequência de amino ácido selecionada do grupo que consiste nas sequências de aminoácido apresentadas em SEQ ID NOS: 4, 6, 8, e variantes conservadoras destes.

10 44. Composição, de acordo com a reivindicação 43, **CHARACTERIZADA** pelo fato de que o dito polipeptídeo da variante conservadora aminoacila um RNAt ortogonal cognate (O-RNAt) com um aminoácido não natural com uma eficiência que é pelo menos 50 % da eficiência observada para um sistema de tradução que compreende o dito O-RNAt, o dito aminoácido não natural, e um RNAt aminoacil sintetase que compreende a sequência de aminoácido de SEQ ID NO: 4, 6 ou 8.

15 45. Composição, de acordo com a reivindicação 43, **CHARACTERIZADA** pelo fato de que a dita composição compreende uma célula que compreende o polipeptídeo.

46. Polinucleotídeo, **CHARACTERIZADO** pelo fato de que codifica o polipeptídeo, de acordo com a reivindicação 43.

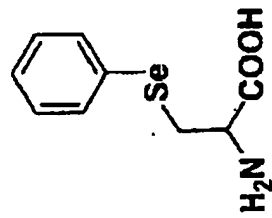
20 47. Polinucleotídeo, de acordo com a reivindicação 46, **CHARACTERIZADO** pelo fato de que o dito polinucleotídeo compreende a sequência de nucleotídeo de SEQ ID NO: 5, 7 ou 9.

48. Vetor, **CHARACTERIZADO** pelo fato de que compreende um polinucleotídeo de acordo com a reivindicação 46.

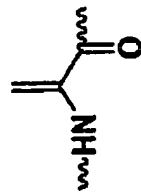
49. Vetor de expressão, **CHARACTERIZADO** pelo fato de que compreende um polinucleotídeo de acordo com a reivindicação 46.

25 50. Célula, **CHARACTERIZADA** pelo fato de que compreende um vetor, o vetor que compreende um polinucleotídeo de acordo com a reivindicação 46.

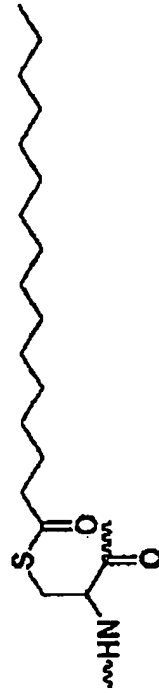
51. Composição, **CHARACTERIZADA** pelo fato de que compreende um polinucleotídeo que compreende uma sequência de nucleotídeo apresentadas em SEQ ID NO: 5, 7 ou 9.



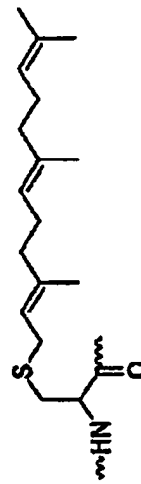
Fenilselenocistina
(1)



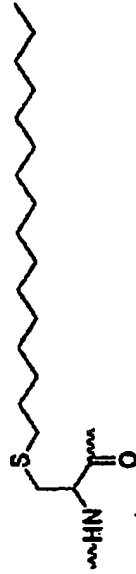
Deidroalanina
(2)



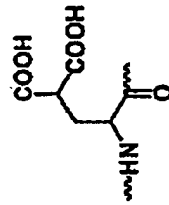
Palmitoilcisteína
(3)



Farnesilcisteína
(4)



S-hexadecilcisteína
(5)



Ácido γ-carboxiglutâmico
(6)

Fig. 1

Sequência de nucleotídeo e aminoácido

| SEQ ID NO: | Descrição | Sequência |
|------------|--|---|
| 1 | supressor <i>Methanococcus jannaschii</i> -tirosil-RNAt _{CUA} aka MjRNA-tyr(CUA) ou mutRNA ^{tyr} _{CUA} | CCGGCGGUAGUUCAGCAGGGCAGAACGGCGGACUCUAAAUCCG CAUGGCGCUGGUUCAAUCCGGCCCGCGGACCA |
| 2 | <i>Methanococcus jannaschii</i> tirosil-RNAt sintetase tipo selvagem (MjTyrS) Sequência de aminoácido | MDEFEMIKRNTSEIISEEELREVLKKDEKSAYIGFEPGSKIHL GHYLQIKKMIDLQNAFGDI IILLADLHAYLNQKGELDEIRKIG DYNKKVFEAMGLKAKYVYGSEFQLDKDYTLNVYRLALKTTTLKR ARRSMELIAREDENPKVAEVIYPIMQVNDIHYLGVDVAVGGME QRKIHMLARELLPKKVVCIHNPVLTGLDGEKGMSSSKGNFI AVDSDPEEIRAKIKKAYCPAGVVEGNPIMEIAKYFLEYPLTIKRP EKFGGDLTVNSYEELESLEFKNKELHPMDLKNVAEELIKILEP IRKRL |
| 3 | <i>Methanococcus jannaschii</i> tirosil-RNAt sintetase tipo selvagem (MjTyrS) Sequência de nucleotídeo | ATGGACGAATTTGAAATGATAAAGAGAAACACATCTGAAATTA TCAGCGAGGAAGAGTTAAGAGAGGTTTTAAAAAAGATGAAAA ATCTGCTTACATAGGTTTTGAACCAAGTGGTAAAATACATTTA GGGCATTATCTCCAAATAAAAAAGATGATTGATTTACAAAATG CTGGATTTGATATAATTATATTGTTGGCTGATTTACACGCCTA TTTAAACCAGAAAGGAGAGTTGGATGAGATTAGAAAAATAGGA GATTATAACAAAAAGTTTTTGAAGCAATGGGGTTAAAGGCAA AATATGTTTTATGGAAGTGAATTCAGCTTGATAAGGATTATAC ACTGAATGTCTATAGATTGGCTTTAAAAACTACCTTAAAAAGA GCAAGAAGGAGTATGGAACCTATAGCAAGAGAGGATGAAAATC CAAAGGTTGCTGAAGTTATCTATCCAATAATGCAGGTTAATGA TATTCATTATTTAGGCGTTGATGTTGCAGTTGGAGGGATGGAG CAGAGAAAAATACACATGTTAGCAAGGGAGCTTTTACCAAAAA AGGTTGTTTGTATTACAAACCTGTCTTAACGGGTTTGGATGG AGAAGGAAAGATGAGTTCTTCAAAGGGAATTTTATAGCTGTT GATGACTCTCCAGAAGAGATTAGGGCTAAGATAAAGAAAGCAT ACTGCCCCAGCTGGAGTTGTTGAAGGAAATCCAATAATGGAGAT AGCTAAATACTTCCTTGAATATCCTTTAACCATAAAAAGGCCA GAAAAATTTGGTGGAGATTTGACAGTTAATAGCTATGAGGAGT TAGAGAGTTTATTTAAAAATAAGGAATTGCATCCAATGGATTT AAAAAATGCTGTAGCTGAAGAACTTATAAAGATTTTAGAGCCA ATTAGAAAGAGATTA |
| 4 | Clone 1 (SD) Fenilselenocistina aminoacil-RNAt sintetase Clone SD (PhSeRS-SD) sequência de aminoácido (derivado de <i>Methanococcus jannaschii</i> tirosil RNAt-sintetase tipo selvagem) | MDEFEMIKRNTSEIISEEELREVLKKDEKSAWIGFEPGSKIHL GHYLQIKKMIDLQNAFGDI IIELADLGAYLNQKGELDEIRKI GDYNKKVFEAMGLKAKYVYGSEFQLDKDYTLNVYRLALKTTTLK RARRSMELIAREDENPKVAEVIYPIMQVNQIHYSGVDVAVGG MEQRKIHMLARELLPKKVVCIHNPVLTGLDGEKGMSSSKGNFI AVDSDPEEIRAKIKKAYCPAGVVEGNPIMEIAKYFLEYPLTIK RPEKFGGDLTVNSYEELESLEFKNKELHPMDLKNVAEELIKIL EPIRKRL |

Fig. 2

| SEQ ID NO: | Descrição | Sequência |
|------------|---|--|
| 5 | <p>Clone 1 (SD)</p> <p>Fenilselenocistina aminoacil-RNAt sintetase Clone SD (PhSeRS-SD) sequência de nucleotídeo</p> | ATGGACGAATTTGAAATGATAAAGAGAAACACATCTGAAATTA TCAGCGAGGAAGAGTTAAGAGAGGTTTTAAAAAAGATGAAAA ATCTGCTTGGATAGGTTTTGAACCAAGTGGTAAAATACATTTA GGGCATTATCTCCAAATAAAAAAGATGATTGATTTACAAAATG CTGGATTTGATATAATTATAGAGTTGGCTGATTTAGGGGCCTA TTTAAACCAGAAAGGAGAGTTGGATGAGATTAGAAAAATAGGA GATTATAACAAAAAAGTTTTTGAAGCAATGGGGTTAAAGGCAA AATATGTTTTATGGAAGTGAATTCCAGCTTGATAAGGATTATAC ACTGAATGTCTATAGATTGGCTTTAAAACTACCTTAAAAAGA GCAAGAAGGAGTATGGAACCTTATAGCAAGAGAGGATGAAAAATC CAAAGGTTGCTGAAGTTATCTATCCAATAATGCAGGTTAATCA AATTCATTATAGTGGCGTTGATGTTGCAGTTGGAGGGATGGAG CAGAGAAAAATACACATGTTAGCAAGGGAGCTTTTACCAAAAA AGGTTGTTTTGTATTACAAACCCTGTCTTAACGGGTTTGGATGG AGAAGGAAAGATGAGTTCTTCAAAGGGAATTTTATAGCTGTT GATGACTCTCCAGAAGAGATTAGGGCTAAGATAAAGAAAGCAT ACTGCCCAGCTGGAGTTGTTGAAGGAAATCCAATAATGGAGAT AGCTAAATACTTCTTGAATATCCTTTAACCATAAAAAGGCCA GAAAAATTTGGTGGAGATTTGACAGTTAATAGCTATGAGGAGT TAGAGAGTTTATTTAAAAATAAGGAATTGCATCCAATGGATTT AAAAATGCTGTAGCTGAAGAACTTATAAAGATTTTAGAGCCA ATTAGAAAGAGATTATAA |
| 6 | <p>Clone 2 (K4)</p> <p>Fenilselenocistina aminoacil-RNAt sintetase Clone K4 (PhSeRS-K4) sequência de aminoácido (derivado de <i>Methanococcus jannaschii</i> tirosil RNAt sintetase tipo selvagem)</p> | MDEFEMIKRNTSEIISEEELREVLKKDEKSAWIGFEPGSKIHL GHYLIQIKMIDLQNAFGDIIHSLADLGAYLNQKGELDEIRKI GDYNNKKVFAMGLKAKYVYVYSENSLDKDYTLNVYRLALKTTL KRARRSMELIAREDENPKVAEVIYPIQVNSIHYEVDVAVG GMEQRKIHMLARELLPKVVCIHNPVLTGLDGEKMSSSKGNF IAVDDSPEEIRAKIKKAYCPAGVVEGNPIMEIAKYFLEYPLTI KRPEKFGGDLTVNSYEELESFLKKNELHPMDLKNVAEELIKI LEPIRKRL |
| 7 | <p>Clone 1 (SD)</p> <p>Fenilselenocistina aminoacil-RNAt sintetase Clone K4 (PhSeRS-K4) sequência de nucleotídeo</p> | ATGGACGAATTTGAAATGATAAAGAGAAACACATCTGAAATTA TCAGCGAGGAAGAGTTAAGAGAGGTTTTAAAAAAGATGAAAA ATCTGCTTGGATAGGTTTTGAACCAAGTGGTAAAATACATTTA GGGCATTATCTCCAAATAAAAAAGATGATTGATTTACAAAATG CTGGATTTGATATAATTATACATTTGGCTGATTTAGGGGCCTA TTTAAACCAGAAAGGAGAGTTGGATGAGATTAGAAAAATAGGA GATTATAACAAAAAAGTTTTTGAAGCAATGGGGTTAAAGGCAA AATATGTTTTATGGAAGTGAATAATCTCTTGATAAGGATTATAC ACTGAATGTCTATAGATTGGCTTTAAAACTACCTTAAAAAGA GCAAGAAGGAGTATGGAACCTTATAGCAAGAGAGGATGAAAAATC CAAAGGTTGCTGAAGTTATCTATCCAATAATGCAGGTTAATTC TATTCATTATGAGGGCGTTGATGTTGCAGTTGGAGGGATGGAG CAGAGAAAAATACACATGTTAGCAAGGGAGCTTTTACCAAAAA AGGTTGTTTTGTATTACAAACCCTGTCTTAACGGGTTTGGATGG AGAAGGAAAGATGAGTTCTTCAAAGGGAATTTTATAGCTGTT GATGACTCTCCAGAAGAGATTAGGGCTAAGATAAAGAAAGCAT ACTGCCCAGCTGGAGTTGTTGAAGGAAATCCAATAATGGAGAT AGCTAAATACTTCTTGAATATCCTTTAACCATAAAAAGGCCA GAAAAATTTGGTGGAGATTTGACAGTTAATAGCTATGAGGAGT TAGAGAGTTTATTTAAAAATAAGGAATTGCATCCAATGGATTT AAAAATGCTGTAGCTGAAGAACTTATAAAGATTTTAGAGCCA ATTAGAAAGAGATTATAA |

Fig.2 (continuação)

| SEQ ID NO: | Descrição | Seqüência |
|------------|---|--|
| 8 | Clone 3 (K5) Fenilselenocistina aminoacil-RNAt sintetase Clone K5 (PhSeRS-K5) de seqüência de aminoácido (derivado de <i>Methanococcus jannaschii</i> tirosil RNAt-sintetase tipo selvagem) | MDEFEMIKRNTSEIISEEELREVLKKDEKSAWIGFEPGSKIHL GHYLQIKKMIDLQAGFDIIIEHLGDLGAYLNQKGELDEIRKI GDYNKKVFEAMGLKAKYVYGSEKSLDKDYTLNVYRLALKTTL KRARRSMELIAREDENPKVAEVIYPIMQVNEIHYEGVDVAVG GMEQRKIHMLARELLPKKVVCIHNPVLTGLDGEKMSSSKGNF IAVDDSPFEEIRAKIKKAYCPAGVVEGNPIMEIAKYFLEYPLTI KRPEKFGGDLTVNSYEELESLEFKNKLHPMDLKNAVABELIKI LEPIRKRL |
| 9 | Clone 3 (K5) Fenilselenocistina aminoacil-RNAt sintetase Clone K5 (PhSeRS-K5) seqüência de nucleotídeo | ATGGACGAATTTGAAATGATAAAGAGAAACACATCTGAAATTA TCAGCGAGGAAGAGTTAAGAGAGGTTTTAAAAAAGATGAAAA ATCTGCTTGGATAGGTTTTGAACCAAGTGGTAAAATACATTTA GGGCATTATCTCCAAATAAAAAAGATGATTGATTTACAAAATG CTGGATTTGATATAATTATACATTTGGGTGATTTAGGCGCCTA TTTAAACCAGAAAGGAGAGTTGGATGAGATTAGAAAAATAGGA GATTATAACAAAAAGTTTTTGAAGCAATGGGGTTAAAGGCAA AATATGTTTATGGAAGTGAAAAGTCTCTTGATAAGGATTATAC ACTGAATGTCTATAGATTGGCTTTAAAAACTACCTTAAAAAGA GCAAGAAGGAGTATGGAACCTATAGCAAGAGAGGATGAAAATC CAAAGGTTGCTGAAGTTATCTATCCAATAATGCAGGTTAATGA GATTCATTATGAGGGCGTTGATGTTGCAGTTGGAGGGATGGAG CAGAGAAAAATACACATGTTAGCAAGGGAGCTTTTACCAAAAA AGGTTGTTTGTATTACAAACCCTGTCTTAACGGGTTTGGATGG AGAAGGAAAGATGAGTTCTTCAAAAGGGAATTTTATAGCTGTT GATGACTCTCCAGAAGAGATTAGGGCTAAGATAAAGAAAGCAT ACTGCCCAGCTGGAGTTGTTGAAGGAAATCCAATAATGGAGAT AGCTAAATACTTCCTTGAATATCCTTTAACCATAAAAAAGGCCA GAAAAATTTGGTGGAGATTTGACAGTTAATAGCTATGAGGAGT TAGAGAGTTTATTTAAAAATAAGGAATTGCATCCAATGGATTT AAAAAATGCTGTAGCTGAAGAACTTATAAAGATTTTAGAGCCA ATTAGAAAGAGATTATAA |

Fig.2 (continuação)

RESUMO**"EXPRESSÃO DE PROTEÍNAS GENETICAMENTE PROGRAMADA CONTENDO O AMINOÁCIDO NÃO NATURAL FENILSELENOCISTEÍNA"**

A invenção diz respeito a pares ortogonais de RNAs e RNAt aminoacil sintetase que podem incorporar o aminoácido não natural fenilselenocisteína em proteínas produzidas em células hospedeiras eubacterianas, tal como E. coli. A invenção fornece, por exemplo, mas sem limitações, RNAt ortogonal aminoacil sintetase inédito, polinucleotídeos que codificam os sistemas de moléculas de sintetase inéditos, fenilselenocisteína e tradução. A invenção fornece adicionalmente métodos para produzir proteínas modificadas (por exemplo, proteínas lipidadas) por meio de modificação programada do resíduo de fenilselenocisteína em uma proteína.